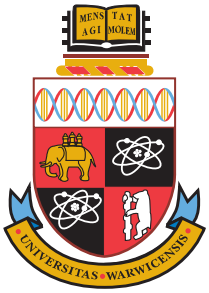


Directed evolution of bacteriophage T7 receptor tropism

by

Paul Richard MacDonald

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Supervisors: Prof. Alfonso Jaramillo and Prof. Nick Waterfield

MOAC Doctoral Training Centre

September 2018

THE UNIVERSITY OF WARWICK

What I learned on my own I still remember

— Nassim Nicholas Taleb

Acknowledgements

I would firstly like to thank Prof. Alfonso Jaramillo for introducing me to the world of phages, and an idea so beautiful that I will never quite forget it.

I'd like to thank my lab, and it's many members over the years, without whom I never would have held a pipette, let alone been able to re-wire my disorganised brain for lab work. In particular, I'd like to thank Rui for mentoring me via his existence; Manish for fascinating conversations long into the routine nights of lab work; Will for the initial advice I did not heed (we all made the same error together) and for generally being cool (plus for providing me with this Latex template); Eduardo for putting me up (and putting up with me); and lastly for Aurelija, who held me up when everything otherwise would have crumbled apart. And thank you also to the many others who I forgot to add in that extended sentence... like Satya - congrats again on the kid! :) and Jack for the fun we had in the early days; and John for guiding the design of one of the key experiments.

I'd also like to thank my parents, for being the nicest parents one could ever have, and my brother, for exceeding me in every way (except in sports, as it will hopefully remain).

P. M.

Declaration

All the data within is from my own work, within the lab of Prof. Alfonso Jaramillo, unless otherwise specified. No part of this work has been submitted for a previous degree.

pSB3T5-HRg5-cmk-trxA (pAJ57) was made by Matthew Tridgett. pSB6A1-gp17 (pAJ148) was made by Antonia Sagona under Aurelija Grigonyte's supervision. pML-HRtrxA gifted by Michal Legiewicz. pET30a+g17_371-467 was a gift from the lab of Mark van Raaij, and the protein gel shown in Figure 2.10d was performed by, and is the data of, Marta Sanz Gaitero as part of said lab. pMMB207-Ail was provided by the Krukonis Lab. T7-ReRb was a gift from the Qimron Lab, as was BW25113 Δ*cmk*::*tetA* Δ*trxA*::*kanR* (referred to as BW25113 Δ*cmk* Δ*trxA*). pET24a-gp5-kan was provided by the Richardson Lab. BW25113 Δ*cmk* Δ*trxA* was transformed with pET24a-gp5-amp by George Kimberley, who both constructed that plasmid, and performed the transformation. T7 Δ*g5*::(Δ*cmk*-*trxA*) and T7-ReRb Δ*g5*::(Δ*cmk*-*trxA*) were constructed by Matthew Tridgett. pSB4K5-ER-g5 was provided by the Jaramillo lab and pAJ216 was constructed by Michal Legiewicz.

Abstract

Bacteriophage (phage) typically infect a narrow range of hosts. In this thesis, we try to control phage-host specificity through the directed evolution of phage receptor-binding domains. As a proof of concept, we aim to re-adapt the binding affinity of T7 phage so that it no longer requires lipopolysaccharide (LPS) for the infection of *E. coli* and instead utilises only an outer membrane protein (OmpF). To this end, we remove an essential gene from T7 and complement it in the host, thus linking phage propagation with the presence or absence of the *in trans* essential gene. By then providing *E. coli* with a receptor to be targeted by T7, we can positively select for those phage that bind to strains with that receptor, or negatively select against phage that bind strains without that receptor. We begin by demonstrating that we can amend the T7 genome *in vivo* and, moreover, can engineer a chimeric tail fibre fused from two different phage species (T7 and Yep-phi phage). We then identify an issue with this engineering process, namely, the appearance of individually deficient T7 that can propagate through co-infection, and both model the phenomena and take steps to address it. To inform the design of the directed evolution experiments, we develop a system of delay differential equations for modelling different strategies. For our first set of evolution experiments, we show that we can effect the binding affinity of T7 phage to different LPS phenotypes, and investigate the impact of different strategies on the outcome. We subsequently make use of this knowledge to coerce T7 affinity towards the OmpF receptor and find indirect evidence of evolved T7 infecting with the OmpF receptor, in the absence of a full LPS phenotype. Finally, we model the stochastic transfer of phage between phenotypes of different receptor binding affinities.

Contents

List of abbreviations

AMR : Anti-microbial resistance

bp : Base pairs

CFU : Colony Forming Unit

Cmk / *cmk*: Cytidine Monophosphate Kinase

CRISPR : Clustered Regularly Interspaced Short Palindromic Repeats

DDE : Delay differential equation

DDEs : Delay differential equations

DNA : Deoxyribonucleic acid

DNAP : DNA polymerase

dsDNA : Double stranded DNA

dsRNA : Double stranded RNA

gX : gene X

gpX : gene product X

GMP : Good manufacturing practice

E. coli : *Escherichia coli*

IPTG : β-D-1-thiogalactopyranoside

LB : Lysogeny broth

LPS : Lipopolysaccharide

NEB : New England Biolabs

NG : N-methyl-N9-nitro-N-nitrosoguanidine

PACE : Phage Assisted Continuous Evolution

PAGE : Polyacrylamide Gel Electrophoresis

PCR : Polymerase Chain Reaction

PFU : Plaque Forming Unit

ODE : Ordinary differential equation

ODEs : Ordinary differential equations

OMP : Outer membrane protein

OMPs: Outer membrane proteins

RBS: Ribosome Binding Site

RNA : Ribonucleic acid

RNAP : RNA polymerase

SD : Standard deviation

SNPs: Single Nucleotide Polymorphisms

WT : Wild type

Y. pestis : *Yersinia pestis*

Glossary

PFU - Plaque forming units. Sufficiently few and dispersed phage infecting a dense population of bacteria that has been spread over the surface of an agar petri dish, will create isolated zones of lysis through replication in proximal bacteria. These zones of lysis are referred to as plaques, and should they not overlap, a singular phage infection is tentatively assumed to have formed each plaque.

CFU - Colony forming units. Sufficiently few and dispersed bacteria spread over the surface of an agar petri dish will give rise to individual, non-overlapping, colonies. Each colony is tentatively assumed to have grown from a singular bacteria.

geneX - Gene names are italicised in the text.

GeneX - Protein names are non-italicised and have their leading letter capitalised.

Δ*geneX* - notation indicating that a gene has been functionally deleted from the organism genome.

Δ*geneX*::*geneY* - indicates that geneX has been replaced by geneY.

Δ*waaC* - A lipopolysaccharide mutant of E. coli BW25113.

gpX - gene protein X.

gp17 - T7 phage tail fibre.

gp12 - T7 tail tube.

gp11 - T7 tail adaptor, or gatekeeper, protein. gp17 tail fibres dock into this protein.

g10 - Encodes the T7 capsid. gp10A is the major capsid protein, gp10B is the minor capsid protein.

gp5 - T7 DNA polymerase. T7 requires gp5 to complex with TrxA to successfully form plaques.

TrxA - Thioredoxin A protein. Used by *E. coli* for redox reactions. Aids T7 polymerase processivity and is essential for T7 plaque formation.

Cmk - Cytidine monophosphate kinase. Phosphorylates cytidine monophosphate to cytidine diphosphate. Used by T7 to enable DNA replication. The absence of Cmk severely hampers T7 productivity.

Washout - Used in reference to a phage population that is diluted out of a culture over repeated passages. Derives from the same phenomena observed in phage evolution experiments in bioreactors, where phage populations that fail to maintain their numbers upon replication are physically flowed out of the cellstat.

List of Figures

List of Tables

1 Introduction

Anti-microbial resistance (AMR) has rapidly evolved into an ongoing concern for governments and forward-looking health authorities such as the WHO. It is predicted that by 2050, there will be 10 million deaths (according to the AMR review [1]) resulting from antibiotic-resistant bacteria worldwide (though this veracity of this number has been questioned [2]). As of September 2018, resistance to all treatments for *Klebsiella pneumoniae* has been observed worldwide (WHO). While the pipeline of new antibiotics slows (only 3 new classes of antibiotics have been brought to market from 1980-2013 [3][1]) and with industry bringing to market redevelopments or combinations of known active molecules [4], the international community has found itself looking for alternative solutions. Bacteriophage, and in particular phage therapies, are making something of resurgence outside of Russia, Georgia and Poland, where research into phage therapy continued during the Cold War. As natural killers of bacterial pathogens, bacteriophage are an obvious candidate for the control of rogue infections. Nevertheless, several arguments against phage therapy still loom large. Detractors point to the low concentrations of phage versus traditional antibiotics. Typical titres for phage manufactured under good manufacturing practices (GMP) are on the order of $10^{10} - 10^{12}$ PFU/ml, and probably can not exceed 10^{13}/ml [5] (compared with, for instance, $\approx 10^{18}$ molecules/ml in a 2.5 mg/ml vancomycin treatment). Phages *tend*[2] to have a narrow range of action, infecting a particular species of bacteria, or even a particular strains and phenotype variants of said species [8]. There are difficulties overcoming regulations for engineered phage products (which are, quite understandably, classed as genetically modified organisms) and in patenting WT phage products. There are questions also on whether different phage cocktails (the amalgamation of more

[1] and up until the time of writing, as far as this author can tell

[2] phages that display broader efficacy have been isolated [6, 7]

than one phage genotype into a given product) will need to be patented separately, as phage therapy joins the wavefront of other medicines pushing for regulation reform in the wake of a personalised medicine revolution. Safety and efficacy are also a concern. Though many bacteriophages live among us in our gut microbiota, liver tissue, lungs, skin and urinary tracts [9], our understanding of their interaction with humans is limited, and an area of research still in its juvenile stages. It appears that bacteriophages are able to exchange genetic material with eukaryotes (see bacteriophage WO [10]) which should give pause to those tempted to apply 'complete' synthetic biology approaches to antibiotic resistance [11, 12].

In defense of the therapeutic application of bacteriophage, phage can replicate at the site of interest, increasing the concentration of phage locally [13]. When compared with the safety of current standard, antibiotic treatments can come with a myriad side effects (cardiotoxicity, nephrotoxicity and hepatotoxicity amongst others [14]) which too have to be considered when assessing alternatives. Where phage specificity does create issues for drug design and application in the clinic, it also offers an opportunity to avoid non-pathogenic members of the microbiome [15], whose disruption can lead to long-lasting side effects (such as type I diabeties, and effects on immune system regulation [16]). In fact, as of September 2018, we are fast approaching a point where sequencing and analysing a patient's microbiome could be possible in a short timeframe. For instance, the US-based company Aperiomics claims to be able to identify all know bacteria and viruses within a sample using next-generation sequencing, and to provide results in 2-3 weeks (with aims to reduce this to 1-2 days in the coming years).

In light of this, it would be useful to engineer a molecular machine that could be

redesigned to adsorb to different pathogens selectively, based on their genomic sequence, without requiring many alterations to that underlying machinery. When infecting Gram-negative bacteria, phages typically rely on cell surface structures to grab a stable foothold for DNA ejection, or fusion with the outer membrane [17]. Of these cell surface structures, antigens are those most plentiful (occupying about 75% of total surface area [18]). That being said, the pathways producing these antigens are complicated (typically involving more than 10 genes) and are difficult to manipulate (though not completely infeasible [19]). Outer membrane proteins (OMPs) are less abundant on the cell surface [20, 21], but the regions which could bind to phage are better understood. Typically these are loop structures that protrude extracellularly [22–24] and that can be manipulated by mutating individual genes. Subsequently, they are more amenable to DNA cloning in *E. coli* and are also rather tolerant to changes in those structures [25, 26]. We thus propose a general methodology for tackling pathogenic infections. 1) Identify and analyse the pathogen genome; 2) Identify OMPs that could serve as phage receptors; 3) Design or evolve phage that infect the pathogen using the identified receptor.[3] In this thesis, we will attempt to deliver on this last part of this methodology. We will use *E. coli* as a rudimentary model for a given pathogen, by expressing the chosen receptor heterologously, and then design phage which infect *E. coli* using solely that receptor. This strategy therefore does not necessitate the culturing of these pathogens for success (or that the pathogen is indeed culturable) with the obvious benefits that entails.

[3] A limitation of this methodology is that extracellular capsules, such as O-antigen structures, restrict access to outer membrane proteins. Further work would also be needed to ensure productive replication in the target pathogen, or at least lysis.

1.1 Bacteriophage overview

Known bacteriophage are categorised into 4 varieties: tailed - which constitute the vast majority of phages categorised ($\approx 96\%$ in 2007 [27]); filamentous - long and rod-like; pleomorphic - with eclectic, tapered shapes; and polyhedral. Phages largely consist of an external structure to transport a given nucleic acid cargo within, so that when inside a complementary host they can reproduce those components, and pass sustainably from generation to generation [28]. The nucleic acid is double-stranded DNA (dsDNA) in tailed and pleomorphic phage. Whereas filamentous phages with either single or double-stranded nucleic acids (DNA or RNA) have been isolated [29]. The external structure is usually formed of protein capsomeres which assemble to encapsulate the internal structure in a protein capsid. In rarer cases the external structure also consists of lipids that derive from whichever host that phage infects (be they bacterial or archaeal [30]).

1.1.1 Bacteriophage life cycles

Phage infection strategies can be classified into 4 distinct types: 1) lytic and non-temperate; 2) chronic and non-temperate; 3) lytic and temperate; and 4) chronic and temperate [31]. In addition to which phage undergo 2 varieties of life cycle, known as a lytic cycle or a lysogenic cycle. The term 'lytic' refers to phages which at some point lyse (break open) the host cell from the inside out. Those phages that can undergo lysogenic life cycles, but which are not strictly limited to doing so, are called 'temperate' phages. In a lysogenic cycle DNA transduced into the host is either integrated into the host genome, or exists as an extrachromosomal plasmid (or rarely, as linear DNA [32, 33]). Phage DNA in this form, referred to

as a prophage, is replicated with host DNA during cell division. There are then 2 ways for phage to release progeny from inside their hosts. The typical strategy entails entering a lytic cycle, producing mature virions within the host, followed at some stage by lysis of the cell membrane from within. The less common strategy, which is characteristic of what are known as 'chronic' infections is virion extrusion, in which immature virions are transported across the cell wall before maturation extracellularly. A good example of this strategy is employed by the filamentous phage M13 [34]. The tailed phage class (or otherwise, the order *Caudovirales*) possesses phages that make use of either or both of these life cycles. *Caudovirales* is made up of 3 families: *Myoviridae*, *Siphovirdae* and *Podoviridae* all of which harbour linear dsDNA genomes. Siphovirdae have long, flexible tails which do not contract upon-infection, Myovirdae have long, straight tails that do contract, and Podoviridae tails are non-contractile and short (see Figure 1.1) [35].

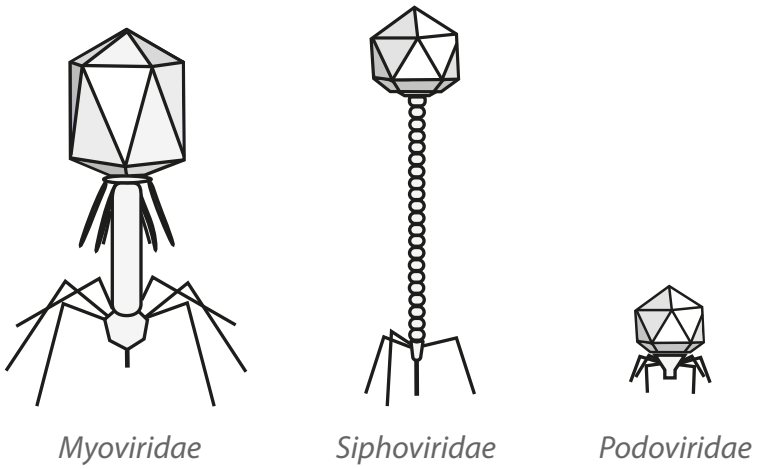


Figure 1.1: Idealised illustrations of *Caudovirales* phages. Adapted from Nobrega *et al.* [36].

1.1.2 Bacteriophage which interact with prokaryotic outer membrane proteins

Since the aim here is to develop a phage which utilises only outer membrane proteins when infecting bacteria, it is apparent that we should discuss phages which are already known to make use of outer membrane proteins in an exclusive way. λ phage of the *Siphovirdae* family uses the maltose porin LamB for infection, and has been experimentally evolved to use OmpF instead [37]. The gp65 tail sheath of *Podoviridae* N4 phage interacts with the approximately 5 copies of NfrA in the outer membrane of *E. coli* [38]. The T7-like phage Yep-phi has been shown to interact with Ail and OmpF outer membrane proteins, in addition to the LPS during infection of *Yersinia pestis*. There are several cases of *Myoviridae* phage which have been established to utilise outer membrane proteins for infection (T4, M2, Ox2 [39]). In particular, the T2-like *Myoviridae* phage Ox2 phage uses OmpA as a receptor [40, 41] and successive mutants isolated from plates of appropriate *E. coli* mutants indicated that it could flexibly switch its binding from OmpA to OmpC and even the LPS [42].

1.2 The case for a T7 bacteriophage scaffold

From this stand point, any of these phages, especially Ox2, could be good candidates for establishing a phage scaffold that could adapt to a given outer membrane protein. There are, however, a few additional constraints that we apply when choosing a phage to work with. We would like that the candidate can be thoroughly understood from the literature, so that we can be sure of the components that would be applied in the clinic. The best understood phages are λ (*Siphoviridae*), T4 (*Myoviridae*)

and T7 (*Podoviridae*). Each of these phages infects *E. coli,* which is the model organism for Gram-negative, and comes with useful resources, such as single-gene knockouts for every non-essential gene [43, 44]. λ phage is a temperate phage, and integrates its DNA into the *E. coli* chromosome post-infection, making it a riskier option for genetic modification. The lysogenic cycle can be prevented with a frameshift mutation to the *cI* repressor (denoted *cI26* [37]). Nevertheless, recombination in the wild can not be discounted. T4 already requires OmpC (in addition to the LPS) for the infection of *E. coli* K12 [45]. However, using the numbers as presented in De Paepe *et al.* [46], and the Bull *et al.* equation [47] for phage growth rate (with a constant bacterial concentration of 10^8 CFU/ml, and with cell and phage death neglected) T7 is asymptotically ≈ 2.6 times more productive than T4 (i.e. produces 2.6 times more phage over a sufficiently long period of time). In addition, since the latent period for T7 is roughly twice that of T4, T7 phage undergoes more frequent cycles of genome diversification and selection. The T4 genome is roughly 169 kb vs 40 kb for T7 (bringing with it a substantial increase in complexity) which means that the tail genes make up a smaller proportion of the genome. T7 is known to be able to tolerate high mutational loads [48]. T7 also lends itself better to replication in continuous culture systems where the phage replication rate has to exceed the bacterial growth (determined by the dilution rate) in order to hinder co-evolution in the cellstat [49]. Because of these advantages, we decided to determine if T7 could be engineered to move from using LPS as a receptor, to outer membrane proteins, as a starting point for a phage therapeutic.

However, with the benefit of hindsight, we should note that Cryo-EM imaging of T7 tail fibres prior to and during infection show no evidence of flexibilty at the

"knee" [50, 51] which limits the potential for binding to receptors that are less frequently found on the surface of *E. coli*. This can be contrasted with T4 long tail fibres which are composed of proximal and distal parts, each totalling roughly 70 nm in length, [52, 53] separated by hinge protein which does appear to flex to a near right angle, as well as extend outwards [54]. As a rough guide, there are $\approx 4 \times 10^4$ copies of the outer membrane protein OmpF [20] on the surface of *E. coli* versus 10^6 copies of LPS [21]. Even accounting for the fact that these OMPs are concentrated at the poles of *E. coli* [55] in the early stages of the cell cycle, we can expect a significantly reduced likelihood of T7 being bound to enough OMPs to stabilise the cell-phage interaction and initiate infection, if that interaction was mediated by the gp17 tail fibres. Were T7 able to find another functional domain for binding to cell surface molecules, for instance, by using the central gp12 tail tube, this issue could be overcome.

1.3 T7 bacteriophage

T7 bacteriophage was first officially reported by Demerec and Fano in 1944, following isolation on *E. coli* B [56]. It is widely studied, and a standard pedaological example for bacteriophage. This podovirdae phage fits a dsDNA genome of ≈40 kb into a $\approx 56nm^3$ icosahedral protein capsid assembled from 415 gp10 molecules, roughly 95% of which are gp10A, and the remainder from the frameshifted gp10B [57]. The capsid is connected to the associated tail elements with the suitably named head-tail connector gp8 (see Figure 1.2). Attached to this is the gatekeeper protein gp11, into which the gp17 tail fibre trimers dock. The central tail tube gp12 is sighted just below, providing part of the conduit for DNA release.

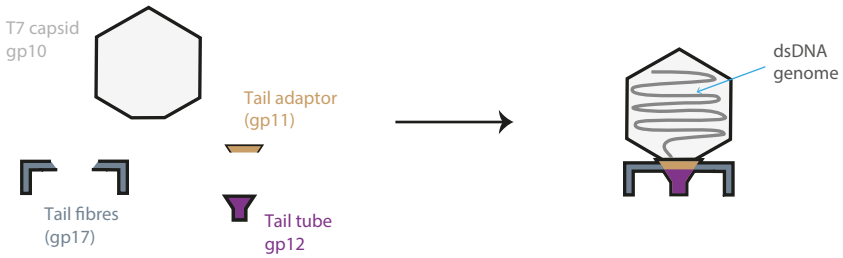


Figure 1.2: A 2D depiction of the pertinent, and externally visible, structural genes of bacteriophage T7. Note in particular that each T7 has six tail fibres, as opposed to the two shown here. Moreover, the T7 capsid is composed of 415 separate proteins.

1.3.1 T7 infection

Prior to attachment to *E. coli*, T7 phage diffuse randomly in the extracellular space. The tips of the tail fibres extend out and away from the capsid in a stochastic manner. About 50% of the population will extend a single tail fibre (gp17 trimer); very few will have either all tail fibres fully extended, or all of them left bound to the capsid [50]. After colliding with the host cell, Molineux speculates that T7 moves from a 3D to a quasi-2D diffusion process, reversibly binding along the cell surface until a suitable site for infection is found. How T7 irreversibly binds to the host prior to DNA ejection is not fully understood [58] though we do know that lipopolysaccharide (LPS) is required for DNA ejection *in vitro* [51]. Succeeding a stable attachment of T7 to the LPS, an internal protein core consisting of gp14, 15 and 16 is released. Gp14 spans the outer membrane; gp15 and 16 span the periplasm and inner membrane, creating a channel for DNA entry [59–61]. Expression of T7 genes are broadly arranged into 3 chronological classes. Class I genes enter the host first, with *E. coli* RNA polymerase (RNAP) pulling in the T7 genome through rapid transcription of the A1, A2 and A3 promoters [62] (see

Figure 1.3).[4] It takes roughly 10 minutes for the entirety of the T7 genome to be pulled into the cytoplasm [62] with each class of genes entering the host sequentially. All class I genes are transcribed by the host machinery, including the synthesis of T7 RNAP (gp1) mRNA, which proceeds to transcribe the remainder of the genome (the class II and III genes). Gp2 inhibits the host RNAP, which otherwise prevents the formation of productive progeny through unarrested transcription into the middle and late regions of the T7 genome [64]. It has been posited that this leads to pausing of the T7 RNAP (when polymerisation is interrupted by a slower moving host RNAP) resulting in double-stranded breaks to the DNA [64–67]. Class II genes are predominately involved in T7 DNA replication and degradation of the host genome. *E. coli* encodes thioredoxin, TrxA, which participates in numerous redox reactions [68]. T7 hijacks TrxA as processivity factor for DNA replication, binding to gp5, the T7 DNA polymerase, in the thioredoxin binding domain. This complex then binds to hexameric gp4 (see Figure 1.4) to unwind the dsDNA genome, and synthesise daughter strands. gp2.5 protects ssDNA during replication [69] and improves the rate of lagging strand DNA synthesis [70]. The class III genes are responsible for progeny formation and host lysis. To initiate the latter, holins (gp17.5) create pores in the inner membrane of *E. coli* which allow the passage of endolysins (gp3.5) to degrade the peptidoglycan. Permeabilisation of the inner membrane also enables complexed spanins (gp18.5 and gp18.7) to disrupt the outer membrane, resulting in lysis of the host [71–74].

[4] T7 RNAP, expressed by the host, or by some other means, can also enable DNA entry in absence of *E. coli* RNAP [63].

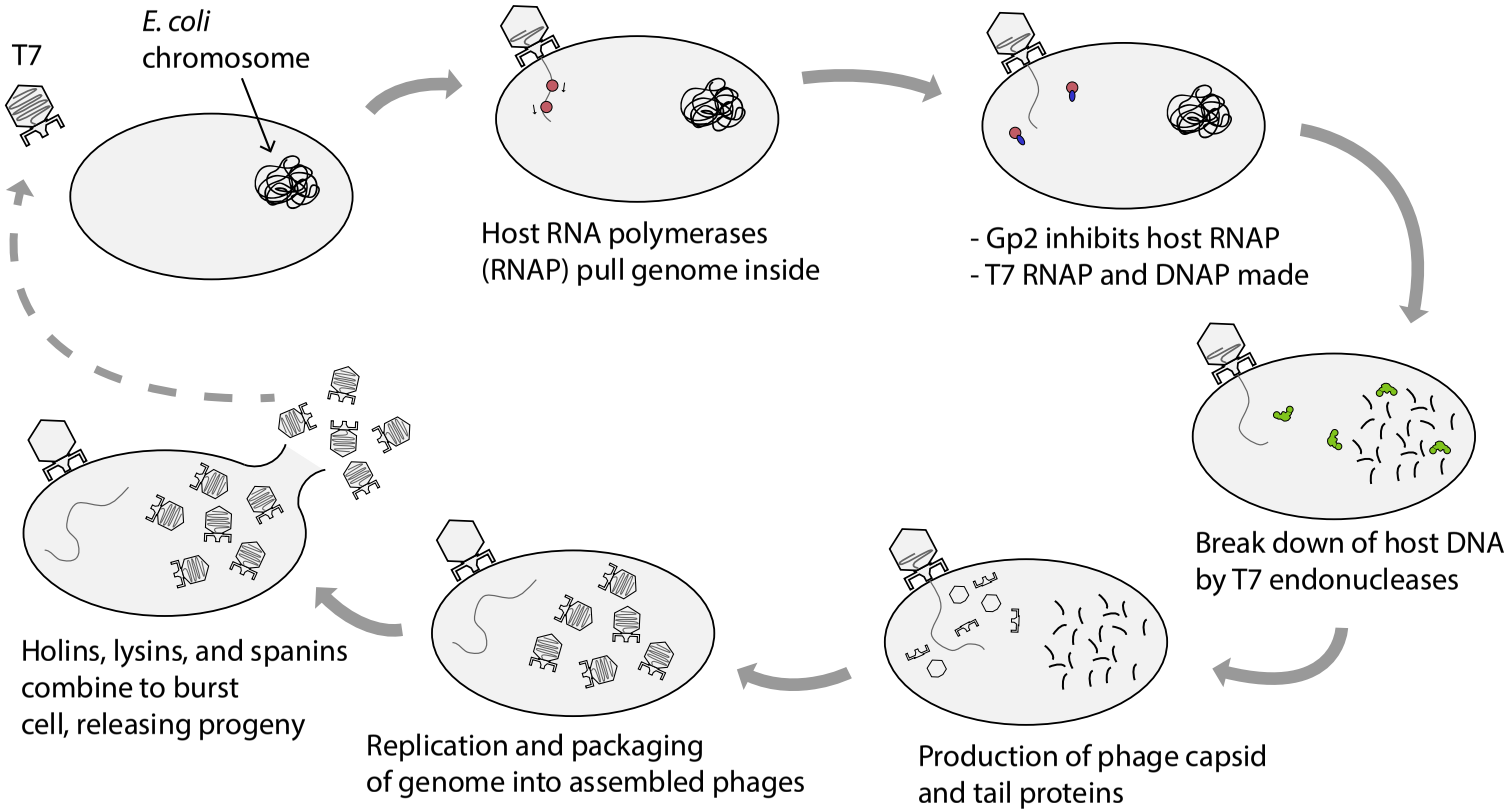


Figure 1.3: A simplified overview of the T7 phage life cycle. T7 phage adsorbs to lipopolysaccharide (LPS) exposed strains of *E.coli* bringing about a concomitant ejection of proteins which puncture a hole in the host cell membrane, and allow the passage of T7 dsDNA into the cytoplasm. The T7 genome encodes various proteins which inhibit host machinery and construct new phages for subsequent generations.

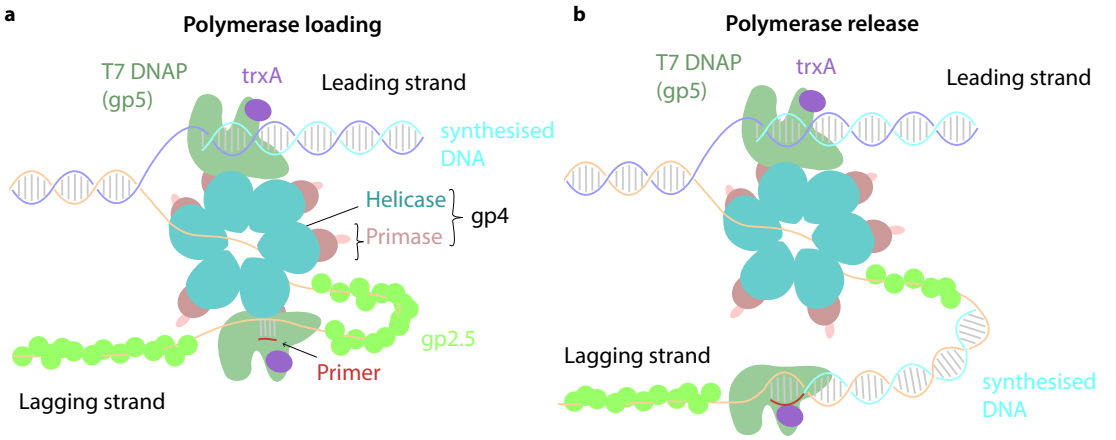


Figure 1.4: Leading and lagging strand synthesis of T7 DNA. a) Four proteins, T7 DNA polymerase (T7 DNAP, or gp5) the processivity factor thioredoixin (trxA), gp4, a hexameric protein with combined helicase and primase activity, and ssDNA binding protein gp2.5, work in tandem to replicate T7 dsDNA. Leading strand synthesis occurs whilst the lagging strand is parsed for RNA primer sites. Following RNA primer synthesis by the primase, complexed gp5-trxA is loaded onto the lagging strand at that site. b) Once loaded, concomitant release of the polymerase from the helicase occurs, and leading to synthesis of Okazaki fragments on the lagging strand. In rare instances where few polymerases are available, the polymerase will remain bound to the helicase while synthesising the Okazaki fragments. Figure and text adapted from [75–78]

1.4 Non-rational design: directed evolution

The drawback of using T7 phage is that it is not known which alterations to the genome might be required in order to change binding affinity from LPS to OMPs, and that the tail genes are too large to make practical libraries from (gp11, 12 and 17 totalling ≈ 4.5 kb in length). In such instances, directed evolution offers a solution, providing we can suitably design an environment which is advantageous for the desired phage phenotype. One such example can be in seen in the work of Lenski *et al.* [37]. The authors found that *E. coli B* grown in glucose-limited conditions with λ $cI26$ would generate mutants that down-regulated the expression of the phage receptor, LamB. λ $cI26$ would still persist in these conditions, however, seemingly because the mutants would spontaneously produce the receptor. This environment was hypothesised to favour the emergence of λ $cI26$ phages that could utilise a receptor other than LamB for infection, and serial passaging of phage and cells in those conditions demonstrated that λ $cI26$ could be adapted to OmpF, whilst retaining the ability to infect via LamB. T7 phage has been subjected to numerous directed evolution experiments [48, 79–82]. Of particular interest are experiments which impacted its host range. Heineman *et al.* passaged T7 in mixtures of permissive *E. coli* C strain with either non-permissive *E. coli* B or K12 so that T7 would adapt to avoid *E. coli* B or K12 respectively [83]. Permissiveness was determined by the presence or absence of the host factor *trxA* - necessary for the production of T7 progeny. In a similar set of experiments, Benmayor *et al.* learnt that when varying ratios of susceptible and resistant bacteria in co-cultures with a lytic *Psuedomonas* phage, that larger proportions of susceptible bacteria could be associated with increased opportunity for mutation in the phage population, and that higher proportions of resistant mutants increased selection pressure to infect

resistant hosts [84].

The procedure of PACE (Phage-Assisted Continuous Evolution) though not strictly phage evolution, invokes the concept of *in trans* essential gene complementation to direct fitness increases in a linked gene circuit element [85, 86]. We will do the same with T7, leaning on that work, and the work of Heineman *et al.* [83] by removing the essential gene *g5* from T7, and complementing it in strains that we want T7 to generate increased fitness for.

1.5 Characterising LPS-independent T7 infection

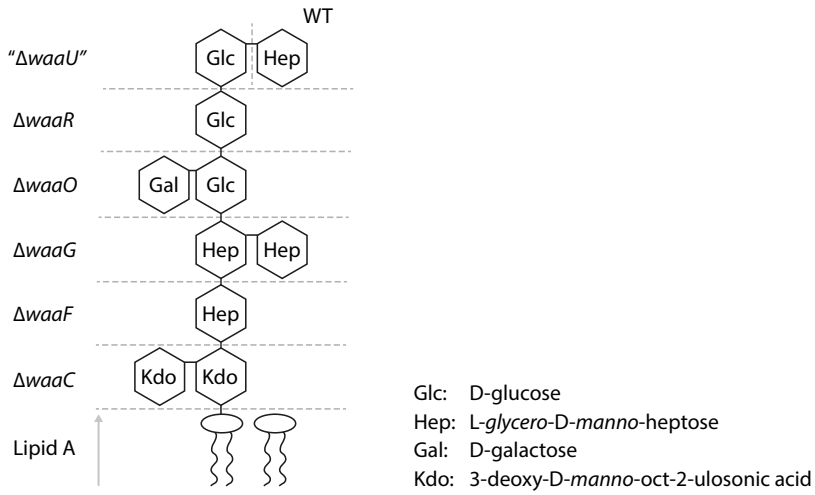


Figure 1.5: Generalised LPS structure of *E. coli* BW25113 for various LPS mutants. Note that *waaU* is an essential gene, and so is not represented in the Keio collection. Adapted from [87–89].

Qimron *et al.* demonstrated that T7 infecting BW25113 $\Delta waaC$, which possesses a truncated LPS phenotype (see Figure 1.5) produces at least 100-fold fewer plaques on plating than when challenged with BW25113. In particular, the authors

postulate that T7 binds to either the terminal heptose sugar of the BW25113 LPS, or the first or second glucose [87]. We will exploit this phenomena to indirectly determine whether an engineered T7 can infect *E. coli* without interacting with the LPS.

1.6 Aims of this work

The overarching objective of this work is to establish a process for changing the receptor of T7 phage from the lipopolysaccharide of *E. coli* to a desired outer membrane protein. In chapter 2 the first aim is to establish a reliable method for engineering T7 bacteriophage. Following this, I clone a novel tail fibre chimera between T7 and a T7-like phage that is known to interact with *Yersinia pestis* OMPs during infection, with a view to engineering a T7 phage that requires outer membrane proteins for infection of *E. coli*. In chapter 3, I prepare the constructs required for directed evolution of T7 phage receptor tropism, and in chapter 4, I design and perform those experiments, using mathematical modelling in parallel to simulate different evolution strategies *in silco*.

2 Engineering a T7 Yep-phi chimeric tail fibre

2.1 Introduction

2.1.1 Phage tail fibre chimeras

Phage tail fibres typically show variation in chunks between conserved sequences [90–92], indicating that recombination between tail proteins is common. Subsequently, nature hints at a clear path for the rational design of novel tail fibre fusions. One such example is Tétart *et al.* where the authors exchanged parts of the hypervariable regions of T4 tail fibres with that of the SV76.3 and Mi phages, permitting robust growth on *Yersinia pseudotuberculosis* [93]. More recently, Ando et al. [15] demonstrated the modularity of T7, and the T7-like phage T3, by generating a variety of different phage fusions through the exchange of regions of the tail fibre, gp17. In the region of T7 gp17 known to dock with a dodecamer of gp11, the N-terminal 149 aa [94] T7 and T3 tail fibres share 97% amino acid homology. T7 and T3 phage generated with each other's tail fibres effectively exchanged host ranges, and this was similarly true when only the C-terminal gp17 regions (the last 410 aa for T3, and the last 405 aa for T7) were swapped [15].

2.1.2 Yep-phi phage

Yep-phi is a T7-like *Podivridae* phage isolated by Zhao *et al.* [24] which infects *Yersinia pestis* while using the LPS as a receptor. The authors determined that linear peptides of Yep-phi tail fibres bound to the outer membrane proteins Ail and OmpF. Amino acids 518N, 519N, 522C and 523S were essential for the interaction with Ail and residues 518-519 and 523 were similarly essential for OmpF. Ail, attachment invasion locus (also known as OmpX) is a beta-barrel protein with 4

extracellular loops (see Figure 2.1) which mediates adherence and internalisation of Y. *pestis* into monocytic (THP-1) and epithelial (HEp-2) cells [95–97]. OmpF is a beta-barrelled porin which is ≈ 60% homologous to *E. coli* OmpF. Though Yep-phi interacts with Ail and OmpF, deletion mutants indicate that the proteins are not required for Yep-phi plaque formation (an ≈ 3-fold reduction in adsorption rate is seen in the absence of Ail [24]).

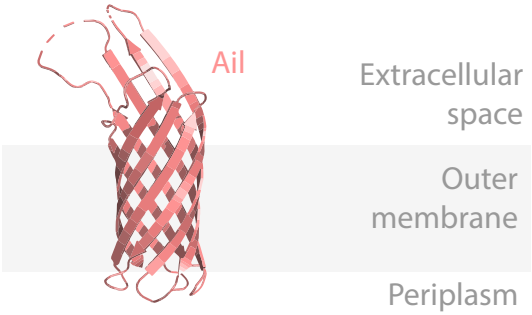


Figure 2.1: Ail from *Yersinia pestis*. Visualised in PyMOL (PDB: 3QRA [98]).

2.1.3 Motivation for a T7 tail fibre fusion

As outlined in chapter 1, we are aiming to construct a T7 phage whose binding is mediated by outer membrane proteins (OMPs). Subsequently, we considered whether it would be viable to graft this capability onto T7, using a receptor binding domain from another phage that had been shown to bind to OMPs during infection. For such a fusion to be successful, there were two properties that it would need to possess. First, the tail fibre chimera would have to correctly trimerise in vivo and dock with gp11 [91]. Second, the tail fibre would need to able to signal when stable adsorption to the host has occurred, and induce the delivery of the T7 genome [51]. It was not known, for instance, if changing the length of gp17 would effect how this signal was transduced. In the preceding sub-section, we saw that Yep-phi, a

T7-like phage, does interact with OMPs during infection. The T7 and Yep-phi tail fibres share a 38% protein sequence identity, with particularly high homology in the N-terminal domain that docks with the tail adaptor protein in both phage (gp11 and gp32 respectively). It therefore seemed possible that a fusion with Yep-phi gp17 could result in a tail fibre with the required properties, in addition to utilising OMPs for infection.

2.1.4 T7 engineering strategies

The first alterations made to the T7 phage genome came about through random mutagenesis and selection [99, 100]. In the time since, three general strategies for have emerged for precisely controlling the insertion (or deletion) of a given sequence into (or from) T7: *in vivo* homologous recombination in *E. coli* followed by marker [101, 102] or non-marker-based selection [103, 104]; and recombineering in yeast (*Saccharomyces cerevisiae*) followed by rescue in *E. coli* 10G cells [15]. For marker-based selection of recombinant T7, the essentiality of the host factors *trxA* [105, 106] and *cmk* [87] to T7 production is exploited. As mentioned briefly in section 1.3.1, TrxA complexes with gp4, gp2.5, and gp5 during T7 replication. In the absence of TrxA, there is a 25-fold reduction *in vitro* T7 DNA synthesis [105] and *trxA*-deficient *E. coli* are not permissive to T7 (no plaques are observed) [105–107]. Cmk, or cytidine monophosphate kinase, transfers a phosphate group from ATP to cytidine monophosphate, CMP, and deoxy-CMP [108]. Qimron *et al.* performed a genomewide screen of *E. coli* BW25113 to find genes which inhibited T7 growth [87]. In that screen *cmk* emerged as another potential marker for T7 recombineering. As a proof of principle, the authors cloned pGp5-cmk, a vector with *cmk* flanked by 400 bp of upstream homology and 200 bp of downstream

homology to *g5* on the T7 genome. Following *in vivo* recombination with T7, the resulting lysate was plated on *cmk* deficient cells and successfully screened for recombinant plaques. For non-marker based selection, a similar procedure was enacted, except with an *in vivo* CRISPR Type I-E system obviating phage genomes that did not pick up the desired gene sequence [103]. Yeast-based T7 engineering mimicks Gibson Assembly in that the phage DNA to be assembled in a replicative yeast plasmid (yeast artificial chromosome) is transformed into *S. cerevisiae* with overlapping homology regions of at least 30 bp [15]. The yeast gap repair machinery then assembles the transformed DNA, and the resulting plasmid is extracted and transformed into *E. coli* 10G cells. Plasmids which contain functional phage genomes will generate plaques in the bacterial lawn, enabling the selection of viable recombinants. Each of these methods has their limitations and strengths. Yeast-based engineering requires the culturing and maintenance of yeast cells, though, as with CRISPR-based selection, it does not scar the recombined genome. Selection using *E. coli* host factors does however leave an undesired gene product (which, with some scarring, can potentially be removed using FRT flippase system [109]) but the degree of selection is orders of magnitude more effective than with Type I-E CRISPR (which also suffers from escape mutants that avoid the CRISPR gRNA).

2.1.5 Rescue of non-functional engineered genes

So long as *in silico* protein folding remains unsolved [110, 111], the rational engineering of fusion proteins will continue to be a challenging affair. To ensure that the desired DNA sequences have indeed recombined into the phage, we can express a functional copy of the WT gene in the host, to allow phages which

have picked up the desired gene sequence to continue to be productive, even if the recombined sequence encodes a non-functional phenotype (see Figure 2.2). Specifically, we will express gp17 of T7 in *E. coli* so that should the recombinant tail fibre have misfolded, or not be capable of binding to the outer membrane structures of *E. coli*, then the recombinant sequence will be retained in the genome. This sequence could then be further engineered, through directed evolution, or the recombinant phage could be presented with a host which did possess the necessary outer membrane structures for productive infection.

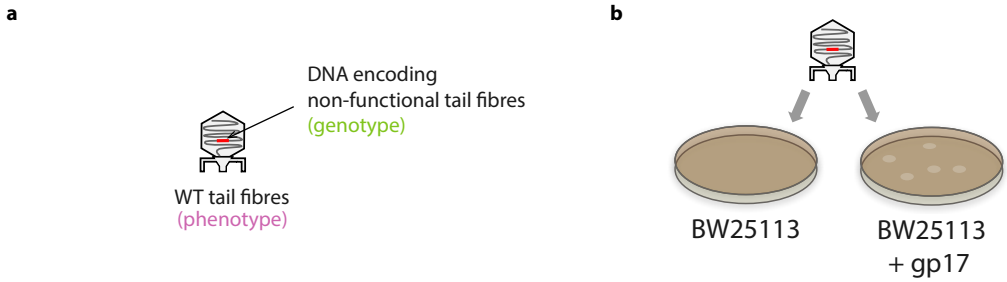


Figure 2.2: a) T7 phage which phenotypically has WT tail fibres, but which genotypically has DNA which only encodes for non-functional tail fibres. b) Rescue of this phage mutant can be achieved complementing the WT tail fibre (gp17) in the host. This allows the initial phage infection to produce phage progeny capable of infecting proximal cells, resulting in observable zones of lysis in bacterial lawns (called phage plaques).

2.1.6 Aims for this chapter

In this chapter, I aim to generate a chimeric T7 phage which interacts with outer membrane proteins as part of its infection cycle. The tail fibre of T7-like virus Yep-phi is known to interact with OmpF and Ail of *Yersinia pestis* and so my intention is to take the interacting domain and swap it into the homologous domain of T7, in an attempt to impart its properties on T7. To do so, I will need to 1) establish a reliable method for engineering T7 phage; 2) determine whether the new

chimera is capable of infecting *E. coli*; 3) determine whether infection is dependent on the interaction with heterologously expressed proteins from *Y. pestis*.

2.2 Results

At the point of construction, there was no evidence in the scientific literature of T7 bacteriophage having been engineered as late into the genome as *g17* (that DNA being in the last 15% ejected into the host). Therefore, before constructing a tail fibre chimera of T7 and Yep-phi phage, it seemed sensible to demonstrate that the engineering methodology (selection using the *trxA* marker) could work in our hands. In addition, we wanted to reduce the likelihood of this gene recombining with an *in trans* complementation of T7 gp17 - present to rescue phages with non-functional tail fibres.

2.2.1 Design and cloning of a T7 *g17* sequence with reduced homology to the WT sequence

Seeing as any alteration to the *g17* sequence of T7 could result in a non-functional tail fibre phenotype (and subsequently, an unproductive phage) we thought it could be useful to provide a copy of *g17 in trans* to rescue phages with DNA genomes encoding non-functional tail fibres. Given that a WT copy of the T7 tail fibre would provide plenty homology for recombination events with the phage genome, we[1] sought to reduce the likelihood of a phage genome recombining with an *in trans* complementation of *g17* by designing a version of *g17* (herby referred to as *g17**) where stretches of homology between and *g17** and the WT sequence were limited to at most 8 nucleotides (see Appendix Figure A.1). The DNA for this sequence was synthesised as a gBlock (GB002, see Methods 5.3.2) between a

[1]Preliminary sequence design was by Marc Guüell Vila-Ferran and Antonia Sagona. I later re-designed and re-synthesised the gene to correct 4 aberrant mutations in the protein sequence

synthetic T7 promoter (SBa_000446) and terminator (SBa_000587 [112]) with RBS BBa_0064 immediately preceding the gene (see Figure 2.3a). SBa_000446 has at most 16 bp base pairs of homology to promoters found in T7 (increasing to 19 bp if one base pair mismatch is allowed) and 15 bp of homology to the consensus promoter sequence. SBa_000587 has a 12 bp homology with the T7 terminator Tϕ. The SEVA vector pSEVA551 [113] was chosen as the backbone and amplified in 2 overlapping sections using the primer pairs oPM013/14 and oPM015/16 (lanes 1 and 2 in Figure 2.3b). Separately GB002 was amplified using oPM017/18 (visualised in lane 3) before cloning of all 3 fragments by Gibson Assembly (section 5.3.6). Colony PCR of the 12 transformants gave rise to only 1 amplicon (in lane 3) that was of the expected size (Figure 2.3c). Sanger sequencing of the extracted plasmid (see section 5.3.8) indicated that there was a single base pair mutation in *g17**, T566G, generating a S189I amino acid substitution (Figure 2.3d).

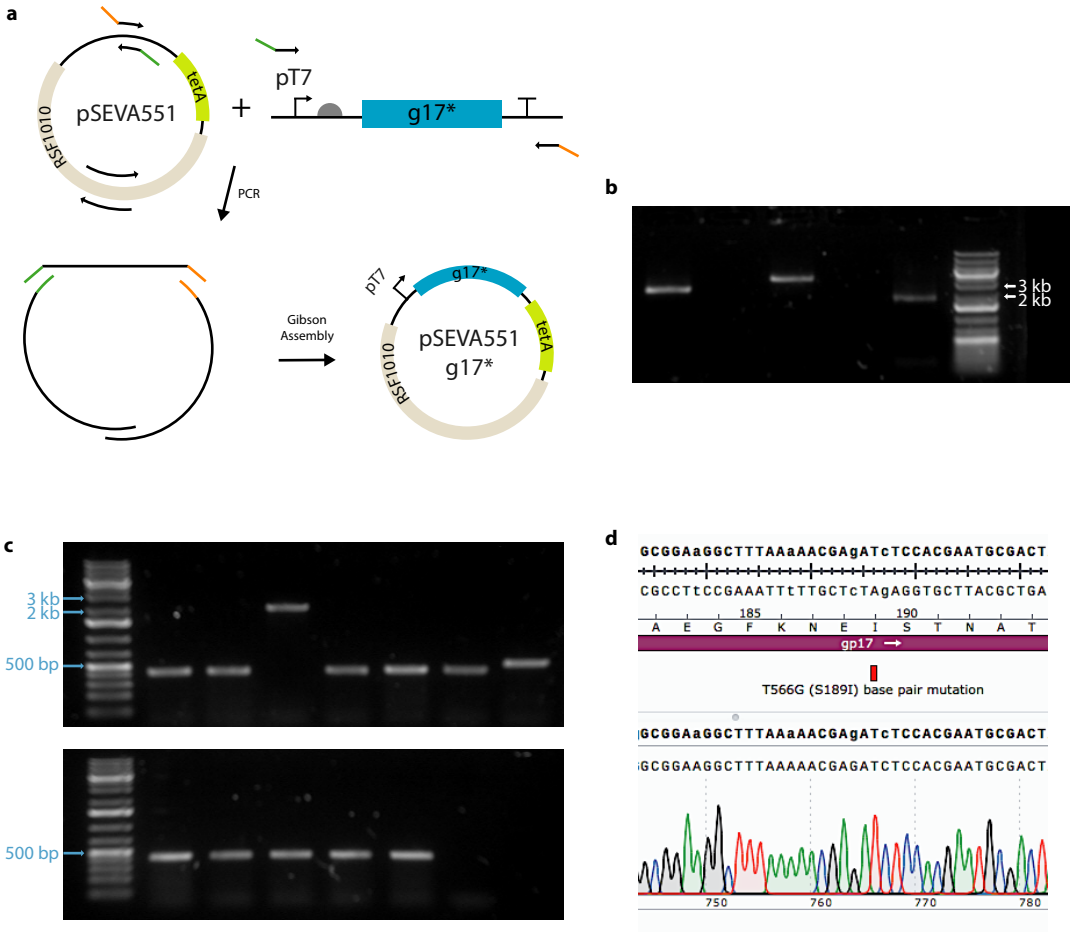


Figure 2.3: Construction of a vector to complement g17-deficient T7 *in trans*. a) Vector backbone was cloned from pSEVA551 [tetR, RSF1010 origin of replication]. The insert was synthetic DNA consisting of a T7 promoter, an RBS (BBa_00064 [114]), codon-altered g17 DNA, and a synthetic T7 terminator (SBa_000587 [112]). Primers were designed to produce 3 overlapping fragments. For simplicity, single lines represent dsDNA (primers excluded). b) Gel of resulting PCR fragments. c) PCR amplicons from sampled transformants. d) Relevant frame of sequencing data for transformant with correct PCR amplicon size. Sequence contains a T566G base-pair mismatch (leading to a S189I amino acid substitution).

2.2.2 Making T7 Δ*g17*::*trxA*

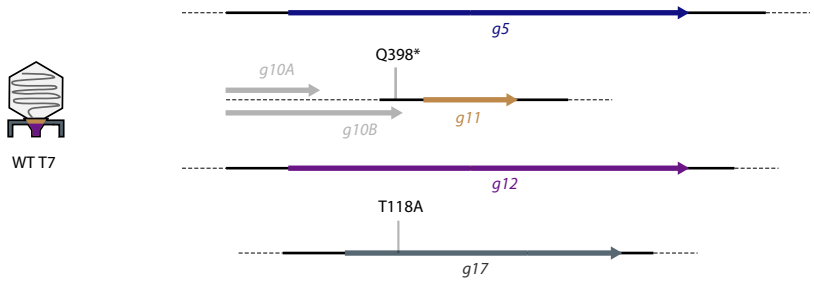


Figure 2.4: Sanger sequencing of *g5*, *g11*, *g12* and *g17* gene regions for WT T7 (see section 5.3.8). The sequence differed from the canonical NC_001604 sequence in 2 places. The nonsense mutation Q398* replaces the last amino acid of the non-essential, minor capsid protein gp10B with a stop codon.

Genes 5, 11, 12 and 17 of T7 phage were PCR amplified and Sanger sequenced (Figure 2.4). Two non-canonical mutations were observed: one in the non-essential minor capsid protein gp10B, and one in the N-terminal region of the T7 tail fibre.

The Gibson assembly method was used to clone pSB3T5-HRg17-trxA (see Figure 2.5). The amplicons used in this assembly came from the backbone of pSB3T5-HRg5-cmk-trxA using primers oPM023/24, and the insert was generated by amplifying GB004 with primers oPM025/26 (see section 5.3.2). pSB3T5-HRg17-trxA was then electroporated into BW25113 Δ*trxA* competent cells along with pSB6A1-gp17 (see Figure 2.6a). pSB6A1-gp17 possesses an ampicillin resistance cassette and a pUC origin of replication, making the plasmid compatible with the tetracycline resistance and p15A origin of pSB3T5-HRgp17-trxA. BW25113 Δ*trxA* pSB3T5-HRg17-trxA + pSB6A1-gp17 was refreshed to log-phase and challenged with WT T7. The resulting phage population was a mixture of WT phage and recombinant phage that had traded *g17* for *trxA*. Plating this against BW25113 Δ*trxA* pSB6A1-gp17 selected

against WT T7 that lacked the *trxA* gene required to replicate, and complemented gp17 to those phage that had lost *g17* in obtaining *trxA* (Figure 2.6b). 2 plaques were picked from the plate shown in Figure 2.6c, and were screened for recombinant and WT phage. As depicted in Figure 2.6d, primers flanking *g17* (oPM046/47) and primers that bind upstream, and the inside of *g17* (gp17fwd/AG061rev) were used to amplify both the plaques and a WT phage control (Figure 2.6e). Bands in the WT lanes indicate that the primers were functional, and the gp17fwd/AG061rev primer pair did not produce amplicons for the recombinant plaques.

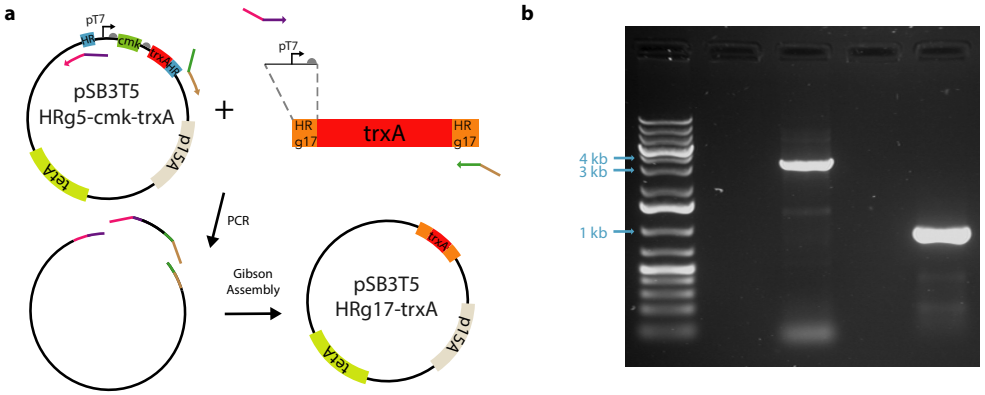


Figure 2.5: Construction of HR plasmid for deleting g17 from T7 phage. a) A vector, pSB3T5 HRg5-cmk-trxA, previously used to remove g5 from T7 phage through the insertion of *cmk* and *trxA* is re-designed to do the same for g17 with *trxA*. The new insert flanks the *trxA* gene with 90 bp DNA sequences homologous to areas immediately upstream and downstream of g17. The location of the upstream homology arm necessitates the inclusion of the corresponding T7 promoter and RBS for g17. The colours on each primer are to indicate where the necessary overlaps occur for Gibson assembly. b) A gel of the PCR fragments used for the Gibson assembly (left - backbone, right - insert).

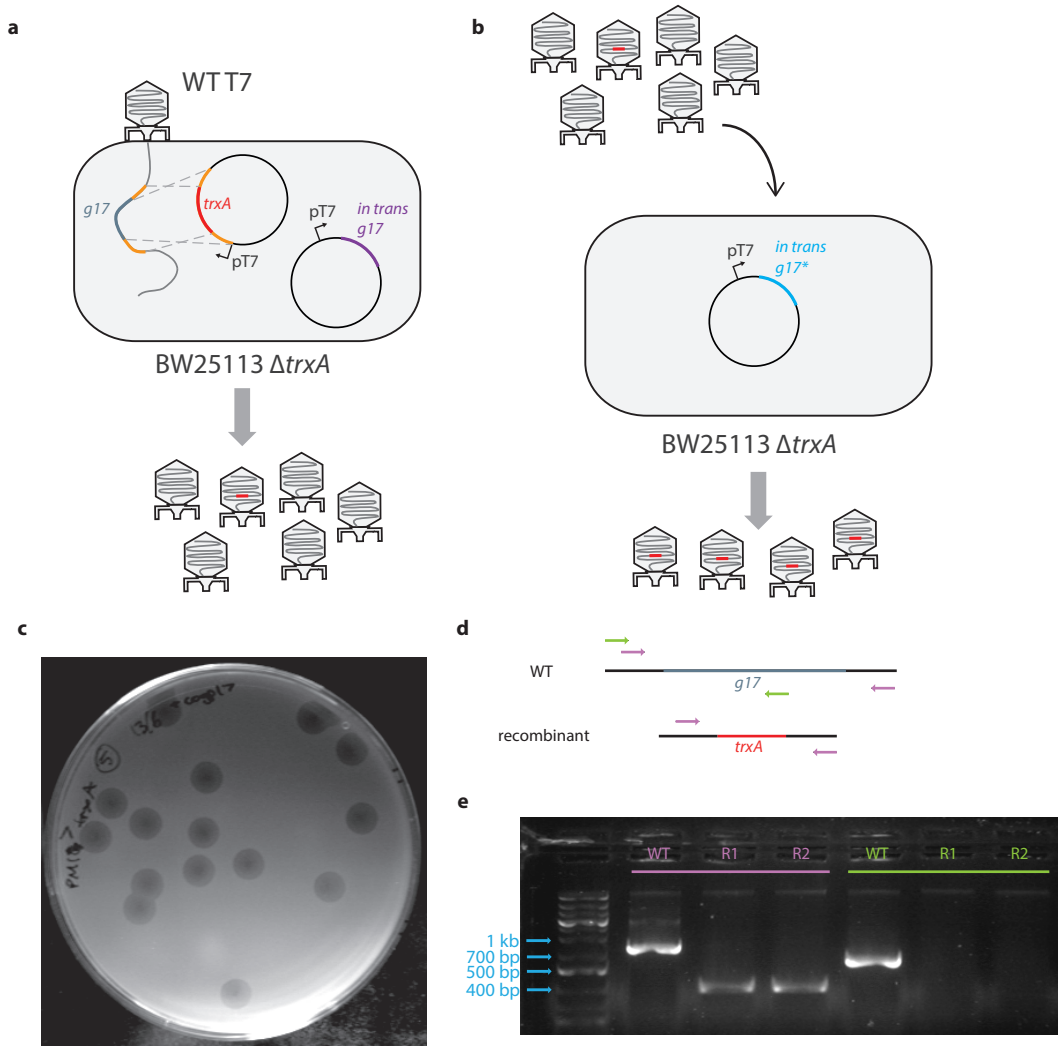


Figure 2.6: Engineering a *g17* deficient T7. a) WT T7 phage infect Δ*trxA* cells containing a plasmid with homology arms designed to replace *g17* with *trxA* on the T7 genome. The T7 promoter upstream of the plasmid-borne *trxA* ensures some expression of TrxA, and hence enables T7 replication. Additionally, gp17 is complemented *in trans* to functionalise *g17*-deficient phage. b) Resulting phage, a mix of WT and phage that picked up the *trxA* insertion (recombinant), are plated on Δ*trxA* complemented with *in trans* gp17. The absence of TrxA ensures that only recombinant phage can replicate (superinfection effects excluded). c) The relevant plaque assay plate from sub-panel b. d) A diagram indicating where the primers used to characterise the phage plaques from sub-panel c bind to on the recombinant and WT genomes. e) Gel of recombinant plaques (R1 and R2) and a WT control following the described PCRs (the colours specify which primer pair was used).

2.2.3 Design and construction of a functional T7/Yep-Phi tail fibre chimera

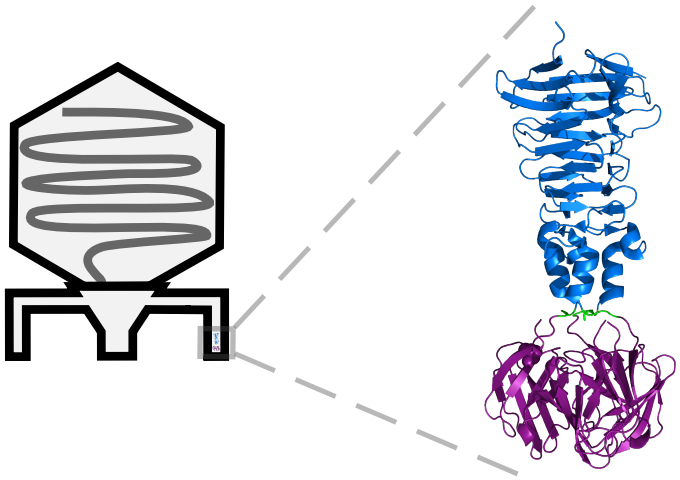


Figure 2.7: Close-up of C-terminal end of a T7 tail fibre (visualised in PyMol [115]). The pyramid and tip domains are coloured in blue and purple respectively. Sandwiched between those domains, in green, is a small flexible loop (Val464-Lys466, i.e. VAK) which was left in tact when replacing the tip domain of the T7 tail fibre, with one from the Yep-phi phage. The crystal structure (PDB ID: 4A0U) is from Garcia-Doval & van Raaij [91].

Having succeeded in inserting the *trxA* selection marker into the tail fibre region of T7, the next objective was to design and engineer a functional fusion of the tail fibre regions of T7 and Yep-phi. Garcia-Doval & van Raaij [91] suggested that the short linker between the pyramid and tip domain (Val464-Lys466) is somewhat flexible (see Figure 2.7). Following this observation, and based on homology between the T7 and Yep-phi tail fibres (Figure 2.8) a fusion between the two tail fibres was designed (see Figure 2.9a) which agreed with the design made by Antonia Sagona independently. The design takes the first 466 amino acids of the T7 tail fibre and pairs them with amino acids 485-569 from Yep-phi gp17. The homology regions upstream and downstream of the sequence to be replaced (aa 467-553 of T7 gp17) were retained and that DNA was synthesised

```
T7_17  1    MANVIKTVLTYQLDGSNRDFNIPFEYLARKFVVVTLIGVDRKVLTINTDYRFATRTTISL  60
            MAN I TV TY L+GS  +F I FEYLARKFV+VTLIG DRK L +N DYRF  +T I+
YP_17  1    MANKISTVRTYPLNGS-VNFTITFEYLARKFVLVTLIGKDRKELVLNQDYRFTAKTQITT  59

T7_17  61   TKAWGPADGYTTIELRRVTSTTDRLVDFTDGSILRAYDLNVAQIQTMHVAEEARDLTTDT  120
             +AW  ADGY  IE+RR TS TDRLVDF DGSILRAYDLN++QIQT+HVAEEARDLT DT
YP_17  60   ARAWTAADGYEMIEIRRFTSATDRLVDFADGSILRAYDLNISQIQTIHVAEEARDLTADT  119

T7_17  121  IGVNNDGHLDARGRRIVNLANAVDDRDAVPFGQLKTMNQNSW----QARNEALQFRNEAE  176
            IGVNNDG+LDARGR+IVNLA A  D DAVP  Q+     + W    +A  +A +   EA
YP_17  120  IGVNNDGNLDARGRKIVNLAFATSDYDAVPLKQITDRESSVWNAVTKASEQADRSNKEAN  179

T7_17  177  TFRNQAEGFKNE----------SSTNATNTKQWRD-------ETKGFRDEAKRFKNTAGQ  219
              R++A+  K E          S+T A   K+  D        +KG+ D        A
YP_17  180  RSRDEADRAKREADRSTQQAGVSATQAVEAKKQADRSNSEANRSKGYADSMTASVEAAKG  239

T7_17  220  YATSAGNSASAAHQSEVN----------------------AENSATASANSAHLAEQQAD  257
            +A SA   A+ + ++E N                       A+N A  S+  A+ A+ +AD
YP_17  240  HAESASKEANRS-RAEANRAADEVTKAAAEVSKAAAHVASAKNQADRSSTEANRAKSEAD  298

T7_17  258  RAEREADKLENYNGLAGAIDKVDGTNVYWKGNIHANGRLYMTTNGFDCGQYQQFFGGVTN  317
            RA+ EADKL N N  AG ++KV+G    +K  I             G +    G  T
YP_17  299  RAKTEADKLGNINEFAGTLEKVEGVTPTFKSGIK-----------LRSGDFYAESGSFTK  347

T7_17  318  RYSVMEWGDENGWLMYVQRREWTTAIGGNIQLVVNGQIITQGGAMTGQLK--LQNGHVLQ  375
            +    +W             +WT A+  ++Q  + GQ         G +   L+    L
YP_17  348  KTQGGDWS------------QWTGALTPDVQHDIEGQ---------GNIAYLLERATTLS  386

T7_17  376  LESASDKAHYILSKDGNRNNWYIGRGSDNNNDCTFHSYVHGTTLTLKQDYAVVNKHFHVG  435
             E    K  Y    +G   N++I              Y+      L+ D A+++     G
YP_17  387  KEPLYAKCLY---NNGVGENFFI------------RQYLRDAHFDLRADGALISSS---G  428

T7_17  436  QAVVATDGNIQGTKWGG----------------------KWLDAYLRDS-----FVAKSKA  469
             ++ A DGN+   K+G                      K +D  L        +  KS
YP_17  429  WSIPA-DGNLYIKKYGSNLDTWVNKRLSAHAYSKGEVNKMVDGLLTTEQGDARYARKSSG  487

T7_17  470  WTQVWSGSAGGGVSVTVSQDLRFRNIWIKCANNSWNFFRTGPDGIYFIASDGGWLRFQIH  529
            WT+VW GSAGGGVSV++SQD+R+R IWI  ANN     + G D  YF+   GGWL+F I
YP_17  488  WTEVWQGSAGGGVSVSLSQDVRWRTIWI-LANNGMCSVQIGADATYFMVVMGGWLKFTIS  546

T7_17  530  SNGLGFKNIADSRSVPNAIMVENE  553
            +NG  F+N  D  +VP  I+V N
YP_17  547  NNGRTFRNDQDRNTVPEQILVRN-  570
```

Figure 2.8: NCBI blastp [116] protein alignment of T7 gp17 and Yep-phi gp17 (GenBank ID: CAA24435.1 and ADQ83192.1 respectively). In this alignment, the proteins share $\approx 38\%$ sequence identity. The orange dashed line indicates the location of the cut site, from which point the protein fusion utilises Yep-Phi amino acids, instead of ones from T7.

between XbaI and SpeI restriction sites (Figure 2.9b). This, in turn, was cloned into the biobrick plasmid pSB6A1 by restriction digest to make pSB6A1-HR-YP (Figure 2.9c). Using the strategy employed to create T7 Δgp17::*trxA*, this plasmid was further amended to include *trxA,* preceded by an RBS (Figure 2.10a) in order to facilitate the selection of the recombined phage. pSB6A1-HR-YP-trxA was made by Gibson assembly (section 5.3.6). Primers oPM058/59 were used to amplify pSB6A1-HR-YP (minus the downstream homology region), and primers oPM060/61 amplified the RBS-*trxA* insert from pML-HRtrxA (in addition to the missing downstream homology region). The bands generated by those PCRs can be seen in Figure 2.10c (labelled A1 and A2 respectively) and match the expected sizes.

Additionally, to determine whether the protein fusion was soluble, and therefore likely to be folding correctly, a vector containing amino acids 371-467 of T7 gp17 and an N-terminal hist-tag was amended to express amino acids 371-466 of T7 and 486-569 of Yep-phi (Figure 2.10b).[2] This vector (pET30a+T7-YP) was cloned by Gibson assembly (section 5.3.6). pET30a+gp17_371-467 was amplified with oPM054/55 and the insert (GB001) was amplified with the primer pair oPM056/57. The resulting bands are seen in Figure 2.10c, in lanes labelled B1 and B2 respectively. The unexpected band at 200 bp in lane B1 was presumed to be due to a mis-priming reaction, and subsequently ignored when gel extracting the band at ≈ 6 kb. Figure 2.10d is an SDS protein gel; protein expressed using the pET30a+T7-YP plasmid migrates roughly in line with the 25 kDa marker.

[2]Note that T7 aa 467 is a serine, as is Yep-phi aa 485, consequently, this cloning decision has led to a different codon sequence for that serine, compared to the pSB6A1-HR-YP-trxA plasmid

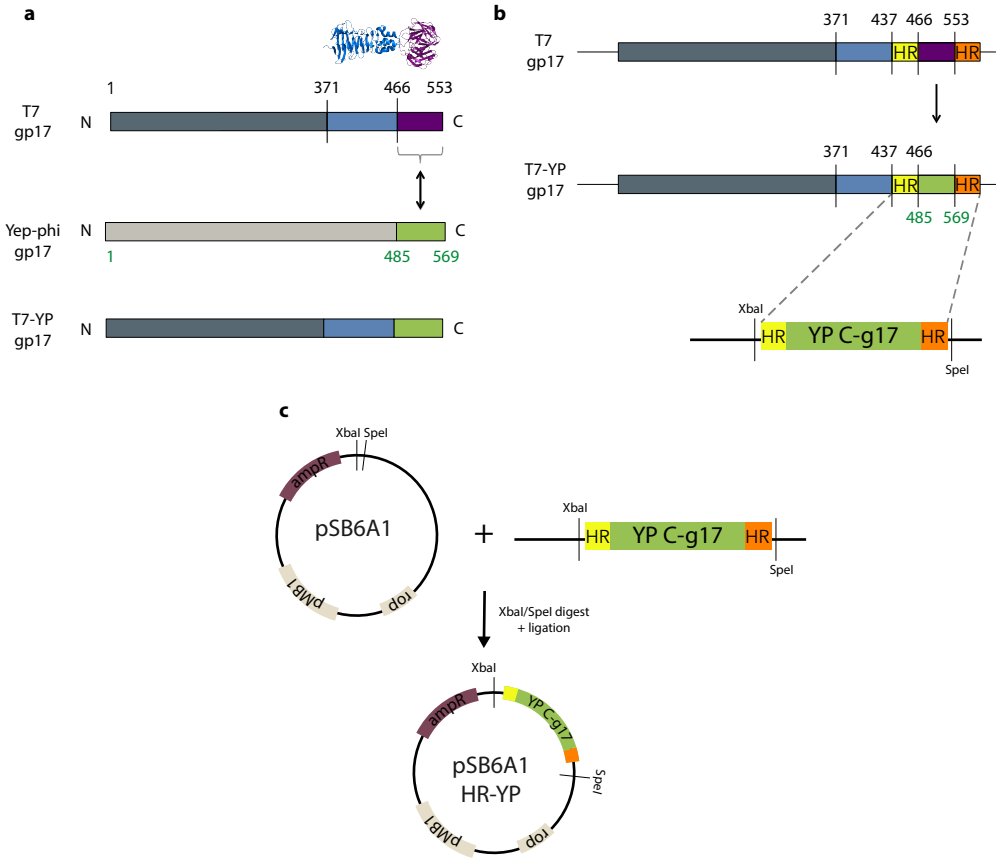


Figure 2.9: a) Exchange of the distal domain of the T7 tail fibre (in purple) with the receptor-binding domain of Yep-phi (in green). b) In order to facilitate the exchange of DNA, 90 bp homology arms are copied from the T7 genome downstream of g17, and upstream of aa 466. Amino acids 485-569 of the Yep-phi gp17 C-terminus are then flanked by these homology regions in a custom DNA synthesis order. c) Construction by restriction digest of a HR plasmid for *in vivo* engineering of T7. The homology arms flanking aa 485-569 of Yep-phi are cloned into pSB6A1 [114] by restriction digest.

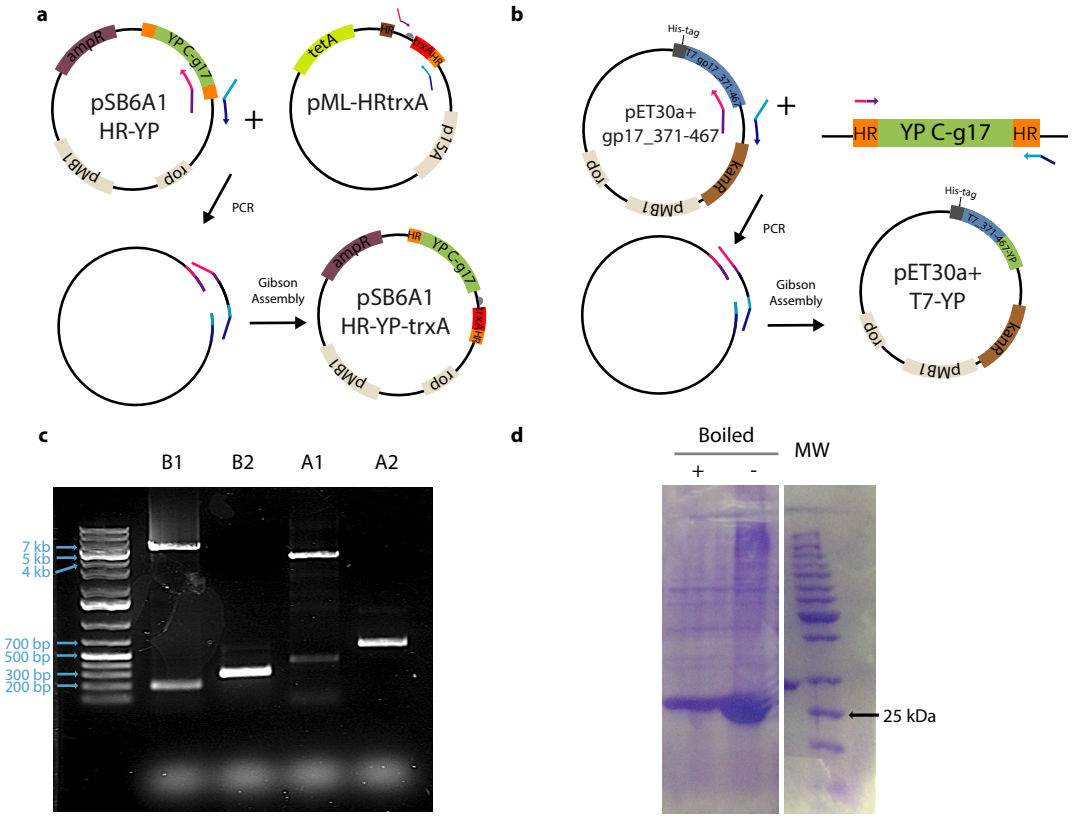


Figure 2.10: a) Amendment of pSB6A1-HR-YP to enable marker-based selection of any engineered phage. Primers are coloured to highlight the homology regions required for Gibson assembly. b) Cloning of the C-terminus of Yep-phi g17 (aa 486-569) into pET30a+gp17_371_467, which contains aa 371-467 of T7 g17. c) Gel of PCR fragments for Gibson assembly. A1 and B1 are the PCRs of the vector backbones from panel A and panel B respectively. A2 and B2 are the inserts. d) SDS gel of purified protein expressed from pET30a+T7-YP (expression, purification, and visualisation performed by Marta Sanz Gaitero, in collaboration with the Raaij Lab). Cut out lanes contained unrelated protein samples.

2.2.4 Engineering of a T7 phage with a T7/Yep-Phi tail fibre chimera

WT T7 phage was amplified in BW25113 Δ*trxA* cells harboring pSB6A1-HR-YP-trxA; plus or minus the vector pSEVA551-g17* (see Figure 2.11a). The resulting phage population formed plaques, irrespective of whether gp17 was provided *in trans* during the amplification step (see Figure 2.11b). 2 plaques from each phage population challenged on BW25113 Δ*trxA* plates (with or without a complementing gp17 plasmid) were PCR amplified and Sanger sequenced (see Figure A.3). Recombined phage genomes were observed in all 4 plaques. In addition, the phage population that had not received gp17 *in trans* when plated on BW25113 only (PM6, see section 5.2) outnumbered that which had been plated on BW25113 Δ*trxA* (3.6×10^{10} vs 3×10^{6} PFU/ml respectively). An isolated plaque from the BW25113 Δ*trxA* plate challenged with PM6 was then amplified in BW25113 Δ*trxA* cells (the amplified phage population was denoted PM12F - see section 5.2). Figure 2.12 describes the SNPs observed in PM12F following next-generation sequencing (see section 5.3.9). Protein structure homology-modelling (see Figure 2.13) highlights where the E545K mutation might sit in the chimeric tip domain.

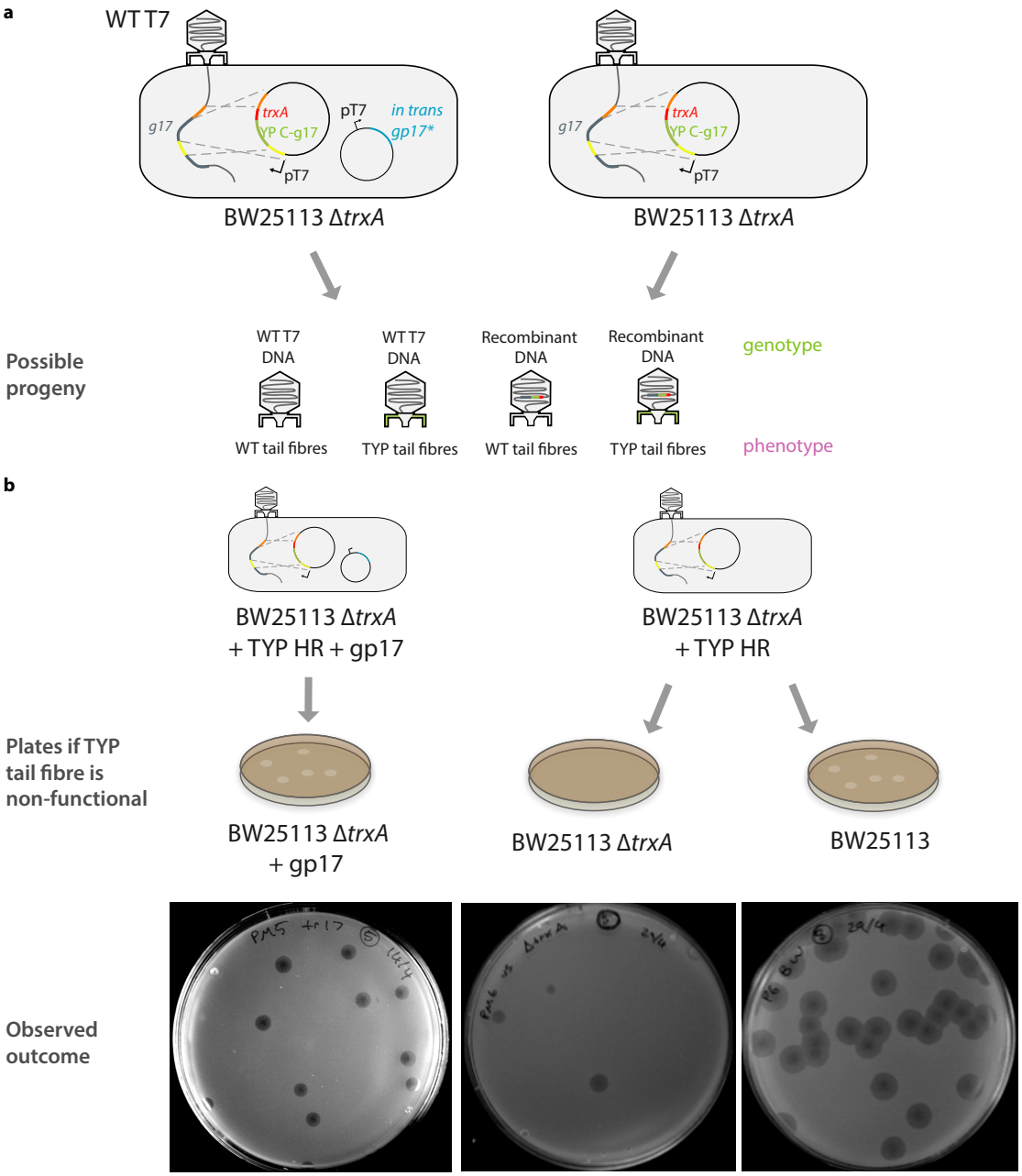


Figure 2.11: Construction of TYP phage: homologous recombination (HR) of WT T7 to exchange the C-terminus of gp17 with that of Yep-phi phage. a) WT T7 are infecting Δ*trxA* pSB6A1 HR-YP-trxA (denoted TYP HR) with or without a plasmid complementing *g17 in trans*. It is possible that the emerging phage populations is a mixture of genotypes that have picked up the Yep-phi *g17* C-terminus (and *trxA*), and a mixture of corresponding tail fibre phenotypes. b) Phage plated on cells which complement T7 gp17 *in trans* can maintain genotypes that do not encode functional tail fibres. Following HR, lysates are plated to screen for recombinants/WT phage. Plates are are of the 5th, 5th, and 8th 10-fold dilutions of phage, reading from left to right.

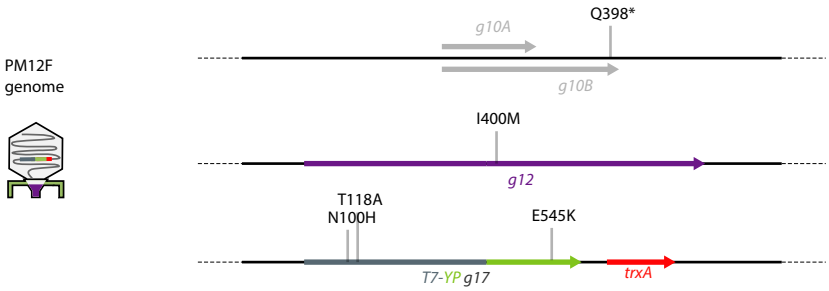


Figure 2.12: Schematic describing the location of mutations in the amplified TYP-phage stock (PM12F) relative to the canonical WT T7 sequence (NC_001604^{+}, see section 5.2) as determined by next-generation sequencing (see Methods 5.3.9). Genome alterations Q398* in gp10B and T118A in gp17 were observed in the progenitor T7 (see Figure 2.4).

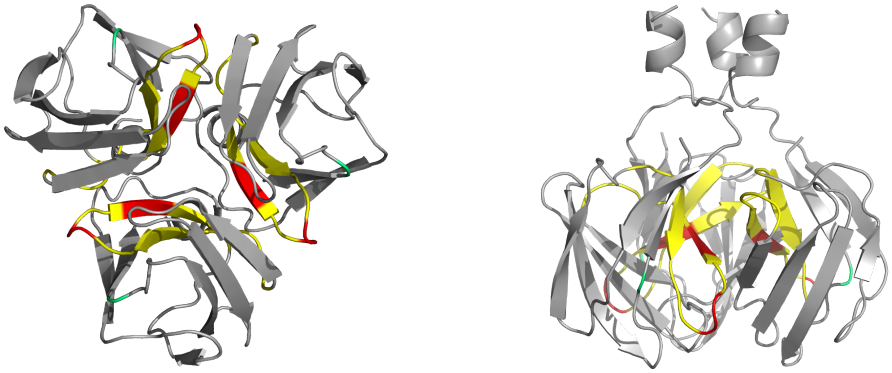


Figure 2.13: Protein structure homology-modelling (SWISS-MODEL [117–120]) of the C-terminus Yep-phi gp17 (aa 474-569) using the homo-trimer PDB structure of T7 gp17 (aa 457-552, PDB ID: 4A0T [91]) as a template. The model is shown viewed from below (left) and from the side (right). Highlighted in red are the amino acids shown to interact with Ail and OmpF of *Y. pesits* [24] using 20 aa linear peptides from Yep-phi (the amino acids from the linear peptide which bound, being displayed in yellow). The location of the E545K mutation in the PM12F tail fibre is coloured in green.

2.2.5 *In vivo* characterisation of TYP phage with a heterlogously expressed co-receptor (Ail)

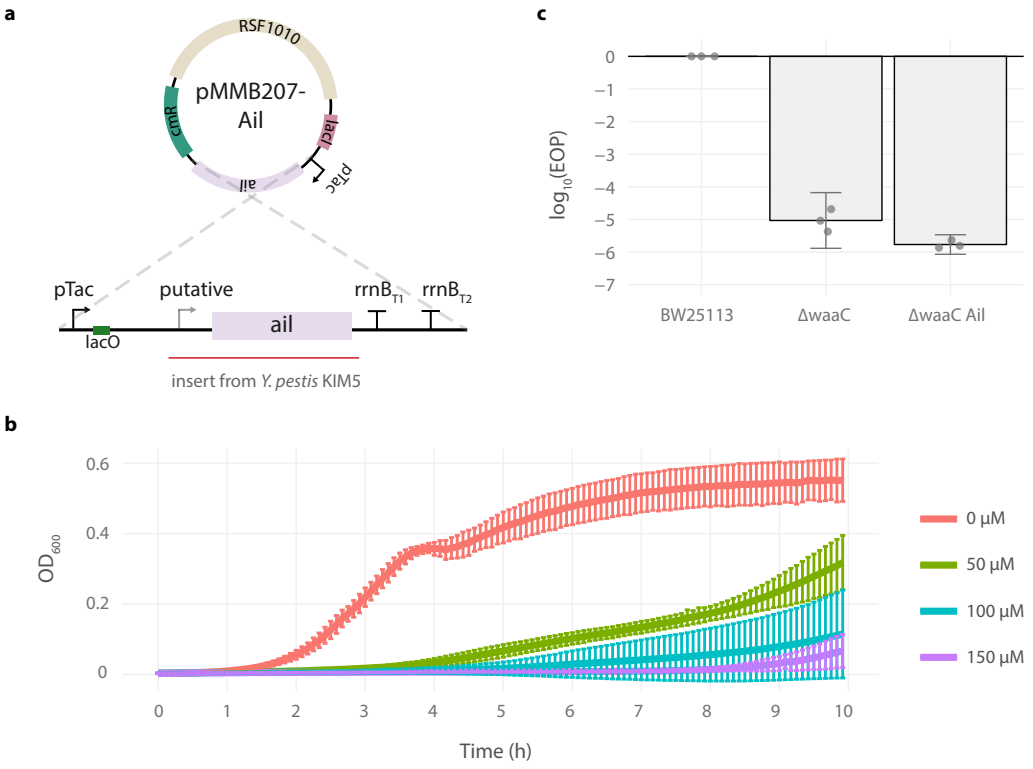


Figure 2.14: a) A plasmid map for pMMB207-Ail (a gift from the Krukonis lab [95]). A tac promoter and lacO site sit upstream of an inserted gene region from *Yersinia pestis*. The BPROM algorithm [121] predicts that an additional promoter sits between *ail* and pTac (labelled putative). b) Growth curves of Δ*waaC* pMMB207-Ail induced with differing concentrations of IPTG. Error bars are given as standard deviations of 3 biological replicates. c) Amplified T7-YP (PM12F) plated against BW25113, Δ*waaC*, and uninduced Δ*waaC* pMMB207-Ail. EOP data are $\log_{10}$ transformed prior to mean and error calculation of 3 biological replicates (each of which is an average of 2 technical replicates). BW25113 serves as the reference strain for the EOP calculations. Error bars are presented as 95% confidence intervals. Raw data for the biological replicates overlay the bar plots.

pMMB207-Ail was a gift from the Krukonis lab [95] (see Figure 2.14a). Cloned from *Y. pestis* KIM5, pMMB207-Ail contains *ail*, its native RBS, and a small

upstream nucleotide sequence. In addition, the BPROM algorithm [121] predicts that there is a constitutive σ_{70} promoter in that upstream region. Following the protocols established in the literature [95, 122] the pTac promoter was expected to regulate the expression of *ail*. pMMB207-Ail was transformed into BW25113 $\Delta waaC$ cells, and picked colonies grown overnight were refreshed in varying concentrations of IPTG in a 96 well plate (Figure 2.14b, see section 5.8.1 for details). Induced cells took several hours (> 8) to reach a satisfactory turbidity, with only uninduced cells growing as expected (mean growth rate 0.0312 OD_{600} min^{-1}, 95% CI $[0.0233, 0.0391]$). Serial dilutions of a PM12F (see section 2.2.4) were plated on BW25113, BW25113 $\Delta waaC$ and BW25113 $\Delta waaC$ pMMB207-Ail cells (Figure 2.14c). PM12F produced 8.8×10^4-fold fewer plaques on the reference strain, BW25113, compared with $\Delta waaC$, and 5.7×10^5-fold fewer on $\Delta waaC$ pMMB207-Ail. A one-sided Welch's t-test to see if the mean EOP against $\Delta waaC$ was larger than mean versus $\Delta waaC$ pMMB207-Ail indicated that the difference was not significant at the 5% level ($p \approx 0.09$).

2.3 Discussion

2.3.1 Design and cloning of a T7 *g17* sequence with reduced homology to the WT sequence

Recombination between T7 phage and the pSEVA551-g17* plasmid involving a single crossover event would result in T7 acquiring the entire pSEVA551-g17* plasmid, which at 7382 bp, is larger than the maximum 2.6% increase ($\approx$ 1 kb) in genome size that can be packaged into T7 capsids [57, 123]. Therefore, we designed the *g17** sequence accounting for research by Jacobus *et al.* [124] suggesting that double crossover recombination events in *E. coli* were possible with as little as 10 bp of flanking homology. We made no attempt to determine from the literature whether the T7 recombination machinery alone could generate recombinants with less homology, and instead opted to change as much as was feasible (in this case allowing at most 8 bp stretches of homology).

The cloning procedure outlined in Figure 2.3a was in fact performed twice; on both occasions, small colonies (and few in number) were seen following transformation of the Gibson reaction. Again, in both cases, only 1 colony was found to contain a correctly sized band, and when sequenced, was determined to be maligned with a single missense mutation (S189I on the first attempt and S402R on the second attempt respectively - data not shown). One explanation for this would be that the codon alterations made to the WT *g17* sequence could have hindered the growth of the host *E. coli* (through translatable, off-frame, cytotoxic protein products, or RNA sequences which encourage ribosome slippage). Another explanation could be

that the PCR primers used to perform Gibson assembly were improperly designed, with too many base pairs (>40) annealing to the template with particular primer pairs (e.g. oPM17/18) leading to low DNA yields (see Figure 2.3b) and possibly impacting the efficacy of the isothermal Gibson reaction. Nevertheless, Grigonyte [125] showed that the *g17** S189I sequence rescued phages with stop codons in *g17* (and so, were without functional tail fibres). For the purposes of this work, retention of the canonical protein sequence was not necessary, and subsequently this construction was utilised regardless.

2.3.2 Making T7 Δ*g17*::*trxA*

Here we have demonstrated that *trxA*-based selection of T7 *g17* recombinants is possible. Since the time of construction, this has also been confirmed in the literature [126]. It would have been more straightforward to have followed previous work [101, 102] and cleanly split recombination, and then selection, into two separate steps. Instead, the chosen method theoretically selected for recombinants at the same time as generating them (Figure 2.6a). However, given that the plasmid used to recombine *trxA* into T7 contained a T7 promoter, this most likely rendered the insertion of these plasmids into a Δ*trxA* strain redundant (especially as TrxA over-expression does not inhibit T7 growth [87]). Therefore, the BW25113 Δ*trxA* pSB3T5 HRg17-trxA + pSB6A1-gp17 strain would have been phenotypically $trxA^+$ after infection by T7.

Design and construction of a functional T7/Yep-Phi tail fibre chimera

We designed and cloned a chimeric fusion of the T7 tail fibre and the Yep-phi tip domain. Yep-phi gp17 and T7 gp17 share a 38% protein sequence identity and a 53% identity in the region of T7 gp17 which was exchanged. For comparison, the T7 tail fibre fusion with the last 410 aa of T3 produced by Ando *et al.* was 79% identical over the whole protein sequence and 72% identical over the region exchanged. We also constructed the pET30a+T7-YP plasmid, with the last 181 amino acids of this fusion. Raaij *et al.* have previously shown that amino acids 371-553 of the *g17* tail fibre trimerise, and form resolvable crystal structures [91]. Protein expressed using the pET30a+T7-YP plasmid migrates to about 25 kDa marker, as expected for the monomeric protein (expected weight for monomeric protein 25.2 kDa, calculated using the average isotopic masses of the amino acid sequence [127] (outlined in Figure A.2) which indicated that the fusion is soluble. That being said, there does not appear to be evidence of stable trimer formation in these conditions, though this need not preclude trimer formation *in vivo*.

2.3.3 Engineering of a T7 phage with a T7/Yep-Phi tail fibre chimera

Unexpectedly, phage that upon recombination, exchange the C-terminus of gp17 for that of Yep-phi phage, were still able to form plaques on *E. coli* lawns (indicating that they are still capable of adsorption, and subsequently DNA ejection - see Figure 2.11). In combination with the protein expression data in Figure 2.10d, these results potentially imply that either Yep-phi phage tail fibres are capable of adsorption to *E. coli*, or that T7 is capable of adsorption to *E. coli* in a manner that

is independent of the tip domain of gp17. In the latter case, we know from the work of Heineman *et al.* [83] that a V544A mutation is associated with the avoidance of *E. coli* K12 $\Delta trxA$. Similarly, Yosef *et al.* [126] found that a D540N mutation lead to a 10^6-fold improvement for T7 transduction of BW25113 with an LPS truncation ($\Delta waaC$, see Figure 1.5). These results alone would suggest that the tip of the T7 tail fibre is responsible for LPS adsorption in *E. coli*. However, data collected by Berbís Moreno [128] indicates that aa 371-553 binds to the dissacharide α-D-glucose-(1→ 2)-α-D-glucose (a component of the LPS, see Figure 1.5) and so an interaction between the LPS and the pyramid domain, or the alpha helix region (see Figure 2.7) can not be ruled out. Finally, a study published after the construction of PM12F by Ando *et al.* [15], points to the N-terminal domain of gp17 (aa 1-150) being involved in T7 adsorption to BW25113. In particular, Ando *et al.* traded the T7 tail fibre for one from the T7-like Enterobacteria phage 13a, and also made a chimera of the two tail fibres, taking the N-terminal domain from 13a. They found that although both phage phenotypes were functional; only the phage which possessed the N-terminal domain of gp17 could infect BW25113. In summary, the current literature does not preclude the possibility of T7 adsorption to BW25113 in a manner independent of the aa 446-553 region of T7 gp17.

As to whether Yep-phi phage tail fibres are capable of adsorption to the LPS of *E. coli*, we note here that the LPS of BW25113 and *Yersinia pestis* CO92 (the host strain of Yep-phi [24]) may share similarities in their inner core. Though I could only locate data on the LPS structure of *Y. pestis* for strains KM260 and I-2377 [96, 129] the common structure of Kdo and heptose molecules, as seen in *E. coli* BW25113 (Figure 1.5) appears to conserved in *Y. pseudotuberulosis* as well [129]. Similarly, *Y. pestis* strains, like *E. coli* BW25113, do not posses O-antigens [122].

Next-generation sequencing of amplified TYP phage (PM12F) indicated that there were several unanticipated mutations (Figure 2.12) that were not observed in the progenitor phage. In particular, there is one mutation, E545K, in the chimeric *g17* that was not observed in the Sanger sequencing of the plaque PM6-1 (Figure A.3b). Whether these mutations were the result of adaptation to the new chimeric tail fibre (and of that fusion protein to a supposedly new host (*E. coli*)) or simplify an artifact of the procedure, can not be determined from the evidence obtained.

2.3.4 *In vivo* characterisation of TYP phage with a heterlogously expressed co-receptor (Ail)

pMMB207-Ail was shown previously to express Ail heterologously in *E. coli* AAEC185 [122]. Comparing the genotypes for both strains, one significant difference is the absence of type I fimbriae in AAEC185 (mutation Δ*fimB-fimH*). Looking at the growth curve data in Figure 2.14b, it is apparent that even low concentrations of IPTG negatively effect growth, which was not seen in the work of Tsang *et al.* [122]. One possibility then is that type I fimbriae and Ail are somehow incompatible, though, as far as this author can tell, there is no literature on the subject. Another explanation could be that the truncated LPS of Δ*waaC* impacted Ail expression. On visual inspection of Figure 2.14c, one might be tempted to conclude that there was a difference between TYP phage plated against Δ*waaC* and Δ*waaC* pMMB207-Ail cells, however, a Welch's t-test found that they were insufficiently different to be statistically significant at the 5% level. Nevertheless, the aim was to see whether TYP phage would produce more plaques in the presence of cells expressing the co-receptor Ail, which the data clearly indicates was not the

case. Reasons for this are numerous: firstly, Ail expression and localisation to the outer membrane has not been demonstrated in BW25113 $\Delta waaC$ cells; secondly, PM12F contains a mutation (E545K) in the C-terminal Yep-phi domain, which may impact on the binding to Ail (though, the mutation does not fall in the region known to interact with Ail, see Figure 2.13); finally, the protein sequence for Ail used by Tsang *et al.* [122] is one amino acid (Y46D) different from the Ail sequence used in the work of Zhao *et al.*, and is located in the extracellular loops of Ail, which interact with the external world [24].

2.4 Conclusion

We've shown that it's possible to engineer tail fibre chimeras of T7 bacteriophage using the *trxA* marker-based selection method, and that, unexpectedly, this phage is capable of infecting *E. coli*. However, our aim was to create a phage which bound to an outer membrane protein expressed by *E. coli* (in this case, heterologously expressed Ail) which we did not manage to accomplish. Difficulties in demonstrating that Ail could be heterologously expressed in *E. coli* meant that future work in this thesis would instead concentrate on whether T7 could be encouraged to bind to, and infect *E. coli* using, endogenously expressed outer membrane proteins.

3 Preparations for directed evolution

3.1 Introduction

3.1.1 T7-ReRb phage

The Keio collection systematically provides single-gene deletions for every non-essential gene in *E. coli* BW25113 [43, 44]. Qimron et al. exploited this collection to generate a variant of T7 by plating this phage on a lawn of BW25113 Δ*waaC* (an 'Re' LPS mutant) and selecting a plaque for sequential plating on Δ*waaR* (Rb) and subsequent plaque isolation. They denoted this variant T7-ReRb, and observed that it plaques with greater efficiency on more LPS mutants of *E. coli* BW25113 than WT T7 phage (see Figure 1.5 and Table 3.1).

Table 3.1: Plating efficiency of T7 compared with T7-ReRb when challenged against a series of BW25113 LPS mutants. EOP ≥ 0.1 is denoted by a '+' and < 0.1 by a '−'. Taken from [87].

	***E. coli* LPS structure**					
Phage	**BW25113**	**Δ*waaR***	**Δ*waaO***	**Δ*waaG***	**Δ*waaF***	**Δ*waaC***
WT T7	+	−	+	−	+	−
T7-ReRb	+	+	+	−	+	+

3.1.2 T7 phage interaction with outer membrane proteins

A recent paper by González-García *et al. et al.* [130] indicated that T7 phage might also interact with outer membrane proteins of *E. coli* via gp17, in addition to the well-characterised LPS receptor. The authors used a virus overlay protein binding assay (VOPBA) [131] and mass spectrometry to determine those proteins to be OmpA and OmpF. However, T7 could still infect an *E. coli* $\Delta ompA$ $\Delta ompF$ $\Delta ompC$ mutant, suggesting that the OMPs are non-essential for T7 infection. *E. coli* OmpA is involved in biofilm formation and functions as an adhesion and invasin of mammalian tissue [132]. Roughly 2.5×10^5 copies are found per cell in BW25113 growing exponentially in M9 minimal media [20]. The crystal structure of OmpA in its 8-stranded β-barrelled narrow pore conformation and 4 extracellular loops [133] is depicted in Figure 3.1a. However, how it folds and sits natively in the outer membrane is currently disputed [23, 132, 134–141]. A 16-stranded β-barrelled large pore conformation has been observed *in vitro* [23, 134–141] but as of 2012 (and to this author's current knowledge) there is no spectroscopic or crystallographic evidence for this conformation [132]. OmpF, on the other hand, is known to exist in a 16-stranded β-barrelled porin natively (see Figure 3.1a). In that same paper by Soufi *et al.* there are estimated to be roughly 4.3×10^4 OmpF proteins in exponentially growing BW25113 [20]. Each one provides a water-filled channel for hydrophilic molecules to pass into the *E. coli* [142]. The L2 loop is necessary for OmpF trimerisation in the outer membrane (see Figure 3.1b) [143].

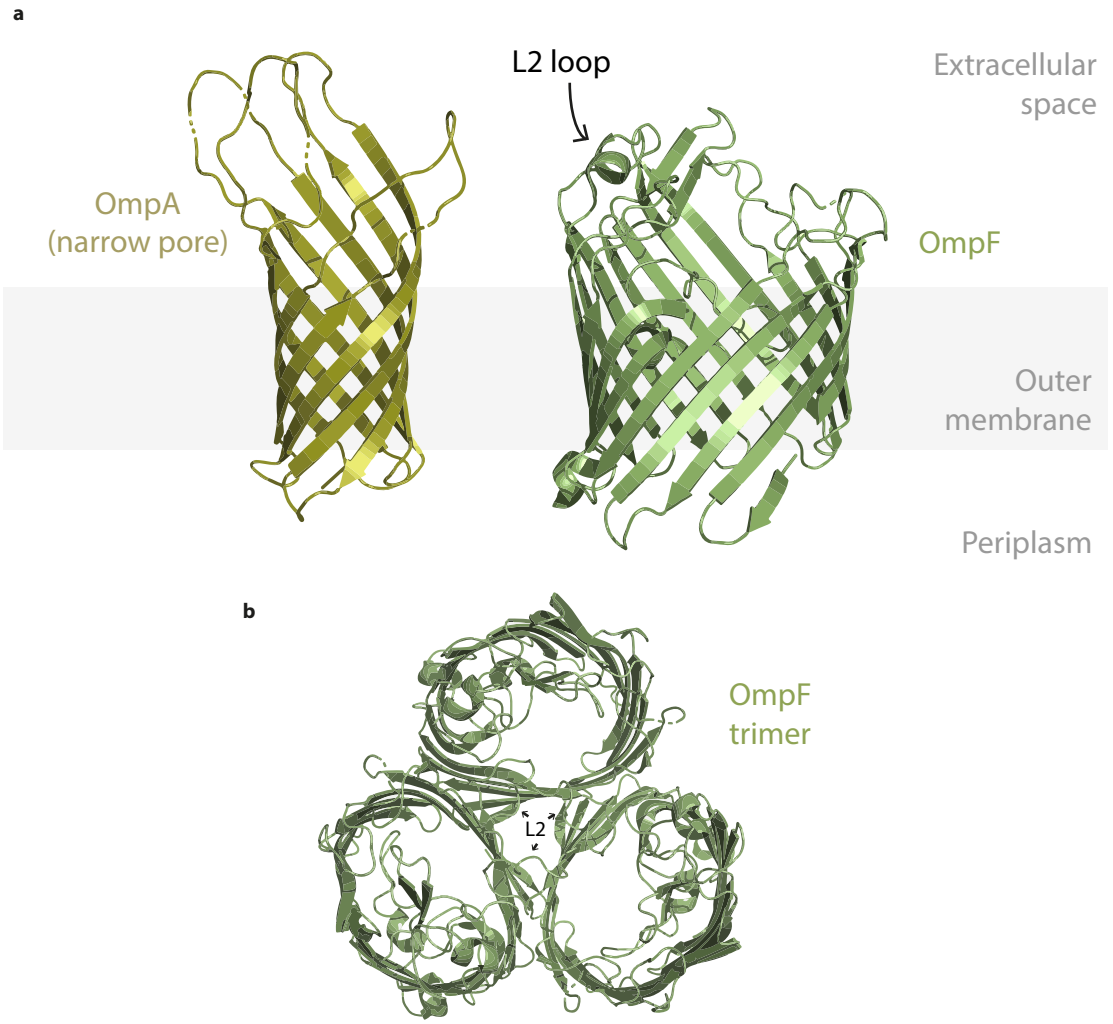


Figure 3.1: a) The narrow pore conformation of the crystallised amino acids 21-192 from the 346 residue long OmpA (PDB ID: 1BXW [133]) shown adjacent to a monomeric OmpF porin (PDB ID: 2ZFG [144]). It has been speculated that OmpA may exist as in a conformation closer to the 16-stranded β-barrelled OmpF natively. b) OmpF forms trimers in the outer membrane of *E. coli*, the stability of which are associated with the WT LPS structure (view from above, PDB ID: 2ZFG [144]). The L2 loop is required for trimer formation. PDB files visuali sed in PyMOL [115].

3.1.3 Design of evolution experiments

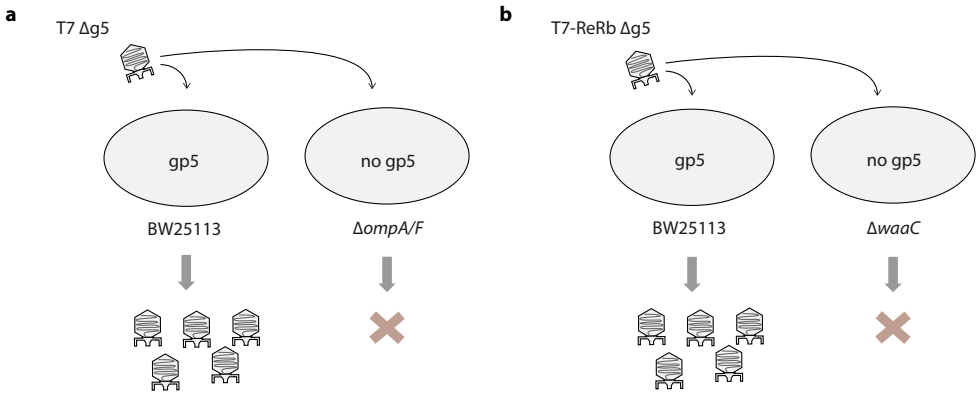


Figure 3.2: a) The proposed directed evolution strategy. T7 $\Delta g5$ requires gp5 to be complemented *in trans*, otherwise phage DNA replication does not occur. Subsequently, T7 $\Delta g5$ phage will be passaged in a co-culture of strains permissive (BW25113 gp5^{+}) and non-permissive (BW25113 $\Delta ompA/F$ gp5^{-}) to phage replication, with the aim of penalising phage adsorption to outer membrane components of BW25113 other than ompA (or, alternatively, ompF). b) T7-ReRb infects both BW25113, and BW25113 with a truncated LPS ($\Delta waaC$). As an initial proof-of-concept, we take T7-ReRb and seek to penalise adsorption to $\Delta waaC$.

As described previously (in section 1.6) our aim is to generate a T7 phage that infects *E. coli* using a given outer membrane proteins as a receptor. We will attempt this by passaging T7 in a mixture of *E. coli* strains that are either permissive or non-permissive to replication. Those strains which contain the targeted receptor will be permissive, whereas strains without that receptor (provided by the Keio collection) will be non-permissive. Thus, we hypothesise that this process should encourage adaptation to the desired component of the outer membrane. To implement this, I will first impart control on the replication of T7 by complementing *g5* (encoding T7 DNAP) *in trans* in the designated permissive strains, whilst removing *g5* from the phage. I will then allow T7 to replicate in cells containing the target outer membrane protein (OmpA or OmpF) and prevent T7 from replicating in

strains without those outer membrane protein ($\Delta ompA$ and $\Delta ompF$, see Figure 3.2a). The *in vivo* mutation rate of the phage can also be increased through the use of mutant T7 DNA polymerases, as will be discussed further in the next section.

To ensure that our methodology is sound, and that T7 can be encouraged to adapt in this way, we will perform a preliminary, control experiment, that tests different strategies for directed evolution in a situation where the desired outcome (the phage adsorption phenotype) of the evolution experiment is known to exist in the mutational landscape of T7 (Figure 3.2b). To this end, I chose to regress T7-ReRb phage back to its progenitor phenotype (more efficient infection of the WT LPS phenotype than with LPS truncations, see Table 3.1).

3.1.4 Mutant T7 DNA polymerases

Evolution is a filter, which sifts out things that are insufficiently fit in a given context. Objects, memes, or organisms that can pass the generator of that thing to next generation, and do so more effectively than others, are considered fitter. In biological contexts, the generator is an organism's genotype, but it is the phenotype that selection acts on. In the case of T7 phage, the template that generates these phenotypes is its dsDNA genome.[1] There are 2 established methods for diversifying the T7 genome *in vivo*. The first is to use the chemical mutagen N-methyl-N'-nitro-N-nitrosoguanidine (NG) which methylates nitrogen and oxygen atoms in DNA [147]. This results in a strong bias for GC→ AT transitions upon DNA replication [148] and in T7, generated roughly 4 nonlethal mutations per genome per generation *in vivo* ($\approx 10^{-4}$ mutations per bp) [48]. The second involves using

[1]The T7 RNA polymerase has an error rate per base pair of 5×10^{-5} [146], and may also diversify the T7 phenotype, but we choose not to effect this.

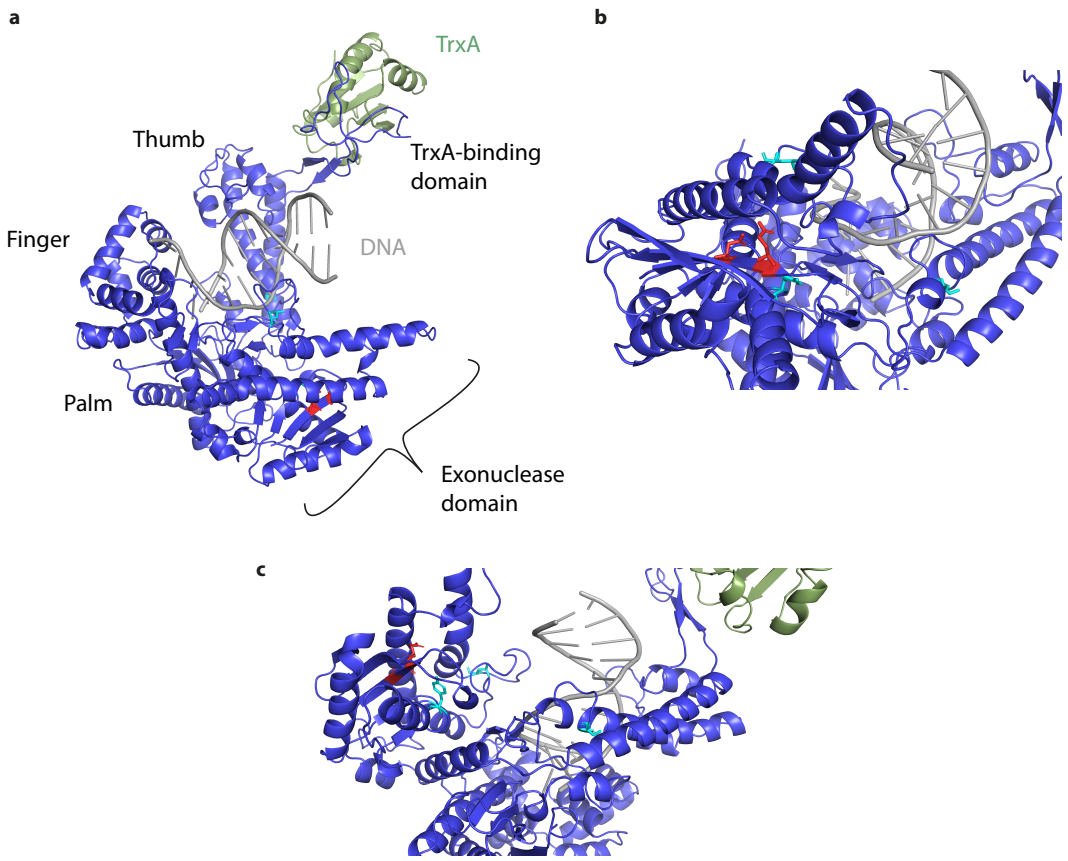


Figure 3.3: a) Crystal structure of T7 gp5 (DNAP) bound to trxA (PDB 1T7P). b) Close-up of D5A and D7A mutant locations (in red). c) Close-up of Y64C, F120L, S399T mutant locations (cyan blue). Figure adapted from Tran *et al.* [145].

error-prone T7 DNA polymerases to yield errors in replicating genomes. Wong et al. [149] characterised a D5A, E7A mutant of T7 DNAP which led to a complete inactivation of the exonuclease activity that repairs DNA mismatches. This and subsequent works suggests that the mutations produce a 10-210 fold decrease in polymerase fidelity [149–151]. Söte *et al.* [152] employed directed evolution to generate a Y64C, F120L, S399T mutant which increases the error rate of gp5 by roughly 55-fold when compared with the WT polymerase. That being said, relative to the exonuclease-deficient polymerase (D5A E7A mutant) there is only a 1.7-fold improvement. The locations of both sets of mutations are presented in Figure 3.3.

In the work that follows, we use a T7 gp5 with all 5 of these mutations (developed previously in the Jaramillo Lab and referred to as ER-gp5) in the directed evolution experiments.

3.1.5 Mathematical modelling

Throughout this work, I will refer to mathematical formulations of phage-bacteria systems, and their associated dynamics, in an attempt to better understand their interactions and to inform the experiments employed herein. Though no model ever gives a completely accurate account of a biophysical system, some may be helpful [153, 154] and we apply this philosophy in developing qualitative models to build on, and to help extend, the knowledge derived from experimentation.

Ordinary differential equations

Ordinary differential equations (ODEs) are commonly used to describe dynamical systems, and will form the basis of most of the mathematical models described in this thesis. In order to develop a better grounding of what an ODE is,[2] let us start with a basic model of an exponentially growing population, N. We call $N(t)$ the population density at a time t, so that $N(t + \delta t)$ represents the population density at a time $t + \delta t$ (where δt is a positive amount of time). We write that,

$$N(t + \delta t) = N(t) + \underbrace{b\ \delta t\ N(t)}_{\text{growth over time period dt}} \tag{3.1}$$

[2]A reader familiar with ODEs and mass action kinetics can skip over the following sections until the introduction of multiplicity of infection (MOI)

where b is the growth rate of the population. We see that the population of N at time $t + \delta t$ is dependant on the population of N at time t. Table 3.2 shows how N increases as we move forward in time by increments of 1, for a fixed growth rate.

Table 3.2: Equation (3.1) incremented in time steps (δt) of length 0.1, with $b = 1$ and $N = 100$ at time $t = 0$. This procedure is an estimate of the true solution. Better approximations can be found by incrementing in time steps of smaller length (i.e. using values of δt that are closer to zero).

t	N
0	100
0.1	$100 + 0.1 \times 100 = 110$
0.2	$110 + 0.1 \times 110 = 121$
..	..

Equation (3.1) tells us how the population N changes as we take steps forward in time of size δt. By rearranging that equation, we can see how the population changes as the time step, δt, takes up infinitely small quantities close to zero - and so, becomes effectively continuous,

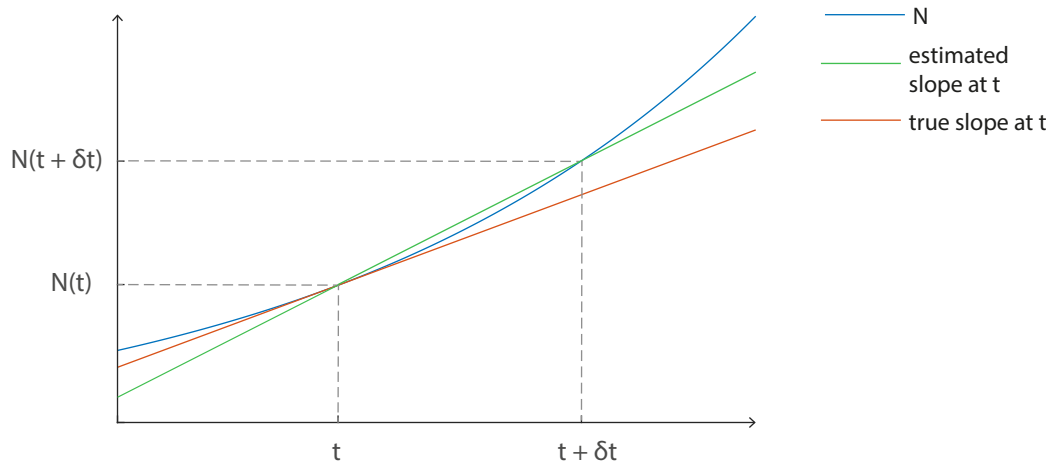
$$\frac{N(t + \delta t) - N(t)}{\delta t} = b\ N(t).$$

Notice that the left hand side is estimating the slope of $N(t)$ at t (see Figure 3.4) and that this estimation improves as $\delta t \to 0$. We denote the slope of $N(t)$ at t to be $\mathrm{d}N(t)/\mathrm{d}t$ (the derivative of N with respect to t) and so we write,

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = b\ N(t). \tag{3.2}$$

We refer to equation (3.2) as an ordinary differential equation, and its solution, in this case, is given by $N(t) = N_0 e^{bt}$, where N_0 denotes the initial population size. It is common to not explicitly state what each variable is a function of where it is

unambiguous to do so (i.e. instead of $N(t)$, we write N). As such equation (3.2) will often be written as $\mathrm{d}N/\mathrm{d}t = b\,N$.



$$\text{estimated slope at t} = \frac{N(t+\delta t) - N(t)}{(t+\delta t) - t} = \frac{N(t+\delta t) - N(t)}{\delta t}$$

Figure 3.4: The estimated slope between the two points (t and $t + \delta t$) approaches the true slope (the tangent of N at t) as $\delta t \to 0$.

Mass action kinetics

If we have substance A that combines with another substance B at a rate k to yield a product C i.e. $A + B \xrightarrow{k} C$, then the law of mass action asserts that the rate of the reaction is given by kAB, which is to say,

$$\underbrace{\frac{\mathrm{d}C}{\mathrm{d}t}}_{\text{gain of C}} = kAB = \underbrace{-\frac{\mathrm{d}A}{\mathrm{d}t}}_{\text{loss of A}} = \underbrace{-\frac{\mathrm{d}B}{\mathrm{d}t}}_{\text{loss of B}}. \tag{3.3}$$

Figure 3.5 shows this diagrammatically, and additionally provides a basic example with bacteria and phage.

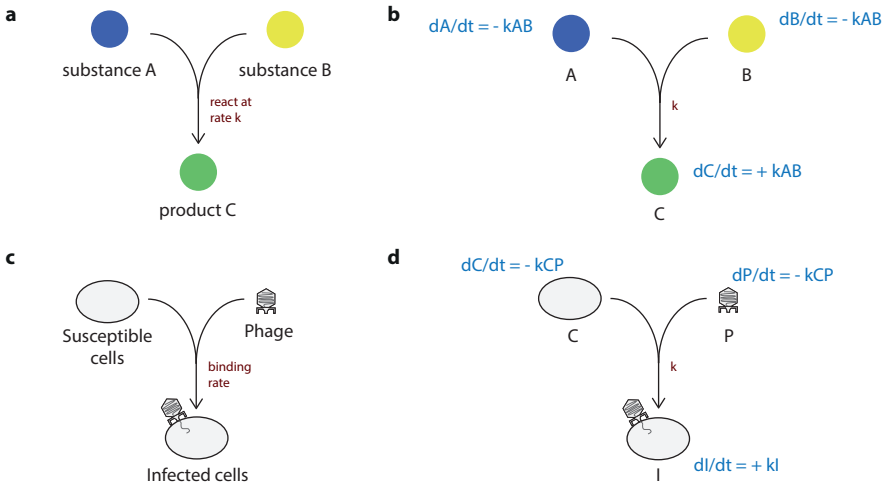


Figure 3.5: a) 2 substances combine to make a product C. b) According to the law of mass action, each substance or product is lost or gained according to the rate of reaction kAB. c & d) Phage binding to susceptible cells can be modelled in the same fashion.

3.1.6 Multiplicity of Infection (MOI)

A frequently used concept in phage and viral literature is that of MOI. The modern interpretation of this [13] is that it is a measure of the number of phages or viruses per cell, in a given setting, and it's formulation reflects this,

$$\text{MOI} = \frac{PFU}{CFU}. \tag{3.4}$$

The MOI is thus the expected number of phage that will be bound to each cell. Patently, not every cell will be bound by this number of phage; instead the numbers bound to each cell will follow a distribution, with fewer phage bound to some cells, and more phage bound to others. Treating the distribution as Poisson, as is typical [13, 155–158], we can say that the probability of n phage infecting one cell is given by

$$P(n \text{ phage infecting one cell}) = \frac{e^{-m}\, m^n}{n!}$$

where m denotes the MOI. Therefore, the probability of at least 1 phage entering a given cell is simply,

$$P(\text{at least 1 phage infects a cell}) = 1 - P(0 \text{ phages infect a cell}) = 1 - e^{-m}.$$

By considering the expected number of phage-cell encounters to be the MOI (denoted $\text{MOI}_{\text{input}}$ by Kasman *et al.* [159]) the implicit assumption is that every phage that can infect a given cell will do so in an allowed time period. However, if the binding rate between phage and cells is small, or if there are fewer cells in a given volume, this assumption can break down. Kasman *et al.* [159] present an alternative formulation of MOI to circumvent this issue. In that work, the authors retain the assumption that there is no reduction in the population of cells (i.e. no cell lysis) over the considered time period. It follows then from the law of mass action kinetics that

$$\frac{\mathrm{d}P}{\mathrm{d}t} = -kS_0P$$

where S_0 is the initial density of phage susceptible cells (CFU/ml), k is the adsorption or binding rate (ml/min) between phage and cells, and P is the PFU/ml of phage. This equation is separable, and can be integrated to obtain $P_t = P_0e^{-kS_0t}$, where P_0 is the density of phage at $t = 0$ (and similarly, P_t is the density at time t). Rearranging, we have,

$$\underbrace{1 - \frac{P_t}{P_0}}_{\text{fraction of phage bound at time } t} \quad = \quad 1 - e^{-kS_0t} \quad := \quad f_b(t)$$

which allows us to derive the $\mathrm{MOI}_{\mathrm{actual}}$ as formulated by Kasman *et al.* [159],

$$\mathrm{MOI}_{\mathrm{actual}} = \underbrace{\frac{P_0}{S_0}}_{\mathrm{MOI}_{\mathrm{input}}} \times \quad f_b(t)$$

where $\mathrm{MOI}_{\mathrm{input}}$ is effectively the maximum possible MOI (which we then adjust by the fraction of bound phage at time t). Then, as before, we assume that the frequency of encounters between phage and cells is Poisson distributed. However, $\mathrm{MOI}_{\mathrm{actual}}$ varies with time, so the procedure for calculating the associated probability is a little more involved (see Appendix B.2). Nevertheless, the end result follows similarly,

$$P(\text{at least 1 phage infects a cell}) = 1 - P(0 \text{ phages infect a cell}) = 1 - e^{-\mathrm{MOI}_{\mathrm{actual}}}.$$

3.1.7 Aims for the chapter

In this chapter, we start preparing the materials to evolve T7 phage towards the use of OMPs for the infection of *E. coli*. Additionally, we aim to set up a proof of concept experiment for our directed evolution strategies, which will determine whether T7-ReRb, a phage which infects both BW25113 and a truncated LPS phenotype ($\Delta waaC$) can revert back to the binding affinities exhibited by WT T7. To do so, we will characterise inherited constructs, such as T7 and T7-ReRb phage with *g5* deletions, as well as plasmids for the exogenous expression of gp5. We will also construct vectors for diversifying the T7 genome *in vivo*.

3.2 Results

3.2.1 Preparation of inherited constructs

The main aim of this chapter is to prepare the phage needed for the evolution experiments described in Section 3.1.3 Figure 3.2a and 3.2b. Given that T7/T7-ReRb Δ*g5*::(Δ*cmk-trxA*) were previously constructed in our lab, it remained to characterise, the phage, and to make stocks. These phages were separately plated on BW25113 Δ*cmk* Δ*trxA* pET24a-gp5-amp cells, and isolated plaques were amplified in that strain too prior to sequencing and characterisation. The sequenced T7 Δ*g5*::(Δ*cmk-trxA*) is identical to WT T7 (see Section 2.2.2 Figure 2.4) in the gene regions sequenced, with the exception of the *cmk-trxA* insertion into *g5*, and an N777H mutation in gp12. Similarly, T7-ReRb Δ*g5*::(Δ*cmk-trxA*) is identical in the regions observed to the T7-ReRb we received from Qimron, barring the *cmk-trxA* insertion (Figure 3.6).

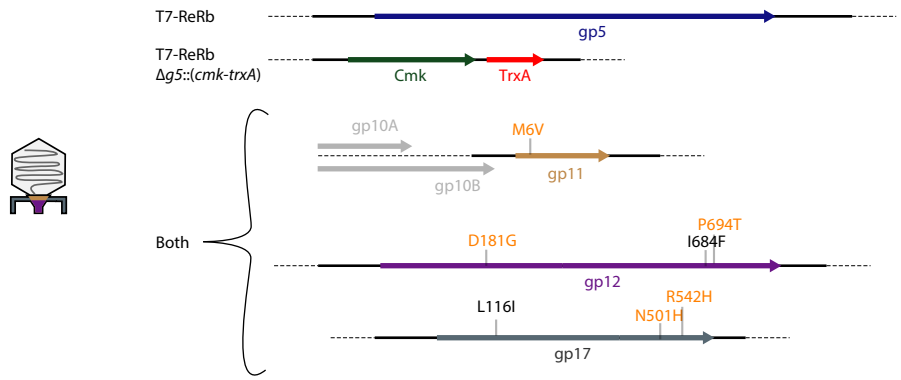


Figure 3.6: Sanger sequencing of T7-ReRb phages. Mutations cited in the Qimron *et al.* paper are coloured in orange [87]. All other mutations observed are non-canonical.

3.2.2 Observation of putatively co-infecting T7

As was routine to do so, phages inherited from others were screened for carry over of T7 *gp5*$^{+}$ that had avoided cells non-permissive to replication during construction. The expected outcome of this screen for these phages is depicted in Figure 3.7a. It was anticipated that in the absence of a complementing gp5, T7/T7-ReRb $\Delta g5$::(Δ*cmk-trxA*) would not form plaques on BW25113. However, this is not what occurred (see Figures 3.7a and 3.7b) where instead small plaques were observed on BW25113. Plaques generated by WT T7 phage (the common culprit in these instances) tend to be of at least the same size as is seen in the complimented T7 *gp5*$^{-}$ plaques (data not shown). Apart from the appearance of small plaques where phage replication should not be possible, another peculiarity is that tenfold dilutions of the phage do not bring about a corresponding dilution in the number of plaques (see Fig 3.7d). Subsequently, I hypothesised that a recombined phage which has gained *g5*, but lost an essential gene in the process, such as *g10*, could persist through co-infection with T7/T7-ReRb $\Delta g5$::(Δ*cmk-trxA*) (which similarly can not replicate by itself - see Figure 3.8).

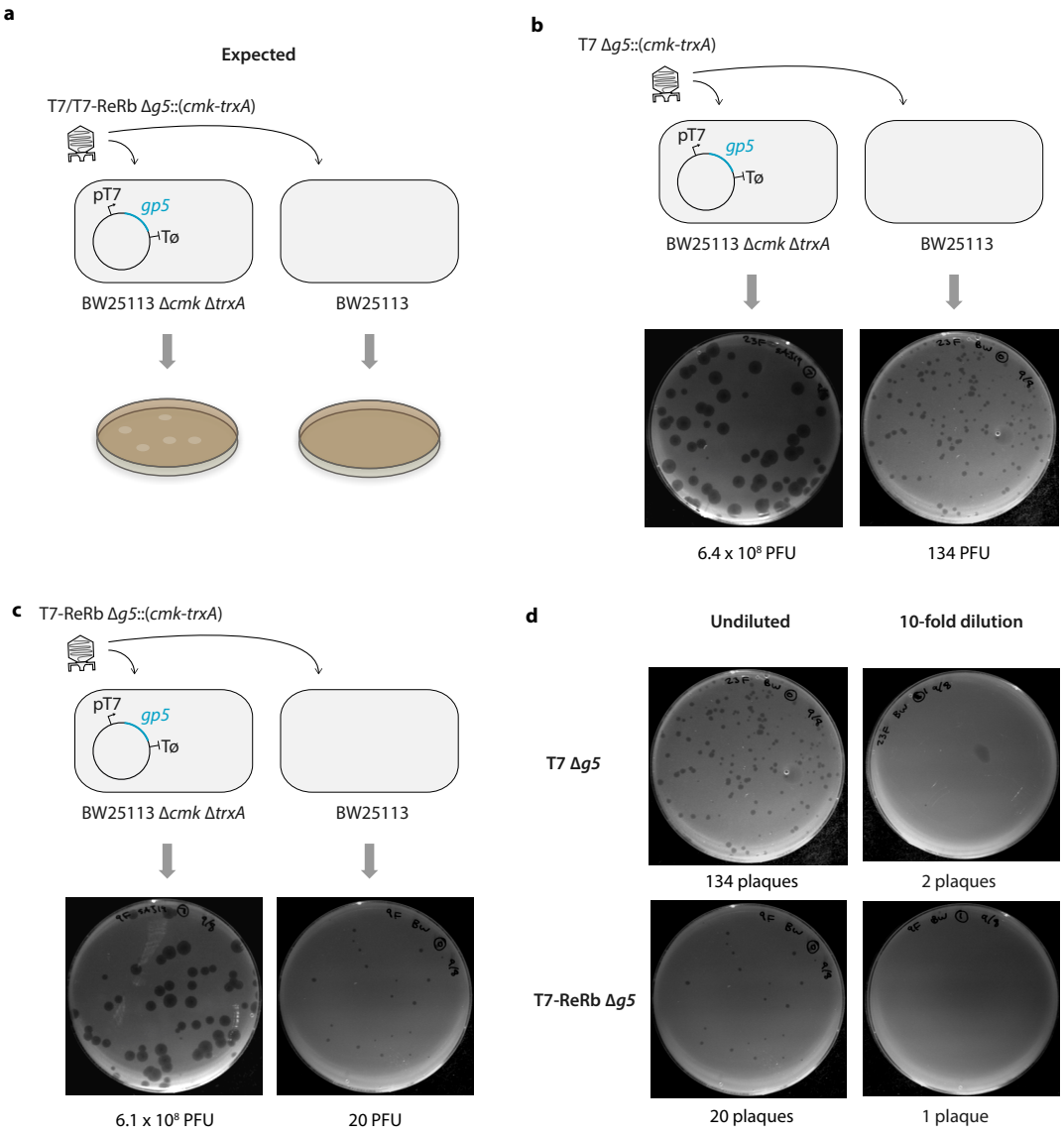


Figure 3.7: a) Expected outcome of plating T7 $\Delta g5$::(Δ*cmk-trxA*) and T7-ReRb $\Delta g5$::(Δ*cmk-trxA*) against cells with and without gp5 complemented *in trans* (expressed by pET24a-gp5-amp). In the absence of gp5, T7 is not expected to replicate, and hence no plaques should be observed. b & c) T7/T7-ReRb $\Delta g5$::(Δ*cmk-trxA*) generate plaques in lawns of cells devoid of gp5. d) Dilution of the phage stock does not produce a corresponding dilution in the number of plaques.

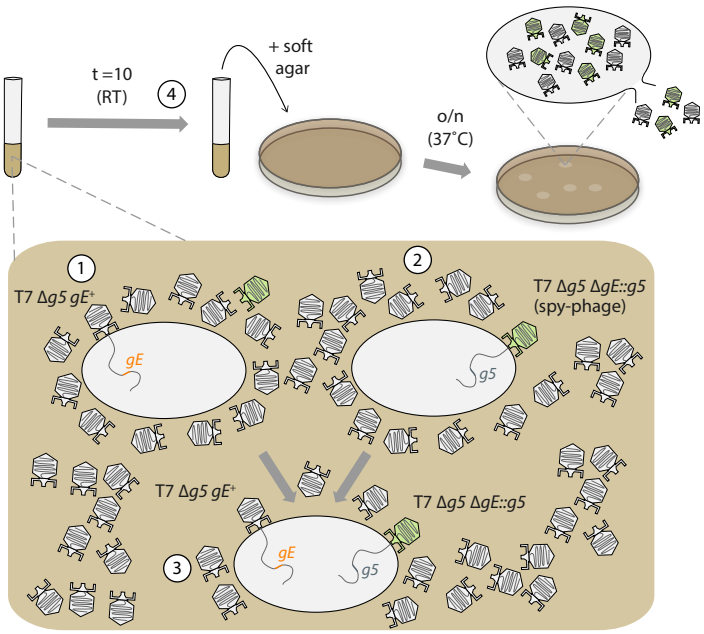


Figure 3.8: Putative mechanism for the appearance of small plaques on bacterial lawns without a complementing *g5* plasmid, after incubation with *g5*-deficient phage. Let *gE* denote the essential gene that has been lost in the spy-phage (in order to attain *g5* through recombination) and is retained in T7 $\Delta g5$ phage (referred to as T7 $\Delta g5$ gE^+). (1) T7 $\Delta g5$ gE^+ is deficient in *g5* and so can not replicate by itself. (2) Similarly, spy-phage (T7 $\Delta g5$ $\Delta gE::g5$) is deficient in the *gE* and so can not produce progeny by itself. (3) Should both phage genotypes infect the same cell, phage production is possible, and upon plating (4), will yield plaques containing both genotypes.

3.2.3 Modelling T7 co-infection

$\mathbf{MOI_{input}}$ model

By formalising this mechanism, we can simulate how the reduction in plaques seen in Figure 3.7d may have materialised. Let us give the name spy-phage to the phage that has recombined to gain *g5,* and in the process become deficient in another essential gene. Now, if we knew the PFU of the spy-phage (denoted PFU_{spy}) in our mixed phage population, we could estimate the percentage of spy-phage infected cells from the MOI. To do so, we follow the traditional assumption that the frequency of encounters between phage and cells are Poisson distributed, so that

$$\left(1 - e^{-\frac{PFU_{spy}}{CFU}}\right) = \% \text{ spy-phage infected cells.} \tag{3.5}$$

Similarly, we could estimate the number cells infected by T7 $\Delta g5$, which we could then combine with equation 3.5, to determine the number of cells infected simultaneously (we assume that each co-infected cell will give rise to a plaque). Hence, the number co-infected plaques produced by plating the undiluted sample of phage (the 0th dilution) is given by,

$$\begin{aligned} \left(1 - e^{-\frac{PFU_{T7}}{CFU}}\right) \times \% \text{ spy-phage infected cells} &= \% \text{ co-infected cells} \\ CFU \times \% \text{ co-infected cells} &= \# \text{ co-infected plaques} := \mathcal{P}_0 \end{aligned} \tag{3.6}$$

where $\mathcal{P}_i$ is the number of co-infected plaques seen on the ith 10-fold diluted plaque assay plate. In multiplying the percentage of T7 and spy-phage infected cells together, we have implicitly assumed that T7 and spy-phage infect cells

independently of one another (so that the infection of a given cell by either T7 or spy-phage does not impact the infection process of the other phage for that cell).

Given that the number of co-infected plaques (on the undiluted plate) is known, we can can rearrange equation 3.6 to obtain the unknown PFU_{spy}. Restating equation 3.6, we have,

$$CFU \times \left(1 - e^{-\frac{PFU_{T7}}{CFU}}\right) \times \left(1 - e^{-\frac{PFU_{spy}}{CFU}}\right) = \mathcal{P}_0 \tag{3.7}$$

which can be written as,

$$1 - e^{-\frac{PFU_{spy}}{CFU}} = \frac{\mathcal{P}_0}{CFU \times \left(1 - e^{-\frac{PFU_{T7}}{CFU}}\right)}, \text{ given that } CFU,\ PFU_{T7}\ > 0.$$

Finally, we continue to rearrange to produce,

$$-\frac{PFU_{spy}}{CFU} = \ln\left(1 - \frac{\mathcal{P}_0}{CFU \times \left(1 - e^{-\frac{PFU_{T7}}{CFU}}\right)}\right)$$

which then becomes,

$$PFU_{spy} = -CFU \times \ln\left(1 - \frac{\mathcal{P}_0}{CFU \times \left(1 - e^{-\frac{PFU_{T7}}{CFU}}\right)}\right). \tag{3.8}$$

Note that because the natural logarithm is defined on the interval $(0, \infty)$, a consequence of equation 3.8 is that we must have,

$$1 - \frac{\mathcal{P}_0}{CFU \times \left(1 - e^{-\frac{PFU_{T7}}{CFU}}\right)} > 0$$

which we can write as,

$$CFU \times \left(1 - e^{-\frac{PFU_{T7}}{CFU}}\right) > \mathcal{P}_0$$

or, more straightforwardly,

$$\text{number of T7 } \Delta g5 \text{ infected cells } > \text{ number of co-infected plaques.}$$

However, this constraint existed in any case as a consequence of the assumption that T7 and spy-phage infect cells independently of one another. Equation 3.8 appears too involved to be insightful, nevertheless, it is possible to find the following more comprehendible approximation by repeated first order expansions of the Taylor series (see Appendix Section B.1) which is given by

$$PFU_{spy} \approx \frac{\mathcal{P}_0}{\text{MOI}_{\text{T7}}} \tag{3.9}$$

where $100 \times \mathcal{P}_0 \leq 10 \times PFU_{T7} \leq CFU$. The intuition now becomes more straightforward. For a fixed number of co-infected plaques ($\mathcal{P}_0$) as the number of T7 $\Delta g5$ phage per cell (MOI_{T7}) decreases, more spy phage are required to generate the same number of co-infected plaques (i.e. PFU_{spy} increases - see Figure 3.9a). This is also what we observe when we simulate equation 3.8 in Figure 3.9b, except that the predicted PFU_{spy} plateaus at the number of co-infected plaques ($\mathcal{P}_0$) as the number of T7 $\Delta g5$ phage per cell (MOI_{T7}) increases (Figure 3.9c).

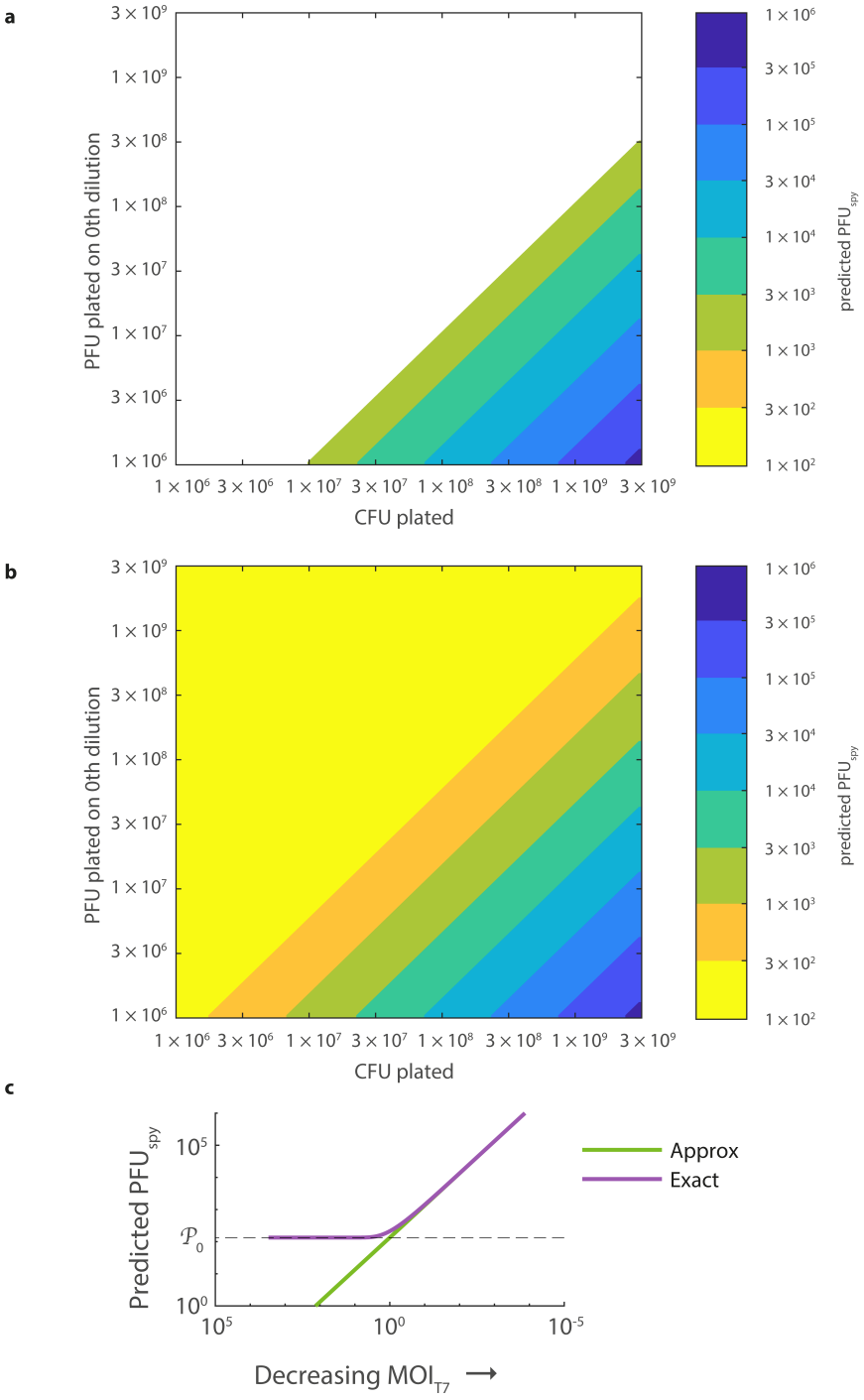


Figure 3.9: a) Heatmap of $PFU_{spy} \approx \mathcal{P}_0/\mathrm{MOI}_{\mathrm{T7}}$. The initial CFU plated and the initial PFU of T7 $\Delta g5$ plated on the undiluted plate (as measured against a reference strain) are varied. The number of co-infected plaques on the undiluted plate ($\mathcal{P}_0$) is fixed to be 134. $\mathrm{MOI}_{\mathrm{T7}}$ values greater than 0.1 are not plotted. Numbers rounded down to the nearest integer. b) As before, except with PFU_{spy} estimated more precisely using equation 3.8. c) Comparison of the approximation and equation 3.8 along the diagonal $PFU = -CFU + 3 \times 10^9$ for varying $\mathrm{MOI}_{\mathrm{T7}}$.

Now that we know how to determine PFU_{spy}, we can predict how many plaques will be seen on the 1st tenfold dilution plate. The equation follows the same logic used to derive $\mathcal{P}_0$ (equation 3.7) but now PFU_{spy} and PFU_{T7} are diluted tenfold:

$$\begin{aligned}\%\text{ co-infected cells} &= \%\text{ T7 } \Delta g5 \text{ infected cells} \times \%\text{ spy-phage infected cells} \\ &= \left(1 - e^{-\frac{PFU_{T7}\times 0.1}{CFU}}\right) \times \left(1 - e^{-\frac{PFU_{spy}\times 0.1}{CFU}}\right)\end{aligned}$$

and so, the number of co-infected plaques on the 1st 10-fold dilution plate ($\mathcal{P}_1$) is given by,

$$CFU \times \left(1 - e^{-\frac{PFU_{T7}\times 0.1}{CFU}}\right) \times \left(1 - e^{-\frac{PFU_{spy}\times 0.1}{CFU}}\right). \tag{3.10}$$

In other words, the predicted the number of co-infected plaques on the 1st tenfold dilution plate is given by equation 4.3, which is obtained by plugging in our estimate for PFU_{spy} from equation 3.8. Figure 3.10a plots equation 4.3 for differing PFU_{T7} and CFU, while fixing number of co-infected plaques ($\mathcal{P}_0$) at 134, and Figure 3.10b shows the fold reduction in plaques seen after plating the 0th and 1st 10-fold dilutions of phage, i.e.

$$\frac{\mathcal{P}_0}{\mathcal{P}_1} = \frac{\left(1 - e^{-\frac{PFU_{T7}}{CFU}}\right) \times \left(1 - e^{-\frac{PFU_{spy}}{CFU}}\right)}{\left(1 - e^{-\frac{PFU_{T7}\times 0.1}{CFU}}\right) \times \left(1 - e^{-\frac{PFU_{spy}\times 0.1}{CFU}}\right)}. \tag{3.11}$$

We can approximate this too for small values. Let $\text{MOI}_{\text{T7}} = \frac{PFU_{T7}}{CFU}$, $\text{MOI}_{\text{spy}} = \frac{PFU_{spy}}{CFU}$. By again employing the Maclaurin series, equation 3.11 can be simplified to,

$$\begin{aligned}\frac{\mathcal{P}_0}{\mathcal{P}_1} &= \frac{\left(1 - e^{-\text{MOI}_{\text{T7}}}\right) \times \left(1 - e^{-\text{MOI}_{\text{spy}}}\right)}{\left(1 - e^{-0.1\times\text{MOI}_{\text{T7}}}\right) \times \left(1 - e^{-0.1\times\text{MOI}_{\text{spy}}}\right)} \\ &\approx \frac{\text{MOI}_{\text{T7}} \times \text{MOI}_{\text{spy}}}{0.1 \times \text{MOI}_{\text{T7}} \times 0.1 \times \text{MOI}_{\text{spy}}} = 100\end{aligned}$$

when $\mathrm{MOI_{T7}}, \mathrm{MOI_{spy}} \leq 0.1$ (so PFU_{T7} and PFU_{spy} are at least 10 times smaller than the CFU plated) which tallies with the fold-reduction seen in Figure 3.10b.

Taken together, the simulations shown in Figure 3.10 suggest that altering the number of T7 $\Delta g5$ phage, or the CFU, that is plated, can effect whether a 10-fold dilution in that phage stock results in an equivalent fold reduction in plaques across the two 10-fold diluted plaque assay plates. In Figure 3.7d, we plated $\approx 6.4 \times 10^8$ PFU of T7 $\Delta g5$ on the undiluted plate. The CFU plated is unknown, but is likely to be in excess of 2×10^7 CFU, based on data for BW25113 pSB4G5 ER-gp5 (discussed in chapter 4). Now, this does not match up particularly well with the simulated data (Figure 3.10), however, the simulations do suggest that the experimental results observed might be consistent with the hypothesis that co-infection is occurring.

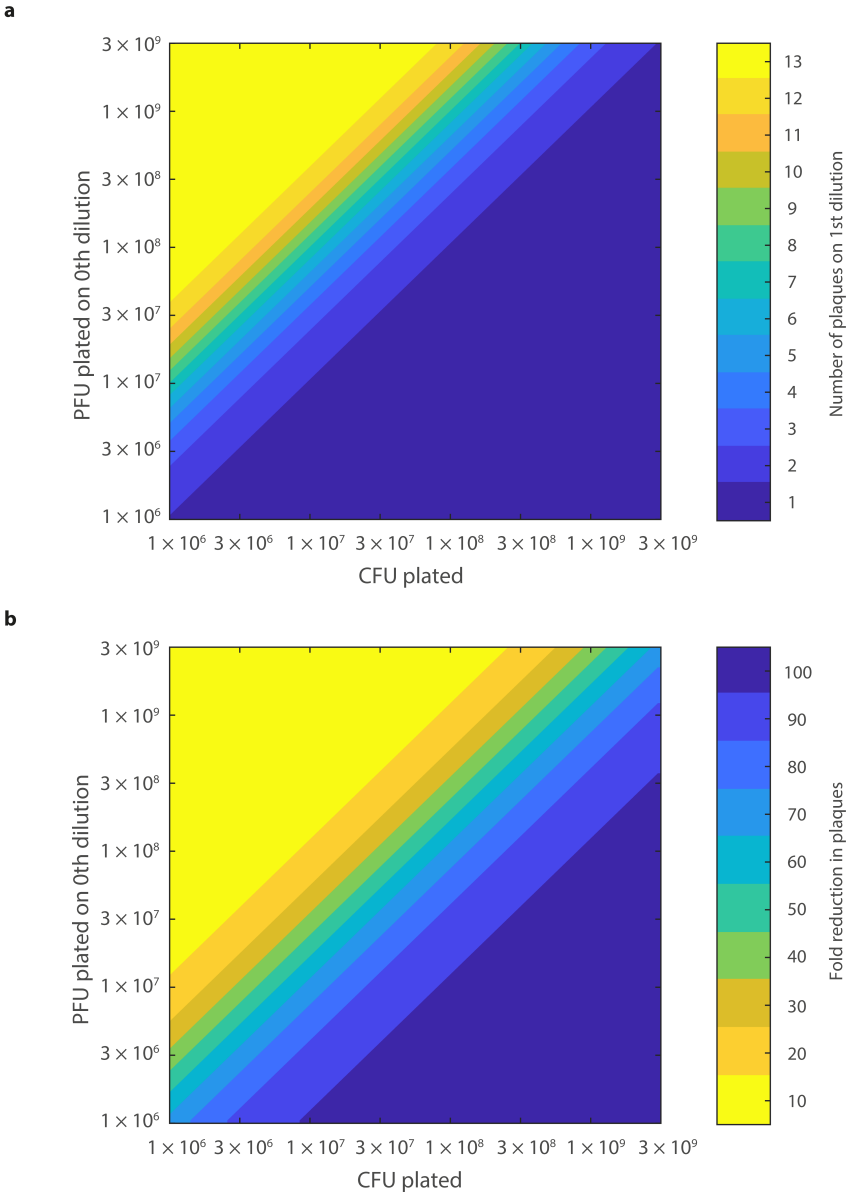


Figure 3.10: a) Simulated experiments predicting the number of co-infected plaques seen on the 1st tenfold diluted plaque assay plate from: the initial CFU plated; the number of co-infected plaques observed on the undiluted plaque assay plate; and the PFU of T7 $\Delta g5$ plated on the undiluted plate (as measured against a reference strain). The number of co-infected plaques on the undiluted plate is fixed to be 134. b) As before, but highlighting the fold reduction in the number of co-infected plaques from the 0th to the 1st tenfold phage dilution. Numbers in both panels are rounded down to the nearest integer.

$\mathbf{MOI_{actual}}$ model

One assumption in the $\text{MOI}_{\text{input}}$ model is that all phages will bind to and infect cells within the initial 10 minute window at room temperature, or when suspended in the soft agar mixture (recall Figure 3.8). A consequence of this simplifying assumption is that the process is independent of the adsorption rate. However, the adsorption rate of phage to cells is inversely proportional to the square of their density. This means that when the 400 µl of phage and cells is diluted with ≈ 3 ml of soft agar, the rate of adsorption decreases by roughly two orders of magnitude. We can vary this assumption of independence by deriving the equations based on MOI (referred to in the literature as $\text{MOI}_{\text{input}}$ [13] and reformulating them with $\text{MOI}_{\text{actual}}$ as defined by Kasman *et al* [159]. In that work, the authors assume that there is no reduction in the population of cells (which is true in our case, so long as we limit the time period to be no more than 10 minutes (i.e. before cell lysis [46, 81, 160, 161]). They then derive the $\text{MOI}_{\text{actual}}$ to be

$$\text{MOI}_{\text{actual}} = \underbrace{\frac{P_0}{S_0}}_{\text{MOI}_{\text{input}}} \times 1 - e^{-kS_0t} := \frac{P_0}{S_0} \times f_b$$

where P_0 is the initial PFU/ml and S_0 the initial CFU/ml. Note that $\text{MOI}_{\text{actual}}$ has dimensions phage per cell, and so, represents the number of phage bound per cell. Recall that we can consider the frequency of encounters between phage and cells to be Poisson distributed. In this case, the frequency varies with time, which means that the associated probability varies according to a non-homogeneous Poisson model (see Appendix B.2 for more detail). Nevertheless, as we saw in Section 3.1.6,

the probability that at least 1 phage infects a cell is given by,

$$\begin{aligned} P(\text{at least 1 phage infects a cell}) &= 1 - P(0 \text{ phages infect a cell}) \\ &= 1 - e^{-\text{MOI}_{\text{actual}}}. \end{aligned} \tag{3.12}$$

We can use (3.12) to re-derive equation (3.5) as

$$\left(1 - e^{-\frac{PFU_{spy}}{CFU} \times f_b}\right) = \% \text{ spy-phage infected cells.} \tag{3.13}$$

The equations to estimate PFU_{spy} and calculate the number of co-infected plaques on 1st 10-fold plate dilution ($\mathcal{P}_1$) follow as before,

$$PFU_{spy} = -\frac{CFU}{f_b} \times \ln\left(1 - \frac{\mathcal{P}_0}{CFU \times \left(1 - e^{-\frac{PFU_{T7}}{CFU} \times f_b}\right)}\right) \tag{3.14}$$

$$\mathcal{P}_1 = CFU \times \left(1 - e^{-\frac{PFU_{T7} \times 0.1}{CFU} \times f_b}\right) \times \left(1 - e^{-\frac{PFU_{spy} \times 0.1}{CFU} \times f_b}\right) \tag{3.15}$$

where we assume that the adsorption rate is the same for T7 $\Delta g5$ and spy-phage. Figure 3.11 shows the predicted PFU_{spy} and fold reduction for the $\text{MOI}_{\text{actual}}$ model with $\mathcal{P}_1$ fixed at 134; the adsorption rate, k, set at 3×10^{-9} ml/min [81]; and $t = 10$ (recall that $1 - e^{-kS_0t} = f_b$). Compared with the $\text{MOI}_{\text{input}}$ model (Figure 3.10) plating fewer CFU results in higher predictions for PFU_{spy} (adsorption to cells is less frequent within the 10 minute time frame, so more phage are needed to produce the same number of co-infected plaques). For similar reasons, the fold reduction in plaques increases also for lower CFU counts. However, the changes in fold reduction still do not align with the results seen in Figure 3.7d. By noting that k and t are constant parameters, we can define a new parameter $\nu = kt$, with

dimensions [ml/min]×[min] = [ml]. One can think of decreasing ν as effectively, decreasing the volume of space that a given phage can search for a cell in, and so, decreasing the frequency of encounters between phage and cells. Figure 3.12 highlights what happens if we decrease ν by 10-fold to 3×10^{-9}. Decreasing ν further increases the required CFU count to obtain a given number of co-infected plaques with the same PFU_{spy} (now ≈ 5000 for 2×10^7 CFU and 6.4×10^8 PFU). Additionally, the fold reduction in plaques is now more in line with the results observed in Figure 3.7d (an ≈ 30 fold reduction for 2×10^7 CFU, and 6.4×10^8 PFU).

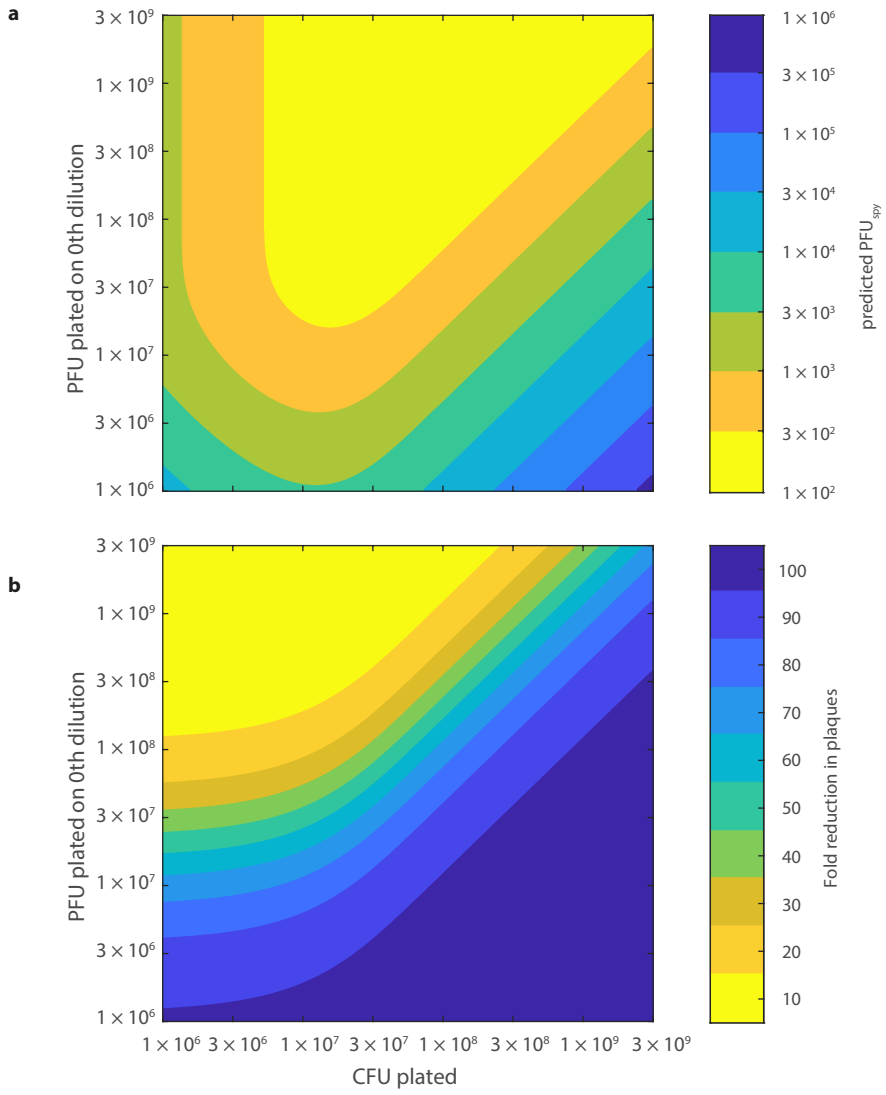


Figure 3.11: Simulated experiments with the $\mathrm{MOI}_{\mathrm{actual}}$ model predicting the number of co-infected plaques seen on the 1st tenfold diluted plaque assay plate. The adsorption rate of phage to cells is set to 3×10^{-9} ml/min and the time allowed for the phage to adsorb is set at 10 minutes (i.e. $\nu = kt = 3 \times 10^{-8}$ ml). The initial CFU and PFU of T7 $\Delta g5$ plated are varied as before, . The number of co-infected plaques on the undiluted plate is fixed at 134. Panel a) shows the predicted PFU_{spy} and panel b) the predicted fold reduction in plaques.

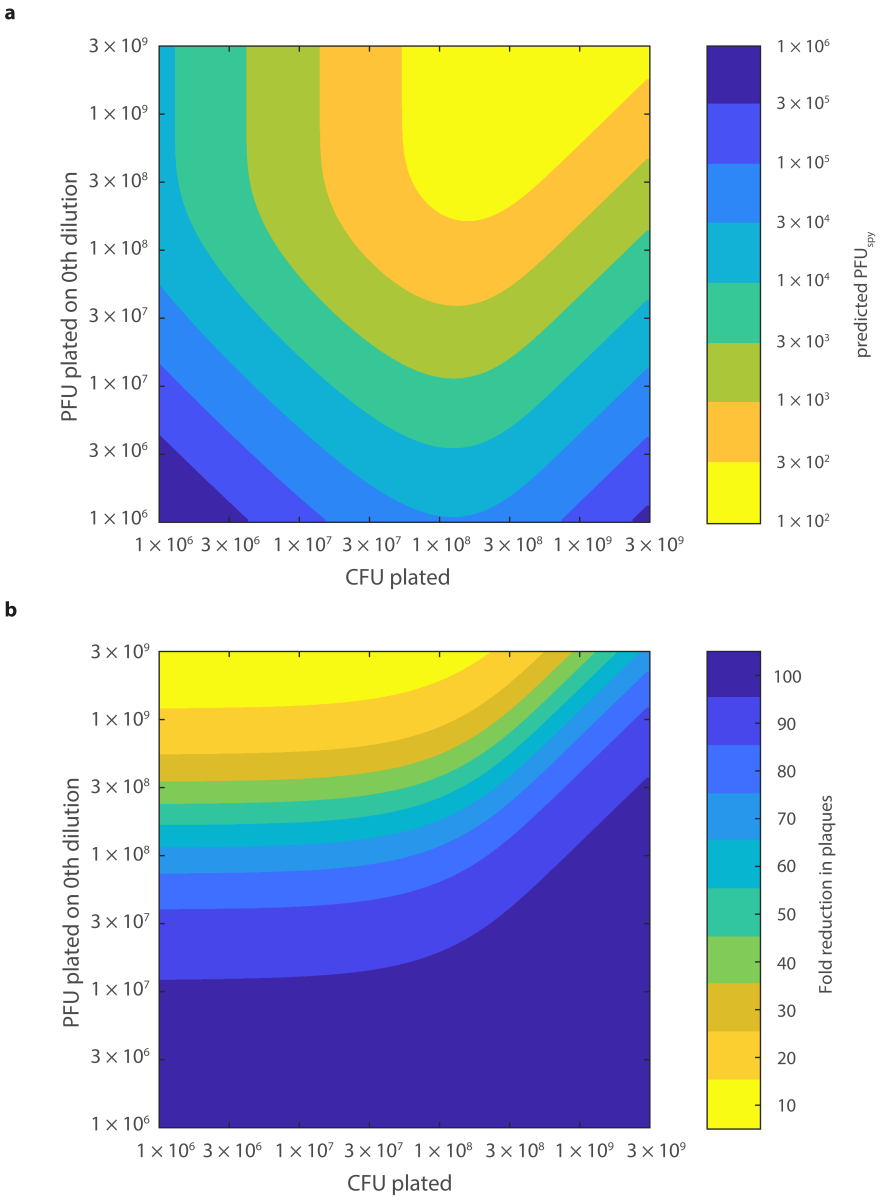


Figure 3.12: The details are the same as for Figure 3.11, except ν is 10-fold smaller (i.e. $\nu = 3 \times 10^{-9}$ ml).

3.2.4 Characterisation of T7 co-infection

It was hypothesised that amplification of the T7 phage in the presence of the pET24a-gp5-amp plasmid might result in recombination between the two elements (see Figure 3.13a). The backbone pET24a vector consists of a T7 ϕ10 promoter followed by the *g10* RBS, and a T7 Tϕ terminator, together which flank the multiple cloning site that *g5* had been cloned into. In total this results in 26 bp, 45 bp and 124 bp of homology immediately flanking *g10* in the T7 genome (see Figure 3.13b). To determine whether the putatively co-infected plaques contained phage that recombined with pET24a-gp5 to acquire *g5* (i.e. whether they contained T7 Δ*g5* Δ*g10*::*g5* recombinants) an isolated plaque was PCR amplified using the primer pair oPM262/265 which bind outside of *g10* on the T7 genome, and primer pairs oPM262/263 and oPM264/265 which both amplified the putative recombined sequence (see Figure 3.13c). Figure 3.13d presents the gel of these PCR products, along with controls (reactions with and without WT T7 phage). Two differently sized amplicons were observed when amplifying the co-infected plaque (see lane 3). One amplicon of the WT genome and another for T7 Δ*g5* Δ*g10*::*g5* recombinant genomes. In WT lane 3, only an amplicon for the WT phage genome is seen. Bands in WT lanes 1 and 2 are presumed to be from mis-priming reactions. Consequently, the largest homology region, Tϕ, was removed from vectors pET24a-gp5-amp and pET24a-gp5-kan along with the *lacI* gene, using primer pairs oPM195/6 and oPM197/8 (see Methods 5.5 and Figure 3.13e).

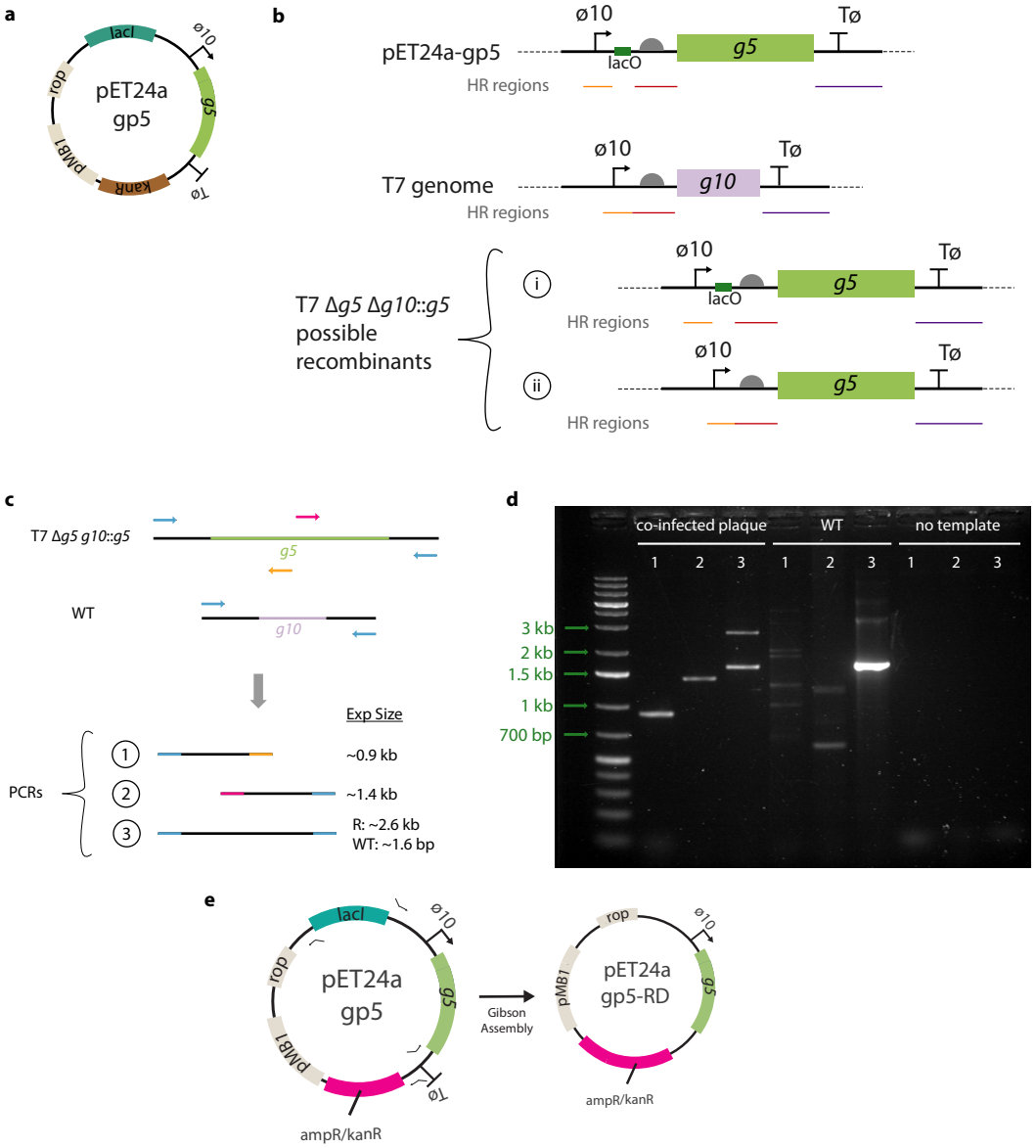


Figure 3.13: a) A plasmid map for pET24a-gp5-kan from which pET24a-gp5-amp was cloned. ϕ10 is the T7 promoter immediately upstream of *g10*. b) A diagram highlighting the shared homology between the pET24a-gp5 plasmid and the regions upstream and downstream of *g10* on the T7 genome. There are 2 sections of homology upstream of the *g5* - the ϕ10 promoter (26 bp long) and a 45 bp section (coloured red) that contains the *g10* RBS - which are separated by a LacO operator. Downstream, in and around the Tϕ terminator, is a 124 bp match to the T7 genome. The two likely double crossover events are also delineated. c) Primers flanking *g10* amplify that region as well as forming primer pairs with oligos binding from within *g5*. d) A gel of said PCRs amplifying phage from WT T7 and a putatively co-infected plaque. e) Cloning of pET24a-gp5 to prevent recombination with T7. pET24a-gp5-kan and pET24a-gp5-amp were PCR amplified in 2 parts, removing both the Tϕ terminator and the lacI gene from the each vector.

3.2.5 Re-construction of T7 Δ*g5* and T7-ReRb Δ*g5* phage

BW25113 Δ*cmk* competent cells were transformed with pSB3T5 HRg5-cmk + gp5-RD-amp. WT T7 was amplified in the aforementioned cells and diluted 100-fold before being plated twice on different biological replicates of BW25113 Δ*cmk* gp5-RD-amp (Figure 3.14a). 4 large plaques were seen in total (average PFU/ml = 2000). An isolated plaque was picked and amplified in BW25113 Δ*cmk* gp5-RD-amp before plating on said cells and BW25113 (see Figure 3.14b). Only high MOI killing was observed on BW25113. Similar to before, an isolated plaque from the BW25113 Δ*cmk* gp5-RD-amp plate was amplified and screened for recombinant phage (Figure 3.14c). Bands observed correspond with the expected sizes. The amplified phage (hereto referred to as T7 Δ*g5*::*cmk*) was plated again on BW25113 Δ*cmk* gp5-RD-amp and screened for WT phage on BW25113 (Figure 3.14d). Again, only high MOI killing was observed on BW25113 (on undiluted and tenfold diluted plates) and so, the limit of detection was 10^3 PFU/ml for this assay. Next-generation sequencing of T7 Δ*g5*::*cmk* indicated that it's genome matches the reference sequence NC_001604$^+$ except for the *cmk* insertion removing *g5*.

In a similar fashion, T7-ReRb Δ*g5*::*cmk* was constructed by first amplifying the progenitor phage T7-ReRb in BW25113 pSB3T5-HRg5-cmk + Δ*cmk* pET24a-gp5-RD-amp cells, and diluting 100-fold before plating on BW25113 Δ*cmk* pET24a-gp5-RD-amp (Figure 3.15a). The isolated plaque was screened for recombinant phage (Figure 3.15b) and then was amplified in BW25113 Δ*cmk* pET24a-gp5-RD-amp. Sanger sequencing of the gene regions 5, 11, 12 and 17 (Figure 3.15c) show that T7-ReRb Δ*g5*::*cmk* differs from the progenitor phage. The Qimron mutations remain [87], as does the I684F mutation in gp12, but the L116I mutation in gp17 has reverted and the G479R mutation in gp17 was not previously seen.

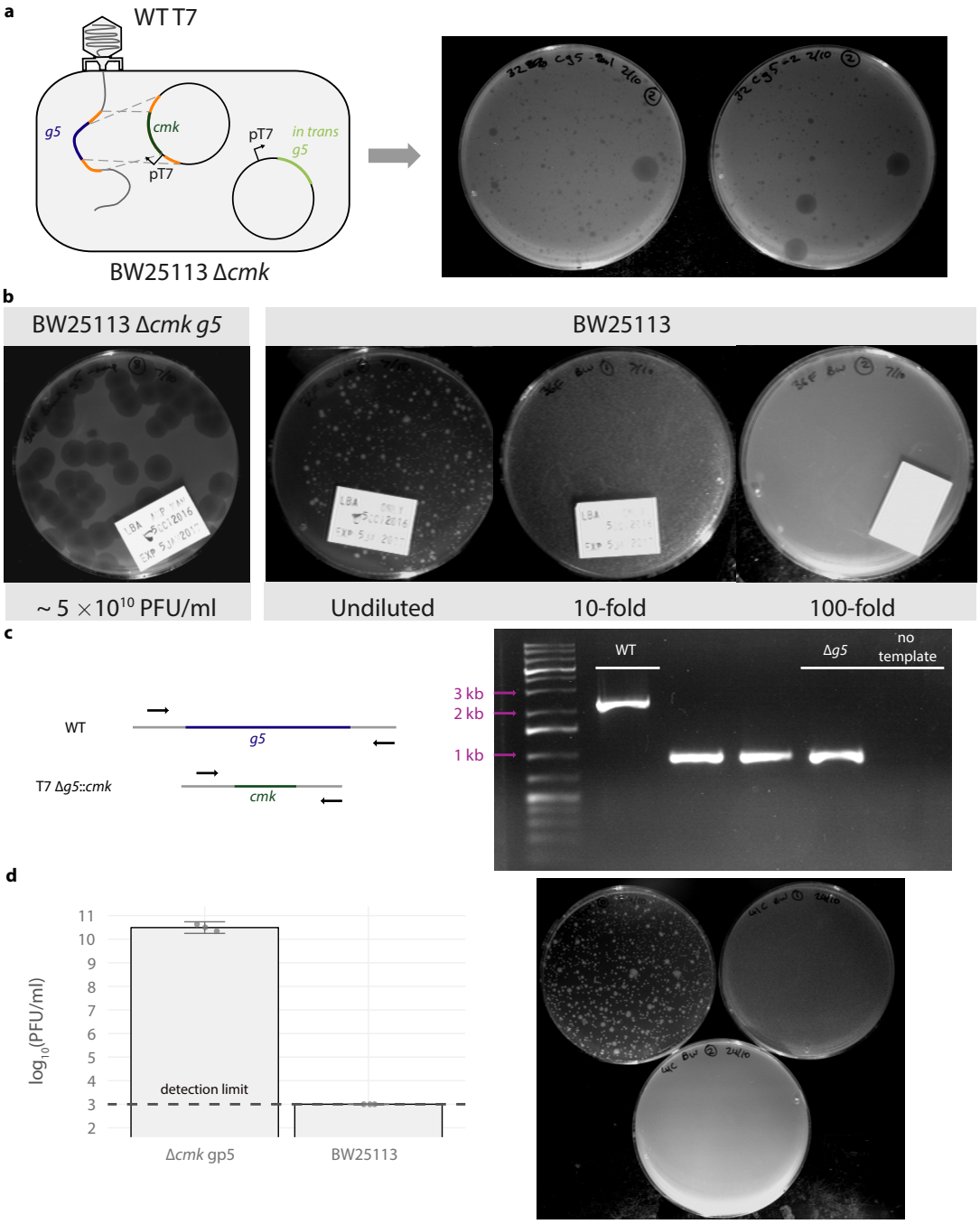


Figure 3.14: a) WT T7 amplified in Δ*cmk* pSB3T5-HRg5-cmk + pET24a-gp5-RD-amp cells are plated on Δ*cmk* pET24a-gp5-RD-amp (i.e. Δ*cmk g5*). Large plaques contain recombinant phage [87]. b) One such plaque is amplified in Δ*cmk g5* and re-plated as shown. No evidence of WT T7 observed. c) A plaque picked from the Δ*cmk g5* plate in panel b, is amplified and screened for T7 Δ*g5*::*cmk* phage. d) T7 Δ*g5*::*cmk* plated on Δ*cmk g5* and BW25113. Bar plots of geometric mean from 3 biological replicates. Errors are given as 95% confidence intervals. Raw data overlay the plot. Only high MOI killing was observed on BW25113 (i.e. no plaques - see representative replicate on the right).

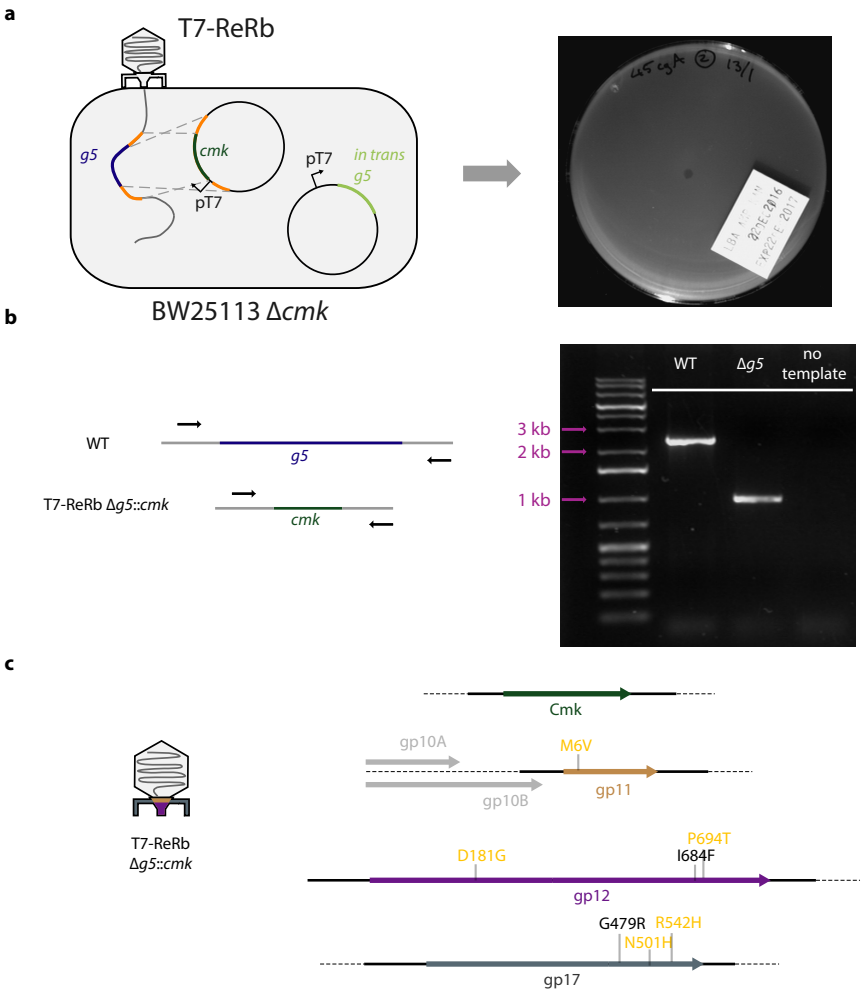


Figure 3.15: a) T7-ReRb amplified in Δ*cmk* pSB3T5 HRg5-cmk + pET24a-gp5-RD-amp cells are diluted 100-fold and plated on Δ*cmk* pET24a-gp5-RD-amp (i.e. Δ*cmk g5*). One small plaque was observed. b) This plaque is screened for T7-ReRb Δ*g5*::*cmk* phage by PCR. c) Sanger sequencing of gene regions 5, 10, 11, 12 and 17 following amplification of the plaque in Δ*cmk g5* cells. Expected Qimron mutations are present [87] and coloured in orange. The L116I mutation in gp17 has reverted from the progenitor phage T7-ReRb. The G479R mutation was not observed previously.

3.2.6 Vectors used to increase the mutation rate of T7 Δ*g5* phage *in vivo*

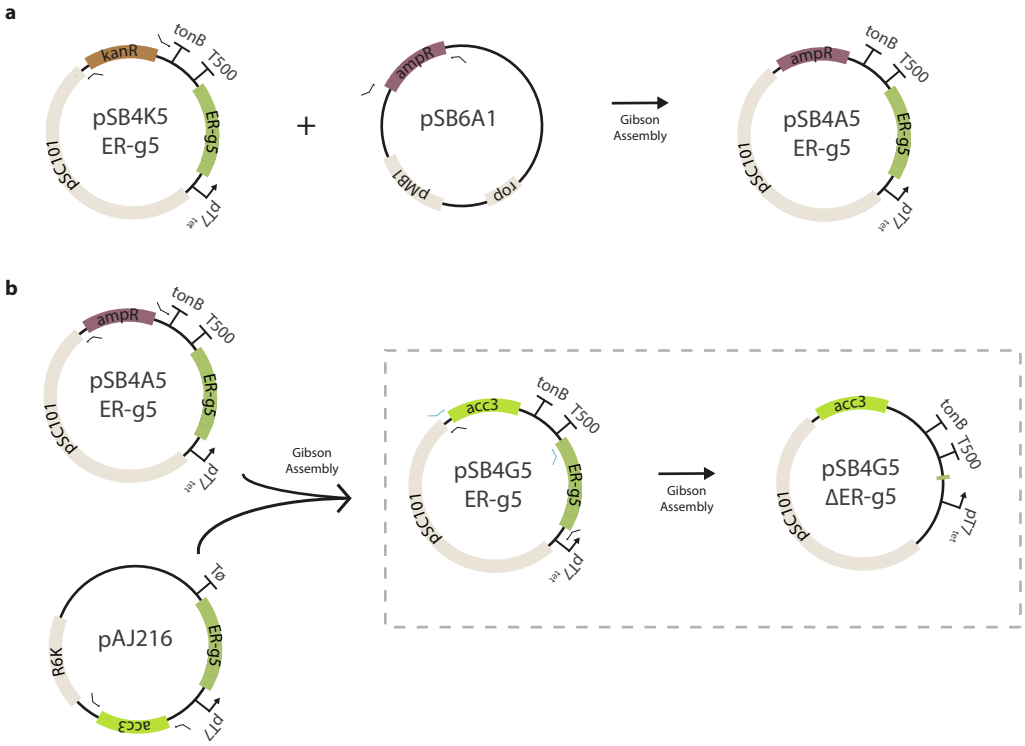


Figure 3.16: a) Construction of error prone g5 plasmids for mutagenising T7 phage during the directed evolution experiments. a) The ampR cassette from pSB6A1 is used to replace the kanR cassette of the pSB4K5 ER-g5 plasmid. b) In turn, the ampR cassette was replaced with *acc3*, conferring gentamicin-resistance, to construct pSB4G5 ER-g5. The ER-g5 gene in this plasmid was severely truncated (leaving only 6 amino acids) to yield pSB4G5 ΔER-g5. The boxed vectors appear frequently in chapter 4.

All the constructs depicted in Figure 3.16 were cloned using Gibson assembly (see Methods 5.3.6). pSB4A5 ER-g5 was constructed by amplifying the backbone of the inherited pSB4K5 ER-g5 and the ampR cassette from pSB6A1 using the primer pairs oPM203/4 and oPM205/6 respectively. pSB4G5 ER-g5 was made by amplifying the backbone of pSB4A5 ER-g5 and *acc3* from pAJ216 using oPM247/8 and oPM249/50 respectively. Finally, pSB4G5 ΔER-g5 was made by amplifying

pSB4G5 ER-g5 in 2 fragments (primer pairs oPM251/52 and oPM253/54) and retaining only 6 amino acids of the ER-g5 gene.

3.3 Discussion

3.3.1 Observation of putatively co-infecting T7

Co-infection of T7 phage has been previously observed in the literature, most of these experiments arising from simultaneous or sequential infection of *E. coli* with T7 amber mutants (phage with premature TAG stop codons in particular genes) [162]. Primary infections of T7 absent one gene, followed by secondary infection of T7 possessing that gene, will result in co-infection, so long as the delay between primary and secondary phage addition is not equal to or in excess of 7 minutes ([163, 164], see Table 1 in both papers). Nevertheless, secondary infection of a cell by T7 is impacted by the primary infection. For instance, *in vitro* alkylation of the T7 DNA has been shown to inhibit *E. coli* RNAP, which transcribes the early genes of T7. If the primary infecting T7 is alkylated first, then the number of infective centers[3] recovered (following secondary phage addition after 5 minutes) increases by $\approx$ 50-fold [164][4]. Similarly, the addition of chloramphenicol at sublethal concentrations inhibits protein synthesis by disrupting *E. coli* ribosomes. Infusing cells with chloramphenicol 5 minutes prior to the addition of primary infecting, and after another 5 minutes, secondary infecting, T7 phage, increases the number of infective centers by $\approx$ 21-fold [163]. Taken together, these results suggest that early T7 protein synthesis in the primary infecting phage plays a role in preventing some

[3]Infective centers are a measure for the number of productive phage infections
[4]Table 1, average for Line 2 divided by average for Line 1

secondary phage infections resulting in co-infection [164]. Karska-wysocki *et al.* [164] acknowledge that there are studies which contradict this claim, but posits that their results differed since they were generated with a shorter pre-incubation of chloramphenicol (2-3 minutes [165, 166] vs 5 minutes). Aksiyote Benbasat *et al.* [163] also argue that, when excluded, secondary phage still adsorb to primary infected cells, but are not capable of successful DNA ejection. *E. coli* RNAP is understood to be involved in ratcheting the T7 genome into the host, so although this is not inhibited by T7 until the expression of *g2*, it is probably true that there is some window early in the infection process in which co-infection of *E. coli* by T7 is more likely.

3.3.2 Modelling T7 co-infection

$\mathbf{MOI_{input}}$ model

The model based on $\mathrm{MOI_{input}}$, and in turn, the Poisson distribution, makes the following assumptions:

1. phage infecting cells do so independently of one another (i.e. without impacting the probability of other phage infection events)

2. phage infection events do not occur at the same time;

3. the average rate at which infection events take place is constant (but unspecified) over the duration;

4. more than one phage can infect a given a cell;

5. the probability of an infection event is proportional to the length of time

allowed for it to occur;

6. the CFU count does not increase or decrease over the time period (so no cell growth or death).

The first assumption to note is that phage infection events clearly can occur at the same time (violating assumption 2). Nevertheless, the assumption is only retained to make the probability calculations more straightforward, and is unlikely to effect the dynamics at such short timescales. Also, assumption 4 is the superinfection/co-infection assumption, which we have to assume to be true for otherwise this model would be redundant.

Most of these assumptions are reasonably justifiable whilst the phage is incubated with the cells at room temperature (RT) for 10 minutes (see Methods 5.4.1 and Figure 3.8). We consider first the viability of the assumptions during that initial incubation period. It is true that T7 adsorption to cells, especially at high MOIs, can deform the surface of the cell membrane [167], and subsequently could have an impact on the rate successive binding events (effecting both assumptions 1 and 3). Assumption 5 is often seen in phage ODE modelling when it is assumed that the rate of infection is proportional to the rate of adsorption of phage to cells [47, 168–173]. A consequence of this assumption is that for fixed CFU/ml, the average expected time taken for a given phage to bind to a cell is $\ln(2)/(kS_0)$, where S_0 is the initial CFU/ml.[5] Regarding assumption 6, during the initial 10 minute incubation at RT, the CFU count is unlikely to change (either due to cellular growth, or lysis by phage, as a result of the temperature and the time

[5]This result is acquired by finding t when 50% of the phage population is bound to cells i.e. solving $1 - e^{-kS_0t} = 0.5$ for t.

allowed for lysis).

Past the initial incubation, however, when the mixed phages and cells are plated with soft agar for overnight incubation at 37°C, cells will have both the time and conditions to replicate, and so will form lawns if not otherwise impeded by predators (i.e. phage). This of course means that phage have further opportunities to co-infect cells, once plated, though the soft agar would hinder diffusion, and subsequently reduce the frequency of collisions (rendering assumptions 3,5 and 6 untrue). This model can not take this into account, though the MOI_{actual} model does attempt to force a resolution to it.

MOI_{actual} model

Since MOI_{actual} can vary with time (leading to a non-homogenous Poisson process - see Appendix B.2) the average rate at which phage infection events occur is no longer constant (contrary to assumption 3 of the MOI_{input} model). The replacement of MOI_{input} for MOI_{actual} also necessitates some additional assumptions, namely:

- the rate of adsorption and infection of cells by phage is proportional to the product of their concentrations (law of mass action kinetics) and can be derived empirically
- phage only infect cells during initial RT incubation period (for $0 < t \leq 10$).

The first of these is already implicitly assumed for the MOI_{input} model, when considering the frequency at which phage infect cells. That being said, the MOI_{actual} model does allow the rate of adsorption to be varied in accordance with the experimental data collected by others [46, 81, 172]. The second assumption is not true.

That being said, the number of additional co-infected cells found during extended 37°C incubation will quickly diminish with time, given that phage after this period will be incubated with growing cells - thus increasing the number of phage-cell encounters, but likely decreasing the number of phage that end up infecting the same cell. Of course, burst cells that produce phage progeny during this period will continue to infect, and co-infect, other growing cells during this period, but these instances will still arise within a single plaque (and so, do not need to be considered in the modelling).

It should be noted that the 67-fold reduction in plaques seen in Figure 3.7d is strongly impacted by the small number of plaques in the denominator (in this case 2). Observation of even a single additional plaque would bring this down to roughly 45-fold reduction (and 2 additional plaques, to roughly 30-fold reduction). In order to decrease ν by 10-fold, one equivalently needs to find justification for reducing the product of kt by 10-fold. To this author's understanding, there does not exist literature testing T7 adsorption rates for various temperatures. However, observations of the Podoviridae phage δ infecting *Pseudomonas aeruginosa* showed a small linear reduction in adsorption rate with decreasing temperature (roughly 1.3-fold reduction between 37 °C and 23 °C (RT) [174]). This would leave a 10/1.3, so roughly 7.7-fold reduction, needed from elsewhere, i.e. $t = 1.3$ minutes, reflecting a scenario where T7 would be able to infect *E. coli* within the first 1.3 minutes of co-culturing only. In this timeframe, $\approx 18\%$ of the phage population would still have enough time to infect a 6.7×10^7 CFU/ml culture (with $k = 2.3 \times 10^{-9}$). Reasons for a curtailed window for infection could include some sort of time-limited superinfection exclusion mechanism [163, 164] (as mentioned in section 3.3.1). These mechanisms though appear to follow a 3-step process, where initially co-

infection seems to be unrestricted, followed by a period of proportional exclusion, and then complete exclusion after 7 minutes. An accurate appraisal of this stepped exclusion process would require modelling of infected cells at an individual level (since each cell's openness to secondary phage infection would depend on how long that cell had been primary infected) such as in agent-based modelling. This being said, later work by Grigonyte in her doctoral thesis [125] found that individually unproductive T7 phage could reliably co-infect a population of cells grown in a plate reader even if the secondary phage addition was as long as 90 minutes after the initial infection - casting some doubt as to whether co-infection is ever really truly restricted. With this in mind, it's possible that alternative mechanisms, like those invoking the spatial structure of bacterial lawns [175, 176] are needed to appropriately model this phenomena. Alternatively, in Appendix B.3, we explore the theoretical distribution of the results we could have observed if we did further experiments. In these simulations, the probability of observing at least a 67-fold reduction in plaques is $\approx 12\%$ if we assume that the true reduction in plaques is 30-fold. Further experimentation could then determine whether the model is in fact correct, and that the observed reduction in plaques recorded was more extreme than would be expected given that co-infection was occurring.

3.3.3 Characterisation of T7 co-infection

In hindsight, the observation of phage genomes which had recombined *g5* into *g10* was none too surprising. The homology regions found on the pET24a-gp5 vector for such a recombination event were sizable (and not too much shorter in total than the lengths used when deliberately engineering phage). Lee *et al.* as part of the Richardson Lab, which provided our lab with pET24a-gp5, published

work complementing T7 Δ*g3* with either pET24a-gp3 or pET28-gp3 (the pET28 backbone has the same homology regions) [177]. Despite not being the same constructs used here (where we complemented T7 Δ*g5* phage with pET24a-gp5) the pET vector system uses a consistent backbone design, with the same ϕ10 promoter and Tϕ terminator flanking the multiple cloning site, and would have undoubtedly led to the same co-infection phenomena. Table 2 of that work [177] notes the observation of small plaques when T7 Δ*g3* infects *E. coli* in the absence of a complementing gp3, which could have been the result of co-infecting phage particles.

By removing the 124 bp homology region in the Tϕ terminator from the pET24a-gp5 vectors, the only recombination pathway remaining was a single crossover event with the ϕ10 promoter or upstream RBS region. This would incorporate the entire 7.37 kb plasmid into T7 Δ*g5* ($\approx$ 38.9 kb long) replacing the *g10* region in the process ($\approx$ 1.3 kb long) which exceeds the maximum amount of DNA that T7 can package ($\approx$ 41.2 kb [57]).

3.3.4 Re-construction of T7 Δ*g5* and T7-ReRb Δ*g5* phage

Obviously, the appearance of phage plaques in the absence of complementing gp5 (whether due to a co-infection or not) is of concern to those attempting to build a evolution experiment that relies on being able to control the replication of the host. Even if said phages existed in low titre, and with decreased fitness (i.e. the ability to produce progeny in a timely fashion) the planned evolution experiments would encourage the adaptation of phages that could utilise gp5. Subsequently, it was critical to make attempts to prevent such occurrences. No evidence of co-infected

plaques was observed in the absence of complementary gp5 in the construction of both T7 Δ*g5* phage and T7-ReRb Δ*g5* phage engineered and amplified with the pET24a-gp5-RD-amp plasmid. This further lends weight to the conclusion that the Tϕ homology to the downstream of *g10* (in combination with the ϕ10/RBS homology) was for responsible both the T7 Δ*g5 g10::g5* recombinants, and the subsequent co-infection phenomena.

3.3.5 Vectors used to increase the mutation rate of T7 Δ*g5* phage *in vivo*

The ER-g5 gene product (ER-gp5) has 5 amino acid mutations when compared with the WT gp5. D5A and E7A mutations occur in the exonuclease domain [178] and the Y64C, F120L, S399T mutations are from Söte *et al.* [152]. The effect of these mutations has been characterised separately, but never explicitly characterised in combination. ER-gp5 with all 5 mutations does successfully complement T7 deficient in *g5 in vivo* (compare Figure 3.14d with Figure 4.5a) and since the mutations occur in isolated domains in the protein, it is possible that they have an additive epistatic effect on the error rate. The base error rate was therefore conservatively estimated to be 55-fold improvement on WT (i.e. no gain from D5A D7A mutations) with a best-case scenario being a 93-fold improvement. That being said, no attempt was made to characterise the error rate of this polymerase.

The inherited pSB4K5-ER-g5 plasmid confers kanamycin resistant (see Figure 3.16), making it incompatible with the Keio collection strains (Δ*waaC*, Δ*ompF*, etc). Subsequently, the cassette was replaced an ampR cassette, and later, a gentamicin cassette (owing to concerns with beta-lactamase build up in continuous

culture systems). pAJ216 was not cloned until some time after pSB4A5-ER-g5 was made, but also crucially contained both of the regulatory elements shown to recombine with T7 in chapter 2 (pT7 and the Tϕ terminator). These circumstances justified the cloning procedure followed.

3.4 Conclusion

Attempts to use the provided $\Delta g5$ phage and complementing gp5 plasmid to start the designed evolution experiments were hindered by the observation of small plaques in cells that did not possess gp5 *in trans*, and which reduced in number more than 10-fold following a 10-fold dilution in the phage sample. A model was developed to explain how these reductions might have been possible, and the parameters under which different fold-reductions could be seen. Further work could have separately propagated each of these individually replication-deficient phage genomes and tested whether the model would stand up to more experimental evidence. Perhaps even varying the initial numbers of phage and cells, to test the models derived. In any case, the likely mechanism generating these small plaques was identified, and new vectors were cloned to obviate that mechanism. Subsequently, $\Delta g5$ phage were generated that did not demonstrate this phenomena. That, and the construction of vectors for diversifying the T7 genome *in trans*, laid the groundwork for the planned evolution experiments to follow in the proceeding chapter.

4 Directed evolution of T7 host-specificity

4.1 Introduction

In this final results chapter, we will seek to affect the receptor-tropism of T7 phage through directed evolution. As touched upon in section 3.1.3, we will subject T7 to serial passages of both positive and negative selection strains in an attempt to effect this change (see Figure 4.1). However, there are many parameters to consider in the design of these experiments. Leaving aside the effect of phage-bacteria kinetic parameters, the duration of incubation, the proportion of the phage population transferred to the next passage, and the ratio of the positive and negative selection strains, could each determine whether the desired phage phenotype evolves (see Figure 4.2). One approach to explore the parameter space is through empirical experimentation in the lab (using factorial designs [154] to hasten the search) and another is to simulate outcomes using mathematical modelling. In this work, we opt to use the results from the latter to guide the design of the experiments, and test those predictions experimentally.

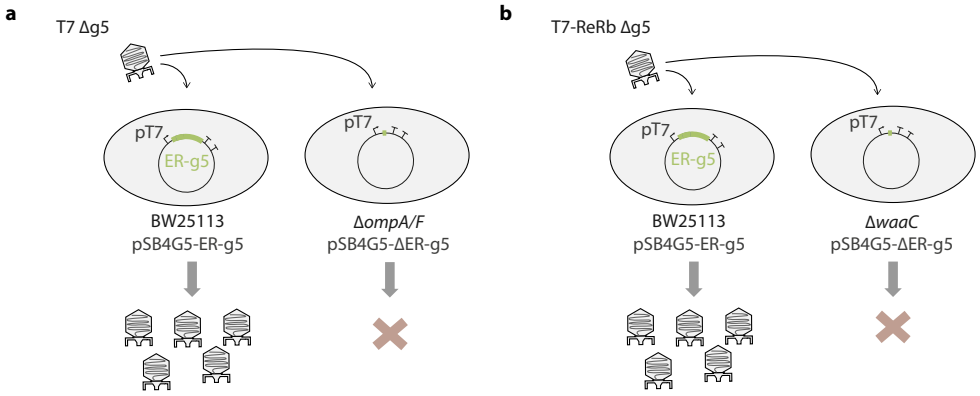


Figure 4.1: A schema presenting the planned directed evolution experiments. Panel a) is the main experiment where we will attempt to adapt T7 so that it solely utilises the OmpF receptor for infection. Panel b) describes the preliminary proof of concept experiment in which T7-ReRb will be evolved towards the WT T7 binding phenotype (that being predominately LPS-based).

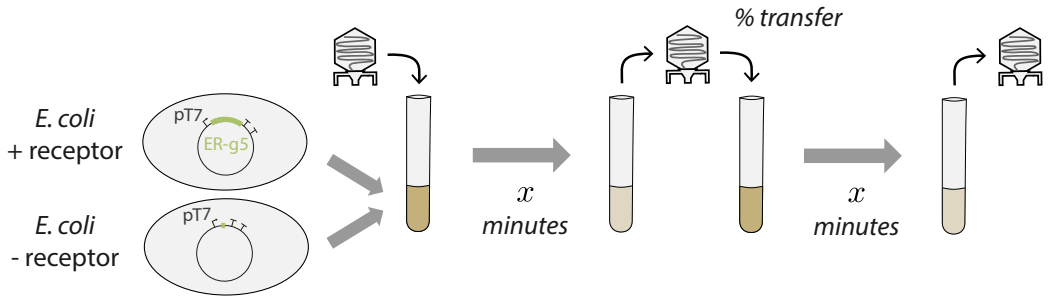


Figure 4.2: The serial passaging procedure for the planned directed evolution experiments. Phage are mixed with a pre-determined proportion of positive and negative selection cells. After incubation at 37 °C for a set period of time, the total phage population is harvested (using chloroform to break open any non-lysed cells) and a percentage of that population is transferred to a freshly grown co-culture of positive and negative selection cells. This procedure is repeated, with the aim of adapting the phage to the desired phenotype (see Methods 5.6 for details).

4.1.1 Mathematical modelling of phage

Phage kinetic parameters

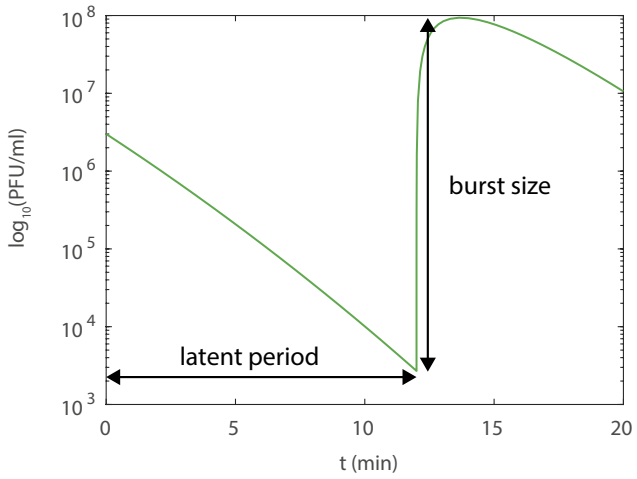


Figure 4.3: A simulated one-step phage growth curve. The phage kinetic parameters for latent period and burst size can be estimated from the curve as shown. In addition, the phage adsorption rate, k, can be estimated from the initial slope by the equation $k = \ln(P_0/P_t)/(C_0 t)$, where P_0 and C_0 are the initial PFU/ml and CFU/ml respectively, and P_t is the PFU/ml at some time t prior to cell lysis.

There are 3 main kinetic parameters estimated for phage which define their dynamics

over short timescales (see Figure 4.3). The adsorption rate, which determines the rate of reversible binding between phage and bacteria; the latent period, which describes the length of time from phage infection to cell lysis; and the burst size, which is the number of phage progeny released per lysed cell. It is typically assumed when modelling that T7 phage has a 1-step binding process, though it is thought that there are in fact 2 stages of T7 phage adsorption: reversible and irreversible binding [59]. Reversible binding is limited by the rate of diffusion of phages and cells, whereas irreversible binding occurs after reversible binding has brought the phage close to the cell surface, and initiates the process of DNA ejection. Choosing to describe adsorption as a 1-step process is both of mathematical and experimental convenience. Phage adsorption assay protocols typically measure the number of phage in the supernatant and subtract that from the initial PFU [81, 83, 171, 172, 179] subsequently producing a measure for the reversible adsorption rate. Though measuring the irreversible adsorption rate directly, is possible [170, 180, 181] we choose not to do so here. Over longer timescales, a 4th parameter, that describes the decay rate of phage (from the breakdown of their constituent proteins, for instance) can be become relevant as well, but this will not be considered when designing the models herein.

4.1.2 Dynamic phage-bacteria models

The interaction of phage and bacteria in a closed system lends itself nicely to mathematical modelling, in particular, modelling using differential equations whose solutions change with time. These equations avoid considerations about space, and so assume that particles move in a well-mixed environment so that they are as likely to interact with one particle as with another. This assumption is reasonable for

phage and bacteria in appropriately agitated liquid culture, but is violated in many other domains where such differential equation modelling is employed (such as in the gene regulatory networks in crowded and compartmentalised cells [182–184]). Differential equations have been used to model the dynamics of phage and bacteria, to this author's knowledge, since at least 1960 when Campbell formulated equations for a continuous culture system (chemostat) [168]. Given the breadth of literature, we will focus on those systems of direct relevance.

ODE systems

The idea of using non-replicative or non-permissive cells to negatively select against certain phage phenotypes is developed through the terminology of ecological traps or sinks in ecology. Dennehy *et al.* [185] present a model for infections of the RNA bacteriophage Φ6 in WT *Pseudomonas phaseolicola* and mutants that over expressed the pili phage receptor, but did not retract those pili to allow infection. They start with the assumption that cells are neither replicate or die, and derive the following differential equation for total phage population, P, in liquid culture,

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \frac{k_1 C_1}{k_1 C_1 + k_2 C_2} gP - \phi P.$$

Here C_1 is the number of phage-permissive cells and C_2 is the number of non-permissive (trap) cells; k_i represents the phage adsorption rate to strain C_i; and g and ϕ are the maximum phage growth rate and decay rate respectively. Dennehy *et al.* use this to find the proportion of trap cells which keep the phage population in equilibrium (i.e. when $\mathrm{d}P/\mathrm{d}t = 0$, so that phage are neither increasing or declining

in number). Said proportion of trap cells is then,

$$\frac{C_2}{C_1 + C_2} = \frac{g - \phi}{g - \phi\left(1 - \frac{k_2}{k_1}\right)}.$$

If the proportion of trap cells is greater than the right-hand side of this equation, then phage extinction (washout) results.

Delay differential equations (DDEs)

For explanatory purposes, let us envisage a phage which both infects hosts instantaneously, and releases phage progeny from a host every 10 minutes, without lysing the host. Mathematically, we could model the number of phage over time by,

$$\frac{\mathrm{d}P(t)}{\mathrm{d}t} = g\ P(t - 10). \tag{4.1}$$

where g is the phage reproduction rate and $P(t)$ is the number of phage at minute t (so that $P(t-10)$ is the number of phage at minute $t-10$). We call equation (4.1) a delay differential equation. Note that when $t = 0$, we need to know the value of $P(-10)$. This means that we have to provide information on the population of phage in the 10 minutes prior to the phage population at $t = 0$ in order to solve the equation for increasing t.

DDE systems

Cairns *et al.* [186] produce a simple DDE system to model the dynamics of *Campylobacter jejuni* cultured with an infecting virulent phage. In this model cells can be either susceptible or resistant to phage infection. The equations for

phage-susceptible cells, C, phage-resistant cells, R, phage-infected cells, I, and phage, P, are shown below,

$$\begin{aligned}
\frac{\mathrm{d}C}{\mathrm{d}t} &= aC \quad - rC \quad \underbrace{- kCP}_{\text{infection of cells by phage}} \\
\frac{\mathrm{d}R}{\mathrm{d}t} &= aR \quad + rC \\
\frac{\mathrm{d}I}{\mathrm{d}t} &= kCP \quad \underbrace{- kC(t-\tau)P(t-\tau)}_{\text{lysis of infected cells at time } t-\tau} \\
\frac{\mathrm{d}P}{\mathrm{d}t} &= \beta kC(t-\tau)P(t-\tau) - kCP - \phi P.
\end{aligned}$$

Here a is growth rate of phage-susceptible and resistant cells, r is the rate at which susceptible cells become resistant, τ is the phage latent period, β is the phage burst size (the number of progeny produced) and k and ϕ operate as before. This model forms the basis of the models we will establish in the results of this chapter, though we neglect the emergence of resistant mutants, as well as the rate of phage decay, due to the short passaging periods employed. As is the case in our experiments, Bull [47] develops a system to explore the evolution of phage in 2 heterogenous hosts populations (C_1 and C_2). First assuming that the cell population does not vary in size, he formulates,

$$\begin{aligned}
\frac{\mathrm{d}P}{\mathrm{d}t} = {} & \underbrace{\beta_1 k_1 C_1 P(t-\tau_1)}_{\text{phage production in } C_1} e^{-\phi_C \tau_1} - \underbrace{k_1 C_1 P}_{\text{phage infection of } C_1} \\
& + \underbrace{\beta_2 k_2 C_2 P(t-\tau_2)}_{\text{phage production in } C_2} e^{-\phi_C \tau_2} - \underbrace{k_2 C_2 P}_{\text{phage infection of } C_2} - \phi_P P
\end{aligned} \tag{4.2}$$

where the subscripts refer to the phage parameters as they relate to host C_1 or host C_2 , and ϕ_C and ϕ_P refer the rate of cell death and phage decay respectively. The $e^{-\phi_C \tau_i}$ term accounts for the loss of cells from the population to non-phage

related causes in the fated period between phage infection and phage lysis (see Beretta and Kuang for more detail [187]). As only P changes with time, Bull notes that the solution must be of the form $P(t) = P(0)e^{gt}$, where g is the phage reproduction (growth) rate. Substituting $P(t)$ into equation (4.2), he derives the following equation for g,

$$g = k_1 C_1 \left(\beta_1 e^{-\tau_1(g+\phi_C)} - 1\right) + k_2 C_2 \left(\beta_2 e^{-\tau_2(g+\phi_C)} - 1\right) - \phi_P.$$

Phages in this environment that avoid the host C_i are then selected for if $\beta_i e^{-\tau_i(g+\phi_C)} < 1$. Following from this work, Bull models the impact of serial passaging for singular phage and cell phenotypes in the presence of growing bacteria [172]. We effectively extend his work to multiple phage and cell phenotypes, though choose to not allow the lysis time to vary.

Mutation models

Levin *et al.* [173] use DDEs to investigate the effect of CRISPR-*cas* immunity to the coevolution of phage and bacteria. Of particular interest is that they add an element of stochasticity to their models by adding a mutant phage or bacteria to the population when a mutational threshold has been met. For instance, a CRISPR-immune phage mutant is produced in a given time interval δt with probability $(\beta - 1)\mu\,\delta t\, I(t - \tau)$, where μ is the mutation probability per phage per infection, and $I(t - \tau)$ is the number of infected cells τ minutes prior. We will use a similar strategy to allocate phage to new phage phenotypes in the directed evolution modelling to come.

4.1.3 Aims for this chapter

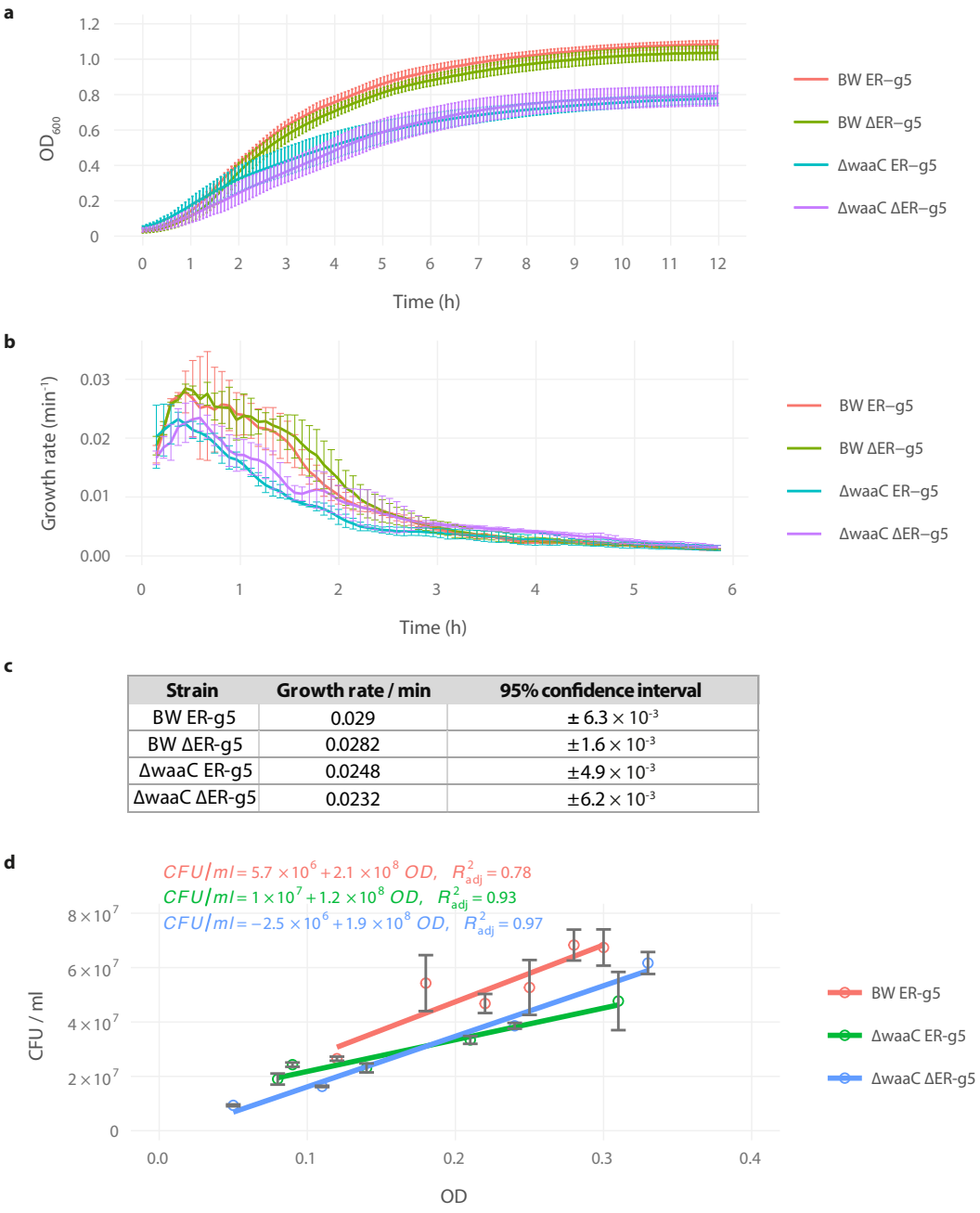
In this chapter, we aim to set up and perform directed evolution experiments to effect T7 receptor tropism. This will involve characterising both the strains and the phage to be used in the experiments. We aim to then mathematically model different directed evolution strategies to inform our experiments in the lab. We will start with a proof of concept experiment to show that we can effect the affinity of T7 to different LPS phenotypes, and then attempt to coerce an exclusive affinity to a chosen outer membrane protein.

4.2 Results

4.2.1 Characterisation of strains used in Δ*waaC* directed evolution experiments

BW25113 and Δ*waaC* were characterised following the transformation of pSB4G5 ER-g5 and pSB4G5 ΔER-g5 (see Figure 4.4). Growth curves for the Δ*waaC* strains plateau earlier than for the BW25113 strains, and do so consistently (Figure 4.4a). The mean growth rates for the BW25113 strains appear higher than in Figures 4.4b and 4.4c, but are within the 95% confidence intervals of the Δ*waaC* strains. BW25113 ER-g5 ranged from 4.0-6.6 $\times 10^7$ CFU/ml and Δ*waaC* ΔER-g5 from 2.9-5.3 $\times 10^7$, within an OD_{600} of 0.17-0.3.

Presented in Figure 4.5 are the data for T7-ReRb Δ*g5*::*cmk* challenged against BW25113 and Δ*waaC* transformed with pSB4G5 ER-g5. A Welch 2-sided t-test suggests that the geometric mean PFU/ml vs Δ*waaC* is significantly different from that observed on BW25113 as determined by plaque assay (significance measured at the 5% level, $p = 0.01$). The log-transformed EOPs determined using plaque assay and spot assays were both found to differ significantly at the 5% level from the null difference hypothesis (p values 0.01 and 0.017 respectively) as did the 2 methods from each other ($p = 0.04$).



Strain	Growth rate / min	95% confidence interval
BW ER-g5	0.029	$\pm 6.3 \times 10^{-3}$
BW ΔER-g5	0.0282	$\pm 1.6 \times 10^{-3}$
ΔwaaC ER-g5	0.0248	$\pm 4.9 \times 10^{-3}$
ΔwaaC ΔER-g5	0.0232	$\pm 6.2 \times 10^{-3}$

Figure 4.4: a) Growth curves for BW25113 and Δ*waaC* transformed with pSB4G5 ER-g5 and pSB4G5 ΔER-g5. Error bars are SD of 3 biological replicates. b) The growth rate per minute for each strain over the first 6 hours. Error bars are SD of 3 biological replicates. c) The calculated mean growth rate (see Methods 5.5.5) and 95% confidence intervals for each strain d) CFU assays for BW25113 ER-g5 and Δ*waaC* with either pSB4G5 ER-g5 or pSB4G5 ΔER-g5. The means of 3 technical replicates at varying OD values are linearly regressed for each strain. Error bars are the SD of 3 technical replicates.

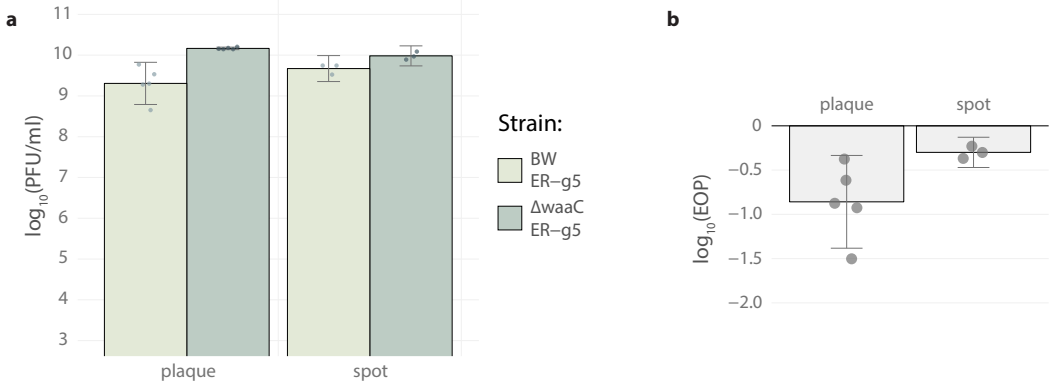


Figure 4.5: a) Plaques and spot assays for T7-ReRb $\Delta g5$::*cmk* challenged against BW25113 and $\Delta waaC$ cells transformed with pSB4G5-ER-g5. b) Here efficiency of plating (EOP) represents the PFU/ml recovered on BW25113 ER-g5 plates, divided by that recovered on $\Delta waaC$ ER-g5 plates. Bar plots are presented as the geometric mean of at least 3 biological replicates with 95% confidence intervals. Individual biological replicates are overlaid.

4.2.2 Simplified models for $\Delta waaC$ directed evolution experiments

Model design

The aim here is to model the behaviour of the desired phage in pairwise competition with another phage-binding phenotype in differing passage conditions. In the first instance, we will define the desired phage phenotype for the evolution experiment to be a T7 phage with a maximum affinity for LPS (chosen to be 10^8 ml/min) and no affinity for $\Delta waaC$ cells. Passage conditions will be varied by altering the relative amounts of BW25113 ER-g5 cells and $\Delta waaC$ ΔER-g5 cells.

We start with the simplest instance where the competing phage with varying binding affinities is simulated in the absence of the desired phage. A schematic of this model is depicted in Figure 4.6 and the equations governing the system and

presented below,

$$\frac{\mathrm{d}C_1}{\mathrm{d}t} = a_1 C_1 - k_1 C_1 P$$
$$\frac{\mathrm{d}C_2}{\mathrm{d}t} = a_2 C_2 - k_2 C_2 P$$

$$\frac{\mathrm{d}I_1}{\mathrm{d}t} = k_1 C_1 P - k_1 C_1(t-\tau) P(t-\tau)$$
$$\frac{\mathrm{d}I_2}{\mathrm{d}t} = k_2 C_2 P - k_2 C_2(t-\tau) P(t-\tau)$$

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \beta k_1 C_1(t-\tau) P(t-\tau) - k_1 C_1 P - k_2 C_2 P - k_1 I_1 P - k_2 I_2 P\,.$$

See Table 4.1 for a description of the variables and parameters for these equations.

Table 4.1: Variable/parameter notation for the $\Delta waaC$ pairwise competition model

Notation	Description	Units
C_i	Phage-susceptible cell strain i	ml^{-1}
I_i	Phage-bacteria complex (infected cell strain i)	ml^{-1}
P	Free Phage	ml^{-1}
a_i	Growth rate of cell strain i	min^{-1}
k_i	Binding rate of phage to susceptible cell strain i	$\mathrm{ml\ min}^{-1}$
τ	Latent period of phage	min
β	Burst size of phage (number of phage progeny)	[dimensionless]

We list here some modelling assumptions:

- The culture of cells and phage is well-mixed
- At most one phage can infect a given cell (i.e. no superinfection/coinfection)
- Neither the latent period or the burst size adapt to the environment, and adaptation affects the rate of phage binding to cells only

- The process of cell lysis does not prevent further phage adsorption to the cell.

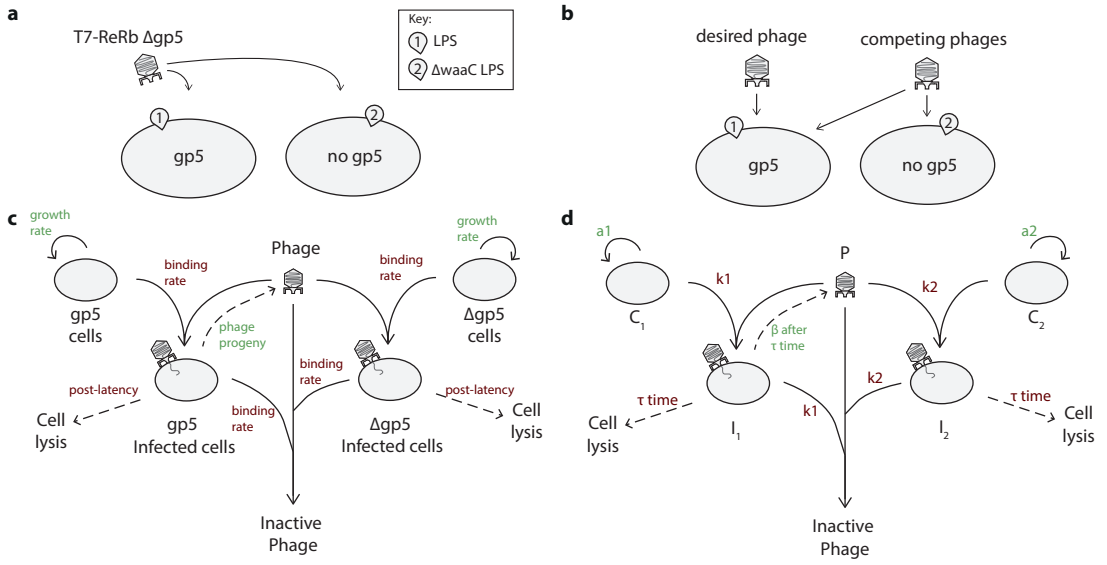


Figure 4.6: a) Abstraction of the $\Delta waaC$ evolution experiment. Phage binds to both the WT LPS and the $\Delta waaC$ truncated-LPS cells at rates k_1 and k_2 respectively, but only produces progeny when infecting cells complemented with gp5. b) In our simulations, we match the desired phage which has affinity for LPS only and a competing phage which binds, at different rates, to both WT and $\Delta waaC$ LPS. c) Schematic for the $\Delta waaC$ phage pairwise competition model in the absence of the desired phage (a.k.a. the washout model). Note that phage can also bind to already infected cells, further depleting the pool of free phage. d) Again, but using the same notation as the model.

In the presence of the desired phage, so when competition between 2 phage phenotypes is being simulated, we need to add in terms and equations to track the dynamics of both phage populations. This means that there are now 4 separate infected cell states to consider: those that arise when the competing phage infects either BW25113 or $\Delta waaC$, and similarly when the optimal phage does so as well. We notate cell strain i being infected by phage phenotype j by I_{ij} . Similarly, each phage phenotype has a separate binding rate to each cell strain. We denote this as k_{ij} i.e. phage phenotype j adsorbs to cell strain i at rate k_{ij} . The equations for this system are then,

$$\frac{\mathrm{d}C_1}{\mathrm{d}t} = a_1 C_1 - k_{11} C_1 P_1 - k_{12} C_1 P_2$$
$$\frac{\mathrm{d}C_2}{\mathrm{d}t} = a_2 C_2 - k_{21} C_2 P_1 - k_{22} C_2 P_2$$

$$\frac{\mathrm{d}I_{11}}{\mathrm{d}t} = k_{11} C_1 P_1 - k_{11} C_1(t-\tau) P_1(t-\tau)$$
$$\frac{\mathrm{d}I_{12}}{\mathrm{d}t} = k_{12} C_1 P_2 - k_{12} C_1(t-\tau) P_2(t-\tau)$$
$$\frac{\mathrm{d}I_{21}}{\mathrm{d}t} = k_{21} C_2 P_1 - k_{21} C_2(t-\tau) P_1(t-\tau)$$
$$\frac{\mathrm{d}I_{22}}{\mathrm{d}t} = k_{22} C_2 P_2 - k_{22} C_2(t-\tau) P_2(t-\tau)$$

$$\begin{aligned}\frac{\mathrm{d}P_1}{\mathrm{d}t} &= \beta k_{11} C_1(t-\tau) P_1(t-\tau) - k_{11} C_1 P_1 - k_{21} C_2 P_1 \\ &\quad - k_{11} I_{11} P_1 - k_{11} I_{12} P_1 - k_{21} I_{21} P_1 - k_{21} I_{22} P_1 \\ \frac{\mathrm{d}P_2}{\mathrm{d}t} &= \beta k_{22} C_2(t-\tau) P_2(t-\tau) - k_{12} C_1 P_2 - k_{22} C_2 P_2 \\ &\quad - k_{12} I_{11} P_2 - k_{12} I_{12} P_2 - k_{22} I_{21} P_2 - k_{22} I_{22} P_2 \,.\end{aligned}$$

Note that the equations for P_1 and P_2 are defined symmetrically, so it does not matter which is chosen to be the desired phage, and which one is chosen to be the competing phage In the simulations, we arbitrarily take P_1 to be the desired phage, and set the binding rate, k_{21} equal to zero (k_{21} being the binding rate of P_1 to the $\Delta waaC$ strain, C_2).

Qualitative predictions

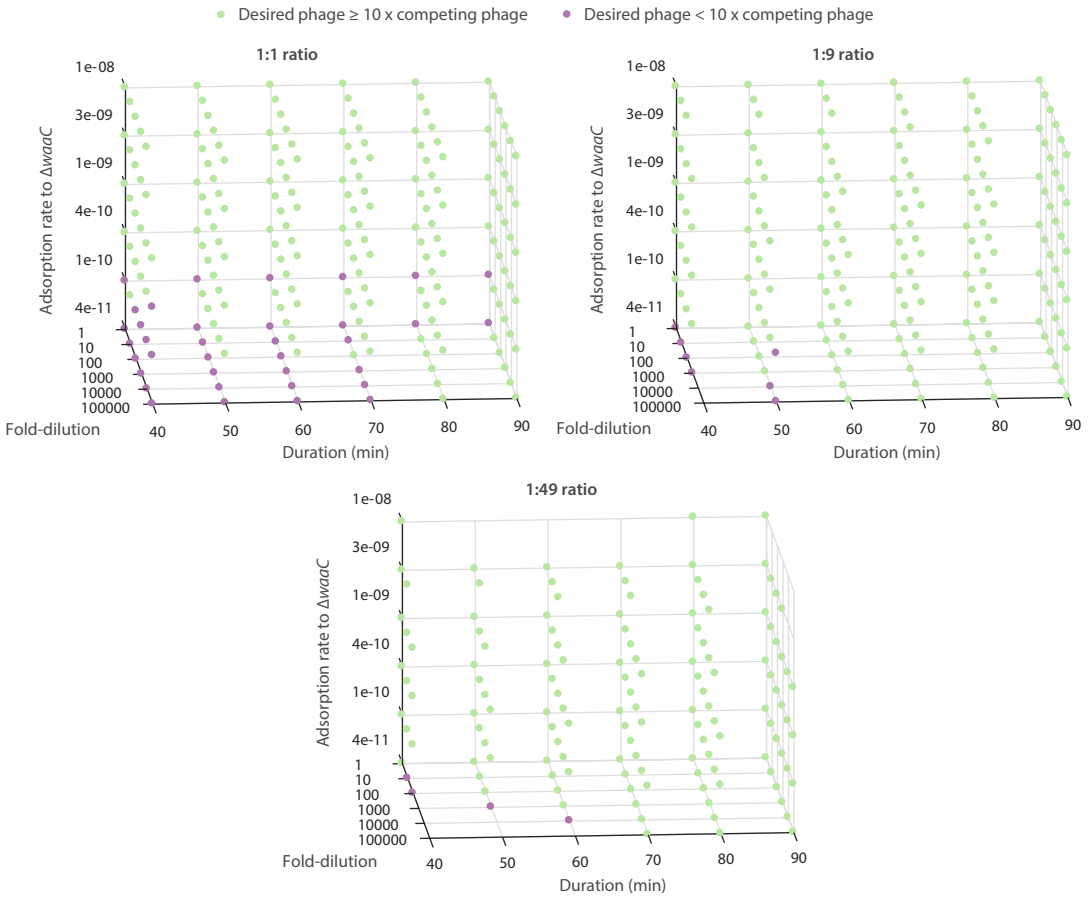


Figure 4.7: Effect of passage duration, and the fold-dilution between passages, on the pairwise competition model for the Δ*waaC* directed evolution experiment. Pairwise competition is between the desired phage (LPS affinity 10^{-8} ml/min and no affinity for Δ*waaC*) and competing phage phenotypes (LPS affinity 10^{-8} ml/min, but Δ*waaC* afinity varies). Passage conditions are notated as the ratio of (BW25113 ER-g5):(Δ*waaC* ΔER-g5). Data where the competing phage washes out in the absence of the challenging desired phage (the washout model) is not plotted. Simulations are performed for 20 passages with $a = 0.026$ min^{-1}, $\beta = 100$, $\tau = 12$ min, initial PFU/ml $= 3 \times 10^6$, initial CFU/ml $= 5 \times 10^7$. Washout is defined as the PFU/ml dropping below 10^4 after 20 passages.

The effects of varying the incubation time (duration) for each passage and of varying the fold-dilution applied when transferring the phage population to the next passage, are simulated in Figure 4.7. The duration of each passage does not

appear to have much impact on the whether the desired phage occupies more than 90% of the total phage population. Increasing the fold-dilution can result in higher likelihood of phage washout, as might be expected. Considering these results, combined with the an inclination to avoid co-evolution of bacteria within each passage, and a desire to avoid washout of the total phage population, the duration of each passage and the fold-dilution was fixed at 60 minutes and 100-fold respectively.

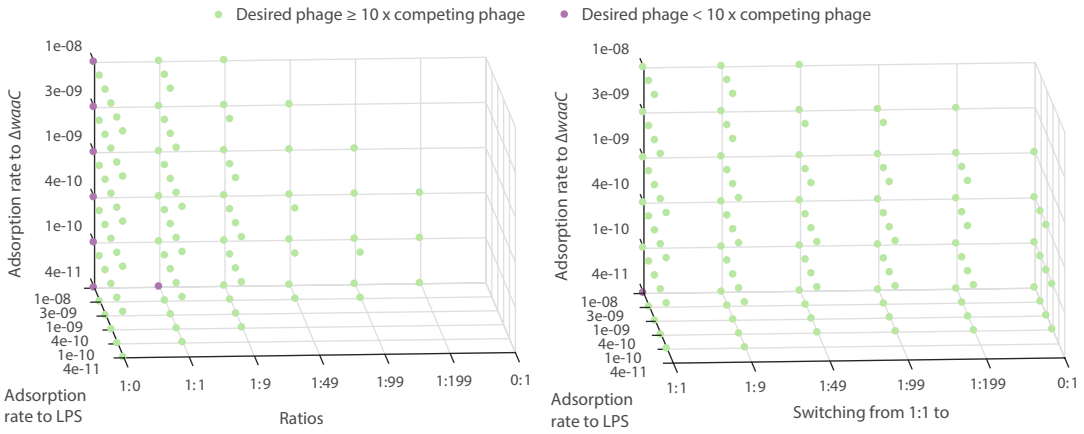


Figure 4.8: Pairwise competition between the desired phage (LPS affinity 10^{-8} ml/min and no affinity for $\Delta waaC$) and competing phage phenotypes. Passage conditions notated as the ratio of (BW25113 ER-g5):($\Delta waaC$ ΔER-g5). Data where the competing phage washes out in the absence of the challenging desired phage (the washout model) is not plotted. Simulations performed for 20 passages with $a = 0.026$ min^{-1}, $\beta = 100$, $\tau = 12$ min, initial PFU/ml $= 3 \times 10^6$, initial CFU/ml $= 5 \times 10^7$, a passage duration of 60 minutes and 1% of the phage population transferred between each passage. Washout is defined as the PFU/ml dropping below 10^4 after 20 passages.

BW25113 ER-g5 only cells show no ability to select for the desired phage (one with affinity for LPS only, see Figure 4.8). A 1:1 ratio of BW25113 ER-g5 to $\Delta waaC$ ΔER-g5 is able to select for the desired phage over a competing phage, and washes out few of the possible competing phage phenotypes. Hence, the 1:1

ratio will be the first that we run the directed evolution experiment with. The model also suggests that a 1:9 ratio should further select for a phage with weaker binding affinity to $\Delta waaC$, so long the phage population does not washout. It also predicts that the switching conditions will succeed too (Figure 4.8b) and that increasing the stringency of selection will be less detrimental to competing phage phenotypes as in the pure ratio conditions.

4.2.3 Regression of T7-ReRb $\Delta g5::cmk$ to WT T7 cell-binding affinities

EOP characterisation

First a note, as these are evolution experiments, we would expect there to be an associated variation between independent replicates passaged in separate conditions. However, only single EOP replicates were recorded for individual serial passaging replicates. This means that there will also be variation associated with the spot assay data appearing in these measurements, which can not be uncoupled from the variation associated with the passaging procedure. Nevertheless, we can make attempt to make general assertions where the fold change is sizeable enough to be confident of a difference.

The procedures used to passage T7-ReRb $\Delta g5::cmk$ are explained in detail in Methods 5.6. We briefly overview the key aspects here for clarity. Phages were passaged for $\approx$ 60 minutes and 1% of the phage population was transferred between each passage. In the main assay of the 1st directed evolution experiment, a 1:1 volume by volume ratio of BW25113 ER-g5 and $\Delta waaC$ ΔER-g5 was used

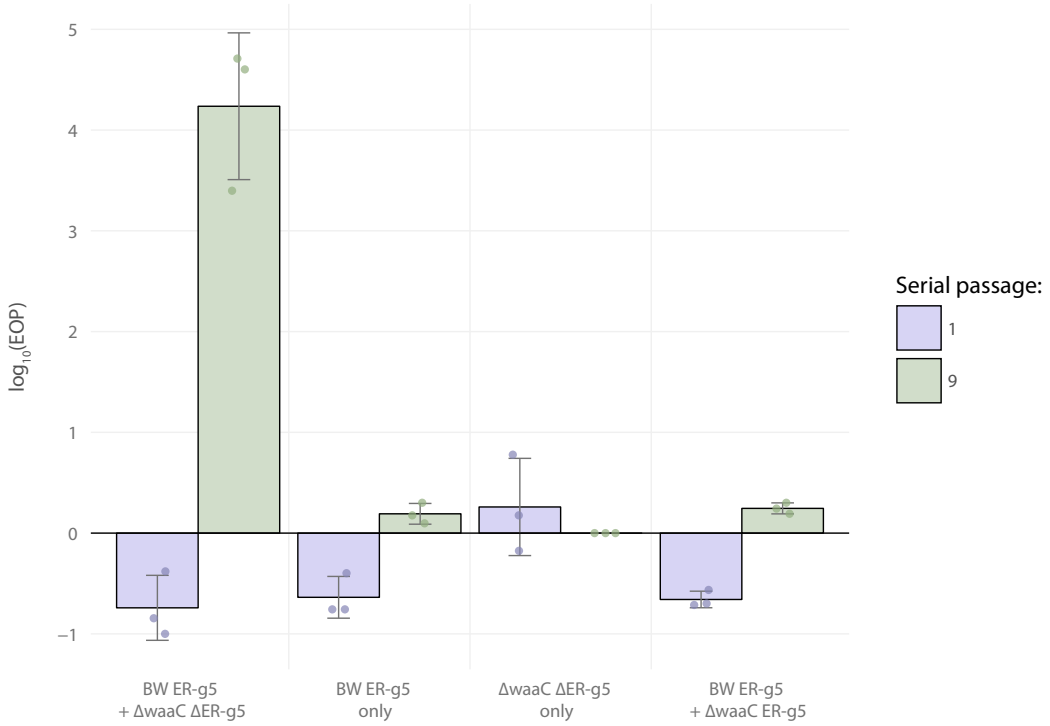


Figure 4.9: Spot assay EOP data from the first 9 passages of the Δ*waaC* directed evolution experiment. T7-ReRb Δ*g5*::*cmk* was passaged independently in 4 separate 1:1 ratios: BW25113 ER-g5 to Δ*waaC* ΔER-g5 cells; 1:1 ratio of BW25113 ER-g5 to LB media; 1:1 ratio of Δ*waaC* ΔER-g5 to LB media; and a 1:1 ratio of BW25113 ER-g5 to Δ*waaC* ER-g5 cells. Data points are independent passaging replicates and are overlaid. EOP is defined as PFU/ml recovered on BW25113 ER-g5 plates, divided by that recovered on Δ*waaC* ER-g5 plates. The bars and errors represent, respectively, the means and standard deviations of the log-transformed EOPs.

throughout. The control assays were built around this. One containing a 1:1 volume of BW25113 ER-g5 and LB + gentamicin, another with Δ*waaC* ΔER-g5 and LB + gentamicin, and a final control with a 1:1 ratio of BW25113 ER-g5 and Δ*waaC* ER-g5. Spot assay EOP data for the endpoint phage populations from the 1st and 9th passages are presented in Figure 4.9. The main assay showed a clear difference in the phage phenotype after the 9th passage compared with the controls (even the replicate with the least fit EOP from the main assay being 1000-fold more

fit than the best replicates from the controls). Note also that the Δ*waaC* ΔER-g5 only control appeared to select against binding to Δ*waaC* (when compared to the other 1st passage EOPs) prior to phage extinction by the end of the 9th passage (no plaques were observed on either strain).

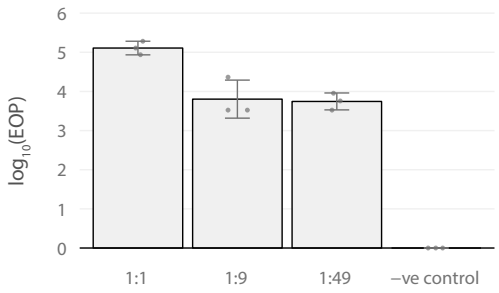


Figure 4.10: The 2nd Δ*waaC* directed evolution experiment. Passage 9 phage evolved in the 1:1 BW25113 ER-g5 to Δ*waaC* ΔER-g5 condition are passaged a further 9 times in 4 different conditions: 1:1, 1:9, 1:49, and a negative control which only Δ*waaC* ΔER-g5 cells. The data shown here is from the endpoint PFU population of passage 18. The (overlaid) data points are independent passaging replicates, bars represent the geometric mean, and the error bars are SD of the log-transformed EOPs.

In the 2nd round of Δ*waaC* directed evolution experiments, phage extracted from each of the independent replicates of passage 9 (denoted 1:1 (9-1), 1:1 (9-2), and 1:1 (9-3)) were subjected to a further 9 passages in 3 different conditions in addition to a negative control. In these conditions, the ratio was either kept at 1:1, or was 1:9, or 1:49 for each of the 9 passages. In Figure 4.10, we see that after the 18th passage the phage population in the 1:1 condition appears to have further increased its fitness, over the 9th passage (Figure 4.9). However, the 1:9 and 1:49 ratios seem to have led to a fitness stagnation, perhaps even decreasing in fitness.

In the 3nd round of Δ*waaC* directed evolution experiments, 1:1 (9-1), 1:1 (9-2), and 1:1 (9-3) phage were challenged by a further 9 passages in switching conditions. The switching conditions were: 1:1 to 1:9, 1:1 to 1:199 and 1:1 to

0:1 (i.e. all negative selection cells). EOP data for this experiment are presented in Figure 4.11a. The 1:1 to 0:1 switching condition had the highest geometric mean (1.6×10^5) but not the most fit replicate overall (found in the 1:1 to 1:199 condition, with EOP 1.0×10^6). Figure 4.11b compares the better evolved lines of the experiments thus far with T7 $\Delta g5$::*cmk*, which is at least 3 orders of magnitude fitter than the evolved lines. Enough improvement was observed however to think that evolving T7 towards OmpF might be possible with the strategies developed up to this point.

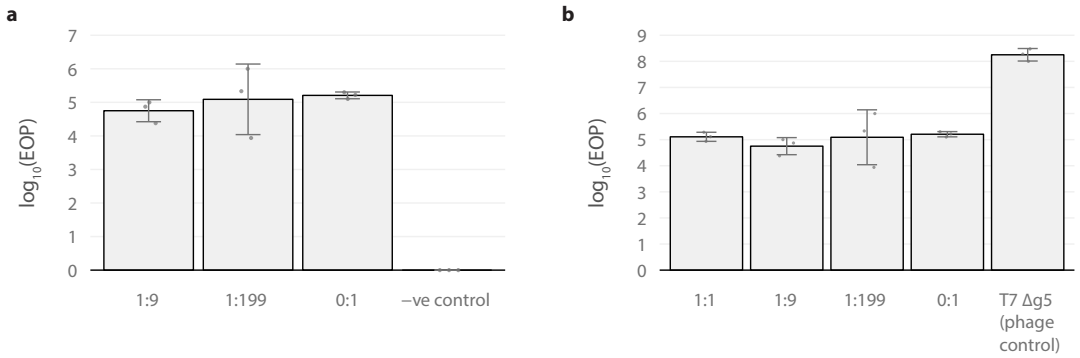


Figure 4.11: The 3rd $\Delta waaC$ directed evolution experiment. a) Passage 9 phage evolved in the 1:1 BW25113 ER-g5 to $\Delta waaC$ ΔER-g5 condition are passaged a further 9 times in 4 different switching conditions. Odd numbered passages where in a 1:1 environment, and even numbered passages were in either a 1:9, 1:199, or a 1:0 environment. The negative control contained only $\Delta waaC$ ΔER-g5 cells. b) The new switching data compared with EOP data from the 1:1 condition and for T7 $\Delta g5$::*cmk*. The data shown here is from the endpoint PFU population of passage 18. Data points (overlaid) are independent passaging replicates (with the exception of those for T7 $\Delta g5$::*cmk*, which are biological replicates) bars represent the geometric mean, and the error bars are SD of the log-transformed EOPs. No plaques were observed on $\Delta waaC$ plates for T7 $\Delta g5$::*cmk*.

Sequencing of tail genes

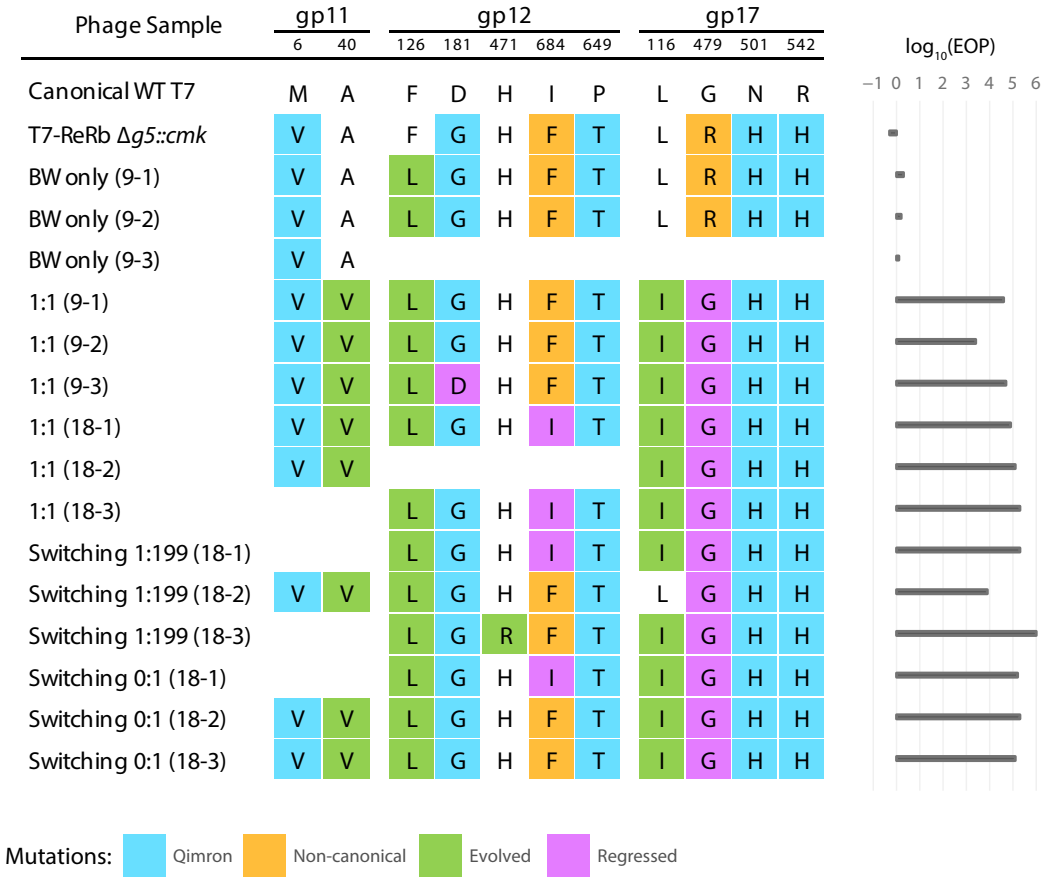


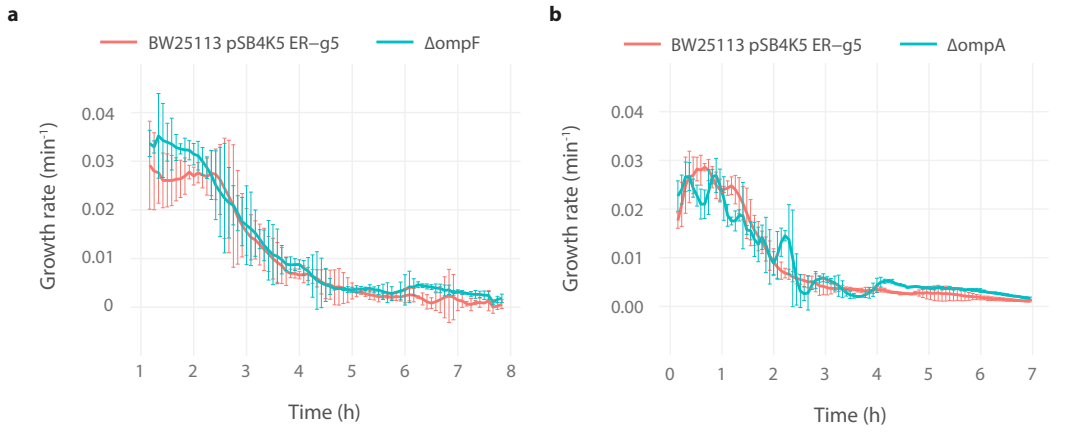
Figure 4.12: Sequencing analysis of the T7-ReRb Δ*g5*::*cmk* receptor tropism regression experiment. Amino acid mutations observed in T7-ReRb by Qimron *et al.* [87] are coloured blue, and mutations that regress back to the canonical WT sequence (NC_001604) are labelled in purple. Non-canonical mutations (those unobserved in the literature) are coloured orange, and mutations observed in the chromatograms of the evolving phage population are coloured green. Phage samples are denoted by '[Condition] ([Passage number] - [Evolution line])', e.g. the 3rd independent evolution line passaged in a 1:1 co-culture for 18 passages would be labelled '1:1 (18-3)'. Gaps in sequencing data are left blank. Efficiency of plating (EOP) is defined as the PFU/ml titre observed on BW25113 lawns divided by the PFU/ml titre observed on Δ*waaC* lawns. The EOP for T7-ReRb Δ*g5*::*cmk* is the geometric mean of 3 biological replicates. All other EOPs presented are from single replicates.

Tail genes known to be responsible for phage-host tropism were sequenced as described in Methods 5.3.8. Observed amino acid mutations are presented in Figure 4.12, alongside the relevant EOP data. The BW25113 control lines match the unevolved (T7-ReRb $\Delta g5::cmk$) with the exception of the F126L mutation in gp12. Since the F126L mutation is observed in all passaged lines, it is presumably an adaption to the general passaging procedure. The Qimron mutations (see section 3.2.5, Figure 3.15c) are roughly retained throughout (barring a gp12-G181D reversion in the 1:1 (9-3) condition). gp11-A40V and the regression gp17-R479G were only found in non-control lines, and appears to be associated with higher EOPs. Similarly, gp17-L116I was also only seen in the non-control lines, with the exception of Switching 1:199 (18-2), which had a markedly lower EOP than the other independent lines for that treatment. 1:1 (9-3) has a moderately higher EOP than lines 1 and 2, which is possibly associated with the reversion of a Qimron mutation in gp12 (reverting from G to D in aa 181). However, this reversion does not appear to be retained in future passages (1:1 (18-3), Switching 1:199 (18-3), Switching 0:1 (18-3)). The evolved line yielding the highest EOP (Switching 1:199 (18-3)) has an additional gp12-H471R mutation in that phage population. The non-canonical gp12-I684F mutation, and it's corresponding reversion, do not appear to have a correspond with higher or lower EOPs.

4.2.4 Directed evolution of T7 towards an outer membrane protein (OmpF)

Preparations for the evolution experiments

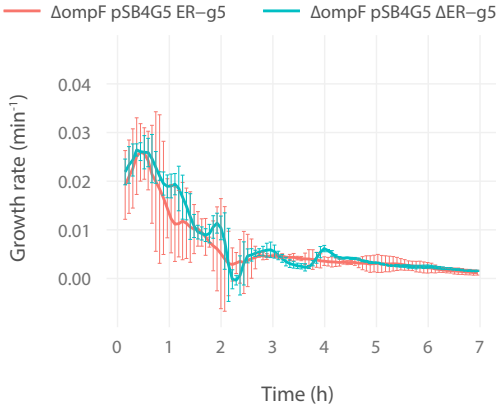
T7 has recently been reported to interact with the outer membrane proteins OmpA and OmpF [130]. The first objective was to decide which of these proteins to direct the tropism of T7 towards. The growth rate of Δ*ompF* and Δ*ompA* were characterised in different plate reader experiments (but the strain BW25113 pSB4K5 ER-g5 was kept common to both) as described in Figure 4.13. Prior to the addition of the pSB4G5 ER-g5 and ΔER-g5 plasmids, Δ*ompF* grew significantly faster than Δ*ompA* (at the 5% level). Transformation with the aforementioned plasmids reduced the growth rate of the Δ*ompF* strains (Figure 4.14) but within the range of the previous characterised BW25113 ER-g5 (Figure 4.4).



Strain	Mean growth rate (min^{-1})	95% confidence interval
BW pSB4K5 ER-g5	0.029	$\pm 8.5 \times 10^{-3}$
ΔompF	0.0353	$\pm 1.0 \times 10^{-2}$

Strain	Mean growth rate (min^{-1})	95% confidence interval
BW pSB4K5 ER-g5	0.0285	$\pm 1.0 \times 10^{-3}$
ΔompA	0.0251	$\pm 2.4 \times 10^{-3}$

Figure 4.13: Growth rate comparison of Δ*ompF* and Δ*ompA* from 2 separate microplate reader experiments which were performed with minor variations in procedure (see Methods 5.9.2 for details). a) Cells were grown overnight, and diluted 1:200 before recording growth in the plate reader. b) Cells grown overnight were diluted 1:200 and incubated for roughly 90 minutes before growth was recorded in the plate reader. Error bars are presented as 95% confidence intervals.



Strain	Mean growth rate (min^{-1})	95% confidence interval
ΔompF pSB4G5 ER-g5	0.0258	$\pm 5.7 \times 10^{-3}$
ΔompF pSB4G5 ΔER-g5	0.0262	$\pm 1.6 \times 10^{-3}$

Figure 4.14: Growth rate analysis of Δ*ompF* transformed with either pSB4G5 ER-g5 or pSB4G5 ΔER-g5. Error bars are presented as 95% confidence intervals.

Plaque and spot assays for T7 Δ*g5*::*cmk* against BW25113 ER-g5 and Δ*ompF* ER-g5 revealed no significant differences either for the method used or the strain

infected (Figure 4.15).

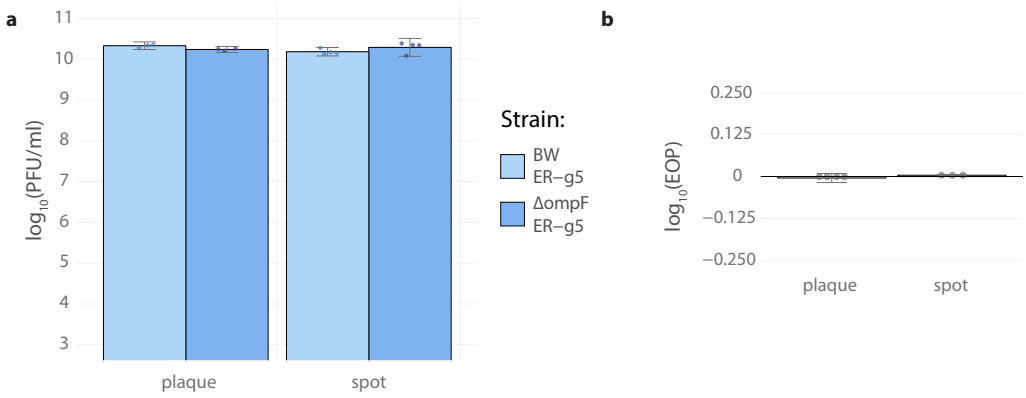


Figure 4.15: a) Plaque and spot assays for T7 $\Delta g5$::*cmk* challenged against BW25113 and $\Delta ompF$ cells transformed with pSB4G5-ER-g5. b) Efficiency of plating (EOP) defined as PFU/ml recovered on BW25113 ER-g5 plates, divided by that recovered on $\Delta ompF$ ER-g5 plates. Bar plots are presented as the geometric mean of at least 3 biological replicates with 95% confidence intervals. Individual biological replicates are overlaid.

Pairwise competition modelling

The equations for the $\Delta ompF$ system follow as before except that the positive selection strain, BW25113 ER-g5, now possesses 2 receptors of interest: OmpF and LPS (see Figure 4.16) and phage phenotype j will adsorb to those receptors at rates k_{1j} and k_{2j} respectively.

$$\frac{\mathrm{d}C_1}{\mathrm{d}t} = a_1 C_1 - \underbrace{(k_{11} + k_{21})}_{\text{OmpF \& LPS}} C_1 P_1 - (k_{12} + k_{22}) C_1 P_2$$
$$\frac{\mathrm{d}C_2}{\mathrm{d}t} = a_2 C_2 - \underbrace{k_{21}}_{\text{LPS}} C_2 P_1 - k_{22} C_2 P_2$$

$$\frac{\mathrm{d}I_{11}}{\mathrm{d}t} = (k_{11} + k_{21}) C_1 P_1 - (k_{11} + k_{21}) C_1(t-\tau) P_1(t-\tau)$$
$$\frac{\mathrm{d}I_{12}}{\mathrm{d}t} = (k_{12} + k_{22}) C_1 P_2 - (k_{12} + k_{22}) C_1(t-\tau) P_2(t-\tau)$$
$$\frac{\mathrm{d}I_{21}}{\mathrm{d}t} = k_{21} C_2 P_1 - k_{21} C_2(t-\tau) P_1(t-\tau)$$
$$\frac{\mathrm{d}I_{22}}{\mathrm{d}t} = k_{22} C_2 P_2 - k_{22} C_2(t-\tau) P_2(t-\tau)$$

$$\begin{aligned}\frac{\mathrm{d}P_1}{\mathrm{d}t} &= \beta k_{11} C_1(t-\tau) P_1(t-\tau) - (k_{11} + k_{21}) C_1 P_1 - k_{21} C_2 P_1 \\ &\quad - (k_{11} + k_{21}) I_{11} P_1 - (k_{11} + k_{21}) I_{12} P_1 - k_{21} I_{21} P_1 - k_{21} I_{22} P_1 \\ \frac{\mathrm{d}P_2}{\mathrm{d}t} &= \beta k_{22} C_2(t-\tau) P_2(t-\tau) - (k_{12} + k_{22}) C_1 P_2 - k_{22} C_2 P_2 \\ &\quad - (k_{12} + k_{22}) I_{11} P_2 - (k_{12} + k_{22}) I_{12} P_2 - k_{22} I_{21} P_2 - k_{22} I_{22} P_2 \,.\end{aligned}$$

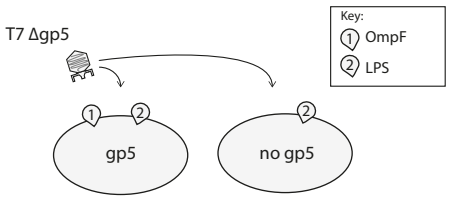


Figure 4.16: Abstraction of the Δ*ompF* directed evolution experiment. Phage j binds to both OmpF and LPS at rates k_{1j} and k_{2j} respectively, but only produces progeny when infecting cells complemented with gp5. The rates are additive, so the total adsorption rate to BW25113 will be $k_{1j} + k_{2j}$ and the adsorption rate to Δ*ompF* will be k_{2j} .

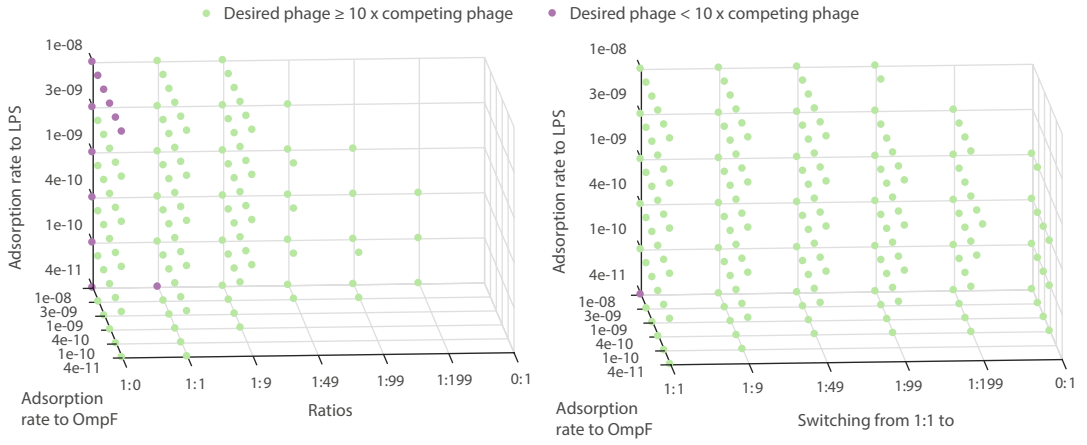


Figure 4.17: Pairwise competition between the desired phage (OmpF affinity 10^{-8} ml/min and no affinity for LPS) and competing phage phenotypes. Passage conditions notated as the ratio of (BW25113 ER-g5):($\Delta ompF$ ΔER-g5). Data where the competing phage washes out in the absence of the challenging desired phage (the washout model) is not plotted. Simulations performed for 20 passages with $a = 0.026$ min^{-1}, $\beta = 100$, $\tau = 12$ min, initial PFU/ml $= 3 \times 10^6$, initial CFU/ml $= 5 \times 10^7$ and 1% of the phage population transferred between each passage. Washout is defined as the PFU/ml dropping below 10^4 after 20 passages.

Since the model treats phage binding affinities for the 2 considered receptors (LPS and OmpF) additively, and there is only selection pressure to avoid LPS, the desired phage (with a 10^{-8} ml/min affinity for OmpF) outcompetes phage with weaker binding affinity for OmpF (Figure 4.17a). Strong adsorption to both LPS and OmpF is preferred in the 1:1 and 1:9 conditions. Since we will most likely begin with a T7 phage (T7 $\Delta g5$::cmk) which interacts strongly with LPS, but minimally with OmpF, the 1:9 ratio is then the least stringent ratio condition that selects for the desired phage, with minimal stress on the competing phages. That said, further increases to that ratio rapidly decrease the pool of phage phenotypes that can sustainably replicate in that environment. Again, switching offers more leeway for selecting the desired phage, whilst retaining competing phage phenotypes that might be stepping stones to increased fitness (Figure 4.17b). Reducing the

binding affinity for OmpF of the desired phage to one more likely to be attainable ($1/9 \times 10^{-8}$ ml/min) alters the state of play somewhat (Figure 4.18). For one, the desired phage does not withstand ratios equal to and greater than 1:49 (Figure 4.17a). The desired phage is only able to outcompete phages with a higher than $1/9 \times 10^{-8}$ ml/min binding affinity when switching from 1:1 to 1:49 or above.

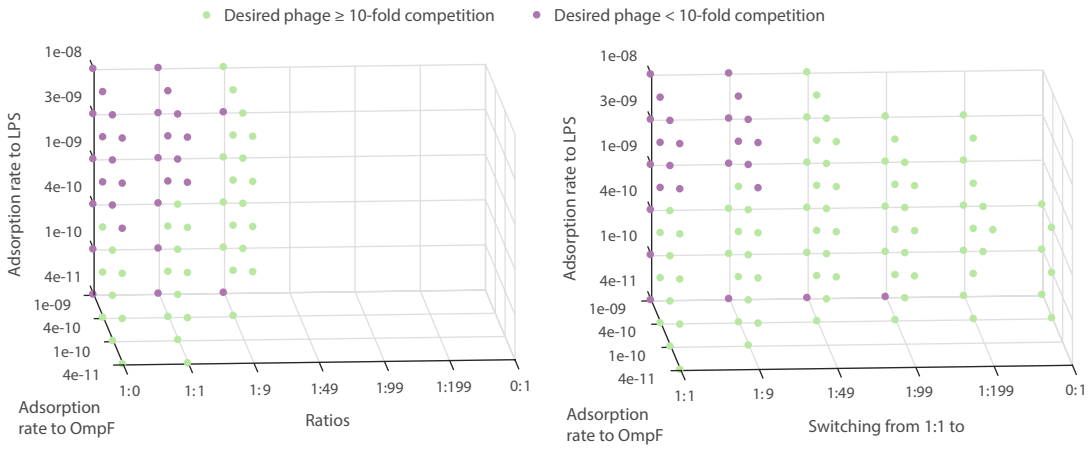


Figure 4.18: Pairwise competition between the desired phage (OmpF affinity $1/9 \times 10^{-8}$ ml/min and no affinity for LPS) and competing phage phenotypes. Passage conditions notated as the ratio of (BW25113 ER-g5):($\Delta ompF$ ΔER-g5). Data where the competing phage washes out in the absence of the challenging desired phage (the washout model) is not plotted. Simulations performed for 20 passages with $a = 0.026$ min^{-1}, $\beta = 100$, $\tau = 12$ min, initial PFU/ml $= 3 \times 10^6$, initial CFU/ml $= 5 \times 10^7$ and 1% of the phage population transferred between each passage.

Determining relative phage fitness from plate reader absorbance data

To compare the relative fitness of unevolved and evolved phages against cells of two different genotypes (here BW25113 and $\Delta ompF$) I constructed a coupled non-linear

ODE system to model the system:

$$\frac{\mathrm{d}C_1}{\mathrm{d}t} = r_1 C_1 \left(1 - \frac{C_1}{K_1}\right) - f_1 C_1 P_1$$
$$\frac{\mathrm{d}C_2}{\mathrm{d}t} = r_2 C_2 \left(1 - \frac{C_2}{K_2}\right) - f_2 C_2 P_2$$
$$\frac{\mathrm{d}P_1}{\mathrm{d}t} = f_1 C_1 P_1$$
$$\frac{\mathrm{d}P_2}{\mathrm{d}t} = f_2 C_2 P_2$$

with notation as described in Table 4.2.

Table 4.2: Variable and parameter notation for the relative phage fitness model

Notation	**Description**	**Units**
C_i	Cells with genotype i (i either 1 or 2) at time t	OD_{600}
P_i	The free phage population infecting the cells population i at time t	OD_{600}
r_i	Growth rate of cells with genotype i	min^{-1}
K_i	Carrying capacity (stationary phase OD of cells population i)	OD_{600}
f_i	"Phage fitness" - rate of phage production from infecting cells i	$OD_{600}{}^{-1}min^{-1}$
t	Time	min

Assumptions:

- The population is well-mixed.
- The correspondence between CFU and OD_{600} is the same for all strains considered.
- That the latent period associated with phage infection of cells can be ignored (i.e. that delay differential equations are not necessary). A consequence of this assumption is that phage genotypes that deplete cellular populations faster, but that produce less phage in the process, could be considered fitter,

if the final phage life cycle results in lysis of most of the cells.

- Phage production is not produced in bursts and instead phage are produced continuously via interactions with cells.

Our aim here is to find an expression for the relative phage fitness, f_1/f_2, that can be determined from plate reader absorbance data. If the initial conditions were known with precision, then we could estimate this proportion through numerically fitting the ODEs to the data. This is not true for our case, so we instead observe first that $\dfrac{\mathrm{d}P_1}{\mathrm{d}t} = f_1 C_1 P_1$ can be solved by separation of variables to obtain,

$$P_1 = P_0\, e^{f_1 \int_0^{t_{min}} C_1(s)\, ds} \approx \; P_0\, e^{f_1 \alpha_1} \tag{4.3}$$

where P_0 denotes initial phage population i.e. $P_i(t = 0)$ for $i = 1$ or 2. Note that C_1 is integrated over a given time interval $[0, t_{min}]$ which will be estimated by calculating the area under the OD_{600} curve observed in the plate reader data. We denote this integral as α_i (i either 1 or 2) for convenience. The procedure to obtain $P_2 = P_0\, e^{f_2 \alpha_2}$ follows similarly, and we note here that both equations are valid over the time interval $[0, t_{min}]$. We can also solve the rate equation for C_1 (or C_2) to obtain an expression in terms of P_1 (or P_2), namely,

$$\begin{aligned} \frac{\mathrm{d}C_1}{\mathrm{d}t} &= r_1 C_1 \left(1 - \frac{C_1}{K_1}\right) - f_1 C_1 P_1 \quad \text{which can be rearranged to,} \\ P_1 &= \frac{r_1 C_1 \left(1 - \frac{C_1}{K_1}\right) - \dfrac{\mathrm{d}C_1}{\mathrm{d}t}}{f_1 C_1}. \end{aligned} \tag{4.4}$$

Since equation (4.3) is true for all times t in the interval $[0, t_{min}]$ we can choose $t = t_{min}$ in equation (4.4). In particular, we choose t_{min} such that it gives the largest timepoint out of the times at which the smallest OD_{600} values for C_1 and C_2

occur (see Figure 4.19). In other words, we choose $t_{min} = \max\{t_{min(C_1)}, t_{min(C_2)}\}$.

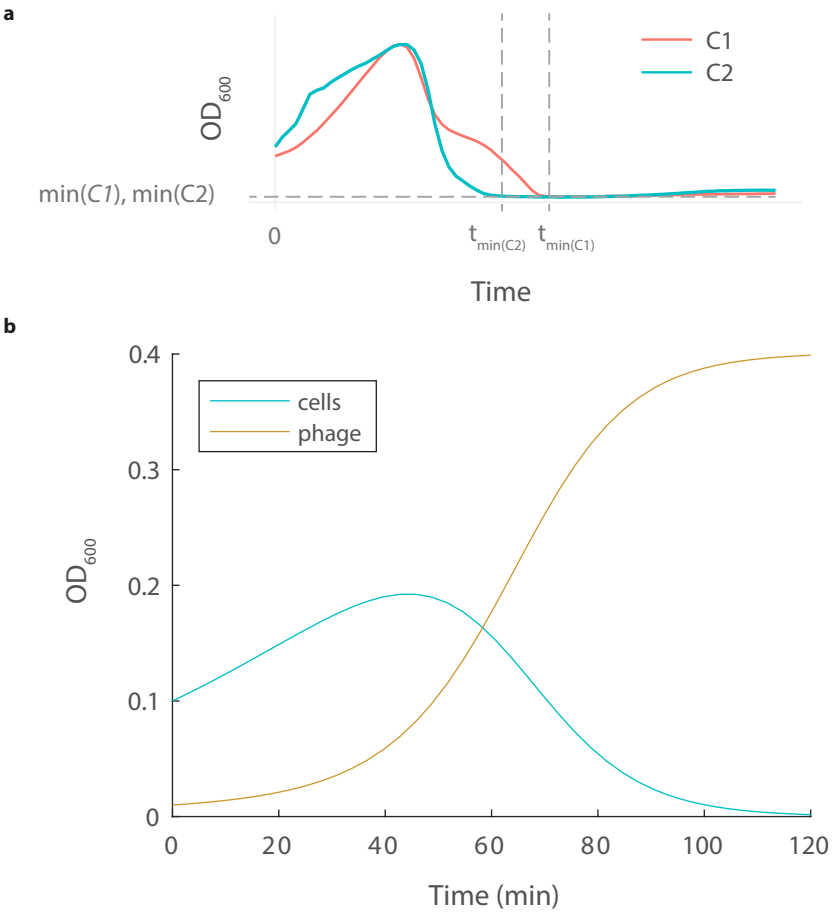


Figure 4.19: Clarification of the plate reader model. a) We choose the time interval on which the model operates to be from $t = 0$ to $t = t_{min}$. For each curve, we determine the time at which the OD is smallest. We denote the largest of these 2 timepoints as t_{min} i.e. $t_{min} = \max\{t_{min(C_1)}, t_{min(C_2)}\}$. b) Example plot of a numerical simulation of the ODEs with dde23 in MATLAB. Parameters used were $r = 0.027$, $K = 1.2$ and $f = 0.3$, with initial conditions of $C(t = 0) = 0.1$ and $P(t = 0) = 0.01$.

A consequence of this choice of t is that $\dfrac{\mathrm{d}C_i}{\mathrm{d}t} \approx 0$ and so,

$$P_1 = \frac{r_1}{f_1}\left(1 - \frac{C_1}{K_1}\right) := \frac{\gamma_1}{f_1} \tag{4.5}$$

$$P_2 = \frac{r_2}{f_2}\left(1 - \frac{C_2}{K_2}\right) := \frac{\gamma_2}{f_2}. \tag{4.6}$$

where, for simplicity, we have defined $\gamma_i = r_i\left(1 - \frac{C_i}{K_i}\right)$ for both $i = 1$ and 2. Combing the two separate equations for P_1 (namely (4.3) and (4.5)) we have,

$$P_0 e^{f_1\alpha_1} = \frac{\gamma_1}{f_1}$$

which after dividing through by P_0 and multiplying both sides by $\alpha_1 f_1$ becomes,

$$\alpha_1 f_1 e^{f_1\alpha_1} = \frac{\alpha_1\gamma_1}{P_0}. \tag{4.7}$$

Recall that our aim here is to find f_1/f_2 in terms of a calculable expression. Note first that the left-hand side of this equation is of the form xe^x. The inverse, f^{-1}, of the function $f(x) = xe^x$ (which would return x, since $f^{-1}(f(x)) = x$) is given by $W(x)$, where W is Lambert W function [188–192]. Applying this function to both sides of equation (4.7), we have,

$$\begin{aligned} W\left(\alpha_1 f_1 e^{f_1\alpha_1}\right) &= W\left(\frac{\alpha_1\gamma_1}{P_0}\right) \quad \text{which since } W(xe^x) = x \text{ is} \\ f_1\alpha_1 &= W\left(\frac{\alpha_1\gamma_1}{P_0}\right) \quad \text{which rearranges to} \\ f_1 &= \frac{1}{\alpha_1} W\left(\frac{\alpha_1\gamma_1}{P_0}\right) \quad \text{and similarly} \quad f_2 = \frac{1}{\alpha_2} W\left(\frac{\alpha_2\gamma_2}{P_0}\right). \end{aligned}$$

Hence, we have,

$$\frac{f_1}{f_2} = \frac{\frac{1}{\alpha_1} W\left(\frac{\alpha_1\gamma_1}{P_0}\right)}{\frac{1}{\alpha_2} W\left(\frac{\alpha_2\gamma_2}{P_0}\right)}. \tag{4.8}$$

Again, this equation would be estimatable, if P_0 could be estimated from experimental data. Recall that in choosing to construct this particular ODE system, we have given the phage population (P) a physically meaningless dimension in OD_{600} (since T7 phage do not detectably perturb 600 nm wavelength light). There are

a few ways to proceed. One way would be to scale a PFU estimated via plaque assay by the initial CFU of the challenged strain (see Appendix C.1 for more detail on how we should interpret the units for P) and then calculate f_1/f_2 from the equation (4.8). Another, simpler, solution could be employed if at least one more, say, 10-fold, dilution of the phage was used in the plate reader assay. In this case,

$$\frac{f_1}{f_2} = \frac{(\alpha_{21} - \alpha_{22}) \ln\left(\frac{0.1(K_1 - C_{11})}{(K_1 - C_{12})}\right)}{(\alpha_{11} - \alpha_{12}) \ln\left(\frac{0.1(K_2 - C_{21})}{(K_2 - C_{22})}\right)} \tag{4.9}$$

where α_{ij} is the area under the curve for phage dilution j challenged against strain i, and we choose the curves such that $\alpha_{i1} > \alpha_{i2}$. Similarly, C_{ij} refers to strain i challenged against phage dilution j. See Appendix C.2 for details of the derivations for both the case where 2 phage dilutions are available, and for when $n > 2$ phage dilutions are utilised. With the experimental data available to us, neither of these methods using either the PFU of the phage, or multiple phage dilutions, is suitable. Instead, we need to find a way to cancel out P_0 when determining the fraction f_1/f_2. With the equation (4.8) in its current form, this is not possible. To proceed, I first determined that,

$$W\left(\frac{\alpha_i \gamma_i}{P_0}\right) \approx 0.01 \ln(10^3 \alpha_i) \ln(10^5 \gamma_i) \ln\left(\frac{2}{P_0}\right) \tag{4.10}$$

over the domain $1 \leq \alpha_i \leq 10^{-3}$, $10^{-3} \leq \gamma_i \leq 10^{-1}$ and $10^{-7} \leq P_0 \leq 1$ (see Methods 5.9.5 for details). The median and mean percent errors for this approximation over this domain are 8% and 116% respectively. The percentage error for fixed $\gamma_i = 2 \times 10^{-2}$ is plotted in Figure 4.20. Note that in practice, since $\left(1 - \frac{A_i(t=t_{min})}{K_i}\right) \approx 1$, that $\gamma_i = r_i(1 - A_i/K_i) \approx r_i$, the growth rate of strain i. Hence γ_i varies in line with expected growth rates for *E. coli* (ranging from 10^{-2} to 3.5×10^{-2}).

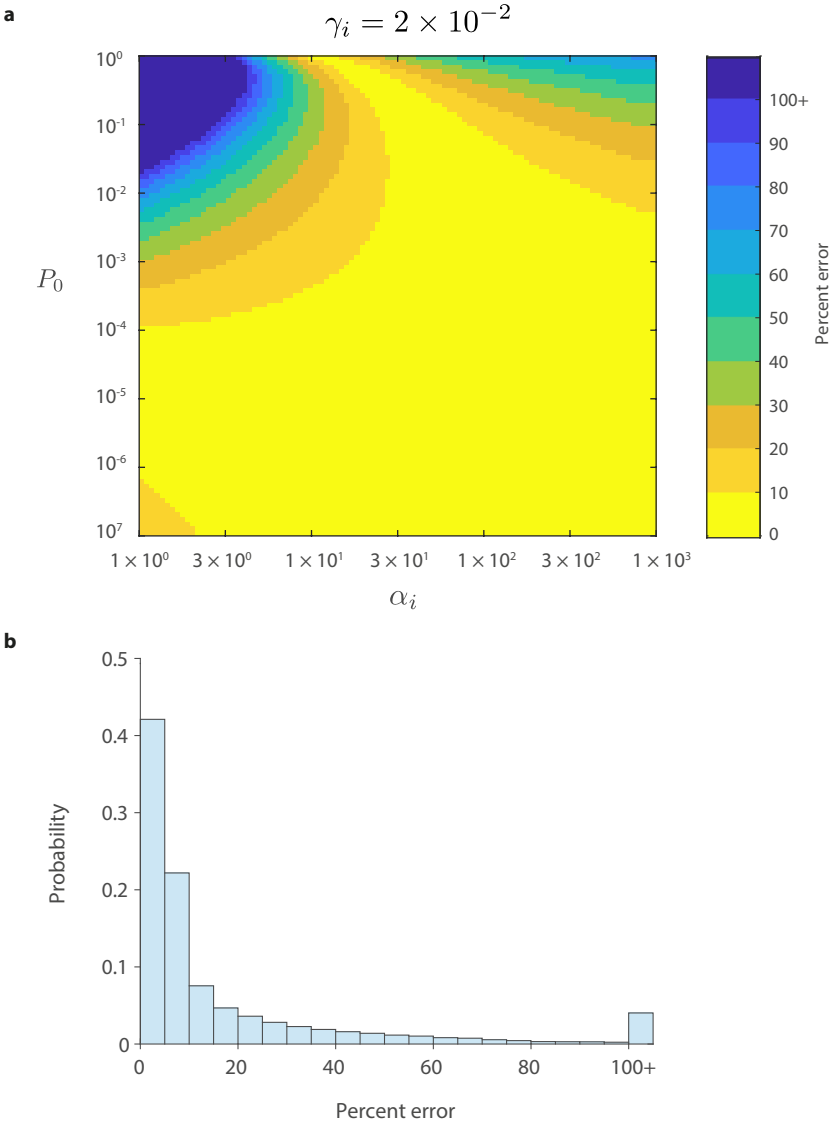


Figure 4.20: Percent error (defined as |true value − approximate|/true value) for the approximation $0.01 \ln(10^3 \alpha_i) \ln(10^5 \gamma_i) \ln\left(2/P_0\right)$ of the function $W\left(\alpha_i \gamma_i / P_0\right)$ over the domain $1 \leq \alpha_i \leq 10^{-3}$, $10^{-3} \leq \gamma_i \leq 10^{-1}$ and $10^{-7} \leq P_0 \leq 1$ (see Methods 5.9.5 for details) and γ_i is fixed at a representative growth rate (2×10^{-2}). P_0 denotes the number of phage on the OD_{600} scale at $t = 0$ (when the plate reader records the first absorbance value) and α_i denotes the area under the lysis curve observed for the time interval $[0, t_{min}]$. b) A histogram of the percent error over whole domain. Errors over 100% are grouped into the final bin.

Now that we have a suitable approximation for $W\left(\frac{\alpha_i\gamma_i}{P_0}\right)$, we can produce an approximate expression for the relative fitness,

$$\begin{aligned}\frac{f_1}{f_2} = \frac{\frac{1}{\alpha_1}W\left(\frac{\alpha_1\gamma_1}{P_0}\right)}{\frac{1}{\alpha_2}W\left(\frac{\alpha_2\gamma_2}{P_0}\right)} &\approx \frac{\frac{1}{\alpha_1}0.11\ln(10^3\alpha_1)\ln(10^5\gamma_1)\ln\left(\frac{2}{P_0}\right)}{\frac{1}{\alpha_2}0.11\ln(10^3\alpha_2)\ln(10^5\gamma_2)\ln\left(\frac{2}{P_0}\right)} \\ &= \frac{\alpha_2\ln(10^3\alpha_1)\ln(10^5\gamma_1)}{\alpha_1\ln(10^3\alpha_2)\ln(10^5\gamma_2)}.\end{aligned}$$

Characterisation of the OmpF directed evolution experiments

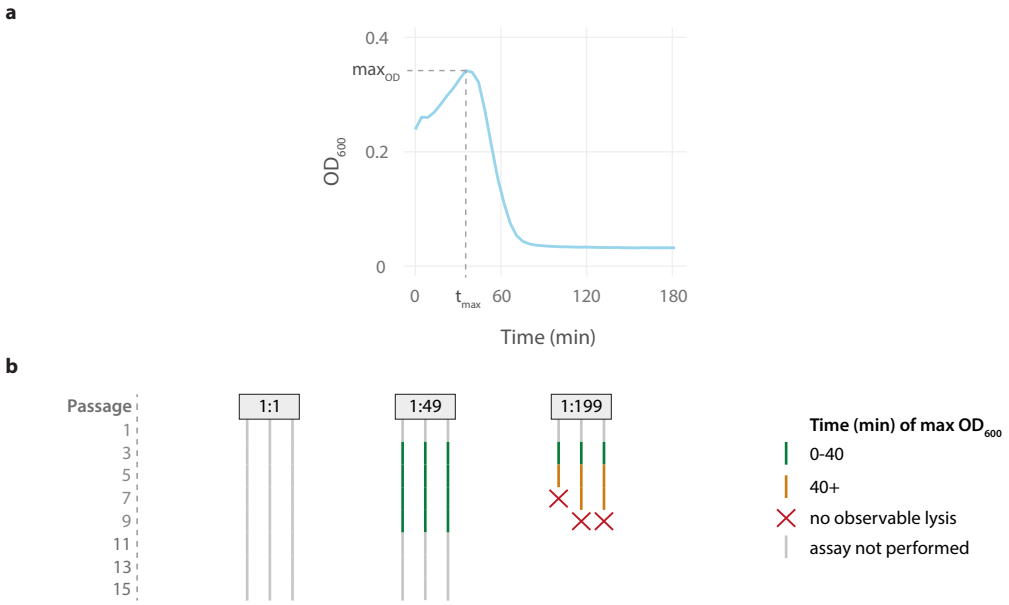
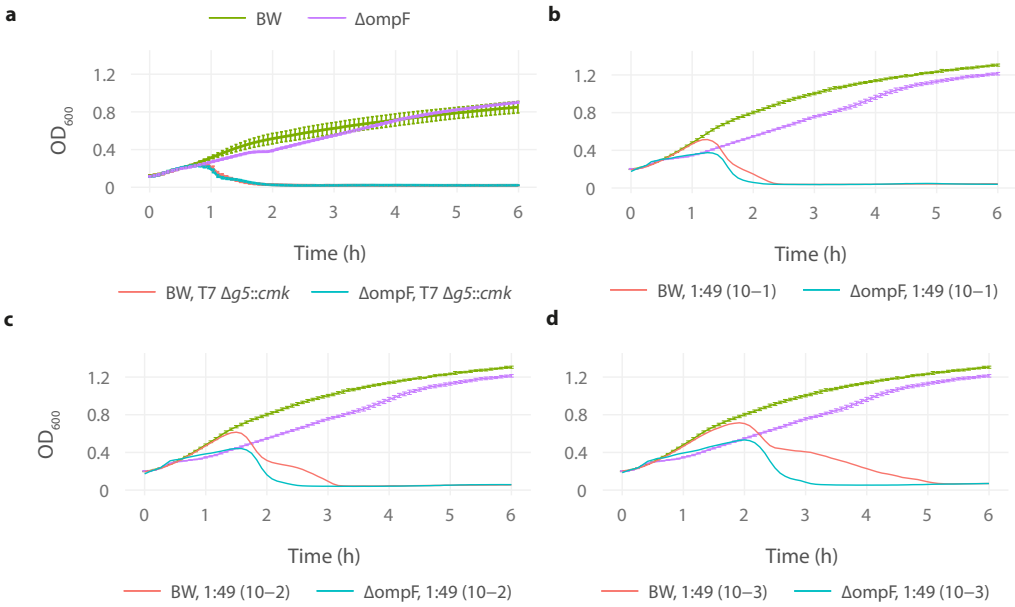


Figure 4.21: a) After incubating the 1:1 passages for 1 hour, a sample of the population is incubated in a microplate reader, and the OD_{600} is recorded (one such curve is depicted). The time associated with maximum OD_{600} value is noted for further analysis (see Methods 5.9.6 for details). b) Maximum OD_{600} analysis of the 1st 15 passages of the ompF directed evolution experiment (see Methods 5.6). Odd numbered passages are under 1:1 conditions and even number passages are under the boxed conditions (in this case, either 1:1, 1:49, or 1:199).

Similar to before, a brief overview of the procedure is provided here; for more details, see Methods 5.6. For each passage T7 $\Delta g5$::*cmk* was incubated with cells for ≈ 60 minutes, and 1% of the endpoint phage population was transferred to the next passage. Odd-numbered passages were 1:1 volume by volume ratios of BW25113 ER-g5 and $\Delta ompF$ ΔER-g5. Even-numbered passages varied by stringency. For the first 15 passages, the even-number passages were the ratios 1:1, 1:49 and 1:199 (see Figure 4.21). After most odd-numbered passages, a sample of the passage was added to a plate reader to observe whether the phage population had survived the passage (Figure 4.21a). Passages where no lysis was observed were additionally

checked by plating for the absence of phage.



e

Phage	BW25113 (PFU/ml)	Δ*ompF* (PFU/ml)
T7 Δ*g5*::*cmk*	2 x 10^{10}	1 x 10^{10}
1:49 (10-1)	2 x 10^{6}	7 x 10^{6}
1:49 (10-2)	6 x 10^{5}	6 x 10^{5}
1:49 (10-3)	2 x 10^{4}	2 x 10^{4}

f

Phage	Relative fitness ($f_{BW}/f_{\Delta ompF}$)
T7 *g5*::*cmk*	1.01
1:49 (10-1)	0.85
1:49 (10-2)	0.77
1:49 (10-3)	0.66

Figure 4.22: a) Lysis curves for BW25113 and Δ*ompF* transformed with pSB4G5 ER-g5 in presence/absence of T7 Δ*g5*::*cmk*. Error bars are the SD of 3 biological replicates. b, c, d) In a separate plate reader experiment, lysis curves for BW25113 and Δ*ompF* transformed with pSB4G5 ER-g5 in presence/absence of evolved lines 1, 2 and 3 from passage 10 of the ompF directed evolution experiment. Error bars for the data in the absence of phage are the SD of 3 biological replicates. Data in the presence of phage were performed on only 1 biological replicate, and so are presented without error bars (as the average of 3 technical replicates). e) Rough PFU/ml of each phage for comparison. f) The relative fitness of unevolved phage (T7 Δ*g5*::*cmk*) and the evolved passage 10 lines challenged with BW25113-pSB4G5 ER-g5 and Δ*ompF* pSB4G5-ER-g5. Relative phage fitness versus BW25113 is estimated using the approximation $f_{BW}/f_{\Delta ompF} \approx [\alpha_{\Delta ompF} \ln(10^3 \alpha_{BW}) \ln(10^5 \gamma_{BW})]/[\alpha_{BW} \ln(10^3 \alpha_{\Delta ompF}) \ln(10^5 \gamma_{\Delta ompF})]$. α_{BW} and $\alpha_{\Delta ompF}$ are estimated areas under the lysis curves for phage challenged with BW25113 and Δ*ompF* respectively. Similarly γ_{BW} and $\gamma_{\Delta ompF}$ are effectively the growth rates for BW25113 and Δ*ompF* respectively (see sections 5.9.4 and 4.2.4 for details).

Table 4.3: p-values (rounded to 4 d.p.) observed for two-sample Welch's t-tests on $\log_{10}$(EOP) data from passage 15 (1:49) against $\Delta waaC$ and BW25113 strains (see Figure 4.23b for details)

unevolved vs 15-1	**unevolved vs 15-2**	**unevolved vs 15-3**
0.0003	0.0010	0.0126
15-1 vs 15-2	**15-1 vs 15-3**	**15-2 vs 15-3**
0.0065	0.2208	0.0117

Figure 4.22 presents lysis curve data for unevolved phage (T7 $\Delta g5$::cmk), and phage extracted from the endpoint of passage 10, that had undergone a 1:1 to 1:49 switching regime. As seen in Figure 4.22f, the relative fitness of phage lines switched for each passage between the 1:1 to 1:49 condition for a total of 10 passages, decreased compared to the relative fitness observed for the unevolved phage (T7 $\Delta g5$::Δcmk). This lends some weight to the suggestion that the phage population was in fact adapting to the negative selection strain.

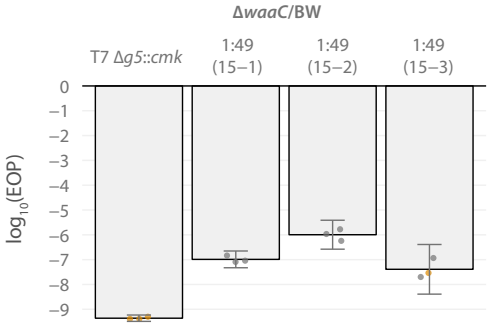
Single replicate EOP data (BW25113 ER-g5/$\Delta ompF$ ΔER-g5) for passage 13, switching condition 1:1 to 1:49, is presented in Figure 4.23a. There is no evidence to suggest that the plating efficiency differs between the two strains. EOP data collected for the condition in passage 15 is shown in Figure 4.23b. All three evolved lines plate significantly more efficiently (at the 5% level) on $\Delta waaC$ ER-g5 strains, relative to BW25113 ER-g5, when compared with the unevolved phage (see Table 4.3). Similarly, the geometric mean EOP for 1:49 (15-2) was significantly different to both 15-1 and 15-3. 3 single plaques were picked from the $\Delta waaC$ plates challenged with 1:49 (15-1) and 1:49 (15-2) phage (Figure 4.23c). These plaques were plated on BW25113 ER-g5 and $\Delta ompF$ ΔER-g5 plates, and the EOP measured by spot

assay (Figure 4.23d). 2 of the plaques found to have higher EOPs were subjected to plaque assay repeatedly for a total of 3 biological replicates. Subsequently, no evidence of a positive bias towards BW25113 over Δ*ompF* was determined from these assays.

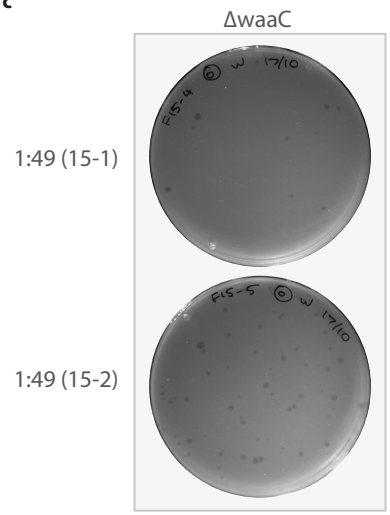
a

Phage	$\log_{10}$(EOP) (BW/Δ*ompF*)
T7 Δ*g5*::*cmk*	-0.15
1:49 (13-1)	0.00
1:49 (13-2)	0.23
1:49 (13-3)	0.18

b



c



d

Phage	Spot assay $\log_{10}$(EOP) (BW/Δ*ompF*)	Plaque assay $\log_{10}$(EOP) (BW/Δ*ompF*)
T7 Δ*g5*::*cmk*	0.19	0.09 ± 0.02
1:49 (15-1) Plaque 1	0.48	-0.16 ± 0.31
1:49 (15-1) Plaque 2	0.53	-0.07 ± 0.08
1:49 (15-1) Plaque 3	0.19	
1:49 (15-2) Plaque 1	-0.30	
1:49 (15-2) Plaque 2	0.20	
1:49 (15-2) Plaque 3	-0.90	

Figure 4.23: a) EOP of 3 independent evolution lines from passage 13 of condition 1:49 compared with the unevolved phage (T7 Δ*g5*::*cmk*). EOP is determined as the PFU against BW25113 divided by the PFU observed with Δ*ompF*. Data are from a single replicate. b) 3 independent evolution lines from passage 15 of condition 1:49, again compared with the unevolved phage. EOP = (PFU vs Δ*waaC*)/(PFU vs BW25113). Bar plots represent geometric mean of 3 biological replicates and the error bars are 95% confidence intervals. The individual data points are overlaid. c) Representative plates from panel b. d) Isolated plaques picked from Δ*waaC* and plated against BW25113 and Δ*ompF*. Logged spot assays EOPs are single replicates; logged plaque assay EOPs show the mean with SD error bars from 3 biological replicates. In all relevant panels EOP replicates for which no plaques were observed on Δ*waaC* are treated as if a single plaque was observed for the mean and error calculations. The individual data points are overlaid, and points where no plaques were observed on Δ*waaC* are coloured orange.

Phage Sample	Cmk			gp10B	gp11	gp12							gp17			
	5	20	184	398	99	137	174	444	490	613	712	723	118	180	203	420
T7 Δ*g5::cmk*	A	T	R	Q	R	N	V	D	S	T	N	Q	T	N	T	T
Switching 1:1 (15-1)	A	T	R	*	R	N	V	D	S	T	N	Q	A	N	T	T
Switching 1:1 (15-2)	A	T	R	*	R	N	V	D	S	T	N	Q	A	N	T	T
Switching 1:1 (15-3)	A	T	R	*	R	N	V	D	S	T	N	Q	A	N	T	T
Switching 1:49 (15-1)	A	T	R	*	R	N	V	A	S	T	N	Q	A	D	T	T
Switching 1:49 (15-2)	A	T	R	*	R	N	A	D	S	T	N	Q	A	N	T	T
Switching 1:49 (15-3)	V	S	H	*	Q	D	V	D	T	I	S	L	A	N	A	A

Figure 4.24: Sequencing analysis of the OmpF directed evolution experiment. Mutations the T7 Δ*g5*::*cmk* sequence observed in the chromatograms of the evolving phage population are coloured green. Bracket notation: (Passage number - line).

Phage from passage 15 of the directed OmpF evolution experiment were subjected to sequencing for key gene regions (see Methods 5.3.8). Figure 4.24 depicts the amino acid mutations observed in the Sanger sequencing data relative to the progenitor T7 Δ*g5*::*cmk* genome. In all evolved lines, the minor coat protein gp10B gained a stop codon in the final amino acid of that protein. The mutation gp17-T118A was also observed in all evolved lines. The switching condition 1:49 (15-3) phage population picked up 13 amino acid changes in the gene regions sequenced, 5 of those occurring in the gp12 tail tube, and 3 occurring in the *cmk* gene.

Phages were passaged in varying conditions for a total of 61 passages (see Figure 4.25). Figure 4.26 highlights single replicate EOP data for two of the evolved lines from that last passage. The data is not particularly resolved, but it is evident that no further improvement in fitness has occurred with respect to relative plating efficiency on Δ*waaC*. Nor is there any evidence of greater efficacy for BW25113 over Δ*ompF*, that would not diminish upon replication.

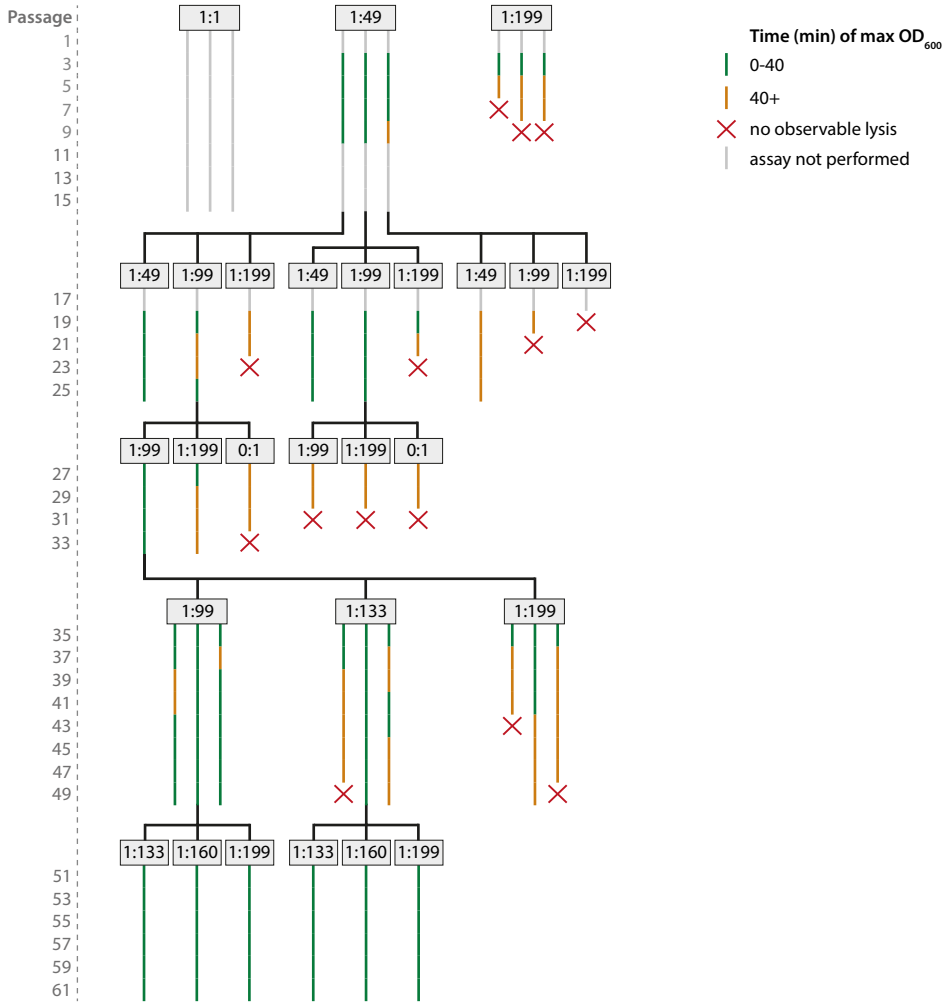


Figure 4.25: Maximum OD_{600} analysis of the 61 passages of the OmpF directed evolution experiment (see Methods 5.9.6). Odd numbered passages are under 1:1 conditions and even number passages are under the boxed conditions. Black lines indicate where an evolution line was carried forward for further passaging. New conditions start with an odd-numbered passage.

Phage	$\log_{10}$(EOP) (BW/Δ*ompF*)	$\log_{10}$(EOP) (Δ*waaC*/BW)
T7 Δ*g5*::*cmk*	-0.46	-7.85
1:49 (15-1)	0.00	-6.00
1:133 (61-1)	-0.52	-5.95
1:133 (61-2)	0.46	-6.30

Figure 4.26: Single replicates from Passage 61 plated with unevolved and 1:49 (15-1) phage for comparison. EOP data for which no plaques were observed on Δ*waaC* are treated as if a single plaque was observed for EOP calculation, and such data points are coloured orange.

4.2.5 Semi-stochastic modelling of the direction evolution experiments

$\Delta waaC$ directed evolution modelling

The tendency of the pairwise competition models to overrate ratios was likely because the potential for more phage mutants to emerge in environments with more positive selection cells was unaccounted for. Here, a semi-stochastic DDE model is developed which introduces a probability of mutation, μ, to an improved phage phenotype (see Methods 5.9.7 for more details). After the passing of a phage latent period in time (loosely considered a phage generation) new phages with the improved phage phenotype are added to the model for each infected cell that wins a virtual coin flip (the probability of winning being $\beta \times \mu$). To simplify the fitness landscape, we assume that fitness improvements are made to one aspect of the phage binding affinity (in the $\Delta waaC$ modelling, this will be be the affinity of the phage to $\Delta waaC$). The parameter λ_j will scale affinity for $\Delta waaC$ in N steps, which in turn represent N distinct phage phenotypes. Otherwise the equations follow in the same manner as before. A quick note though on notation. If an equation reads $\frac{\mathrm{d}A_j}{\mathrm{d}t} = h_j$, this means that equation occurs N times (one for each distinct phage phenotype) so we would have

$$\begin{aligned}
\frac{\mathrm{d}A_1}{\mathrm{d}t} &= h_1 \\
\frac{\mathrm{d}A_2}{\mathrm{d}t} &= h_2 \\
&\cdots \quad \cdots \\
\frac{\mathrm{d}A_N}{\mathrm{d}t} &= h_N \,.
\end{aligned}$$

We then write,

$$\frac{\mathrm{d}C_1}{\mathrm{d}t} = a_1 C_1 - k\sum_{n=1}^{N} C_1 P_n$$
$$\frac{\mathrm{d}C_2}{\mathrm{d}t} = a_2 C_2 - k\sum_{n=1}^{N} \lambda_n C_2 P_n$$

$$\frac{\mathrm{d}I_{1j}}{\mathrm{d}t} = kC_1 P_j - kC_1(t-\tau)P_j(t-\tau)$$
$$\frac{\mathrm{d}I_{2j}}{\mathrm{d}t} = k\lambda_j C_2 P_j - k\lambda_j C_2(t-\tau)P_j(t-\tau)$$

$$\frac{\mathrm{d}P_j}{\mathrm{d}t} = \beta k C_1(t-\tau)P_j(t-\tau) - kC_1 P_j - k\lambda_j C_2 P_j - \underbrace{k\sum_{n=1}^{N} I_{1n} P_j}_{\text{loss of phage that bind to } I_{1j}} - \underbrace{k\sum_{n=1}^{N} \lambda_n I_{2n} P_j}_{\text{loss of phage that bind to } I_{2j}}.$$

Figure 4.27 demonstrates how this model works in practise for a representative example. The simulation begins with one phage phenotype that infects both BW25113 and Δ*waaC* at the same rate (similar to T7-ReRb Δ*g5*::*cmk*). The cellular environment is a 1:1 to 1:49 switching condition. As the simulation proceeds, phage with decreased affinity for Δ*waaC* emerge, and eventually outcompete the predominant phenotype. This simulation is run 100 times for 100 passages. Figure 4.27c is a plot of the percentage of simulations won by a given phage phenotype. The curve for the final phage phenotype with relative affinity to Δ*waaC* of 1/16, referred to as the desired phage phenotype, is then compared with the curves generated by the other passaging conditions. It is in this way that we attempt to determine which strategy might be best for acquiring the desired phage phenotype, for a given set of conditions.

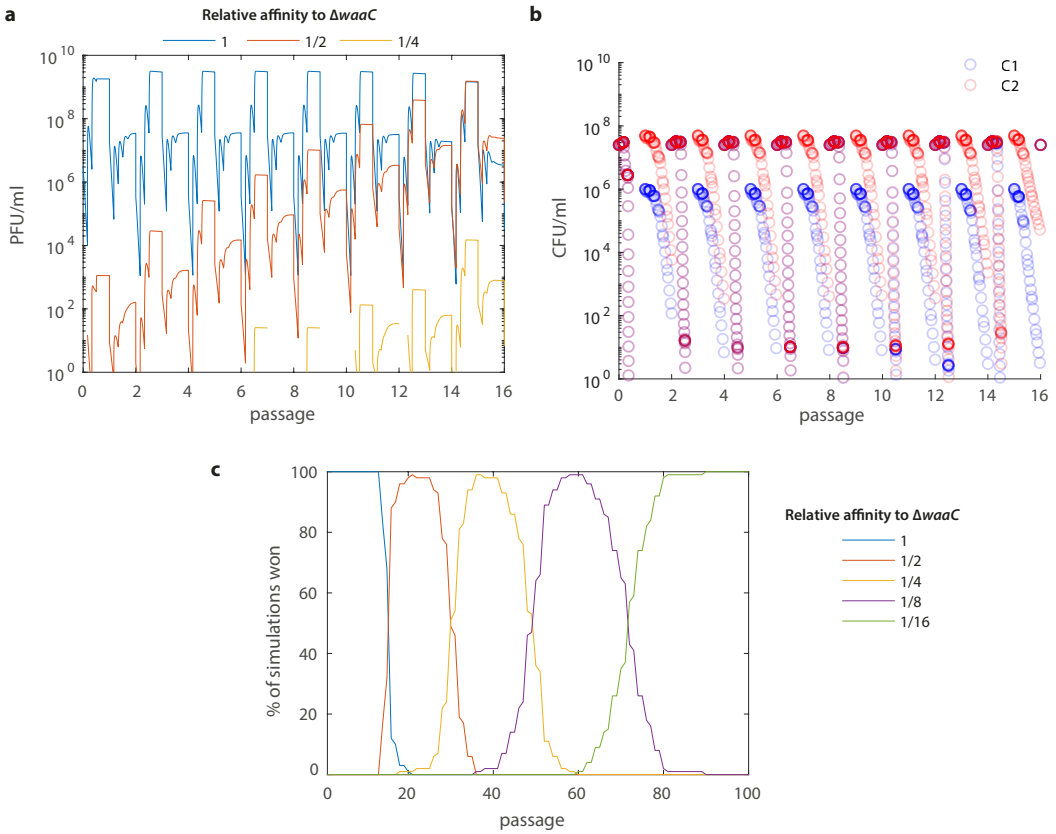


Figure 4.27: A representative example of the mutation model for the $\Delta waaC$ directed evolution experiments. a) PFU/ml for the first 16 passages. Odd numbered passages are a 1:1 ratio of BW25113:$\Delta waaC$. Even numbered passages are a 1:49 ratio. The simulation starts with a phage that binds to both the LPS and $\Delta waaC$ at a set adsorption rate (either 3×10^{-9} or 10^{-8} ml/min). Each phage generation provides an opportunity for a mutant phenotype to emerge which binds half as well to $\Delta waaC$. Note that after each passage $\approx 1\%$ of the phage population is transferred to next passage (the precise number transferred is Poisson distributed). b) CFU/ml for first 16 passages. BW25113 cells, C_1 , coloured blue, $\Delta waaC$ cells, C_2 , coloured red. Overlapping points appear purple and points are plotted discretely for clarity. c) Plot of the number of simulations won by each phage phenotype. Plot is of 100 simulations. Simulations typically look like this, where one phage phenotype outcompetes the previously dominant one, in a sequential fashion, until the optimal phage phenotype (1/16) wins all simulations.

Figure 4.28 shows the results of the semi-stochastic modelling for the $\Delta waaC$ directed evolution experiments. In the case where we set $k = 3 \times 10^{-9}$, $\tau = 12$, $\beta = 100$, switching routines obtain evolved lines with less affinity for $\Delta waaC$ faster than simple ratio conditions (see the 1:1 to 1:99 or 1:1 to 1:133 versus the 1:9 ratio condition in Figure 4.28a). Switching routines are also more robust to changes in phage kinetic parameters (in particular, the binding rate k, see Figure 4.28b). Phage with latent periods of 20 minutes do not evolve the optimal phenotype in either the switching or ratio conditions.

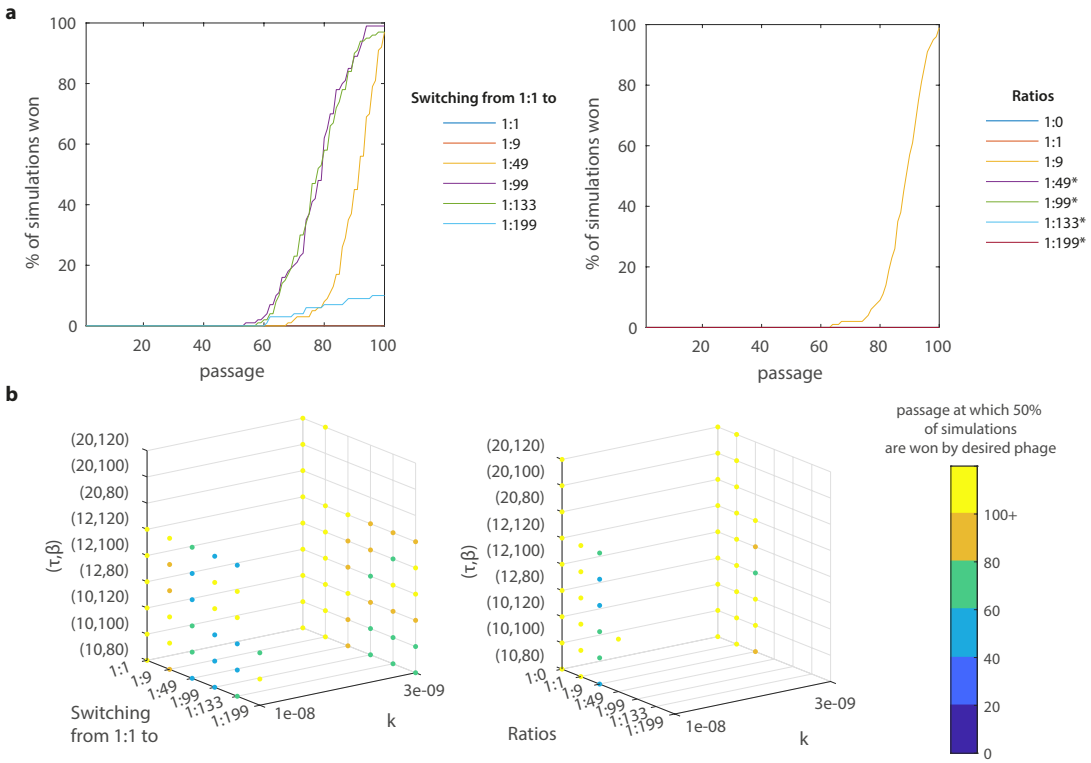


Figure 4.28: Simulations of the mutation model for the $\Delta waaC$ directed evolution experiments. a) The percentage of simulations won by the desired phage phenotype, with relative $\Delta waaC$ affinity $1/16$, for different conditions. In these simulations, initial binding rate, k, is 3×10^{-9}, the latent period, τ, is 12, and the burst size, β, is 100. Asterisks indicate conditions in which the entire phage population washed out. b) Plots presenting the passage at which the desired phage wins 50% of the simulations, for different conditions, and varying k, τ and β. Data points in which all phage phenotypes washed out have been removed. 100 simulations were performed in each instance for 100 passages.

Δ*ompF* directed evolution modelling: increasing OmpF binding affinity

The same idea is applied here, except that now we are looking to increase OmpF binding affinity. We write θ_j for the parameter that scales OmpF affinity by $1/\lambda_j$ in N steps (1 step for each phage phenotype).

$$\frac{\mathrm{d}C_1}{\mathrm{d}t} = a_1 C_1 - \sum_{n=1}^{N}(k + k\theta_n)C_1 P_n$$
$$\frac{\mathrm{d}C_2}{\mathrm{d}t} = a_2 C_2 - k\sum_{n=1}^{N} C_2 P_n$$

$$\frac{\mathrm{d}I_{1j}}{\mathrm{d}t} = (k + k\theta_j)C_1 P_j - (k + k\theta_j)C_1(t-\tau)P_j(t-\tau)$$
$$\frac{\mathrm{d}I_{2j}}{\mathrm{d}t} = kC_2 P_j - kC_2(t-\tau)P_j(t-\tau)$$

$$\frac{\mathrm{d}P_j}{\mathrm{d}t} = \beta k C_1(t-\tau)P_j(t-\tau) - (k + k\theta_j)C_1 P_j - kC_2 P_j$$
$$- \underbrace{(k + k\theta_j)\sum_{n=1}^{N} I_{1n}P_j}_{\text{loss of phage that bind to } I_{1j}} - \underbrace{k\sum_{n=1}^{N} I_{2n}P_j}_{\text{loss of phage that bind to } I_{2j}}$$

Figures 4.29, 4.30 and 4.31 present the results from simulations on evolving phage phenotypes that differ in their affinity for OmpF. Switching routines though successful for many parameter sets and culture proportions, demonstrate a large variability in the number of passages required to generate a dominant optimal phage population. In particular, note that for phages with $k = 10^{-8}$, $\tau = 10$, $\beta = 120$ the 1:1 to 1:9 switching routine requires an excess of 250 passages to generate the desired outcome (Figure 4.31) but that a 1:1 to 1:1 routine, or a 1:1 to 1:49 routine, achieve this in fewer than 200 passages. In contrast, ratio conditions generate better

outcomes with far less variability in the number of required passages. Moreover, the results indicate that a simple but effective heuristic strategy would be to find the least stringent ratio that washes out the phage, and passage at some ratio less stringent than this.

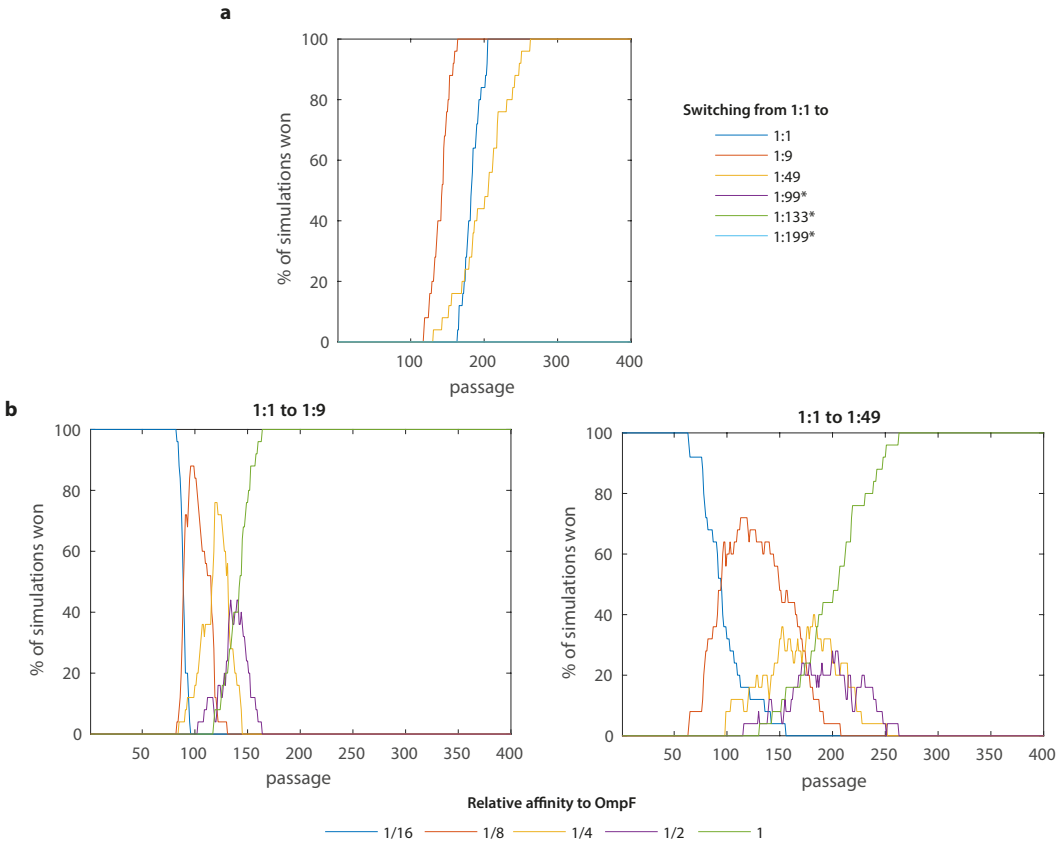


Figure 4.29: Simulations of the mutation model for OmpF positive selection with varying switching conditions in the $\Delta ompF$ directed evolution experiments. a) The percentage of simulations won by the desired phage phenotype, with a relative OmpF affinity of 1, for different conditions. In these simulations, initial binding rate, k, is 10^{-8}, the latent period, τ, is 12, and the burst size, β, is 100. Asterisks indicate conditions in which the entire phage population washed out. b) The percentage of simulations won by each phage phenotype, for the 1:1 to 1:9 (left) and the 1:1 to 1:49 switching conditions. 25 simulations were performed in each instance for 400 passages.

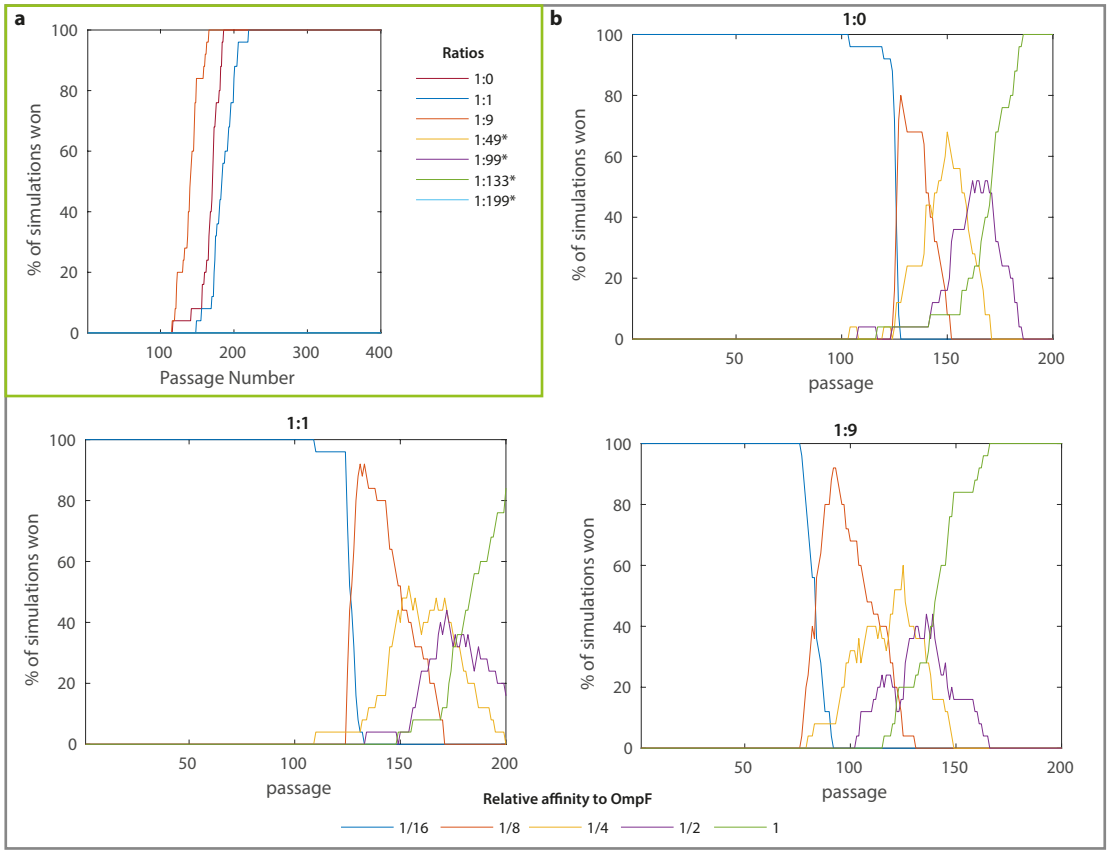


Figure 4.30: Simulations of the mutation model for OmpF positive selection with varying ratio conditions in the $\Delta ompF$ directed evolution experiments. a) The percentage of simulations won by the desired phage phenotype, with a relative OmpF affinity of 1, for different conditions. In these simulations, initial binding rate, k, is 10^{-8}, the latent period, τ, is 12, and the burst size, β, is 100. Asterisks indicate conditions in which the entire phage population washed out. b) The percentage of simulations won by each phage phenotype, for the 1:0 (BW25113 ER-g5 only) 1:1 and 1:9 ratios. 25 simulations were performed in each instance for 400 passages.

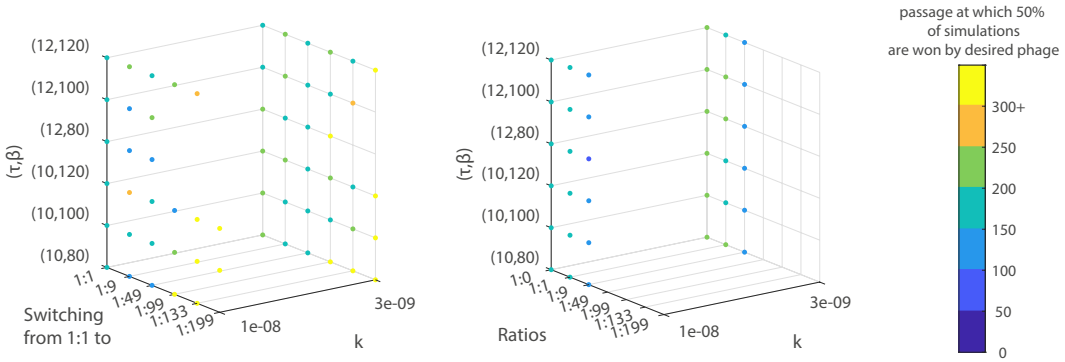


Figure 4.31: Simulations of the mutation model for OmpF positive selection in the Δ*ompF* directed evolution experiments for differing k,τ and β. Plots show the passage at which the desired phage wins 50% of the simulations, for both switching conditions and ratios. Data points in which all phage phenotypes washed out have been removed. 25 simulations were performed for each parameter and condition for 400 passages.

Δ*ompF* directed evolution modelling: decreasing LPS binding affinity

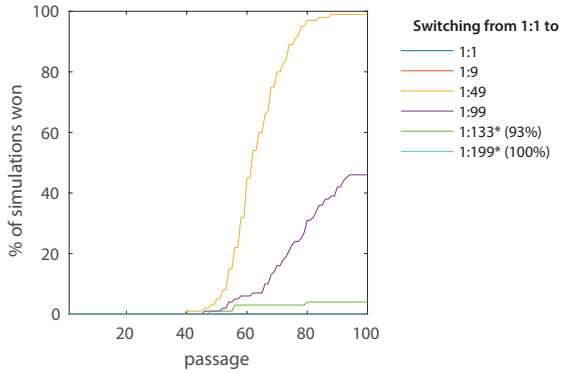


Figure 4.32: Simulations of the mutation model for LPS negative selection in the Δ*ompF* directed evolution experiments for $k = 10^{-8}$, $\tau = 12$ and $\beta = 100$. The percentage of simulations for each condition which washout are provided in brackets if non-zero. No simulations saw a ratio condition which generated a winning desired phage population (a phage with a relative affinity for LPS of 1/16, and an OmpF affinity of k). 100 simulations were performed for each condition for 100 passages.

Here we attempt to decrease the binding affinity of the phage to LPS, now that it has a strong affinity for OmpF. The equations generalise the same way from

the pairwise competition model for OmpF, as the semi-stochastic modelling for $\Delta waaC$ generalises from the $\Delta waaC$ pairwise competition model.

$$\frac{\mathrm{d}C_1}{\mathrm{d}t} = a_1 C_1 - (k + k\lambda_j) C_1 P_j$$
$$\frac{\mathrm{d}C_2}{\mathrm{d}t} = a_2 C_2 - k\lambda_j C_2 P_j$$

$$\frac{\mathrm{d}I_{1j}}{\mathrm{d}t} = (k + k\lambda_j) C_1 P_j - (k + k\lambda_j) C_1(t-\tau) P_j(t-\tau)$$
$$\frac{\mathrm{d}I_{2j}}{\mathrm{d}t} = k\lambda_j C_2 P_j - k\lambda_j C_2(t-\tau) P_j(t-\tau)$$

$$\frac{\mathrm{d}P_j}{\mathrm{d}t} = \beta k C_1(t-\tau) P_j(t-\tau) - (k + k\lambda_j) C_1 P_j - k\lambda_j C_2 P_j$$
$$- \underbrace{(k + k\lambda_j) \sum_{n=1}^{N} I_{1n} P_j}_{\text{loss of phage that bind to } I_{1j}} - \underbrace{k\lambda_j \sum_{n=1}^{N} I_{2n} P_j}_{\text{loss of phage that bind to } I_{2j}}$$

From the simulations performed (see Figure 4.32) we see that switching strategies are required for the timely evolution of the desired phage phentoype. However, more stringent switching conditions do not necessarily result in better outcomes, suggesting that multiple conditions would need to be experimentally tested in parallel to ensure success.

4.3 Discussion

4.3.1 Characterisation of strains used in Δ*waaC* directed evolution experiments

It was important to ensure that the growth rates and CFU of the BW25113 and Δ*waaC* strains were roughly similar, so that one cell population would not overwhelm the other during the course of passage, and influence phage adaptation in an unexpected way (Figure 4.4). Fortunately, both strains were so, with a roughly third more CFU/ml observed for BW25113 ER-g5 compared with Δ*waaC* ΔER-g5, in addition to the recorded mean growth rate presenting as 25% higher for BW25113 ER-g5 over Δ*waaC* ΔER-g5. All passages were performed with cells grown to within an OD_{600} range of 0.17-0.3, which, since cells were mixed in different ratios by volume (see Methods 5.6) meant that there was at most an 2.3-fold deviation from the idealised ratios in any given passage.

Qimron *et al.* [87] observed $\log_{10}(\text{EOP})$ to be between -1 and $1 - \log_{10}(7)$ (approximately 0.15) for T7-ReRb challenged against BW25113 and Δ*waaC*. Both of the means estimated by spot assay and plaque assay are within those bounds,[1] and so are in line with those observations.

[1] Qimron *et al.* calculate EOP using PFU on Δ*waaC* divided by PFU on BW25113, which is the reciprocal of our use.

4.3.2 Simplified models for Δ*waaC* directed evolution experiments

Model design

One of the assumptions we made during the design of the model (see section 4.2.2) was: “Neither the latent period or the burst size adapt to the environment, and adaptation affects the rate of phage binding to cells only." We assume in the modelling that the latent period and burst size do not adapt during the evolution experiments, and that any adaption is limited to the adsorption rate of phage to cells. This assumption is patently untrue. T7 phage almost certainly does adapt its latent period and burst size to the conditions [81, 160] in addition to any changes made to the adsorption constants. It is very likely in fact that in avoiding a given host, T7 might reduce its reversible binding to the cell surface, and may sit less stably on the surface as a result of a weaker irreversible binding interaction. This in turn could lead to slower DNA ejection [179] and possibly a change in the number of progeny produced. That said, altering only the adsorption constant (which approximately accounts for both the reversible and irreversible binding interactions) can still bring the solutions close enough to the true dynamics that we may gain from the enterprise.

Qualitative predictions

Ratios appear to better at selecting for the desired phage when in competition with phages with affinity for Δ*waaC*. Switching however is more tolerant of less fit phages (Figure 4.8b) which, given that we are unlikely to start the experiment

with a phage phenotype resembling the desired phage, may be necessary for the emergence of fitter phages. The pairwise competition and washout models indicate how a given environment might favour a pre-existing phage phenotype, but they do not account for impact the environment may have on the emergence of a given phage phenotype. This is something that we attempt to address with mutation between phage phenotypes in the semi-stochastic models.

4.3.3 Regression of T7-ReRb Δ*g5*::*cmk* to WT T7 cell-binding affinities

EOP characterisation

Molineux and Bull serially passage co-cultures for differing lengths of time, typically before complete lysis of the culture [48, 79, 81, 193, 194]. They do this to prevent the latent period of the phage population syncing with the passage duration [48, 81]. I opted not to follow this approach, figuring that inevitable variations in time at which the passage was removed from the incubator (never greater than 2 minutes) and the time taken to aliquot samples for centrifugation, for subsequent supernatant extraction (at most 6 additional minutes) would be sufficient to prevent synchronisation. They also dilute the population by 10^4 or 10^5 fold (as experience dictated) between each passage. Varying the dilution was not seen to have a marked effect on the outcome of the pairwise competition (Figure 4.7) and so was not explored experimentally. This was likely an oversight on my part. Increasing the dilution rate increases the likelihood of phage washout, and of an individual mutant phage not being transferred to the next passage. However, it also increases the number of generations a phage population will go through in a passage, which in

turn increases the chance of mutant populations emerging, and increasing in number.

The data from the 1st $\Delta waaC$ directed evolution experiment (Figure 4.9) demonstrates that the procedure selects against phage phenotypes that bind to $\Delta waaC$. Phage passaged in a 1:1 ratio of BW25113 ER-g5 and $\Delta waaC$ ER-g5 did not have the same impact on BW25113 plating efficiency over 9 passages compared to that seen on $\Delta waaC$ plates. Nor did passaging in BW25113 ER-g5 cells alone. Whether adaption occurred during the passaging procedure, or whether a pre-existing adaption in the population was selected for can not be known from the data collected, but there is no experimental evidence to suggest that it was pre-existing.

The 2nd round of $\Delta waaC$ directed evolution experiments (Figure 4.10) provided the first confounding evidence that the pairwise competition model might not offer a complete enough picture of the evolution schema. The model (Figure 4.8) predicted that increasing the stringency of the ratios from 1:1 to a ratio of 1:9, or 1:49, would make the procedure more effective at selecting for the desired phage, so long as it did not result in phage extinction. No such phage extinction was observed here, indicating that the model was qualitatively wrong in this aspect. The heuristic prediction which transpired from this insight was that the number of positive selection cells (gp5 complemented cells) impacted the number of opportunities for more fit phage mutants to emerge. This, as it turns out, was not a novel insight, having been alluded to previously by Benmayor *et al.* [84] in their work with mixed ratios co-cultured with phage.

Looking at Figure 4.11, although the variance of the 1:1 to 1:199 switching condition is much larger the 1:1 to 0:1 condition, it did produce a markedly fitter phage

population (Switching 1:199 (18-3)). This was presumed to be because the 1:199 ratio still provided opportunities for T7 replication, in contrast to the 0:1 ratio. Noting that the unevolved phage, T7 $\Delta g5::cmk$, proved so much fitter than the evolved lines (Figure 4.11b) suggests that there was still substantial room for fitness improvements in the phenotypic landscape.

Sequencing of tail genes

Surprisingly, only one of the Qimron mutations (recall Section 3.2.1, Figure 3.6) deemed to be responsible for the T7-ReRb plating efficiency, regressed to the WT T7 sequence during the course of the experiments (see Figure 4.12). The gp12-181G reversion appeared in 1:1 (9-3), but did not appear to substantially improve the EOP (though, the data is not resolved enough to be sure of this). Nor was it further observed in the phage populations that succeeded it. The data instead implies that a different fitness pathway was taken in the genotypic landscape, perhaps because several of the mutations would have needed to occur in combination, in order to increase fitness. The first mutant isolated by Qimron *et al.* [87], T7-Re, picked up 3 mutations in the tail gene proteins (gp11-M6V, gp12-D181G and gp17-N501H) which lends weight to that suggestion. Alternatively, more recent work in the Qimron Lab, indicates that only a single mutation (D540N) is in fact required for T7 to infect $\Delta waaC$ strains [126]. Though this may be at the expense of efficient BW25113 infection, no data is presented on this, and therefore the discrepancy between this research, and that which proceeded it, is is not addressed in that paper. Nevertheless, it does imply that more than one route to $\Delta waaC$ adsorption is possible, and thus, conversely, more than one route through the fitness landscape must exist to avoid $\Delta waaC$ adsorption.

4.3.4 Directed evolution of T7 towards an outer membrane protein (OmpF)

Preparations for the evolution experiments

T7 supposedly interacts with two OMPs during phage infection (OmpA and OmpF [130]) so both were considered as targets for directed evolution. OmpA superficially appears to be a better option. There are roughly 6-fold more copies of OmpA relative to OmpF in exponentially growing BW25113 [20]. However, the growth of *E. coli* appeared to be negatively affected by the removal of OmpA (Figure 4.13) which could complicate the serial passaging procedure. In particular, making it more difficult to culture the strains to a similar OD_{600} prior to their passaging, as well as effecting the relative production rate of T7 in each strain [195]. Moreover, the native conformation of the full OmpA protein sequence is still debated, and it is unknown whether such a confirmation might dimerise (computational work hints that it may well do [196]) or indeed trimerise. These facts taken together influenced the decision to make OmpF the first target protein to be tested. Nevertheless, though $\Delta ompF$ does not have the same impact on the growth rate of *E.coli*, OmpF trimerisation is known to be affected in LPS deficient strains [89, 197, 198]. Further, interaction with LPS moeties is required for trimer formation *in vivo*. The trimerisation is dependent on basic residues in 2 sites on OmpF, and crystal structure data from a *E. cloacae* OmpC ortholog suggests that a $\Delta waaC$ LPS phenotype may only interact with one of those sites [89]. These details will prove important later in determining whether T7 is binding to outer membrane protein(s) independently of the LPS during infection.

Determining relative phage fitness from plate reader absorbance data

This work is not the first to attempt to extract measures of phage productivity from microplate reader optical spectroscopy data [199–206]. Of particular note is the work of Xie *et al.* [204] which uses the following formula for measuring phage virulence:

$$\frac{A_c - A_p}{A_c} = 1 - \frac{A_p}{A_c}$$

where A_p is the background-subtracted estimate for the area under the OD_{600} curve which is generated in the presence of phage, and A_c is the area estimated in the absence of phage. Storms *et al.* use this formula too, but sum the result for different initial MOIs [199]. However, no mechanistic derivation is given for this measure of phage productivity, and so here we derive an ODE model for describing the kinetic relationship between phage and bacteria. This in turn enables comparisons of phage productivity along with a biophysical understanding of the quantity being compared. That being said, the new method described in this thesis (see Figure 4.22) has not been tested over varying MOIs (to determine whether the number of cells, or the number of phage, present, effects the results). So the drop in relative fitness for each evolved line (from 1 to 3) could be due to the correspondingly lower PFUs. However, we would also expect the PFUs observed to be lower if the phage population was less fit, and so in that sense the relative fitness estimated could be operating correctly. A further limitation with this method would be apparent if no lysis was observed for phage-bacteria co-culture. In this case, the lysis curve might not reach a stationary point in the considered time period. One way to resolve this would be to log-transform the underlying data, and allow sufficient time for lysis curve to reach stationary curve. Since the data used here did not possess this

property, no attempt to verify this resolution has been made.

Characterisation of the OmpF directed evolution experiments

As shown in Figure 4.23b, the switching 1:1 to 1:49 passage 15 evolved lines produced a marked difference (an ≈ 100-fold or more improvement) relative to T7 Δ*g5*::*cmk* against the Δ*waaC* ER-g5 strain. This strain had not been presented to the evolved phage up until this point, suggesting that the phage may have adapted to bind to something other than the LPS, in addition to the LPS. However, individual plaques isolated from plates of Δ*waaC* ER-g5 did not perform better on BW25113 than Δ*ompF*(see Figure 4.23d). There are few conclusions one could draw from this. The first is that the evolved phage binds to something other than OmpF, which is not being negatively selected against by Δ*ompF* cells. This could be due to the cell surface of Δ*ompF* presenting fewer of a certain outer membrane protein or structure. Another explanation could be that binding to Δ*waaC* is an artifact of the passaging procedure. Some phage are known, including T7, to produce a residual fraction of the population which binds poorly to the hosts in the current environment, as a way of hedging for changes to its environment [207]. It is not impossible that the stock of phage T7 Δ*g5*::*cmk* has been maintained in a way that has inhibited this feature. Finally, there is a possibility that the evolved phage binds to both LPS and OmpF, but that the LPS binding affinity is more than sufficient in the absence of OmpF. The evidence to hand is too thin to draw definitive conclusions on any of these options though.

15 mutations in total, including silent mutations, were observed in the 1:49 (15-3) passage line over a region of 5310 bp (for genes *cmk*, 11, 12, and 17). If we presume

that there were at most 75 generations in that time (which would be an unlikely 5 generations per passage) a conservative estimate for the mutation rate would be 4×10^{-5} per bp per generation. Note that this estimate does not have access to deleterious mutations (such as frameshifts). Whether these mutations resulted from a true adaption, instead of a population bottleneck, is unclear from the data. The mutations to Cmk are interesting, especially since they suggest that it may be possible to improve the ability of Cmk to aid the T7 life cycle. It was suspected that an adapted tail tube might be the best way to attain T7 infections which bound only to OmpF, since the density of OmpF molecules in the cells surface is far less than for the LPS, and so the likelihood of a sufficient number of OmpF molecules occurring close enough together for T7 to bind to using its tail fibres (gp17) is much smaller. Subsequently, it was encouraging to find 5 mutations in that region for evolved line 1:49 (15-3). The mutation gp17-T118A, which was observed in all evolved lines, has been seen previously in an evolution experiment performed by Heineman & Bull [81], and from its location in the N-terminal domain, is likely involved in docking the tail fibre with gp11. The remaining mutations have not been observed previously in the literature, and so it is difficult to draw conclusions on what their effects may have been. Alanine scanning is a method which replaces individual amino acids in a protein with alanines, to test that amino acid's contribution to the overall structure and functionality of a protein. Alanine is typically chosen as the replacement amino acid because its short side chain (a methyl group, $-CH_3$) is relatively chemically inert in this context. Intriguingly, evolved line 1:49 (15-3) has done this naturally, replacing 3 amino acids in the protein sequence for gp17 with alanine. It may be that these amino acids were involved in reducing the affinity of gp17 to the LPS, but there is no other experimental evidence to support this conclusion. Though 1:49 (15-2) was observed to be significantly different, at

the 5% level, from 1:49 (15-1) and 1:49 (15-3), it should be noted that the p-value for 1:49 (15-2) vs 1:49 (15-3) was 0.0117 (see Table 4.3) and that recent work by Taleb regarding the meta-distribution of p-values suggests that a 10-fold smaller p-value may be required to pass a given significance criteria than traditionally thought [208]. Further elaboration on this is given in Appendix C.3.

4.3.5 Semi-stochastic modelling of the direction evolution experiments

In silico semi-stochastic simulations of the $\Delta waaC$ evolution experiments do show some stark deviations from those recorded in section 4.2.3. For instance, a ratio of 1:9 seems to better select for the desired phage phentoype than a 1:1 ratio (see Figure 4.10). The simulations also indicate that a 1:9 ratio would outperform a 1:1 to 1:9 switching routine, which in turn was unsupported by the experimental evidence (see Figures 4.10 and 4.11). Nevertheless, consistent with the directed evolution experiments, switching routines are qualitatively better at generating dominant phage populations with reduced affinity for $\Delta waaC$ cells.

The simulations presented in Figures 4.29, 4.30 and 4.31 indicate that, in general, the best strategy is find a ratio at which the phage washout, and work back from this until the phage population can tolerate the passages. This came as a surprise, but it perhaps should not have done. It was presumed that the lesson learnt from the $\Delta waaC$ directed evolution experiments would generalise i.e. that conditions which provided more opportunities for the phage to mutate, *whilst also negatively selecting against undesired phenotypes*, would better produce fit phages. That italicised sub-clause is important, and easily misunderstood. In this simulation

(and in the $\Delta ompF$ evolution experiment) we were looking for phage phenotypes which bound with higher affinity to OmpF. The negative selection cells in this experimental set-up, only select against LPS-binding. Since this is not the binding affinity which we directly wish to act on, at least not initially, the heuristic formed from the $\Delta waaC$ experimentation can not be so straightforwardly applied.

Observe also in Figures 4.30a and 4.31 that passaging with BW25113 ER-g5 alone is deemed a sufficient selection pressure to adapt phage to binding more strongly to OmpF. Note that the model here assumes that binding with more affinity to OmpF is not detrimental to LPS binding, as well as assuming that binding affinities to different receptors combine additively (i.e. total adsorption rate = binding rate to OmpF + binding rate to LPS). If these assumptions were correct, then this would suggest that T7 phage should have already found a phenotype which binds strongly to LPS and OmpF when evolving in the wild, if such a phenotype existed.

5 Materials and methods

5.1 Strains

E. coli BW25113 was used throughout for both cloning and characterisation. Gene knockouts referred to herein were derived from BW25113 as part of the Keio collection [43, 44]. The double gene knockout strain BW25113 Δ*cmk* Δ*trxA* was a gift from the Qimron Lab [102].

5.2 Phage

Phages referred to in the text are provided in Table 5.1.

5.3 Cloning

5.3.1 Plasmids

Table 5.2 lists the plasmids used in this work in the order that they appear in the text.

Table 5.1: Phages listed in order that they appear in the text.

Phage	Description
WT T7	A gift from Prof. Andrew Easton which was obtained from the Richardson Lab. The reference genome is NC_001604 with the exception of an additional adenine insertion at position 1897 (which we denote NC_001604^{+}). This insertion is also seen in the **P** isolate from the Hillis *et al.* study [209] as determined through sequencing by Molineux *et al.* (accession number AY264774) [194]. 2 non-canonical mutations to the genome sequence were observed following Sanger sequencing of key T7 gene regions (see section 5.3.8 and Figure 2.4).
T7 Δ*g17*::*trxA*	T7 with genomic tail fibre deletion. WT tail fibres are complemented *in trans* to maintain infectivity.
PM5	A plaque pick from T7 phage that was recombined with the Yep-Phi *g17* C-terminal sequence in the presence of *in trans* T7 gp17 and plated on Δ*trxA* pSEVA551 g17*.
PM6	A plaque pick from T7 phage that was recombined with the Yep-Phi *g17* C-terminal sequence and plated on Δ*trxA* .
PM12F (a.k.a TYP)	Derived from an isolated plaque of PM6 that was plated on Δ*trxA*. Phage from this plaque was amplified in Δ*trxA* and notated PM12F.
T7-ReRb	Provided by the Qimron Lab [87]. Infects both BW25113 and Δ*waaC*. The version used in this study possesses additional mutations gp12-I684F and gp17-L116I.
T7 Δ*g5*::*cmk-trxA*	T7 with *cmk* and *trxA* markers replacing *g5*. Made by Matthew Tridgett as part of the Jaramillo Lab.
T7-ReRb Δ*g5*::*cmk-trxA*	T7-ReRb with *cmk* and *trxA* markers replacing *g5*. Made by Matthew Tridgett as part of the Jaramillo Lab.
T7 Δ*g5*::*cmk*	T7 with the *cmk* marker replacing *g5*.
T7-ReRb Δ*g5*::*cmk*	T7-ReRb with the *cmk* marker replacing *g5*.

Table 5.2: Plasmids used in this work.

Plasmid names	**Replication Origin**	**Antibiotic resistance**	**Description**
pSEVA551	RSF1010	TetR	Cloning vector from SEVA collection [113].
pSEVA551-g17*	RSF1010	TetR	Codon-altered T7 g17 flanked by a synthetic T7 promoter (SBa_000446) [112], a BBa_B0034 RBS, [210] and a T7 terminator (SBa_000587) [112].
pSB3T5-HRg5-cmk-trxA	p15A	TetR	HR plasmid to replace g5 with cmk and trxA in T7.
pSB3T5-HRg17-trxA	p15A	TetR	HR plasmid to replace g17 with trxA in T7.
pSB6A1	pMB1	AmpR	Biobrick cloning vector [114].
pSB6A1-HR-YP	pMB1	AmpR	HR plasmid to replace the C-terminus of g17 of T7 with that of the Yep-phi phage.
pML-HR-trxA	p15A	TetR	Made by Michal Legiewicz as part of the Jaramillo Lab. Used to construct pSB6A1 HR-YP-trxA.
pSB6A1-HR-YP-trxA	pMB1	AmpR	HR plasmid to replace the C-terminus of g17 of T7 with that of the Yep-phi phage, using trxA as a marker.
pET30a+g17_371_467	pMB1	KanR	Plasmid for expression of C-terminal T7 gp17. Made by Marta Sanz Gaitero as part of the Raaij Lab.
pET30a+T7-YP	pMB1	KanR	Plasmid for expression of C-terminal T7 and Yep-phi g17 tail fibre chimera.
pMMB207-Ail	RSF1010	CmR	Plasmid expressing *Y. pestis* KIM5-3001 Ail protein. A gift from the Krukonis Lab [95].

pET24a-gp5-kan	pMB1	KanR	Plasmid for expression of T7 gp5. A gift from the Richardson Lab.
pET24a-gp5-amp	pMB1	KanR	Plasmid for expression of T7 gp5. Constructed from pET24a-gp5-kan by George Kimberley as part of the Jaramillo Lab.
pET24a-gp5-RD-amp	pMB1	AmpR	Plasmid for expression of T7 gp5 with the lacI gene and the Tϕ terminator removed (and so, recombination deficient).
pET24a-gp5-RD-kan	pMB1	KanR	Plasmid for expression of T7 gp5 with the lacI gene and the T*phi* terminator removed (and so, recombination deficient).
pSB3T5-HRg5-cmk	p15A	TetR	HR plasmid to replace g5 with cmk in T7. Constructed by Aurelija Grigonyte as part of the Jaramillo Lab.
pSB4K5-ER-g5	pSC101	KanR	Plasmid expressing error prone T7 DNAP. Constructed by the Jaramillo Lab. Contains D5A, D7A mutations [149], and Y64C, F120L, S399T mutations [152].
pSB4A5-ER-g5	pSC101	AmpR	Plasmid expressing error prone T7 DNAP. Constructed by the Jaramillo Lab. Contains D5A, D7A mutations [149], and Y64C, F120L, S399T mutations [152].
pAJ216	R6K	GentR	Expresses error prone T7 DNAP under the control of a T7 promoter and the Tϕ terminator. Made by Michal Legiewicz as part of the Jaramillo Lab. Used to construct pSB4G5 ER-g5.
pSB4G5-ER-g5	pSC101	GentR	Expresses error prone T7 DNAP expression under the control of a T7 promoter. Used in the directed evolution experiments.

pSB4G5-ΔER-g5	pSC101	GentR	Same as pSB4G5 ER-g5, but with only 6 amino acids of the ER-g5 protein sequence remaining. Used in the directed evolution experiments.

5.3.2 Oligos

gBlocks and primers were synthesised by IDT, details of which are listed in Tables 5.3 and 5.4 respectively.

Table 5.3: List of gBlocks synthesised from IDT.

gBlock	Description
GB002	Codon-altered *g17* (a.k.a. *g17**) under the control of a T7 promoter, an RBS BBa_B0064 from the BioBrick registry, and a synthetic T7 terminator (SBa 000587 [112]).
GB004	*trxA* flanked by homology arms upstream (100 bp long) and downstream (90 bp) of *g17* in the T7 genome.
GB001	Amino acids 486-569 of Yep-phi *g17* flanked by homology arms for replacing amino acids 371-466 of T7 *g17* in the T7 genome.

Table 5.4: Primers used in this work

Primer	**Sequence (5' ⟶ 3')**	**Description**
oPM13	TTTTTTGTCTAGCTAGCTAGGGATCCTCTAGAGTCGACCTGC	Gibson primers to construct pSEVA551-g17*
oPM14	CAAGTGGGAGGCCGTCACCGCTGATGCCATGGTGGTCAAA	
oPM15	TTTGACCACCATGGCATCAGCGGTGACGGCCTCCCACT	
oPM16	CCCATTAGTGAGTCGTATTACCGGGTACCGAGCTCGAAT	
oPM17	AATTCGAGCTCGGTACCCGGTAATACGACTCACTAATGGGAGA	
oPM18	AGGTCGACTCTAGAGGATCCCTAGCTAGCTAGACAAAAAACCC	
oPM23	TTAACAACGAATTGATTTGATTGAACTAGTAGCGGCCGCT	Gibson primers to construct
oPM24	TACGCTCCCTCAAGTTAACAGAGTCACTAAGGGCTAACTAACTAA TTACG	pSB3T5-HRg17-trxA
oPM25	TAGTTAGCCCTTAGTGACTCTGTTAACTTGAGGGAGCGTA	
oPM26	AGCGGCCGCTACTAGTTCAATCAAATCAATTCGTTGTTAAA	
oPM46	CGAGCAGGACTTCATGACTG	Primers for characterising WT T7 and
oPM47	AACATCCGGCTGCAACTTAC	T7 Δ*g17*::*trxA*
gp17fwd	TCGGCTGGCTTTGTGGCTAACG	
AG061rev	CCCCACCGCCAGCACTAC	
oPM58	GGCCTGCAGGAGTCACTACTAGTAGCGGCCGCTG	Gibson primers to construct
oPM59	CGCCGCGCGCCCCGAAGTTAGTTTCGAACTAAGATTTGC	pSB6A1-HR-YP-trxA
oPM60	TCTTAGTTCGAAACTAACTTCGGGGCGCGCGGCG	
oPM61	CAGCGGCCGCTACTAGTAGTGACTCCTGCAGGCCTTAATCAATT CGTTGTT	
oPM54	TCTTAGTTCGAAACTAATAAGCTTGCGGCCGCACTCGAGC	Gibson primers to construct pET30a+T7-YP
oPM55	CATACCTCAGTCCAACCAGAGGACTTCGCAACGAAGCTGTCACG	
oPM56	ACAGCTTCGTTGCGAAGTCCTCTGGTTGGACTGAGGTATG	
oPM57	GCTCGAGTGCGGCCGCAAGCTTATTAGTTTCGAACTAAGATTTG	

Primer	Sequence (5' → 3')	Description
oPM262	GCGAGATGATTAAAGCTATGAG	Coinfection characterisation:
oPM263	TCTTCGATTGCTTTTGTGTC	Coinfection characterisation: primer binding in reverse orientation inside of *g17*
oPM264	GACCAAGTACACCGATAAGG	Coinfection characterisation: primer binding in forward orientation inside of *g17*
oPM265	CACGAGATTGAATCTGTCGG	Coinfection characterisation: primer downstream of *g17*
oPM195	GGCCCCGTGGCCGGGGGACTGCCTTCAACCCAGTCAGCTCCTTC	Gibson primers to construct pET24a-gp5-RD-amp and pET24a-gp5-RD-kan
oPM196	AGCTTGCGGCCGCACTCGAGCCGGATTGGCGAATGGGACG	
oPM197	CGTCCCATTCGCCAATCCGGCTCGAGTGCGGCCGCAAGCTTTCA	
oPM198	GAGCTGACTGGGTTGAAGGCAGTCCCCCGGCCACGGGGCC	
oPM203	AAACAAATAGGGGTTCCGCGACGTGGCTTTGTTGAATAAA	Gibson primers to construct pSB4A5 ER-g5
oPM204	CACTGATTAAGCATTGGTAACAGTTTCATTTGATGCTCGA	
oPM205	TCGAGCATCAAATGAAACTGTTACCAATGCTTAATCAGTGAGGC	
oPM206	TTTATTCAACAAAGCCACGTCGCGGAACCCCTATTTGTTT	
oPM247	CAAGTACCGCCACCTAACAACAGTTTCATTTGATGCTCGA	Gibson primers to construct pSB4G5 ER-g5
oPM248	TCGTTGCTGCTGCGTAACATACTCTTCCTTTTTCAATATTATTGA	
oPM249	AATATTGAAAAAGGAAGAGTATGTTACGCAGCAGCAACGA	
oPM250	TCGAGCATCAAATGAAACTGTTGTTAGGTGGCGGTACTTGG	
oPM251	ATTTAGTGGCAAATCGCCCAGTGATATCTCCTCTGTTTATGG	Gibson primers to construct pSB4G5 ΔER-g5
oPM252	GCTTTGTTGAATAAATCGAACTTTTGCTGAGTTGAAGGAT	
oPM253	ATCCTTCAACTCAGCAAAAGTTCGATTTATTCAACAAAGCCA	
oPM254	ATAAACAGAGGAGATATCACTGGGCGATTTGCCACTAAAT	

5.3.3 Transformations

Electrocompetent cells were made following a protocol from the Krantz lab (UC Berkeley [211]) with some practical alterations (centrifugation steps were performed using a 5920R Eppendorf Centrifuge and 50 ml Falcon tubes). 20-200 ng DNA was then mixed with 50 µl electrocompetent cells that were thawed on ice. This mixture was added to a 2 mm path length electroporation cuvettes that had been chilled on ice, and were singularly pulsed at 2.50 kV in a Bio-Rad MicroPulser on the Ec2 setting, before immediate re-suspension in 950 µl SOC or LB media pre-warmed to 37 °C. The resulting suspension was incubated at 37 °C for 45-60 minutes, and then plated on agar plates infused with the appropriate media/working concentrations of antibiotics.

5.3.4 PCR

PCR was largely performed according to the manufacturer's instructions, with only the concentration of template DNA differing. The thermocycling routine routines used for both NEB Phusion polymerase mastermix and Thermo Scientific Taq polymerase mastermix are provided in Table 5.5 and Table 5.6 respectively. Annealing temperatures were calculated using the NEB Tm Calculator for Phusion reactions and using Promega or ThermoScientific T_m calculators for Taq polymerase. In PCRs involving T7, the phage was treated as if it were DNA template (typically comprising 2% of the total volume for each reaction).

Table 5.5: Phusion PCR thermocycling routine

Cycle step	Cycle(s)	Temperature °C	Time
Initial denaturation	1	98	30 s
Denaturation	25-35	98	10 s
Annealing		45-72	30 s
Extension		72	30 s/kb
Final extension	1	72	7 m
Hold	1	4-12	-

Table 5.6: Taq PCR thermocycling routine

Cycle step	Cycle(s)	Temperature °C	Time
Initial denaturation	1	95	2 m
Denaturation	25-35	95	1 m
Annealing		45-65	1 m
Extension		72	1 m/kb
Final extension	1	72	5 m
Hold	1	4-12	-

5.3.5 Agarose gels

1% (w/v) agarose gels were made with Hi-Res Standard Agarose (Cambridge Reagents) in 1 x TAE and stained with either GelRed (Biotium) or SYBRsafe (Invitrogen). Each gel ran in electrophoresis tanks buffered with 1x TAE, and were visualised using UV (BioDocAnalyse BioMetra) or blue light (IO Rodeo, Large Blue LED transilluminator).

5.3.6 Gibson Assembly

Primers used for Gibson Assembly were designed using the Benchling Gibson Assembly Wizard and screened to ensure that secondary structures (e.g. hairpins) were not predicted to form *in silico* (IDT Oligo Analyzer).

Gibson mastermix was made following the protocol of Hillson *et al.* [212]. Briefly, fragments were designed using the Benchling Gibson Assembly Wizard to have overlapping homology regions (roughly 40 bp) following Phusion PCR amplification in 50 µl volumes. Each reaction was subjected to digestion by 1 µl DpnI for 15 minutes at 37 °C prior to gel extraction of relevant amplicons. Fragments were mixed with 15 µl Gibson mastermix in equimolar amounts up to a total volume of 20 µl, and incubated for 60 minutes. Amplicons that included a resistance cassette which matched the antibiotic resistance selected for were individually treated in the same fashion (to control for vectors carried over from the PCR). For each Gibson reaction, 0.33 µl was then electroporated with 50 µl of competent cells. Colonies were screened by PCR and mini-prepped before Sanger sequencing.

5.3.7 DNA purification

Plasmids were mini-prepped using Thermo Fisher Scientific DNA purification kits. PCR products were column purified using the Wizard SV Gel and PCR Clean-Up System (Promega).

5.3.8 Sequencing

Sanger sequencing was supplied by GATC Biotech AG and samples were provided according to their instructions. T7 was PCR amplified as described and gel-extracted before being sent to GATC. Regions of the T7 genome containing genes 5, 11, 12 and 17 were regularly sequenced. See Table 5.7 for a list of the primer pairs used to amplify those sequences, and Table 5.8 for the oligos that primed the

Sanger sequencing reactions for each gene.

Table 5.7: Forward and reverse primers used for PCR of T7 gene regions for sequencing

Gene region	Name	Sequence (5' → 3')	Annealing temperature (°C)	Extension time
g5	oPM182	GGTGTTACTCCACGCGGTGCAA	72	1 m 11 s
	oPM183	ACCAAGTCCTCTTTACTCTGAGTCAGCA		
g11	oPM049	AACTCGGCTCAATACCAGCAG	65	1 m 40 s
	oPM050	CCTTAACGCCAGCACAGGAG		
g12	oPM048	CGTCTCTGCATGGAGTATGAGATG	68	1 m 21 s
	oPM049	AACTCGGCTCAATACCAGCAG		
g17	oPM046	CGAGCAGGACTTCATGACTG	64	1 m 10 s
	oPM047	AACATCCGGCTGCAACTTAC		

Table 5.8: Primers used to Sanger sequence the T7 gene regions

Gene region	Name	Sequence (5' → 3')
g5	oPM182	GGTGTTACTCCACGCGGTGCAA
	oPM183	ACCAAGTCCTCTTTACTCTGAGTCAGCA
g11	oPM050	CCTTAACGCCAGCACAGGAG
g12	oPM048	CGTCTCTGCATGGAGTATGAGATG
	oPM299	CTGACGGTAATTTCGACTTCAAGTGG
	oPM328	CTCCTAACGGCTACATGGTG
	oPM300	GGAGACCACTCAAGCCACTTGAAG
	oPM049	AACTCGGCTCAATACCAGCAG
g17	oPM021	TGTTAACTTGAGGGAGCGTA
	AG012	GAAGTAGATTCCATCGGGGCC
	oPM047	AACATCCGGCTGCAACTTAC

5.3.9 Next-gen sequencing

DNA was extracted from phage samples and library prepped for sequencing by Aurelija Grigonyte, and sent for Illumina Next-gen sequencing by Andrew Millard. The quality of the sequence data was assessed using fastqc. Poor quality reads were trimmed using sickle, with length threshold 100, and quality threshold 20. Contigs were then assembled using spades.py, and filtered for contigs exceeding 10 kb. Remaining contigs were then indexed and mapped against the trimmed reads using the Burrows-Wheeler Alignment Tool.

5.4 Phage techniques

5.4.1 Plaque assays

Colonies incubated overnight were refreshed 1:200 into LB with appropriate antibiotics at working concentration until observed to be in log phase. Phages were serially diluted and mixed with said cells (100 µl phage with 300 µl cells) at 30 second intervals, and each mixture was incubated at RT for 10 minutes. Molten LB agar diluted 1:1 with LB (soft agar) was then added to each mixture at the 10 minute mark, before promptly being poured on to a hard agar plate, and allowed to set for at least 20 minutes at RT. Plates were then inverted and incubated overnight at 37 °C. All plaque assays were performed with controls for phage and cellular contamination as necessary in the phage stock, media, soft agar and refreshed cells.

5.4.2 Spot assays

Colonies incubated overnight were refreshed 1:200 into LB with appropriate antibiotics at working concentration until observed to be in log phase. 900 µl of culture was mixed with ≈ 9 ml molten soft agar and poured onto pre-made 12 cm × 12 cm hard agar plates and allowed to set at RT for 5 mins. 10-fold serial dilutions of phage were then pipetted onto the spot assay plate 3 µl at time, taking care so as to not introduce bubbles. Each serial dilution was performed in triplicate.

5.4.3 Plaque picks

Clear and isolated plaques were picked from their centres by attempting to aspirate 3 µl with a Gilson P20 pipette and tip. Picked phage were dispensed into at least 1 ml LB, and mixed by pipetting and/or by briefly vortexing. These mixtures were then filtered using a 0.22 µm cellulose acetate filter (Sartorius, MiniSart).

5.4.4 Phage amplification

Colony picks incubated overnight were refreshed 1:200 in LB mixed with a working concentration of antibiotic, and incubated at 37 °C, 200 rpm. During log phase growth, the cells were inoculated with phage (MOI $\leq$ 0.01) and incubated until lysis was observed (typically 90-150 minutes later). Phage/cell mixtures were then spun down at $\geq$ 3000 g for at least 10 minutes, before the resulting supernatant was filtered through a 0.22 µm cellulose acetate filter (Sartorius, MiniSart).

5.4.5 Phage recombineering

Phages were amplified (section 5.4.4) in cells chromosomally deficient in a selection marker (either *trxA* or *cmk*). These cells had been transformed with a plasmid that bore homology arms to a targeted region of the phage genome, and the intended cargo to be transferred to the phage (selection marker included). Following amplification, and phage extraction, phages were plated on either $\Delta trxA$ or Δcmk (with any necessary gene complementation supplied *in trans*) and plaques were isolated for characterisation, and further purification.

5.5 Bacterial culture techniques

5.5.1 Working/stock antibiotic concentrations

Table 5.9: Working and stock concentrations for antibiotics use.

Antibiotic (abbrv.)	**Stock concentration (mg/ml)**	**Working concentration (µg/ml)**
Ampicillin (Amp)	100	100
Chloramphenicol (Cm)	35	35
Gentamicin (Gen)	10	10
Kanamycin (Kan)	50	50
Tetracycline (Tet)	10	10

The working and stock concentrations employed for different antibiotics is provided in Table 5.9. In cases where 3 different antibiotics were added to media, the working concentration for each antibiotic was halved (unless the antibiotic was ampicillin, which was always used at a 100 µg/µl working concentration).

5.5.2 Hard agar plates

20-25 ml molten LBA was mixed with a 1:1000 dilution of stock concentration antibiotic (except where multiple antibiotics were used, see section 5.5.1) and pipetted into 9 cm diameter petri dishes.

For spot assays, 50 ml of molten LBA was mixed with a 1:1000 dilution of stock concentration antibiotic (except where multiple antibiotics were used, see section 5.5.1) and pipetted into 12 cm x 12 cm square petri dishes. This formed the base of the double overlay.

5.5.3 Colony picks

Colonies picked from a plate were mixed into LB media containing a working concentration of antibiotic and typically incubated overnight at 37 °C.

5.5.4 Plate reader assays

Plate reader assays were performed using the Tecan Infinite F500, or the Tecan SpectraFluor. Details varied depending on the particular assay performed, but in general, isolated colonies were picked and grown in media with antibiotics (at appropriate working concentrations) overnight at 37 °C in a shaking incubator. The following day, cultures were refreshed in new media and antibiotics until in log-phase. Cultures were then aliquoted into a 96 well plate and incubated in the plate reader at 37 °C. Blank wells were filled with media + antibiotic. The microplate was shaken orbitally for 20 seconds (amplitude 2 mm) prior to OD_{600} measurements (10 reads of a 2 × 2 square per well) followed by further shaking for

120 seconds (orbital, amplitude 2.5 mm).

For analysis, the background absorbance of sterile medium in control well was subtracted from the recorded test well measurements. Technical replicates were averaged, and biological replicates were represented by equivalent cultures generated from separate colonies.

5.5.5 Growth rate assays

The procedure follows that of Hall *et al.* [213]. The slopes of log-transformed plate reader data were calculated over a 5 point sliding window. For each biological replicate, the time points of the slopes that were at least 95% of the steepest gradient were used to determine a range of time points to consider. The slopes in that range were then averaged, before the biological replicates were further averaged and the standard error of the mean, or the 95% confidence interval was calculated.

5.6 Serial passaging protocol

Table 5.10: Phage and strains used in the directed evolution experiments

Experiment	**Progenitor phage**	**Positive selection strain**	**Negative selection strain**
Δ*waaC*	T7-ReRb Δ*g5*::*cmk*	BW25113 pSB4G5-ER-g5	Δ*waaC* pSB4G5-ΔER-g5
Δ*ompF*	T7 Δ*g5*::*cmk*	BW25113 pSB4G5-ER-g5	Δ*ompF* pSB4G5-ΔER-g5

Three colonies from both the positive selection strain and the negative selection strain were incubated at 37 °C, 200 rpm, with agitation overnight in LB media

infused with a working concentration of gentamicin. The next day, cells were refreshed 1:200 in 10 ml of pre-warmed LB + gentamicin until they reached OD 0.17-0.31. For each condition and replicate tested, biological replicates of positive and negative selection cells were mixed (v/v) according to a pre-determined procedure to a total volume of 2 ml. For the first passage $\approx 2 \times 10^6$ PFU was added to each replicate. Subsequent passages were infected with 1% of the phage population extracted from the previous passage. After phage addition, mixtures were briefly vortexed to ensure sample homogeneity, and then incubated at 37 °C, 200 rpm, for 60 minutes. 1 ml of each replicate was then centrifuged at $\approx 14,000$ g. The removed supernatent was mixed with 1% cholorform, briefly vortexed, and archived at 4 °C for further passaging and/or characterisation.

5.7 Mathematical modelling

All modelling performed during this work was simulated in MATLAB (versions 2017 and 2018). Delay differential equations (DDEs) were evaluated using the dde23 solver.

5.8 Chapter 1 specific methods

5.8.1 Δ*waaC* pMMB207-Ail plate reader characterisation

The procedure follows section 5.9.4 with a few exceptions. Single colonies were grown overnight in LB media mixed with working concentrations of kanamycin and chloramphenicol. Overnight cultures underwent a 1:200 dilution in LB media (+ antibiotics) with varying concentrations of IPTG, before being transferred in

200 µl aliquots to a 96 well plate. As previously described, the plate was then incubated in plate reader at 37 °C for measurement.

5.9 Chapter 3 specific methods

5.9.1 CFU assays

Cells cultured overnight in LB and appropriate working concentrations of antibiotics were refreshed 1:200 in fresh LB + antibiotics until the cells were sufficiently dense ($\geq$ 0.1 OD_{600}). Aliquots were then taken from the incubating cells, and serially diluted, prior to OD measurement. Appropriate serial dilutions were plated immediately and left to incubate overnight at 37 °C. Linear regressions were performed in R using the lm function and the mean OD_{600} to CFU estimates for each biological replicate.

5.9.2 $\Delta ompA$ and $\Delta ompF$ growth rate comparison

The procedure follows section 5.9.4 with a few exceptions. Single colonies were grown overnight in LB media mixed with working concentrations of kanamycin. 50 µl of the overnight cultures were mixed with 10 ml LB media + kanamycin before being transferred to a 96 well plate (200 µl aliquots). As previously described, the plate was then incubated in the plate reader at 37 °C for measurement.

5.9.3 Pairwise competition and phage washout modelling

Models were simulated using the parameters listed in Table 5.11. Data points from the presented simulations were removed in instances where the final PFU/ml of the phage population was less than 10^4 in the absence of the desired phage phenotype.

Table 5.11: Parameters and initial conditions used for the pairwise competition and phage washout models.

Notation	Description	Value	Source
$C_i(t=0)$	Initial concentration of each phage-susceptible cell strain i	5×10^7 ml^{-1}	Rounded-average of experimentally-derived CFU/ml for BW25113 pSB4G5 ER-g5 and Δ*waaC* pSB4G5 ΔER-g5 at 0.25 OD_{600} (see section 4.2.1).
$P(t=0)$	Initial concentration of free phage	3×10^6 ml^{-1}	Initial PFU/ml used for Δ*waaC* and Δ*ompF* directed evolution experiments.
a_i	Growth rate of cell strain i	0.026 min^{-1}	Rounded-average of experimentally-derived growth rates for BW25113 pSB4G5 ER-g5 and Δ*waaC* pSB4G5 ΔER-g5 OD_{600} (see section 4.2.1).
k_i	Maximum adsorption rate of a phage to receptor i	10^{-8} ml min^{-1}	Largest absorption rate observed from adsorption assays between T7-ReRb, BW25113 and Δ*waaC*. Data not shown.
τ	Latent period of phage	12 min	Jaramillo lab experimental data [214].
β	Burst size of phage (number of phage progeny)	100	[215].
N/A	Percentage of total phage population transferred between passages	1%	Used for Δ*waaC* and Δ*ompF* directed evolution experiments.
N/A	Number of passages	20	Upper limit for the number of planned passages for the Δ*waaC* directed evolution experiment

5.9.4 Lysis curve experiment (Relative fitness assay)

The procedure follows section with a few exceptions. Single colonies were grown overnight in LB media mixed with working concentrations of gentamycin. Overnight cultures were refreshed 1:200 in LB media + gentamycin before being transferred to a 96 well plate (200 µl aliquots). 20 µl of each phage sample, or room temperature LB media, was added to the cells in each well. The approximate PFU added to each well, for each phage sample, is tabulated below.

Table 5.12: Approximate PFU added to each well. Phage samples are notated according to which passage, and then passaging line, they were extracted from (i.e. Passage-Line).

Phage sample	**≈PFU added to each well**
T7 $\Delta g5::cmk$	2×10^4
10-1	10^5
10-2	10^4
10-3	5×10^2

The plate was then incubated in plate reader at 37 °C for measurement, as usual.

5.9.5 Relative fitness modelling

To derive an approximate expression for estimating the relative fitness, the "true values" of $W\left(\frac{\alpha_i \gamma_i}{P_0}\right)$ were obtained numerically in Mathematica over the domain $1 \leq \alpha_i \leq\ 10^{-3},\ 10^{-3} \leq \gamma_i \leq 10^{-1}$ and $10^{-7} \leq P_0 \leq 1$ and imported to MATLAB in a h5 data file. The form of the approximate expression was determined by trial and error to be $b_1 \ln(b_2 \alpha_i) \ln(b_3 \gamma_i) \ln\left(b_4/P_0\right)$, where $\gamma_i = r_i \left(1 - \frac{C_i}{K_i}\right)$, and the parameters b_1, b_2, b_3, b_4 were fit to the numeric data using the nlinfit MATLAB

function (with a mean squared error of 0.737) before rounding to nearest significant digit.

Lysis curves were imported into R and the parameter α_i was estimated using the trapezoid method of the AUC (area under the curve) function. The growth rate for each strain, r_i , was estimated using the nls (nonlinear least squares) function and the carrying capacity for each strain, K_i , was taken to be the maximum OD_{600} observed in the cells only assays.

5.9.6 Maximum OD_{600} analysis of the lysis curves assays following $\Delta ompF$ passaging

Approximately 10 minutes after a 1:1 (BW25113 ER-g5:$\Delta ompF$ ΔER-g5) passage was removed from the incubator, 200 µl of each sample was aliquoted into a 96 well plate and incubated at 37 °C in either the Tecan Infinite F500 or the Tecan SpectraFluor plate reader for at least 3 hours. The plate reader conditions followed the methods established in section 5.9.4. The data presented in Figures 4.21b and 4.25 uses the time at which the maximum OD_{600} was observed in each of these assays.

5.9.7 Semi-stochastic mutation models

Each simulation was performed with a different set of a pseudo-random numbers (i.e. a new seed). Within each passage, infected cells generate phage phenotypes with a 2-fold smaller binding affinity to the undesired receptor than the originally infecting phage. After τ time has passed (the length of the phage latent period) the number

of new phage which have mutated to the new phage phenotype is simulated using the binornd function (with parameters $n = \sum_{i=0}^{2} \beta I_{1i}$, $p = \mu$, where $\mu = 10^{-5}$). To minimise the computational load, when $n > 10^6$, the binomial distribution is approximately Poisson and so the number of phage generated is instead estimated by the Poisson distribution using the poissrnd function ($p = \sum_{i=0}^{2} \beta\mu I_{1i}$). The new phages are then cycled back into dde23 solver via the options 'initialY' argument so that the next period of length τ can be simulated. At the end of each passage, $\approx 1\%$ of the total phage population is transferred to the next passage, the precise number being determined by the poissrnd function ($p = 0.01$).

Table 5.13: Parameters and initial conditions used for the semi-stochastic modelling.

Notation	Description	Value(s)	Source
$C_i(t=0)$	Initial concentration of each phage-susceptible cell strain i	5×10^7 ml^{-1}	Rounded-average of experimentally-derived CFU/ml for BW25113 pSB4G5 ER-g5 and $\Delta waaC$ pSB4G5 ΔER-g5 at 0.25 OD_{600} (see section 4.2.1).
$P(t=0)$	Initial concentration of free phage	3×10^6 ml^{-1}	Initial PFU/ml used for $\Delta waaC$ and $\Delta ompF$ directed evolution experiments.
a_i	Growth rate of cell strain i	0.026 min^{-1}	Rounded-average of experimentally-derived growth rates for BW25113 pSB4G5 ER-g5 and $\Delta waaC$ pSB4G5 ΔER-g5 OD_{600} (see section 4.2.1).
k_i	Maximum adsorption rate of a phage to receptor i	3×10^{-9} and 10^{-8} ml min^{-1}	The smaller adsorption rate was taken from [81]. This value was re-estimated by Bull *et al.* [172] and the rounded upper limit of this is 10^{-8}.
τ	Latent period of phage	10, 12 and 20 min	Smallest value incremented by 1 minute from the minimum observed DNA translocation time for T7 (9 minutes, [62]). 12 minute latent period is from experimental data recorded previously in the Jaramillo lab [214]. The largest latent period simulated is 2 times the smallest value used.
β	Burst size of phage (number of phage progeny)	80, 100 and 120	The value used by Endy *et al.* [215] varied by $\pm$20%.
μ	The rate at which mutations that cause increased fitness (better adsorption rates) occur.	10^{-5}	Selected from range. Lower bound for range taken to be the DNA substitution rate of WT T7 DNAP *in vitro* (2.2×10^{-6}) [216]. Upper bound for range is 93-fold larger than this (ER-g5) i.e. $\approx 2 \times 10^{-4}$ (see section 3.3.5).
N/A	Percentage of total phage population transferred between passages	1%	Used for $\Delta waaC$ and $\Delta ompF$ directed evolution experiments.
N/A	Number of passages	100-400	Limited physical meaning can be applied to the values since the evolutionary landscape has been abstracted.

6 Conclusions and further work

6.1 Conclusions

The overriding aim of this PhD was to establish a methodology for controlling the receptor-binding affinity of T7 phage, and to achieve this in a way that could be generalised to other non-temperate lytic phage. This objective was not fully realised, but we nevertheless laid the foundation for further work in this area. In chapter 2, we engineered a phage with tail fibres fused between two different species, and saw that it was still able to infect *E. coli*. Though we could not show that infection was dependent on a heterogeneously expressed receptor, we nevertheless demonstrated the modularity of T7 receptor-binding domains. In chapter 3, we discovered that T7 phage complemented with certain pET vectors (i.e. containing the Tϕ terminator) were liable to undergo homologous recombination with those plasmids. As a result, we unearthed evidence of co-infection between T7 phages which would otherwise be replication-deficient if infecting cells individually. In chapter 4, we proposed a method for a comparing phage fitness on 2 different strains using 600 nm absorbance data, which should be effective so long as the

relationship between OD and CFU is the same for both compared strains. We also developed a platform for controlling T7 phage receptor use through host avoidance. In addition, we attempted to expand this system to evolve T7 phage towards receptors available in the Keio collection (or potentially for those expressed heterogeneously). Finally, throughout this work we modelled the strategies for implementing directed evolution, and learnt that there may be environments in which an otherwise successful strategy is less successful (and vice versa). To conclude, the breakthroughs made in understanding and implementing positive and negative selection when adapting T7 phage to novel receptors should prove useful to those wishing to build on this work. As antibiotic resistance looms on the horizon, developing phage therapies that can be flexibly adapted to the particular pathogen at hand would no doubt be advantageous, and it is my hope that this thesis can be of some benefit to those attempting to do so.

6.2 Further work

To continue this work, the TYP phage would need re-cloning, ensuring that there are no unexpected mutations to the chimeric tail fibre. This phage should then be re-plated on *E. coli* cells to demonstrate conclusively that the chimeric tail fibre is capable of enabling T7 infection. Growth curve analysis of BW25113 transformed with pMMB207-Ail would help determine whether the truncated LPS phenotype ($\Delta waaC$) was contributing to the weakened growth of $\Delta waaC$ pMMB207-Ail under IPTG induction. Further, western blotting of the BW25113 pMMB207-Ail is required to evidence that the heterogeneously expressed protein is translocated to outer membrane in this particular strain of *E. coli* either in the presence or absence of IPTG.

For the work in chapter 3, it would be beneficial to generate more co-infection data to better test and validate the predictions the models make. In addition, an agent-based model, which would be able to look at individual phage-cell interactions, rather than modelling what happens at the population level, might be an improvement on the models described here. To further evidence that it is the pET24-gp5 plasmid which is recombining with the phage to provide g5 in exchange for g10, we could amplify T7 $\Delta g5::cmk$ in 2 separate conditions: 1) in cells containing the pET24-gp5 plasmid, 2) in cells with where that plasmid has been modified so that it does not possess the Tϕ terminator (pET24-gp5-ΔTϕ). Plating of the amplified phage from both conditions on the cells used in the latter would demonstrate the dependence on pET24-gp5 for this particular co-infection phenomenon. Regarding the use of the error prone polymerase (expressed by pSB4G5-ER-g5) to diversify the T7 genome *in vivo*, this thesis would have gained from a characterisation of the error rate per base pair that the polymerase incurred.

Chapter 4 leaves a lot of scope for further investigation. It would have been interesting to vary the duration of the passaging steps, as well as the proportion of phage transferred between passages both in the modelling, and experimentally, to see if better strategies would emerge from such an approach. It also would have been interesting to adapt T7-ReRb to infect $\Delta waaC$ only instead of both BW25113 and $\Delta waaC$. The missing sequence data of the samples collected during the $\Delta waaC$ directed evolution experiments could have provided more evidence that the genotype in those regions was responsible for the changes in EOP. Similarly, next generation sequencing would have indicated whether other genes were involved in this relationship. For the relative fitness assays, the modelling could have been better informed by more experimental data, in particular, if we had assayed

more than one phage dilution against the strains to be compared. Additionally, performing an experiment where only the phage dilutions are varied could help determine whether the relative fitness changes are responding to changes in phage fitness, and not changes in PFU. In the OmpF directed evolution experiments, further work could investigate the failure of the system to adapt T7 solely to the OmpF receptor. Also, further characterisation of the evolved line 'Switching 1:49 (25-3)' might inform us as to whether the experiments failed because of the partial selection imposed by the experimenter (choosing which passage lines to continue based on the max OD plate reader data) or because of other factors (such as a non-proximal solution in the genotype space).

There are few avenues which could be exploited to improve on the directed evolution modelling provided here. One obvious option (which was not discussed due to time constraints) would be the introduction another class of receptors to the model abstractions. For instance, including a third class for all *E. coli* receptors with the exception of the target OMP and the LPS. This would allow us to investigate the impact of cheaters (phage which circumvent the selection procedure by binding to that class of receptors and so still generating progeny). We could also consider simulations involving a third negative selection strain (one with no gene knockout of the outer membrane) to enable the correct selection against certain receptor classes in cases where the gene knockout(s) might have knock-on effects on other components in the outer membrane. Finally, simulating the OmpF semi-stochastic model for the remaining parameter settings would tie up that section, and validate the conclusions drawn.

The directed evolution experiments could also benefit from the use of bioreactors to maintain continuous cultures, which would avoid the manual labour involved with

serial passaging of batch cultures. Furthermore, it would enable more generations of phage to evolve and be selected for over a 24 hour period. The requisite set-up would involve two chemostats (each culturing either the positive or negative selection strain) feeding a cellstat, where the phage and bacteria would reside in co-culture. The flow rate would be set such that the dilution rate of the cellstat would be sufficiently high that T7 phage could at least maintain their numbers, and sufficiently below the doubling time of the co-cultured bacteria to reduce the likelihood of co-evolution phage and bacteria. The modelling work established here can be straightforwardly applied to a continuous culture setting, and follows from the work of Husimi on the subject [49]. Alternatively, high-throughput liquid-handling machines could employ serial passages in an automated fashion which, contamination excepting, could be avoid the possibility of co-evolution all together.

Both rational engineering methods and directed evolution were employed in this thesis, in an attempt to create a T7 phage which would infect using solely OMPs. That directed evolution was employed at all here speaks to our ignorance of a correct solution, rather than to any fault with rational engineering methods. Synthetic libraries of tail fibres and tail tubes were considered, but both the lack of knowledge on which regions within those proteins to diversify, and, at the time of this work, the expense of the procuring those libraries, prevented further exploration of this. Those who read this in the coming years, with a view to investigating this topic may benefit from such an approach. Directed evolution then would only be necessary if initial screenings did not produce satisfactory results.

In this work, we have proposed a method for controlling the host tropism of a strictly lytic and non-temperate phage. We do this with a view to solving the antibiotic resistance crisis - aiming to generate phages which infect hosts using

receptors of our choosing. There are of course, numerous limitations to the espoused course of action. Firstly, T7 phage would not be capable of infecting gram-positive bacteria. Even for the gram-negative *E. coli*, the expression of O-antigens (extended polysaccharide chains) can limit access of phage to molecules that sit closer to surface [217], similar to how low-lying fauna in a rainforest will only receive sunlight which passes through the dense canopy overhead. Then there are those bacteria which posses extracellular capsules (in the case of *E. coli* K1, made up of polysialic acid [218, 219]) which must be enzymatically broken down by before phage infection can take place [220]. Moreover, successful ejection of phage DNA into a cell by no means guarantees a productive infection (i.e. subsequent phage replication and production of progeny). This need not be an impediment to progress; depending on the desired outcome, progeny production might not be necessary. For instance, if only lysing of the pathogenic bacteria is required, as opposed to a local maintenance of the initial phage dose (as is supplied by repeated amplifications upon replication). Irrespectively, phage DNA (or RNA) ejected into a host can be neutralised upon entry (either by CRISPR systems, restriction modification, or nucleases) and so further adaptations to the engineered nucleic acid may be required to circumvent these defences. Transforming mutagenised phage nucleic acid, along with a selection marker, into the targeted host, could be used to select phage genomes which are not excluded by the host.

Beyond the drawbacks of the overriding concept, there are also limitations with the specific methods used herein. The advocated directed evolution method relies on the negative selection strain containing the same composition and stoichometry of cell surface molecules as the positive selection strain (with the exception of the receptor to be targeted). An additional negative selection strain (without gene knockouts)

may help resolve this issue, but would require further investigation. Furthermore, certain parameters of the serial passaging method were not empirically perturbed. Changing the incubation time, and, especially, the proportion of phage transferred between each passage, would likely have had a beneficial effect on the outcome. Varying only the adsorption rates of phage to receptors, and not the phage latent periods, was both a requirement of the modelling methods available to me, and a drawback of the approach involved. In practise, T7 phage latent periods are known to vary within evolution experiments [81, 160]. Also, T7 phage adsorption is believed to consist of two stages: an initial reversible attachment, followed by irreversible adsorption, and concomitant DNA ejection [58, 59]. Both stages were modelled as one in this thesis, and phages, once adsorbed, could not reverse the process. Finally, for the semi-stochastic modelling, new phage mutants were added to the appropriate phage phenotypes at the end of each latent period. Moreover, no attempt was made to avoid double counting of newly mutated phage (the single phage gifted to a new phenotypic state would not be removed from the originating phenotype). Both workarounds are unphysical, and do not precisely reflect the underlying biological process. Though the latter could have been removed with more care (but was unimportant for the final result) the former would require an agent-based approach, which at the time of writing was too computationally intensive to implement for a biophysically realistic number of phages and bacteria.

A Appendix to chapter 2

```
g17*       ATGGCaAACGTgATTAAgACCGTaTTGACcTACCAaTTAGATGGTAGTAACAGgGACTTC
WT         atggctaacgtaattaaaaccgttttgacttaccagttagatggctccaatcgtgatttt
           ***** ***** ***** ***** ***** ***** ********    **  * ** **

g17*       AACATTCCATTTGAaTAcCTAGCgCGTAAaTTCGTtGTGGTtACTCTgATTGGcGTAGAT
WT         aatatcccgtttgagtatctagcccgtaagttcgtagtggtaactcttattggtgtagac
           ** ** ** ***** ** ***** ***** ***** ***** ***** ***** *****

g17*       cGtAAAGTTTTAACAATcAACACGGATTATCGCTTcGCgACgCGcACGACTATCAGtTTA
WT         cgaaaggtccttacgattaatacagactatcgctttgctacacgtactactatctctctg
           ** ** **  * ** ** ** ** ** ******** ** ** ** ** ******  * *

g17*       ACGAAAGCGTGGGGGCCAGCGGATGGCTAtACGACaATCGAaTTACGcCGAGTgACCTCC
WT         acaaaggcttggggtccagccgatggctacacgaccatcgagttacgtcgagtaacctcc
           ** ** ** ***** ***** ******** ***** ***** ***** ***** ******

g17*       ACgACCGACCGAcTGGTAGATTTCACTGACGGcAGTATTTTGcGcGCCTACGATTTAAAT
WT         actaccgaccgattggttgactttacggatggttcaatcctccgcgcgtatgaccttaac
           ** ********* **** ** ** ** ** **    **  * ***** ** **  * **

g17*       GTGGCACAAATcCAGACTATGCACGTTGCCGAGGAAGCGcGtGATTTAACAACCGACACA
WT         gtcgctcagattcaaacgatgcacgtagcggaagaggcccgtgacctcactacggatact
           ** ** ** ** ** ** ******** ** ** ** ** *****  * ** ** ** **

g17*       ATcGGCGTCAACAATGACGGcCATCTCGACGCAcGtGGGcGtcGcATcGTTAATCTcGCG
WT         atcggtgtcaataacgatggtcacttggatgctcgtggtcgtcgaattgtgaacctagcg
           ***** ***** ** ** ** **  * ** ** ***** ***** ** ** ** ** ***

g17*       AACGCgGTcGATGACCGtGATGCCGTACCCTTCGGtCAGCTCAAAACAATGAACCAaAAC
WT         aacgccgtggatgaccgcgatgctgttccgtttggtcaactaaagaccatgaaccagaac
           ***** ** ******** ***** ** ** ** ***** ** ** ** ******** ***

g17*       TCcTGGCAgGCtCGTAATGAgGCCCTCCAATTTcGtAACGAAGCGGAAACATTTAGAAAC
WT         tcatggcaagcacgtaatgaagccttacagttccgtaatgaggctgagactttcagaaac
           ** ***** ** ******** *** * ** ** ***** ** ** ** ** ** ******

g17*       CAgGCGGAaGGCTTTAAaAACGAgAGcTCCACGAATGCGACTAATACCAAACAGTGGAGG
WT         caagcggagggctttaagaacgagtccagtaccaacgctacgaacacaaagcagtggcgc
           ** ***** ******** ******  *   ** ** ** ** ** ** ** ****** *

g17*       gACGAAACCAAGGGcTTCCGAGAtGAGGCAAAAAGATTTAAAAACACAGCAGGCCAGTAT
WT         gatgagaccaagggtttccgagacgaagccaagcggttcaagaatacggctggtcaatac
           ** ** ******** ******** ** ** **  * ** ** ** ** ** ** ** **

g17*       GCAACGAGCGCAGGAAATAGCGCAAGTGCAGCCCACCAGAGTGAAGTTAATGCAGAAAAT
WT         gctacatctgctgggaactctgcttccgctgcgcatcaatctgaggtaaacgctgagaac
           ** **    ** ** **    **    ** ** ** **   *** ** ** ** ** **

g17*       AGCGCGACTGCTAGTGCAAATAGTGCCCACCTAGCTGAGCAACAGGCTGACCGcGCGGAA
WT         tctgccacagcatccgctaactctgctcatttggcagaacagcaagcagaccgtgcggaa
              ** ** **    ** **   *** **  * ** ** ** ** ** ***** ******

g17*       CGcGAGGCcGACAAaCTGGAGAACTATAACGGTCTTGCAGGAGCTATAGATAAGGTcGAc
WT         cgtgaggcagacaagctggaaaattacaatggattggctggtgcaattgataaggtagat
           ** ***** ***** ***** ** ** ** **  * ** ** ** ** ******** **
```

```
g17*        GGAACGAACgTTTATTGGAAGGGTAACATACATGCTAACGGaCGCCTcTACATGACaACA
WT          ggaaccaatgtgtactggaaaggaaatattcacgctaacgggcgcctttacatgaccaca
            ***** ** ** ** ***** ** ** ** ** ******** ***** ******** ***

g17*        AACGGCTTCGATTGCGGACAATACCAGCAATTTTTTGGTGGgGTtACTAATCGcTACAGC
WT          aacggttttgactgtggccagtatcaacagttctttggtggtgtcactaatcgttactct
            ***** ** ** ** ** ** ** ** ** ** ******** ** ******** ***

g17*        GTAATGGAATGGGGTGACGAAAATGGTTGGTTAATGTACGTCCAGAGACGTGAATGGACT
WT          gtcatggagtggggagatgagaacggatggctgatgtatgttcaacgtagagagtggaca
            ** ***** ***** ** ** ** ** *** * ***** ** **  *  * ** *****

g17*        ACTGCCATTGGGGGAAATATTCAACTCGTTGTTAATGGTCAAATTATTACtCAaGGAGGC
WT          acagcgataggcggtaacatccagttagtagtaaacggacagatcatcacccaaggtgga
            ** ** ** ** ** ** ** **  * ** ** ** ** ** ** ** ** ***** **

g17*        GCAATGACTGGCCAATTGAAATTGCAaAATGGtCATGTATTGCAGCTTGAAAGTGCTAGT
WT          gccatgaccggtcagctaaaattgcagaatgggcatgttcttcaattagagtccgcatcc
            ** ***** ** **  * ******** ***** *****  * **  * **    **

g17*        GATAAAGCaCACTATATcCTATCgAAAGAcGGTAACCGAAACAATTGGTATATAGGACGA
WT          gacaaggcgcactatattctatctaaagatggtaacaggaataactggtacattggtaga
            ** ** ** ******** ***** ***** ****** * ** ** ***** ** **  **

g17*        GGAAGCGATAAtAACAAcGACTGcACCTTCCAtTCCTACGTTCACGGCACCACACTCACC
WT          gggtcagataacaacaatgactgtaccttccactcctatgtacatggtacgaccttaaca
            **    ***** ***** ***** ******** ***** ** ** ** ** **  * **

g17*        TTAAAACAGGACTAcGCAGTAGTAAATAAGCATTTTCATGTTGGCCAAGCTGTCGTCGCC
WT          ctcaagcaggactatgcagtagttaacaaacacttccacgtaggtcaggccgttgtggcc
             * ** ******** ******** ** ** ** ** ** ** ** ** ** ** ** ***

g17*        ACTGACGGCAACATCCAGGGAACCAAATGGGGCGGCAAGTGGCTTGACGCCTATTTGCGA
WT          actgatggtaatattcaaggtactaagtggggaggtaaatggctggatgcttacctacgt
            ***** ** ** ** ** ** ** ** ***** ** ** ***** ** ** **  * **

g17*        GATTCTTTTGTCGCTAAAAGTAAaGCATGGACACAAGTTTGGTCgGGTAGTGCaGGCGGT
WT          gacagcttcgttgcgaagtccaaggcgtggactcaggtgtggtctggtagtgctggcggt
            **    ** ** ** **    ** ** ***** ** ** ***** ******** ******

g17*        GGGGTTTCTGTAACAGTCAGTCAAGATCTtCGCTTtCGCAAcATCTGGATcAAATGCGCA
WT          ggggtaagtgtgactgtttcacaggatctccgcttccgcaatatctggattaagtgtgcc
            *****   *** ** **    ** ***** ***** ***** ******** ** ** **

g17*        AATAATAGTTGGAACTTtTTCCGcACTGGCCCaGATGGAATaTAtTTCATAGCaTCTGAc
WT          aacaactcttggaacttcttccgtactggccccgatggaatctacttcatagcctctgat
            ** **   ********* ***** ******** ******** ** ******** *****

g17*        GGTGGcTGGTTgCGTTTTCAGATCCATTCTAATGGCCTTGGCTTCAAGAAcATTGCAGAt
WT          ggtggatggttacgattccaaatacactccaacggtctcggattcaagaatattgcagac
            ***** ***** ** ** ** ** ** ** ** ** ** ** ******** ********

g17*        AGcCGTTCcGTACCcAATGCAATtATGGTaGAGAAtGAGTAA
WT          agtcgttcagtacctaatgcaatcatggtggagaacgagtaa
            ** ***** ***** ******** ***** ***** ******
```

Figure A.1: Nucleotide sequence alignment of *g17** and WT *g17* (Clustal Ω v1.2.1 [221]). Instances of matching base pairs are denoted by an asterisk.

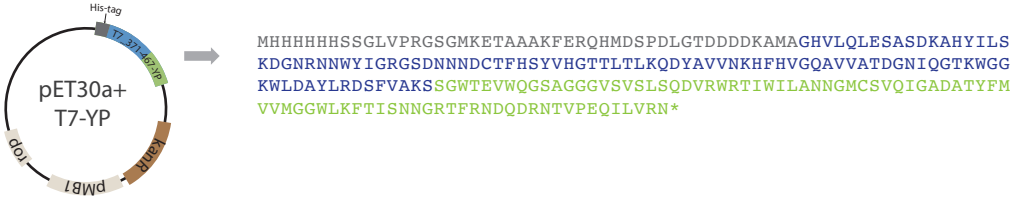


Figure A.2: The relevant protein sequence expressed by pET30a+ T7-YP. In grey are the vector-derived amino acids (note the initial N-terminal hist-tag sequence). In blue are aa 371-467 of T7 gp17. In green are aa 486-569 of Yep-phi gp17.

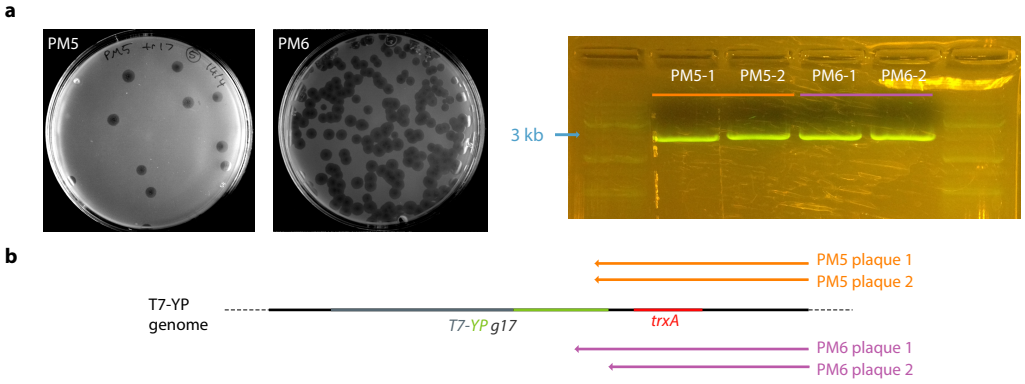


Figure A.3: A gel of four PCR amplified plaques (2 picked from PM5 plated against BW25113 Δ*trxA* pSEVA551-g17*, 2 from PM6 plated on BW25113 Δ*trxA*). b) A diagram indicating the region covered by Sanger sequencing reads. Reads run from downstream of where the Yep-phi sequence and *trxA* selection marker had recombined into the T7 genome.

B Appendix to chapter 3

B.1 Co-infection model approximations

Let $\mathrm{MOI}_{\mathrm{T7}} = \frac{PFU_{T7}}{CFU}$. We expand $1 - e^{\mathrm{MOI}_{\mathrm{T7}}}$ using the Maclaurin series,

$$1 - e^{-\mathrm{MOI}_{\mathrm{T7}}} = \mathrm{MOI}_{\mathrm{T7}} - \frac{(\mathrm{MOI}_{\mathrm{T7}})^2}{2!} + \frac{(\mathrm{MOI}_{\mathrm{T7}})^3}{3!} - \frac{(\mathrm{MOI}_{\mathrm{T7}})^4}{4!} + ...$$

and take the 1st order approximation $1 - e^{\mathrm{MOI}_{\mathrm{T7}}} \approx \mathrm{MOI}_{\mathrm{T7}}$, which is true for small values for $\mathrm{MOI}_{\mathrm{T7}}$ (say, $\mathrm{MOI}_{\mathrm{T7}} \leq 0.1$). Substituting into equation 3.8 we have,

$$\begin{aligned} PFU_{spy} &= -CFU \times \ln\left(1 - \frac{\mathcal{P}_0}{CFU \times \left(1 - e^{-\mathrm{MOI}_{\mathrm{T7}}}\right)}\right) \\ &\approx -CFU \times \ln\left(1 - \frac{\mathcal{P}_0}{CFU \times \mathrm{MOI}_{\mathrm{T7}}}\right) \\ &= -CFU \times \ln\left(1 - \frac{\mathcal{P}_0}{CFU \times \frac{PFU_{T7}}{CFU}}\right) \\ &= -CFU \times \ln\left(1 - \frac{\mathcal{P}_0}{PFU_{T7}}\right). \end{aligned}$$

Now, let us assume further that $0 \leq 10 \times \mathcal{P}_0 \leq PFU_{T7}$ (i.e. that the number of co-infected plaques is at least 10 times fewer than the number of T7 $\Delta g5$ phage - again, a likely occurrence in our system) so that we have $0 \leq \frac{\mathcal{P}_0}{PFU_{T7}} \leq 0.1$. Let $y = \frac{\mathcal{P}_0}{PFU_{T7}}$, then the Maclaurin series for $\ln(1-y)$ is given by,

$$\ln(1-y) = -y - \frac{y^2}{2!} - \frac{y^3}{3!} - \frac{y^4}{4!} - ...$$

and the series converges to the correct values for $-1 < y = \frac{\mathcal{P}_0}{PFU_{T7}} < 1$, of which $0 \leq \frac{\mathcal{P}_0}{PFU_{T7}} \leq 0.1$ is part. Similar to before, we take the first order approximation, $\ln(1-y) \approx -y$, and use this to further simplify PFU_{spy} as follows,

$$\begin{aligned} PFU_{spy} &\approx -CFU \times \ln\left(1 - \frac{\mathcal{P}_0}{PFU_{T7}}\right) \\ &= -CFU \times \ln(1-y) \\ &\approx -CFU \times -y \\ &= CFU \times \frac{\mathcal{P}_0}{PFU_{T7}} \\ &= \frac{\mathcal{P}_0}{\mathrm{MOI_{T7}}} \end{aligned} \tag{B.1}$$

where $100 \times \mathcal{P}_0 \leq 10 \times PFU_{T7} \leq CFU$.

B.2 $\mathbf{MOI_{actual}}$ co-infection model

We assume a non-homogenous Poisson process with intensity rate $\lambda(t) = mkS_0e^{-kS_0t}$, where m is the $\text{MOI}_{\text{input}}$, k is the phage adsorption rate and S_0 is the initial density of cells. Let the number of phage infecting a cell at time t be given by the counting process $\{N(t),\ t > 0\}$, then the probability mass function over the interval $(0, t]$ is,

$$P(N(t) - N(0) = n) = \frac{\Lambda^n e^{-\Lambda}}{n!},\ \text{where } \Lambda = \int_0^t \lambda(\tau)\, d\tau.$$

Since $\lambda(t) = mkS_0e^{-kS_0t}$, $\Lambda = m(1 - e^{-kS_0t}) = \text{MOI}_{\text{actual}}$. Subsequently, the probability of at least 1 phage infecting a cell at time t, is given by,

$$1 - P(N(t) - N(0) = 0) = 1 - e^{-\text{MOI}_{\text{actual}}}$$

B.3 A theoretical distribution for the number of co-infected plaques observed

Assume $\mathcal{P}_0 \sim \text{Poisson}(134)$ and $\mathcal{P}_1 \sim \text{Poisson}(134/30)$. Further, let samples from $\mathcal{P}_1$ be adjusted so that in cases where no plaques are observed, $\mathcal{P}_1 = 1$. We can then draw 10^6 samples from $\mathcal{P}_0$ and $\mathcal{P}_1$. The distribution of fold changes $\mathcal{P}_0/\mathcal{P}_1$ is shown in Figure B.1. In this random sample, there is an $\approx 12\%$ chance of observing at a least a 67-fold change in the number of plaques.

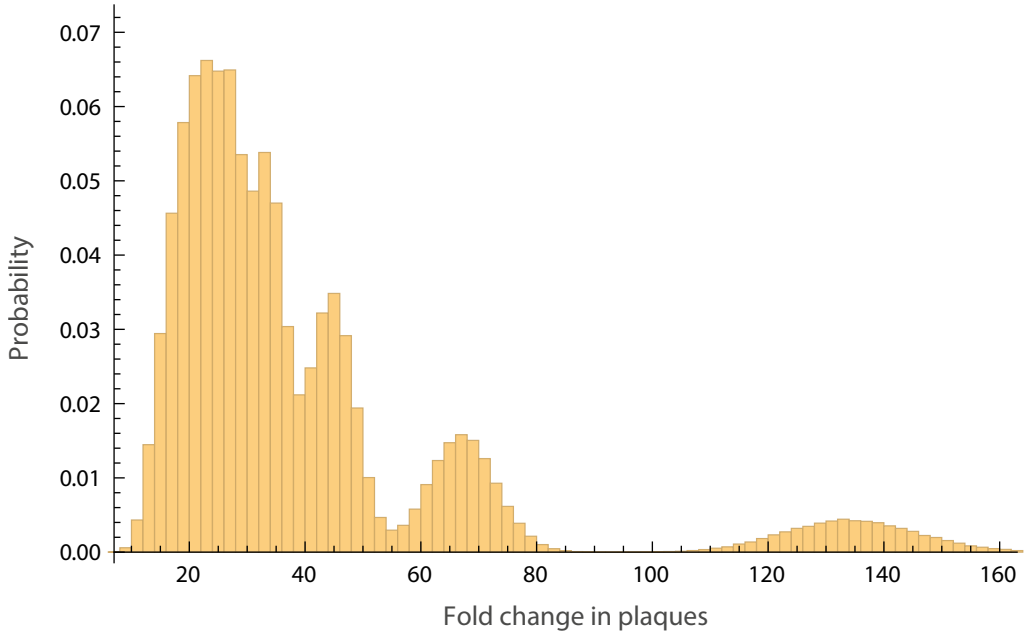


Figure B.1: This histogram derives from a Monte Carlo simulation of 10^6 samples drawn from 2 distributions: the number of co-infected plaques observed on the undiluted plate (assumed to be $\mathcal{P}_0 \sim \text{Poisson}(134)$); and the number of co-infected plaques observed on the first tenfold diluted plate (assumed to be $\mathcal{P}_1 \sim \text{Poisson}(134/30)$). The samples from $\mathcal{P}_1$ be adjusted so that in cases where no plaques are observed, $\mathcal{P}_1 = 1$. The presented histogram is then the fold change between the 10^6 sample pairs $(\mathcal{P}_0/\mathcal{P}_1)$.

B.3.1 Mathematica Code

```
In[1]:=  X = RandomVariate[PoissonDistribution[134],10^6];

In[2]:=  Y = RandomVariate[PoissonDistribution[134/30],10^6];

In[3]:=  Yno0 = Table[If[Y[[i]] == 0,1,Y[[i]]],{i,1,Length[Y]}];

In[4]:=  Length[Select[X/Yno0, # > 67 &]] / Length[X/Yno0] // N

Out[4]=  0.117478
```

C Appendix to chapter 4

C.1 Interpreting the OD_{600} description of free T7 phage in the relative fitness model

Let us first consider how the population of phage changes with time for a typical scale (phage numbers per ml). Let the rate of adsorption, k, be 3×10^{-9} ml/min, let the initial number of cells be $C(0) = 10^8$ CFU per ml [81] and the initial MOI be 0.01 (so that $P(0) = 10^6$). Assume further that MOI is significantly small over one minute that the reduction in cells is negligible. Then, the reduction in free phage density during that 1st minute is given by,

$$\begin{aligned}\frac{\mathrm{d}P(0)}{\mathrm{d}t} \times 1 \text{ min} &= -kC(0)P(0) \times 1 \text{ min} \\ &= -3 \times 10^{-9} \text{ ml min}^{-1} \times 10^8 \text{ ml}^{-1} \times 10^6 \text{ ml}^{-1} \times 1 \text{ min} \\ &= -3 \times 10^4 \text{ ml}^{-1}\end{aligned}$$

On the OD_{600} scale, let us say that 10^8 cells ml^{-1} corresponds to an OD_{600} of 0.1 in the plate reader (the precise relation is unimportant here). We then scale the other quantities accordingly. The density of free phage (10^6 ml^{-1}) was set to a value 100 times less than the initial cell population (10^8 ml^{-1}) and so on the OD_{600} scale we have $P(0) = \frac{10^6 \text{ ml}^{-1}}{10^8 \text{ ml}^{-1}} \times 0.1 \text{ OD}_{600} = 10^{-3}$ free phage OD_{600}. Similarly, we can scale the reduction in free phage in density so that the fraction of phages adsorbing stays the same, i.e.

$$\frac{\mathrm{d}P(0)}{\mathrm{d}t} \times 1 \text{ min} = \frac{-3 \times 10^4 \text{ ml}^{-1}}{10^8 \text{ ml}^{-1}} \times 0.1 \text{ OD}_{600} = -3 \times 10^{-5} \text{ OD}_{600}.$$

We then have,

$$\begin{aligned}\frac{\mathrm{d}P(0)}{\mathrm{d}t} \times 1 \text{ min} &= -fC(0)P(0) \times 1 \text{ min} \\ &= -f \text{ OD}_{600}{}^{-1} \text{ min}^{-1} \times 0.1 \text{ OD}_{600} \times 10^{-3} \text{ OD}_{600} \times 1 \text{ min} \\ &= -f \times 10^{-4} \text{ OD}_{600} \times \text{min} = -3 \times 10^{-5} \text{ OD}_{600}\end{aligned}$$

which implies, $f = 0.3 \text{ OD}_{600}{}^{-1} \text{ min}^{-1}$.

C.2 Relative fitness estimation from n phage dilutions

C.2.1 Estimation using 2 phage dilutions

Similar to before, we define P_{ij} to be the population of phage for phage dilution j challenged against strain i. We let P_{12} be a 10-fold dilution of P_{11} at $t = 0$, though we could pick any non-trivial dilution without loss of generality. Analogously to

equation 4.3 we have $P_{11} = P_0\, e^{f_1\alpha_{11}}$ and because P_{12} is a 10-fold dilution, $P_{12} = 0.1P_0\, e^{f_1\alpha_{12}}$ where we recall that P_0 is the initial phage population on the OD_{600} scale. Similar to equations 4.5 and 4.6 we also have $P_{11} = \frac{r_1}{f_1}\left(1 - \frac{C_{11}}{K_1}\right) := \frac{r_1}{f_1}\eta_{11}$ so symmetrically $P_{12} = \frac{r_1}{f_1}\left(1 - \frac{C_{12}}{K_1}\right) := \frac{r_1}{f_1}\eta_{12}$. Therefore, we can equate our two equations for P_{11}, and do the same for P_{12} to get

$$\begin{aligned}\frac{r_1}{f_1}\eta_{11} &= P_0 e^{f_1\alpha_{11}}\\ \frac{r_1}{f_1}\eta_{12} &= 0.1P_0 e^{f_1\alpha_{12}}.\end{aligned}$$

Dividing both equations through by the η term, and then equating both again, we have,

$$\begin{aligned}P_0\frac{e^{f_1\alpha_{11}}}{\eta_{11}} - 0.1P_0\frac{e^{f_1\alpha_{12}}}{\eta_{12}} &= 0\\ P_0\left(\frac{e^{f_1\alpha_{11}}}{\eta_{11}} - \frac{0.1e^{f_1\alpha_{12}}}{\eta_{12}}\right) &= 0.\end{aligned}$$

Since $P_0 > 0$, we must have that,

$$\begin{aligned}\frac{e^{f_1\alpha_{11}}}{\eta_{11}} - \frac{0.1e^{f_1\alpha_{12}}}{\eta_{12}} &= 0, \text{ which implies,}\\ \ln(\eta_{12}) + f_1\alpha_{11} &= \ln(0.1\eta_{11}) + f_1\alpha_{12}, \text{ which implies,}\\ f_1(\alpha_{11} - \alpha_{12}) &= \ln(0.1\eta_{11}) - \ln(\eta_{12}), \text{ which implies,}\\ f_1 &= \frac{\ln\left(\frac{0.1\eta_{11}}{\eta_{12}}\right)}{(\alpha_{11} - \alpha_{12})}.\end{aligned}$$

Following the same process gives us,

$$f_2 = \frac{\ln\left(\frac{0.1\eta_{11}}{\eta_{12}}\right)}{(\alpha_{21} - \alpha_{22})}.$$

Therefore, we can conclude that,

$$\frac{f_1}{f_2} = \frac{(\alpha_{21} - \alpha_{22}) \ln\left(\frac{0.1\eta_{11}}{\eta_{12}}\right)}{(\alpha_{11} - \alpha_{12}) \ln\left(\frac{0.1\eta_{21}}{\eta_{22}}\right)}$$

This result can be simplified in certain instances. First note that,

$$\frac{\eta_{i1}}{\eta_{i2}} = \frac{1 - \frac{C_{i1}}{K_i}}{1 - \frac{C_{i2}}{K_i}} = \frac{K_i - C_{i1}}{K_i - C_{i2}}.$$

Suppose further that $\frac{K_i - C_{i1}}{K_i - C_{i2}} \approx 1$, which might be the case if the estimated carrying capacities of both strains are similar, and lysis of both strains is complete, i.e. $C_{i1}, C_{i2} \approx 0$. Then we have,

$$\frac{f_1}{f_2} \approx \frac{\alpha_{21} - \alpha_{22}}{\alpha_{11} - \alpha_{12}}.$$

C.2.2 Estimation using n phage dilutions

In the case of more than 2 phage dilutions, we have a system of non-linear equations with 2 unknowns,

$$\begin{aligned}
\frac{r_1 f_2 \eta_{11}}{r_2 f_1 \eta_{21}} - \frac{e^{f_1 \alpha_{11}}}{e^{f_2 \alpha_{21}}} &= 0 \\
\frac{r_1 f_2 \eta_{12}}{r_2 f_1 \eta_{22}} - \frac{e^{f_1 \alpha_{12}}}{D_1 e^{f_2 \alpha_{22}}} &= 0 \\
... &= 0 \\
\frac{r_1 f_2 \eta_{1n}}{r_2 f_1 \eta_{2n}} - \frac{e^{f_1 \alpha_{1n}}}{D_{n-1} e^{f_2 \alpha_{2n}}} &= 0
\end{aligned}$$

where, n denotes the number of lysis curves, and D_j is the relative dilution in initial PFU. Since there are more equations than unknowns, the system is overdetermined, but we can still find numerical solutions for f_1 and f_2 using, for instance, fsolve in MATLAB. Observe that η_{2j} must be not equal to 0 for all j, i.e. that $C_{2j} \neq K_2$ for all j. In practise this would be the case if no phage was added, or if the time period considered was not sufficiently large as to observe cell lysis via OD_{600} measurements.

C.3 Notes on the meta-distribution of p-values

What follows is an explanatory piece on the work of Taleb [208] on the meta-distribution of p-values, and how it relates to the work performed in this thesis.

Consider the log-transformed data from an efficiency of plating (EOP) experiment, where we have two distinct populations whose $\log_{10}$(EOP)s we can consider to have been drawn from normal distributions, with the same mean and variance (i.e. $X_1, X_2 \sim N(\mu, \sigma^2)$). If we took random samples from these populations, and ran two-sample t-tests, every p-value, in the interval $(0, 1)$, is a possible outcome, and uniformly so. In other words, each p-value is equally likely to arise. This is what we might expect from a p-value; it is meant to show us how likely a difference in the means of the two samples was, if we assumed that there was no such difference.

If instead, however, we took two distinct populations, which had means and variances such that the "true" p-value, if calculated, would be 0.5, and we again took random samples from these populations, and ran t-tests, we would find that every p-value is no longer equally likely to arise. In fact, the distribution of p-values

would be right skewed[1] (see Figure C.1). We can not replicate this procedure precisely for the p-values calculated in this thesis (in the case where we can not assume equal variance when performing a t-test - commonly referred to as Welch's t-test) however, we can do so if we relax that assumption. In this instance, and given a "true" p-value of 0.5, approximately 15% of the estimated p-values would pass a significance test at the 0.05 level, and $\approx 3.5\%$ of the values would be found below the 0.01 cut-off. If we did this again starting with a "true" p-value of 0.1, we get $\approx 55\%$ below 0.05, and $\approx 20\%$ below 0.01. From this, it is clear that p-values obtained in this thesis may not be as close to the "true" p-values as one might hope.

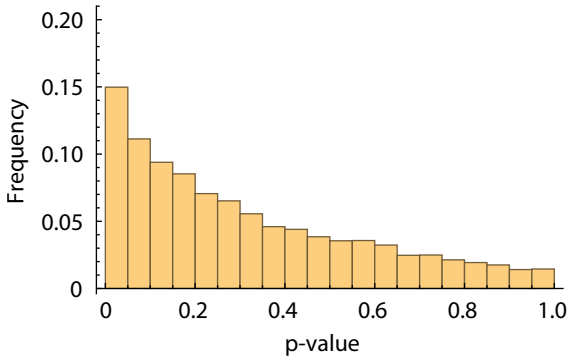


Figure C.1: A representative probability histogram of the meta-distribution of p-values for a "true" p-value of 0.5, made up of $10,000$ t-tests on random samples (with $n = 3$ data points in each sample) taken from normal distributions with carefully chosen means and variances so as to ensure a "true" p-value of 0.5.

C.3.1 Mathematica code for the required Monte Carlo simulations

Code for the two-sample, two-sided case (equal variances). The "true" p-value is set to 0.5.

[1] the term refers to the tail of the distribution, so values found on the left of the distribution occur more frequently

In[1]:=
```
truep = 0.5;
```

In[2]:=
```
n = 3;
```

In[3]:=
```
m=10^4;
```

In[4]:=
```
sol = Solve[CDF[StudentTDistribution[2(n-1)],t] == 1- truep/2,t,  Reals]
```

Out[4]=
```
{{t→0.740697}}
```

In[5]:=
```
t_exact = t == sampleMean / Sqrt[(s1^2+s2^2)/n] /.  sol[[1]]
```

Out[5]=
$$0.740697==\sqrt{\frac{3}{2}}\ \text{sampleMean}$$

In[6]:=
```
sol2 = Solve[t_exact ,sampleMean]
```

Out[6]=
```
{{sampleMean→0.604777}}
```

In[7]:=
```
s1 = 1;

s2 = 1;
```

In[9]:=
```
x = sampleMean /. sol2[[1]]
```

Out[9]=
```
0.604777
```

In[10]:=
```
repetitionsOfprocedure = 5;
```

In[11]:=
```
metaPvalues = Table[

  X1 = Table[RandomVariate[NormalDistribution[x,s1],n],m];

  X2 = Table[RandomVariate[NormalDistribution[0,s2],n],m];

  tstatistics = Table[(Mean[X1[[i]]] -  Mean[X2[[i]]])/Sqrt[(Variance[X1[[i]]]+ Variance[X2[[i]]])/n],{i,1,m}];
  p = Table[1 - CDF[StudentTDistribution[2(n - 1)], tstatistics[[i]]], {i, 1, m}],
repetitionsOfprocedure]
```

In[12]:=
```
below005 = Table[N[Length[Select[metaPvalues[[i]],#<0.05&]]/m],{i,1,
repetitionsOfprocedure}]
```

Out[12]=
```
{0.1457,0.1588,0.1539,0.1552,0.1504}
```

In[13]:=
```
Median[below005]
```

```
Out[13]=  0.1539

In[14]:=  MedianDeviation[below005]

Out[14]=  0.0035

In[15]:=  below001 =
          Table[N[Length[Select[metaPvalues[[i]],#<0.01&]]/m],{i,1,repetitionsOfprocedure}]

Out[15]=  {0.0353,0.0379,0.0383,0.033,0.0377}

In[16]:=  Median[below001]

Out[16]=  0.0377

In[17]:=  MedianDeviation[below001]

Out[17]=  0.0006
```

References

1. O'Neil, J. Review on Antimicrobial Resistance. Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations. https://amr-review.org/sites/default/files/AMR%20Review%20Paper%20-%20Tackling%20a%20crisis%20for%20the%20health%20and%20wealth%20of%20nations_1.pdf (2014).
2. De Kraker, M. E., Stewardson, A. J. & Harbarth, S. Will 10 Million People Die a Year due to Antimicrobial Resistance by 2050? *PLoS Medicine* **13,** 1–6 (2016).
3. Lewis, K. Platforms for antibiotic discovery. *Nature Reviews Drug Discovery* **12.** Review Article, 371–387 (Apr. 2013).
4. Simpkin, V. L., Renwick, M. J., Kelly, R. & Mossialos, E. Incentivising innovation in antibiotic drug discovery and development: Progress, challenges and next steps. *Journal of Antibiotics* **70,** 1087–1096 (2017).
5. Nilsson, A. S. Phage therapy-constraints and possibilities. *Upsala Journal of Medical Sciences* **119,** 192–198 (2014).
6. Schwarzer, D. *et al.* A Multivalent Adsorption Apparatus Explains the Broad Host Range of Phage phi92: a Comprehensive Genomic and Structural Analysis. *Journal of Virology* **86,** 10384–10398 (2012).

7. Hamdi, S. *et al.* Characterization of two polyvalent phages infecting Enterobacteriaceae. *Scientific Reports* **7,** 1–12 (2017).
8. Liu, M. *et al.* Reverse Transcriptase-Mediated Tropism Switching in Bordetella Bacteriophage. *Science* **295,** 2091–2094 (2002).
9. Barr, J. J. A bacteriophages journey through the human body. *Immunological Reviews* **279,** 106–122 (2017).
10. Bordenstein, S. R. & Bordenstein, S. R. Eukaryotic association module in phage WO genomes from Wolbachia. *Nature Communications* **7,** 1–10 (2016).
11. Yosef, I., Manor, M., Kiro, R. & Qimron, U. Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria. *Proceedings of the National Academy of Sciences* **112,** 7267–7272 (2015).
12. Yosef, I., Manor, M. & Qimron, U. Counteracting selection for antibiotic-resistant bacteria. *Bacteriophage* **6,** e1096996 (2016).
13. Abedon, S. T. Phage therapy dosing: The problem(s) with multiplicity of infection (MOI). *Bacteriophage* **6,** e1220348 (2016).
14. Lin, D. M., Koskella, B. & Lin, H. C. Phage therapy: An alternative to antibiotics in the age of multi-drug resistance. *World Journal of Gastrointestinal Pharmacology and Therapeutics* **8,** 162 (2017).
15. Ando, H., Lemire, S., Pires, D. P. & Lu, T. K. Engineering Modular Viral Scaffolds for Targeted Bacterial Population Editing. *Cell Systems* **1,** 187–196 (2015).
16. Yoon, M. Y. & Yoon, S. S. Disruption of the gut ecosystem by antibiotics. *Yonsei Medical Journal* **59,** 4–12 (2018).
17. Daugelavicius, R., Cvirkaite, V., Bakiene, E., Gabrenaite-verkhovskaya, R. & Bamford, D. H. Penetration of Enveloped Double-Stranded RNA Bacte-

riophages ϕ13 and ϕ6 into Pseudomonas syringae Cells. *Journal of Virology* **79,** 5017–5026 (2005).

18. Stick, R. V. & Williams, S. J. Modifications of Glycans and Glycoconjugates. *Carbohydrates: The Essential Molecules of Life,* 343–367 (2009).
19. Price, N. L. *et al.* Glycoengineered Outer Membrane Vesicles: A Novel Platform for Bacterial Vaccines. *Scientific Reports* **6,** 1–9 (2016).
20. Soufi, B., Krug, K., Harst, A. & Macek, B. Characterization of the E. coli proteome and its modifications during growth and ethanol stres. *Frontiers in Microbiology* **6,** 1–11 (2015).
21. Neidhardt, F. C. & Umbarger, H. E. in *Escherichia coli and Salmonella : cellular and molecular biology* chap. 3 (ASM Press, Washington, D.C, 1996). ISBN: 978-1555810849.
22. Koebnik, R., Locher, K. P. & Van Gelder, P. Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Molecular Microbiology* **37,** 239–253 (2000).
23. Nikaido, H. Molecular Basis of Bacterial Outer Membrane Permeability Revisited. *Microbiology and Molecular Biology Reviews* **67,** 593–656 (2003).
24. Zhao, X. *et al.* Outer Membrane Proteins Ail and OmpF of Yersinia pestis Are Involved in the Adsorption of T7-Related Bacteriophage Yep-phi. *Journal of Virology* **87,** 12260–12269 (2013).
25. Hildegard, E., Minh, D. B., Schellack, C., Nagy, E. & Meinke, A. Bacterial phage receptors, versatile tools for display of polypeptides on the cell surface. *Journal of Bacteriology* **183,** 6924–6935 (2001).
26. German, G. J. & Misra, R. The TolC protein of Escherichia coli serves as a cell-surface receptor for the newly characterized TLS bacteriophage. *Journal of Molecular Biology* **308,** 579–585 (2001).

27. Ackermann, H. W. 5500 Phages examined in the electron microscope. *Archives of Virology* **152,** 227–243 (2007).

28. Abedon, S. T. in *Bacteriophages in health and disease* (eds Hyman, P. & Abedon, S. T.) 1–5 (Cabi, 2012). ISBN: 9781845939847.

29. Ackermann, H. W. Bacteriophage observations and evolution. *Research in Microbiology* **154,** 245–251 (2003).

30. Atanasova, N. S. *et al.* in *Advances in Virus Research* 1–61 (Elsevier, 2015).

31. Hobbs, Z. & Abedon, S. T. Diversity of phage infection types and associated terminology: the problem with 'Lytic or lysogenic'. *FEMS microbiology letters* **363,** 1–8 (2016).

32. Shen, M. *et al.* A linear plasmid-like prophage of Actinomyces odontolyticus Promotes Biofilm Assembly. *Applied and Environmental Microbiology* (2018).

33. Casjens, S. R. *et al.* The pKO2 Linear Plasmid Prophage of Klebsiella oxytoca. *Journal of Bacteriology* **186,** 1818–1832 (2004).

34. Foster, J. W. & Slonczewski, J. L. in *Microbiology: An Evolving Science* 2nd ed., eTopic (W.W. Norton, 2017). http://www.wwnorton.com/college/biology/microbiology2/ch/11/etopics.aspx.

35. Salmond, G. P. & Fineran, P. C. A century of the phage: Past, present and future. *Nature Reviews Microbiology* **13,** 777–786 (2015).

36. Nobrega, F. L. *et al.* Targeting mechanisms of tailed bacteriophages. *Nature Reviews Microbiology* **16,** 760–773 (2018).

37. Meyer, J. R. *et al.* Repeatability and contigency in the evolution of a key innovation in phage lambda. *Science* **335,** 428–432 (2012).

38. McPartland, J. & Rothman-Denes, L. B. The tail sheath of bacteriophage N4 interacts with the escherichia coli receptor. *Journal of Bacteriology* **191,** 525–532 (2009).

39. Trojet, S. N., Caumont-Sarcos, A., Perrody, E., Comeau, A. M. & Krisch, H. M. The gp38 adhesins of the T4 superfamily: A complex modular determinant of the Phage's host specificity. *Genome Biology and Evolution* **3,** 674–686 (2011).

40. Morona, R., Klose, M. & Henning, U. Escherichia K12 outer membrane protein (ompA): Analysis of mutant genes expressing altered proteins. *Journal of Bacteriology* **159,** 570–578 (1984).

41. Morona, R. & Henning, U. Host range mutants of bacteriophage Ox2 can use two different outer membrane proteins of Escherichia coli K-12 as receptors. *Journal of Bacteriology* **159,** 579–582 (1984).

42. Drexler, K., Dannull, J., Hindennach, I., Mutschler, B. & Henning, U. Single mutations in a gene for a tail fiber component of an Escherichia coli phage can cause an extension from a protein to a carbohydrate as a receptor. *Journal of Molecular Biology* **219,** 655–663 (1991).

43. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Ntl. Acad. Sci. USA* **97,** 6640–6645 (2000).

44. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: The Keio collection. *Molecular Systems Biology* **2.** arXiv: arXiv:0706.1336 (2006).

45. Washizaki, A., Yonesaki, T. & Otsuka, Y. Characterization of the interactions between Escherichia coli receptors, LPS and OmpC, and bacteriophage T4 long tail fibers. *MicrobiologyOpen* **5,** 1003–1015 (2016).

46. De Paepe, M. & Taddei, F. Viruses' life history: Towards a mechanistic basis of a trade-off between survival and reproduction among phages. *PLoS Biology* **4,** 1248–1256 (2006).

47. Bull, J. J. Optimality models of phage life history and parallels in disease evolution. *Journal of Theoretical Biology* **241,** 928–938 (2006).

48. Springman, R., Keller, T., Molineux, I. J. & Bull, J. J. Evolution at a high imposed mutation rate: Adaptation obscures the load in phage T7. *Genetics* **184,** 221–232 (2010).

49. Husimi, Y. Selection and evolution of bacteriophages in cellstat. *Advances in Biophysics* **25,** 1–43 (1989).

50. Hu, B., Margolin, W., Molineux, I. J. & Liu, J. The bacteriophage T7 virion undergoes extensive structural remodeling during infection. *Science* **339,** 576–579 (2013).

51. González-García, V. A. *et al.* Conformational changes leading to T7 DNA delivery upon interaction with the bacterial receptor. *Journal of Biological Chemistry* **290,** 10038–10044 (2015).

52. Cerritelli, M. E., Wall, J. S., Simon, M. N., Conway, J. F. & Steven, A. C. Stoichiometry and Domainal Organization of the Long Tail-fiber of Bacteriophage T4: A Hinged Viral Adhesin. *Journal of Molecular Biology* **260,** 767–780 (1996).

53. Bartual, S. G. *et al.* Structure of the bacteriophage T4 long tail fiber receptor-binding tip. *Proceedings of the National Academy of Sciences* **107,** 20287–20292 (2010).

54. Hu, B., Margolin, W., Molineux, I. J. & Liu, J. Structural remodeling of bacteriophage T4 and host membranes during infection initiation. *Proceedings of the National Academy of Sciences* **112,** E4919–E4928 (2015).

55. Rassam, P. *et al.* Supramolecular assemblies underpin turnover of outer membrane proteins in bacteria. *Nature* **523,** 333 EP - (June 2015).

56. Demerec, M. & Fano, U. Bacteriophage-Resistant Mutants in Escherichia coli. *Genetics* **30,** 119 (1945).

57. Molineux, I. J. The T7 group. *The Bacteriophages,* 277–301 (2006).

58. Casjens, S. R. & Molineux, I. J. in *Viral Molecular Machines* (eds Rossmann, M. G. & Rao, V. B.) 143–179 (Springer US, Boston, MA, 2012).

59. Molineux, I. J. No syringes please, ejection of phage T7 DNA from the virion is enzyme driven. *Molecular Microbiology* **40,** 1–8 (2001).

60. Chang, C. Y., Kemp, P. & Molineux, I. J. Gp15 and gp16 cooperate in translocating bacteriophage T7 DNA into the infected cell. *Virology* **398,** 176–186 (2010).

61. Leptihn, S., Gottschalk, J. & Kuhn, A. T7 Ejectosome Assembly: A story unfolds. *Bacteriophage* **7081,** 00–00 (2016).

62. Garcia, L. R. & Molineux, I. J. Rate of translocation of bacteriophage T7 DNA across the membranes of Escherichia coli. *Journal of Bacteriology* **177,** 4066–4076 (1995).

63. Moffatt, B. A. & Studier, F. W. Entry of bacteriophage T7 DNA into the cell and excape from host restriction. *Journal of Bacteriology* **170,** 2095–2105 (1988).

64. Savalia, D., Robins, W., Nechaev, S., Molineux, I. & Severinov, K. The Role of the T7 Gp2 Inhibitor of Host RNA Polymerase in Phage Development. *Journal of Molecular Biology* **402,** 118–126 (Sept. 2010).

65. Qimron, U., Kulczyk, A. W., Hamdan, S. M., Tabor, S. & Richardson, C. C. Inadequate inhibition of host RNA polymerase restricts T7 bacteriophage growth on hosts overexpressing udk. *Molecular Microbiology* **67,** 448–457 (2008).

66. DeWyngaert, M. A. & Hinkle, D. C. Bacterial mutants affecting phage T7 DNA replication produce RNA polymerase resistant to inhibition by the T7 gene 2 protein. *Journal of Biological Chemistry* **254,** 11247–11253 (1979).

67. DeWyngaert, M. A. & Hinkle, D. C. Characterization of the defects in bacteriophage T7 DNA synthesis during growth in the Escherichia coli mutant tsnB. *Journal of Virology* **33,** 780–788 (1980).

68. Holmgren, A. Thioredoxin. *Annual Review of Biochemistry* **54,** 237–271 (1985).

69. Lee, S.-J. & Richardson, C. C. Choreography of bacteriophage T7 DNA replication. *Current Opinion in Chemical Biology* **15.** Molecular Machines/-Analytical Techniques, 580–586 (2011).

70. Lee, J., Chastain, P. D., Griffith, J. D. & Richardson, C. C. Lagging strand synthesis in coordinated DNA synthesis by bacteriophage T7 replication proteins. *Journal of Molecular Biology* **316,** 19–34 (2002).

71. Berry, J., Rajaure, M., Pang, T. & Young, R. The Spanin Complex Is Essential for Lambda Lysis. *Journal of Bacteriology* **194,** 5667–5674 (2012).

72. Berry, J. D., Rajaure, M. & Young, R. Spanin function requires subunit homodimerization through intermolecular disulfide bonds. *Molecular Microbiology* **88,** 35–47 (2013).

73. Young, R. Phage lysis: Three steps, three choices, one outcome. *Journal of Microbiology* **52,** 243–258 (Mar. 2014).

74. Kongari, R. *et al.* Phage spanins: diversity, topological dynamics and gene convergence. *BMC Bioinformatics* **19,** 326 (2018).

75. Hamdan, S. M. *et al.* A unique loop in T7 DNA polymerase mediates the binding of helicase-primase, DNA binding protein, and processivity factor. *Proceedings of the National Academy of Sciences* **102,** 5096–5101 (2005).

76. Kulczyk, A. & Richardson, C. in *DNA Replication Across Taxa* (eds Kaguni, L. S. & Oliveira, M. T.) 89–136 (Academic Press, 2016).

77. Geertsema, H. *Single-molecule studies of DNA replication: Visualization of DNA replication by the T7 bacteriophage replisome at a single-molecule level* English. PhD thesis (University of Groningen, 2014). ISBN: 978-90-367-7402-4.

78. Duderstadt, K. E. *et al.* Simultaneous Real-Time Imaging of Leading and Lagging Strand Synthesis Reveals the Coordination Dynamics of Single Replisomes. *Molecular Cell* **64,** 1035–1047 (Dec. 2016).

79. Springman, R., Kapadia-desai, D. S., Molineux, I. J. & Bull, J. J. Evolutionary Recovery of a Recombinant Viral Genome. *G3: Genes, Genomes, Genetics* **2,** 825–830 (2012).

80. Bull, J. J., Badgett, M. R. & Molineux, I. J. A General Mechanism for Viral Resistance to Suicide Gene Expression. *Journal of Molecular Evolution* **53,** 47–54 (July 2001).

81. Heineman, R. H. & Bull, J. J. Testing optimality with experimental evolution: Lysis time in a bacteriophage. *Evolution* **61,** 1695–1709 (2007).

82. Paff, M. L., Stolte, S. P. & Bull, J. J. Lethal mutagenesis failure may augment viral adaptation. *Molecular Biology and Evolution* **31,** 96–105 (2014).

83. Heineman, R. H., Springman, R. & Bull, J. J. Optimal Foraging by Bacteriophages through Host Avoidance. *The American Naturalist* **171,** E149–E157 (2008).

84. Benmayor, R., Hodgson, D. J., Perron, G. G. & Buckling, A. Host Mixing and Disease Emergence. *Current Biology* **19,** 764–767 (2009).

85. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472,** 499–503 (2011).

86. Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A. & Liu, D. R. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nature Chemical Biology* **10,** 216–222 (2014).

87. Qimron, U., Marintcheva, B., Tabor, S. & Richardson, C. C. Genomewide screens for Escherichia coli genes affecting growth of T7 bacteriophage. *Proceedings of the National Academy of Sciences* **103,** 19039–19044 (2006).

88. Nakao, R., Ramstedt, M., Wai, S. N. & Uhlin, B. E. Enhanced Biofilm Formation by Escherichia coli LPS Mutants Defective in Hep Biosynthesis. *PLoS ONE* **7** (2012).

89. Arunmanee, W. *et al.* Gram-negative trimeric porins have specific LPS binding sites that are essential for porin biogenesis. *Proceedings of the National Academy of Sciences* **113,** E5034–E5043 (2016).

90. Pouillot, F., Blois, H. & Iris, F. Genetically engineered virulent phage banks in the detection and control of emergent pathogenic bacteria. *Biosecur Bioterror* **8,** 155–169 (June 2010).

91. Garcia-Doval, C. & van Raaij, M. J. Structure of the receptor-binding carboxy-terminal domain of bacteriophage T7 tail fibers. *Proceedings of the National Academy of Sciences* **109,** 9390–9395 (2012).

92. Pell, L. G., Kanelis, V., Donaldson, L. W., Lynne Howell, P. & Davidson, A. R. The phage λ major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. *Proceedings of the National Academy of Sciences* **106,** 4160–4165 (2009).

93. Tétart, F., Repoila, F., Monod, C. & Krisch, H. M. Bacteriophage T4 host range is expanded by duplications of a small domain of the tail fiber adhesin. *Journal of Molecular Biology* **258,** 726–731 (1996).

94. Steven, A. C. *et al.* Molecular substructure of a viral receptor-recognition protein. *Journal of Molecular Biology* **200,** 351–365 (Mar. 1988).

95. Felek, S. & Krukonis, E. S. The Yersinia pestis Ail Protein Mediates Binding and Yop Delivery to Host Cells Required for Plague Virulence. *Infection and Immunity* **77,** 825–836 (2009).

96. Kolodziejek, A. M. *et al.* Phenotypic characterization of OmpX, an Ail homologue of Yersinia pestis KIM. *Microbiology* **153,** 2941–2951 (2007).

97. Kolodziejek, A. M., Hovde, C. J. & Minnich, S. A. Yersinia pestis Ail: multiple roles of a single protein. *Frontiers in Cellular and Infection Microbiology* **2,** 1–10 (2012).

98. Yamashita, S. *et al.* Article Structural Insights into Ail-Mediated Adhesion in Yersinia pestis. *Structure/Folding and Design* **19,** 1672–1682 (2011).

99. Hausmann, R. & Gomez, B. Amber mutants of bacteriophages T3 and T7 defective in phage-directed deoxyribonucleic acid synthesis. *J Virol* **1,** 779–792 (Aug. 1967).

100. Studier, F. The genetics and physiology of bacteriophage T7. *Virology* **39,** 562–574 (1969).

101. Kim, Y. T. & Richardson, C. C. Bacteriophage T7 gene 2.5 protein: an essential protein for DNA replication. *Proceedings of the National Academy of Sciences of the United States of America* **90,** 10173–7 (1993).

102. Marintcheva, B., Qimron, U., Yu, Y., Tabor, S. & Richardson, C. Mutations in the gene 5 DNA polymerase of bacteriophage T7 suppress the dominant lethal phenotype of gene 2.5 ssDNA binding protein lacking the C-terminal phenylalanine. *Molecular Microbiology* **72,** 869–880 (2009).

103. Kiro, R., Shitrit, D. & Qimron, U. Efficient engineering of a bacteriophage genome using the type I-E CRISPR-Cas system. *RNA Biology* **11,** 42–44 (2014).

104. Manor, M. & Qimron, U. Selection of Genetically Modified Bacteriophages Using the CRISPR-Cas System. *Bio-protocol* **7** (2017).

105. Modrich, P. & Grant, S. Bacteriophage T7 Deoxyribonucleic Acid Replication in Vitro. *The Journal of Biological Chemistry,* 5508–5514 (1975).

106. Mark, D. F. & Richardson, C. C. Escherichia coli thioredoxin: A subunit of bacteriophage T7 DNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **73,** 780–4 (1976).

107. Chamberlin, M. Isolation and Characterization of Prototrophic Mutants of Escherichia coli Unable to Support the Intracellular Growth of T7. *Journal of Virology* **14,** 509–516 (1974).

108. Briozzo, P. *et al.* Structures of Escherichia coli CMP kinase alone and in complex with CDP: a new fold of the nucleoside monophosphate binding domain and insights into cytosine nucleotide specificity. *Structure* **6,** 1517–1527 (Dec. 1998).

109. Kiro, R. *et al.* Gene product 0.4 increases bacteriophage T7 competitiveness by inhibiting host cell division. *Proceedings of the National Academy of Sciences* **110,** 19549–19554 (2013).

110. Dill, K. A., Ozkan, S. B., Shell, M. S. & Weikl, T. R. The protein folding problem. *Annu Rev Biophys* **37,** 289–316 (June 2008).

111. Dill, K. A. & MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science* **338,** 1042–1046 (2012).

112. Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F. & Voigt, C. A. Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Research* **40,** 8773–8781 (2012).

113. Silva-Rocha, R. *et al.* The Standard European Vector Architecture (SEVA): A coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Research* **41** (2013).

114. *iGEM Parts Registry* http://parts.igem.org/.

115. Schrödinger, LLC. *The PyMOL Molecular Graphics System, Version 1.8* Nov. 2015.

116. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25,** 3389–3402 (Sept. 1997).

117. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **46,** W296–W303 (May 2018).

118. Guex, N., Peitsch, M. C. & Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* **30,** S162–S173 (2009).

119. Benkert, P., Biasini, M. & Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **27,** 343–350 (Dec. 2010).

120. Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L. & Schwede, T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports* **7,** 10480 (2017).

121. Solovyev, V. & Salamov, A. Automatic annotation of microbial genomes and metagenomic sequences. *Metagenomics and its applications in agriculture, biomedicine and environmental studies,* 61–78 (Jan. 2011).

122. Tsang, T. M., Wiese, J. S., Felek, S., Kronshage, M. & Krukonis, E. S. Ail Proteins of Yersinia pestis and Y. pseudotuberculosis Have Different Cell Binding and Invasion Activities. *PLoS ONE* **8,** 1–9 (2013).

123. Kemp, P., Gupta, M. & Molineux, I. J. Bacteriophage T7 DNA ejection into cells is initiated by an enzyme-like mechanism. *Molecular Microbiology* **53,** 1251–1265 (2004).

124. Jacobus, A. P. & Gross, J. Optimal cloning of PCR fragments by homologous recombination in Escherichia coli. *PLoS ONE* **10,** 1–17 (2015).

125. Grigonyte, A. *Engineering bacteriophages to enhance their potential use in therapy.* Unpublished. PhD thesis (University of Warwick, Sept. 2018).

126. Yosef, I., Goren, M. G., Globus, R., Molshanski-Mor, S. & Qimron, U. Extending the Host Range of Bacteriophage Particles for DNA Transduction. *Molecular Cell* **66,** 721–728.e3 (2017).

127. Gasteiger, E. *et al.* in *The Proteomics Protocols Handbook* (ed Walker, J. M.) 571–607 (Humana Press, Totowa, NJ, 2005). ISBN: 978-1-59259-890-8.

128. Berbís Moreno, M. Á. *Aplicación y nuevos desarrollos de la espectroscopía por RMN para el estudio de procesos de reconocimiento molecular entre carbohidratos y sus receptores* PhD thesis (Universidad Complutense de Madrid, 2015).

129. Knirel, Y. A. *et al.* New Features of Yersinia Lipopolysaccharide Structures as Revealed by High-Resolution Electrospray Ionization Mass Spectrometry. *Advanced Science Letters* **1,** 192–198 (2008).

130. González-García, V. A. *et al.* Characterization of the initial steps in the T7 DNA ejection process. *Bacteriophage* **5,** e1056904 (2015).

131. Puig, A., Araujo, R., Jofre, J. & Frias-Lopez, J. Identification of cell wall proteins of Bacteroides fragilis to which bacteriophage B40-8 binds specifically. *Microbiology* **147,** 281–288 (2001).

132. Reusch, R. N. Insights into the Structure and Assembly of Escherichia coli Outer Membrane Protein A. *The FEBS Journal* **279,** 894–909 (2012).

133. Pautsch, A. & Schulz, G. E. Structure of the outer membrane protein A transmembrane domain. *Nature Structural Biology* **5.** Article, 1013–1017 (Nov. 1998).

134. Zakharian, E. & Reusch, R. Outer membrane protein A of Escherichia coli forms temperature-sensitive channels in planar lipid bilayers. *FEBS Letters* **555,** 229–235 (2003).

135. Zakharian, E. & Reusch, R. N. Kinetics of Folding of Escherichia coli OmpA from Narrow to Large Pore Conformation in a Planar Bilayer. *Biochemistry* **44,** 6701–6707 (2005).

136. Negoda, A., Negoda, E. & Reusch, R. Oligo-(R)-3-hydroxybutyrate modification of sorting signal enables pore formation by Escherichia coli OmpA. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1798,** 1480–1484 (2010).

137. Negoda, A., Negoda, E. & Reusch, R. N. Resolving the native conformation of Escherichia coli OmpA. *FEBS J* **277,** 4427–4437 (Nov. 2010).

138. Sugawara, E. & Nikaido, H. Pore-forming activity of OmpA protein of Escherichia coli. *Journal of Biological Chemistry* **267,** 2507–2511 (1992).

139. Saint, N., De, E., Julien, S., Orange, N. & Molle, G. Ionophore properties of OmpA of Escherichia coli. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1145,** 119–123 (1993).

140. Sugawara, E. & Nikaido, H. OmpA protein of Escherichia coli outer membrane occurs in open and closed channel forms. *Journal of Biological Chemistry* **269,** 17981–17987 (1994).

141. Arora, A., Rinehart, D., Szabo, G. & Tamm, L. K. Refolded Outer Membrane Protein A of Escherichia coliForms Ion Channels with Two Conductance States in Planar Lipid Bilayers. *Journal of Biological Chemistry* **275,** 1594–1600 (2000).

142. Masi, M. & Pagès, J.-M. Structure, Function and Regulation of Outer Membrane Proteins Involved in Drug Transport in Enterobactericeae: the OmpF/C - TolC Case. *Open Microbiol J* **7,** 22–33 (Mar. 2013).

143. Masi, M., Réfregiers, M., Pos, K. M. & Pagès, J. M. Mechanisms of envelope permeability and antibiotic influx and efflux in Gram-negative bacteria. *Nature Microbiology* **2** (2017).

144. Yamashita, E., Zhalnina, M. V., Zakharov, S. D., Sharma, O. & Cramer, W. A. Crystal structures of the OmpF porin: function in a colicin translocon. *The EMBO Journal* **27,** 2171–2180 (2008).

145. Tran, N. Q., Lee, S. J., Akabayov, B., Johnson, D. E. & Richardson, C. C. Thioredoxin, the processivity factor, sequesters an exposed cysteine in the thumb domain of bacteriophage T7 DNA polymerase. *Journal of Biological Chemistry* **287,** 39732–39741 (2012).

146. Huang, J., Brieba, L. G. & Sousa, R. Misincorporation by wild-type and mutant T7 RNA polymerases: identification of interactions that reduce misincorporation rates by stabilizing the catalytically incompetent open conformation. *Biochemistry* **39,** 11571–11580 (2000).

147. Gordon, A. J. E., Burns, P. A. & Glickman, B. W. N-Methyl-N'-nitro-N-nitrosoguanidine-induced mutation in a RecA strain of Escherichia coli.

Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis **201,** 219–228 (1988).

148. Gordon, A. J., Burns, P. A. & Glickman, B. W. N-Methyl-N'-nitro-N-nitrosoguanidine induced DNA sequence alteration; non-random components in alkylation mutagenesis. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **233,** 95–103 (1990).

149. Wong, I., Johnson, K. A., Patel, S. S. & Johnson, K. A. An Induced-Fit Kinetic Mechanism for DNA Replication Fidelity: Direct Measurement by Single-Turnover Kinetics. *Biochemistry* **30,** 526–537 (1991).

150. Kumar, J. K., Tabor, S. & Richardson, C. C. Role of the C-terminal Residue of the DNA Polymerase of Bacteriophage T7. *Journal of Biological Chemistry* **276,** 34905–34912 (2001).

151. Kunkel, T. A., Patel, S. S. & Johnson, K. A. Error-prone replication of repeated DNA sequences by T7 DNA polymerase in the absence of its processivity subunit. *Proceedings of the National Academy of Sciences of the United States of America* **91,** 6830–6834 (1994).

152. Söte, S., Kleine, S., Schlicke, M. & Brakmann, S. Directed Evolution of an Error-Prone T7 DNA Polymerase that Attenuates Viral Replication. *ChemBioChem* **12,** 1551–1558 (2011).

153. Shewhart, W. & Deming, W. *Statistical Method from the Viewpoint of Quality Control* Reprinted in 1986 by Courier Corporation (Washington, DC: Graduate School of the Department of Agriculture, 1939).

154. Box, G. E. P. Science and Statistics. *Journal of the American Statistical Association* **71,** 791–799 (1976).

155. Ellis, E. L. & Delbrück, M. The Growth of Bacteriophage. *Journal of General Physiology* **22,** 365–384 (Jan. 1939).

156. Delbrück, M. The Growth of Bacteriophage and Lysis of the Host. *Journal of General Physiology* **23,** 643–660 (May 1940).

157. Luria, S. E. & Dulbecco, R. Genetic Recombinations Leading to Production of Active Bacteriophage from Ultraviolet Inactivated Bacteriophage Particles. *Genetics* **34,** 93–125 (Mar. 1949).

158. Benzer, S. Resistance to ultraviolet light as an index to the reproduction of bacteriophage. *Journal of Bacteriology* **63,** 59–72 (Jan. 1952).

159. Kasman, L. M. *et al.* Overcoming the Phage Replication Threshold: a Mathematical Model with Implications for Phage Therapy. *Journal of Virology* **76,** 5557–5564 (2002).

160. Heineman, R. H., Molineux, I. J. & Bull, J. J. Evolutionary Robustness of an Optimal Phenotype : Re-evolution of Lysis in a Bacteriophage Deleted for Its Lysin Gene, 181–191 (2005).

161. Nguyen, H. M. & Kang, C. Lysis Delay and Burst Shrinkage of Coliphage T7 by Deletion of Terminator Tϕ Reversed by Deletion of Early Genes. *Journal of Virology* **88,** 2107–2115 (2014).

162. Studier, F. The genetics and physiology of bacteriophage T7. *Virology* **39,** 562–574 (1969).

163. Aksiyote Benbasat, J., Burck, K. B. & Miller, R. C. Superinfection exclusion and lack of conservative transfer of bacteriophage T7 DNA. *Virology* **87,** 164–171 (1978).

164. Karska-wysocki, B., Racine, J.-f. & Mamet-, M. D. Alkylation of T7 Bacteriophage Blocks Superinfection Exclusion. *Journal of Virology* **44,** 708–710 (1982).

165. Hirsch-Kauffmann, M., Pfennig-Yeh, M.-l., Ponta, H. & Herrlich, P. A virus-specified mechanism for the prevention of multiple infection—T7- and T3-

mutual and superinfection exclusion. *Molecular and General Genetics MGG* **149,** 243–249 (Jan. 1976).

166. McAllister, W. T. & Barrett, C. L. Superinfection exclusion by bacteriophage T7. *Journal of Virology* **24,** 709–11 (1977).

167. Hantke, K. & Braun, V. Fluorescence studies on first steps of phage-host interactions. *Virology* **58,** 310–312 (1974).

168. Campbell, A. Conditions for the Existence of Bacteriophage. *Evolution* **15,** 153 (1961).

169. Bull, J. J., Millstein, J., Orcutt, J. & Wichman, H. A. Evolutionary Feedback Mediated through Population Density, Illustrated with Viruses in Chemostats. *The American Naturalist* **167,** E39–E51 (2006).

170. Moldovan, R., Chapman-McQuiston, E. & Wu, X. L. On kinetics of phage adsorption. *Biophysical Journal* **93,** 303–315 (2007).

171. Shao, Y. & Wang, I. N. Bacteriophage adsorption rate and optimal lysis time. *Genetics* **180,** 471–482 (2008).

172. Bull, J. J., Heineman, R. H. & Wilke, C. O. The Phenotype-Fitness Map in Experimental Evolution of Phages. *PLoS ONE* **6,** 1–9 (2011).

173. Levin, B. R., Moineau, S., Bushman, M. & Barrangou, R. The Population and Evolutionary Dynamics of Phage and Bacteria with CRISPR-Mediated Immunity. *PLoS Genetics* **9** (2013).

174. Knezevic, P. *et al.* Phages of Pseudomonas aeruginosa: Response to environmental factors and in vitro ability to inhibit bacterial growth and biofilm formation. *Journal of Applied Microbiology* **111,** 245–254 (2011).

175. Taylor, B. P., Penington, C. J. & Weitz, J. S. Emergence of increased frequency and severity of multiple infections by viruses due to spatial clustering of hosts. *Physical Biology* **13,** 1–9 (2016).

176. Bull, J. J. *et al.* Phage-Bacterial Dynamics with Spatial Structure : Self Organization around Phage Sinks Can Promote Increased Cell Densities (2018).

177. Lee, S.-J., Chowdhury, K., Tabor, S. & Richardson, C. C. Rescue of bacteriophage T7 DNA polymerase of low processivity by suppressor mutations affecting gene 3 endonuclease. *Journal of Virology* **83,** 8418–27 (2009).

178. Pandey, M. *et al.* Two mechanisms coordinate replication termination by the Escherichia coli Tus-Ter complex. *Nucleic Acids Research* **43,** 5924–5935 (2015).

179. Yosef, I. *et al.* Natural selection underlies apparent stress-induced mutagenesis in a bacteriophage infection model. *Nature Microbiology* **1,** 1–5 (2016).

180. Mackal, R. P. & Kozloff, L. M. BIOCHEMICAL STUDIES OF VIRUS REPRODUCTION: XII. THE FATE OF BACTERIOPHAGE T7. *Journal of Biological Chemistry* **209,** 83–90 (1954).

181. Schwartz, M. The adsorption of coliphage lambda to its host: Effect of variations in the surface density of receptor and in phage-receptor affinity. *Journal of Molecular Biology* **103,** 521–536 (1976).

182. De Jong, H. Modeling and Simulation of Genetic Regulatory Systems: A Literature Review. *Journal of Computational Biology* **9,** 67–103 (2002).

183. Ando, T. & Skolnick, J. Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion. *Proceedings of the National Academy of Sciences* **107,** 18457–18462 (2010).

184. Höfling, F. & Franosch, T. Anomalous transport in the crowded world of biological cells. *Reports on Progress in Physics* **76,** 046602 (2013).

185. Dennehy, J. J., Friedenberg, N. A., Yang, Y. W. & Turner, P. E. Virus population extinction via ecological traps. *Ecology Letters* **10,** 230–240 (2007).

186. Cairns, B. J., Timms, A. R., Jansen, V. A. A., Connerton, I. F. & Payne, R. J. H. Quantitative models of in vitro bacteriophage-host dynamics and their application to phage therapy. *PLoS Pathogens* **5,** 1–10 (2009).

187. Beretta, E. & Kuang, Y. Modeling and analysis of a marine bacteriophage. *Mathematical Biosciences* **149,** 57–76 (Apr. 1998).

188. Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J. & Knuth, D. E. On the LambertW function. *Advances in Computational Mathematics* **5,** 329–359 (Dec. 1996).

189. Barry, D. *et al.* Analytical approximations for real values of the Lambert W-function. *Mathematics and Computers in Simulation* **53,** 95–103 (2000).

190. Corless, R. M., Jeffrey, D. J. & Knuth, D. E. *A Sequence of Series for the Lambert W Function* in *Proceedings of the 1997 International Symposium on Symbolic and Algebraic Computation* (ACM, Kihei, Maui, Hawaii, USA, 1997), 197–204. ISBN: 0-89791-875-4.

191. Bronstein, M., Corless, R. M., Davenport, J. H. & Jeffrey, D. J. Algebraic properties of the Lambert W function from a result of Rosenlicht and of Liouville. *Integral Transforms and Special Functions* **19,** 709–712 (2008).

192. Lehtonen, J. The Lambert W function in ecological and evolutionary models. *Methods in Ecology and Evolution* **7,** 1110–1118 (2016).

193. Rokyta, D., Badgett, M. R., Molineux, I. J. & Bull, J. J. Experimental Genomic Evolution: Extensive Compensation for Loss of DNA Ligase Activity in a Virus. *Molecular Biology and Evolution* **19,** 230–238 (Mar. 2002).

194. Bull, J. J., Badgett, M. R., Rokyta, D. & Molineux, I. J. Experimental evolution yields hundreds of mutations in a functional viral genome. *Journal of Molecular Evolution* **57,** 241–248 (2003).

195. You, L., Suthers, P. F. & Yin, J. Effects of Escherichia coli Physiology on Growth of Phage T7 In Vivo and In Silico. *Journal of Bacteriology* **184,** 1888–1894 (2002).

196. Ortiz-Suarez, M. L., Samsudin, F., Piggot, T. J., Bond, P. J. & Khalid, S. Full-Length OmpA: Structure, Function, and Membrane Interactions Predicted by Molecular Dynamics Simulations. *Biophysical journal* **111,** 1692–1702 (Oct. 2016).

197. Ried, G., Hindennach, I. & Henning, U. Role of lipopolysaccharide in assembly of Escherichia coli outer membrane proteins OmpA, OmpC, and OmpF. *Journal of Bacteriology* **172,** 6048–6053 (1990).

198. Laird, M. W., Kloser, A. W. & Misra, R. Assembly of LamB and OmpF in deep rough lipopolysaccharide mutants of Escherichia coli K-12. *Journal of Bacteriology* **176,** 2259–2264 (1994).

199. Storms, Z. J., Teel, M. R., Mercurio, K. & Sauvageau, D. The Virulence Index: A Metric for Quantitative Analysis of Phage Virulence. *PHAGE* **1,** 17–26 (2019).

200. Niu, Y. *et al.* Host range and lytic capability of four bacteriophages against bovine and clinical human isolates of Shiga toxin-producing Escherichia coli O157:H7. *Journal of Applied Microbiology* **107,** 646–656 (2009).

201. Niu, Y. D., Stanford, K., Ackermann, H.-W. & McAllister, T. A. Characterization of 4 T1-like lytic bacteriophages that lyse Shiga-toxin Escherichia coli O157:H7. *Canadian Journal of Microbiology* **58,** 923–927 (2012).

202. Wang, J. *et al.* Feces of feedlot cattle contain a diversity of bacteriophages that lyse non-O157 Shiga toxin-producing Escherichia coli. *Canadian Journal of Microbiology* **61,** 467–475 (2015).

203. Niu, Y. D., McAllister, T. A., Nash, J. H. E., Kropinski, A. M. & Stanford, K. Four Escherichia coli O157:H7 Phages: A New Bacteriophage Genus and Taxonomic Classification of T1-Like Phages. *PLOS ONE* **9,** 1–11 (June 2014).

204. Xie, Y., Wahab, L. & Gill, J. J. Development and Validation of a Microtiter Plate-Based Assay for Determination of Bacteriophage Host Range and Virulence. *Viruses* **10.** PMC5923483[pmcid], 189 (Apr. 2018).

205. Rajnovic, D., Muñoz-Berbel, X. & Mas, J. Fast phage detection and quantification: An optical density-based approach. *PLOS ONE* **14,** 1–14 (May 2019).

206. Mudgal, P., Breidt, F., Lubkin, S. R. & Sandeep, K. P. Quantifying the Significance of Phage Attack on Starter Cultures: a Mechanistic Model for Population Dynamics of Phage and Their Hosts Isolated from Fermenting Sauerkraut. *Applied and Environmental Microbiology* **72,** 3908–3915 (2006).

207. Storms, Z. J. & Sauvageau, D. Evidence that the heterogeneity of a T4 population is the result of heritable traits. *PLoS ONE* **9,** 1–19 (2014).

208. Taleb, N. N. A Short Note on P-Value Hacking. *arXiv e-prints.* arXiv: 1603.07532 [stat.AP]. https://arxiv.org/abs/1603.07532 (Mar. 2016).

209. Hillis, D., Huelsenbeck, J. & Cunningham, C. Application and accuracy of molecular phylogenies. *Science* **264,** 671–677 (1994).

210. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403,** 335 (Jan. 2000).

211. http://mcb.berkeley.edu/labs/krantz/protocols/electrocomp_cells.pdf.

212. Hillson, N. J., Rosengarten, R. D. & Keasling, J. D. J5 DNA assembly design automation software. *ACS Synthetic Biology* **1,** 14–21 (2012).

213. Hall, B. G., Acar, H., Nandipati, A. & Barlow, M. Growth rates made easy. *Molecular Biology and Evolution* **31,** 232–238 (2014).

214. Papili Gao, N. *Phage therapy: A software system for phage quantification and kinetic model inference.* MA thesis (Università di Bologna, 2015).

215. Endy, D., Kong, D. & Yin, J. Intracellular kinetics of a growing virus: A genetically structured simulation for bacteriophage T7. *Biotechnology and Bioengineering* **55,** 375–389 (1997).

216. Eun, H.-M. in *Enzymology Primer for Recombinant DNA Technology* (ed Eun, H.-M.) 345–489 (Academic Press, San Diego, 1996). ISBN: 978-0-12-243740-3.

217. Letarov, A. V. & Kulikov, E. E. Adsorption of Bacteriophages on Bacterial Cells. *Biochemistry Mosc.* **82,** 1632–1658 (Dec. 2017).

218. Roberts, I. *et al.* Molecular cloning and analysis of genes for production of K5, K7, K12, and K92 capsular polysaccharides in Escherichia coli. *Journal of Bacteriology* **168,** 1228–1223 (1986).

219. Jiménez, N. *et al.* Effects of lipopolysaccharide biosynthesis mutations on K1 polysaccharide association with the Escherichia coli cell surface. *Journal of Bacteriology* **194,** 3356–3367 (2012).

220. Pires, D. P., Oliveira, H., Melo, L. D. R., Sillankorva, S. & Azeredo, J. Bacteriophage-encoded depolymerases: their diversity and biotechnological applications. *Applied Microbiology and Biotechnology* **100,** 2141–2151 (Mar. 2016).

221. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7** (2011).

Integrative Biology

REVIEW ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Integr. Biol.*, 2016, 8, 465

Genetically modified bacteriophages

Antonia P. Sagona,[a] Aurelija M. Grigonyte,[ab] Paul R. MacDonald[ac] and Alfonso Jaramillo*[ad]

Received 25th October 2015,
Accepted 14th February 2016

DOI: 10.1039/c5ib00267b

www.rsc.org/ibiology

Phages or bacteriophages, viruses that infect and replicate inside bacteria, are the most abundant microorganisms on earth. The realization that antibiotic resistance poses a substantial risk to the world's health and global economy is revitalizing phage therapy as a potential solution. The increasing ease by which phage genomes can be modified, owing to the influx of new technologies, has led to an expansion of their natural capabilities, and a reduced dependence on phage isolation from environmental sources. This review will discuss the way synthetic biology has accelerated the construction of genetically modified phages and will describe the wide range of their applications. It will further provide insight into the societal and economic benefits that derive from the use of recombinant phages in various sectors, from health to biodetection, biocontrol and the food industry.

Insight, innovation, integration

In this review, we address the technological advances of synthetic biology that enable the genetic engineering of bacteriophages. Further, we discuss the wide spectrum of applications of genetically engineered bacteriophages.

1. Introduction

1.1 Why use recombinant phages?

Current advances in synthetic biology have facilitated the rational design, modification and construction of recombinant phages – that contain genetically engineered DNA and/or have been through genetic recombination – enabling the extension of their innate phenotypes. The host specificity of phages is evolutionarily refined, with most phages targeting one species. This host recognition specificity is conferred by receptor binding domains (RBDs) that are found in either the tail-spike or tail fiber protein assemblies of the virions. Many researchers have altered the specificity of phages towards non-native hosts.[1–6] In one such example, the host range of fd filamentous phage is altered by fusing a RBD from another filamentous phage (IKe), onto the infection-mediating protein of fd. Moving on to lytic phages, a genetically engineered T4 phage repository, was curated by randomizing the T4 RBD using polymerases with less fidelity to PCR amplify non-conserved regions. The repository was found to propagate in *Yersinia ruckeri* and *Pseudomonas aeruginosa*, indicating that the host range of T4 had been re-directed from its native *Escherichia coli* host.[2] Similarly, T3 (which naturally infects *E. coli*) and T7 phage (which infects *E. coli* and some species of *Yersinia*) was assembled in *Saccharomyces cerevisiae* with exogenous phage protein domains in order to alter its host range. As a proof of concept, it was demonstrated that modified T3 and T7 phage scaffolds could target pathogenic *Yersinia* and *Klebsiella* bacteria respectively. In addition, *Klebsiella* phage scaffolds were retargeted against *Escherichia coli* by swapping their phage tail components.[3]

Recombinant phages also function as vehicles for antimicrobials that are either incorporated into the phage[7] or attached to its surface.[8] For instance, they have been used to deliver light-activated antimicrobial agents (photosensitizers), which are seen as promising alternatives to antibiotics for treatment of common skin infections.[9] Phages can also be engineered to suppress host SOS DNA repair system, enhancing the effectiveness of broad-spectrum antibiotics *in vitro*.[10] Understandably, a natural concern about using engineered organisms is that the balance between natural phages in the environment may be affected. However, a recent study illustrated that phages endowed with gain of function mutations were out-competed by natural phages specific to the same host, suggesting that engineered phage genomes might not persist in the wild.[11] Additionally, one strategy to

[a] *Warwick Integrative Synthetic Biology Centre and School of Life Sciences, University of Warwick, Coventry, CV4 7AL, UK. E-mail: Alfonso.Jaramillo@synth-bio.org*
[b] *Synthetic Biology Centre for Doctoral Training, University of Warwick, Coventry, CV4 7AL, UK*
[c] *MOAC DTC, Senate House, University of Warwick, Coventry, CV4 7AL, UK*
[d] *iSSB, Genopole, CNRS, UEVE, Université Paris-Saclay, Évry, France*

commercialize phages is to re-engineer them to be non-replicative or non-lytic.[12,13]

1.2 Synthetic biology technologies for phage engineering

Genome engineering. A wide range of genome engineering methods have been applied to modify phage genomes and provide the desired characteristics for different applications. For simplicity, we have divided the methods in the text into *in vitro* and *in vivo.*

a. In vitro methods. Restriction endonuclease-based methods have been used to construct recombinant genomes *in vitro*. For instance, the genome of T7 was redesigned to remove overlapping genetic elements among other modifications. The new version of the genome (T7.1) is divided into 73 'parts' belonging to six sections. The first two sections of T7.1 were synthesized and shown to be viable.[14]

An additional, *in vitro* method is reported for the genetic modification of lytic phages, called genome recombineering with electroporated DNA (BRED), where they used the P1vir phage as a proof of concept. BRED is based on the use of recombinases (obtained from bacterial heterologous overexpression) to assemble phage genomes from purified phage genome and given synthetic DNA fragments.[15] BRED has been recently used to genetically modify a P1vir phage. Specifically, a copy of the mobile element IS1 was removed from the genome of P1vir phage, with the use of BRED.[16]

b. In vivo methods. In vivo methods have been used for bacteriophage engineering involving marker-based or marker-less selections of genetically modified bacteriophages. The most

Antonia P. Sagona

Dr Antonia P. Sagona received her PhD (2012) in cancer cell biology in the laboratory of Professor Harald Stenmark, in the Centre for Cancer Biomedicine, Faculty of Medicine, University of Oslo (Norway). She then received an EMBO Postdoctoral long-term fellowship (2013), which was hosted in the Division of Biomedical Cell Biology, Warwick Medical School (UK). She is currently a Research Fellow in the School of Life Sciences, University of Warwick (UK) and her work focuses in establishing an in vitro novel system for phage therapy, consisting of recombinant phages designed to target a specific pathogen in a mammalian cell environment.

Aurelija M. Grigonyte

Aurelija Marija Grigonyte is a PhD candidate in synthetic biology at SynBio Centre for Doctoral Training, University of Warwick, UK. Originally trained in biochemistry she is now working on engineering synthetic bacteriophages. Her main research interests include: understanding the key variants responsible for bacteriophage to its target binding; broadening bacteriophage host range; interactions between modified bacteriophage and mammalian cells; using engineered bacteriophages for treatment against variety of pathogens.

Paul R. MacDonald

Paul MacDonald is a PhD candidate in MOAC (Molecular Organisation and Assembly in Cells) Doctoral Training Centre, University of Warwick, UK interested in developing proof of concepts for phage therapeutics. Trained in mathematics at the University of Birmingham, UK, his research interests include phage genome engineering, receptor binding proteins and computational modeling.

Alfonso Jaramillo

Professor Alfonso Jaramillo holds a PhD in Particle Physics (1999) and a Habilitation in Biology (2007). After postdoctoral appointments with Prof. Wodak (ULB, Brussels) and Prof. Karplus (ULP, France and Harvard, USA), he started his academic career in 2003 as Assistant Professor at the Ecole Polytechnique (France), becoming tenured in 2005. There, he further developed computational synthetic biology. In 2009, after moving to Genopole (France) as CNRS-senior researcher, he started his experimental synthetic biology lab, utilizing directed evolution, microscopy, microfluidics and 3D printing. In 2013, he opened a second lab at the University of Warwick (UK), where he holds the Chair of Synthetic Biology.

commonly used *in vivo* method is homologous recombination (HR), where the sequence to be inserted is cloned into a vector with flanking regions matching upstream and downstream of phage genome sequence. Phage that has undergone HR can then be selected from the phage population, with the methods described below.

Marker-based selection methods exploit genes promoting phage propagation. They could be genes encoded in the phage genome or host factors required for phage propagation. For the former, the gene has to be previously deleted from the genome, where the phage may require a strain expressing such gene to propagate. A first step would consist of the homologous recombination of the insert and the gene marker. Later, a marker-deficient strain could be used for selection. Alternatively, a host factor can be used. For instance, in order to identify a marker for T7 selection, *E. coli* BW25113 is screened for genes that promote T7 phage growth, and these can be used as selection markers when editing T7 phage genome. Two potential genes are identified: *cmk* and *trxA*. Deletion of the *trxA* gene appears to confer phage infection inhibition, whereas deletion of the *cmk* gene in the bacterial host shows a lower efficiency of plating of T7 bacteriophage in comparison to the control bacteria. The reduced efficiency of plating, but not a complete absence of plaques, indicates that despite the gene deletion, some degree of T7 infection is obtained. These false positives can be removed by serial dilutions. In this study the *E. coli* gene *cmk*, which encodes for CMP/dCMP kinase, was inserted instead of T7 gp5 by HR between wild-type T7 and a plasmid containing the *cmk* gene. The plated recombined phage was shown to be negative for growth on cells that were deficient in *cmk*, and which did not contain a plasmid expressing gp5. This expected phenotype was confirmed to be true at the genotypic level by sequencing.[17]

False positives may occur in the marker-less selection method, CRISPR/Cas.[18] In a recent report, CRISPR/Cas system, and specifically type I-E CRISPR/Cas system, was used to select for engineered T7 bacteriophage (see Fig. 1). The T7 phage genome was edited by homologous recombination and the recombinant phages were selected by targeting wild-type phages with the CRISPR/Cas system.[19]

In another report, in addition to using CRISPR/Cas for the selection of engineered phage, CRISPR/Cas II-A system was used to *in vivo* modify phage 2972. Phage genome editing included gene exchange, point mutation and small or large deletions. For the gene exchange, *orf33* in the phage 2972 was replaced with methyltransferase gene of the type II restriction/modification (R/M) system LlaDCHI from *L. lactis*. Since the results showed a successful gene swap and a fully functioning methyltransferase it was concluded that the CRISPR/Cas engineering system can be used for gene insertions into the phage's genome.[20] CRISPR/Cas technology could be adapted to other phage genomes.

In another example, bacteriophage engineering is accomplished using an *in vivo* yeast platform as an alternative host for bacteriophage assembly. This platform is used to engineer phage with novel host ranges by swapping viral tail fibre scaffolds.

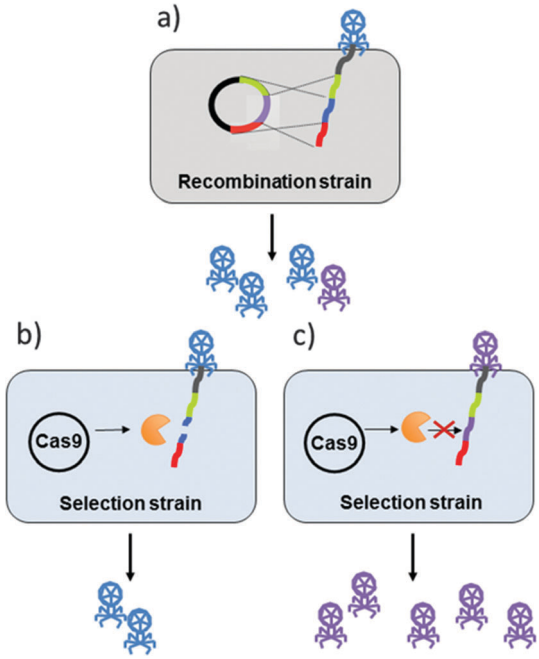


Fig. 1 Summary of CRISPR/Cas system used for recombinant phage selection *via* non-edited genome targeting. (a) The DNA of wild type (WT) bacteriophage (in light blue) undergoes homologous recombination following infection of the recombinant strain. The recombination strain contains a plasmid with a sequence (in violet) with flanking homology regions that replaces the WT phage genome sequence (in blue). This infection results in mixed population of phage progeny, producing recombinant phage (in violet) in addition to the WT phage (in blue). (b) The selection strain contains a CRISPR-Cas system, which targets and cleaves the WT sequence. Phages can still propagate in this strain, either because of inefficiencies associated with the CRISPR machinery, or because they contain one or more SNPs in the region targeted by the Cas-gRNA complex (deemed escape mutants). (c) The CRISPR-Cas system does not complex with the recombinant phage, resulting in relatively more progeny.

Phage genomes are placed in *Saccharomyces cerevisiae*, allow gene modifications and result in generation of engineered phage. Each fragment of the phage genome is first amplified by PCR while retaining a homologous overhang. First and last fragments of the phage genome have a homology region with the yeast artificial chromosome (YAC). All amplified phage genome fragments as well as YAC are then transformed into yeast where gap repair facilitated joining of all the fragments and the YAC according to homology regions. After the purification of the vector the phage can then be initiated to form functional phages when transformed into bacteria. This yeast phage-engineering platform has a great potential because any genomic loci can be modified even by adding genes toxic to *E. coli*.[3] One possible inconvenience is that the phage may have repeats at their ends (as it is the case for T7) and recombination may produce excision of the phage from the vector. This may be overcome by including a selective marker for yeast inside the phage genome.

Directed evolution methodologies. Directed evolution may be used to alter and optimize bacteriophage genomes providing a phenotypic or genotypic advantage in a given environment. It is possible to evolve bacteriophages *via* serial passaging, or

continuous culture in a bioreactor.[21] This platform was used to evolve lambda phage, which normally infects its host *via* LamB protein receptor, to infect *via* a novel OmpF receptor.[4] In another remarkable study, a phage was optimized by serial passages in a living mouse providing a 13 000 fold greater capacity of the phage to evade immune system and remain in the circulatory system. This was a result of a single mutation which led to amino acid substitution in the major lambda phage capsid protein E.[22]

PACE (phage-assisted continuous evolution) combines continuous culture with increased mutagenesis to accelerate the evolution of M13 phagemids (non-replicative phages that require a strain carrying a helper system). PACE is used to evolve regulatory molecules rather than structural proteins in the virion. This is done by removing an essential gene (gpIII) from the phage and placing it, in an inactivated form, in the host strain. The gene responsible for the initiation of the evolution is introduced into the phage. The product of this gene must activate gpIII to produce infectious progeny. It is possible to use a similar methodology to evolve proteins from the phage such as those conferring host specificity or involved in replication.[23] This method has been further expanded to evolve biomolecules with altered or highly specific new activities, using negative selection and modulation of selection stringency.[24]

Phage display. Phage display is based on generating a library of synthetic or natural peptides and then fusing them onto a coat protein of a bacteriophage. Modified phages that bind strongly to the ligand displayed are enriched *via* sequential recovery from the surface and upon that they re-infect bacteria to propagate and increase in number. Filamentous phages M13[25–33] and fd[34–36] are the most commonly used phages for phage display although T4,[37,38] T7[39–43] and lambda[44–47] phages have also been used. Traditionally phage display has been used for antibody production, proteomics,[48] therapeutics, diagnostics (specially for cancer applications), infectious diseases and drug discovery.[49,50] Phage display has also been used for epitope mapping, a method to identify the epitope of the antigen that interacts with an antibody. The identification of epitopes is important for the development of diagnostic tools, vaccines and new therapeutic targets.[51,52] Additionally, phage display has proven useful in targeting membrane receptors *via* the identification of their agonists and antagonists, which present biological applications as drugs for various diseases.[53] Phage display is also an excellent tool for the identification of protein–protein interactions.[54–57] Its biggest advantage, compared to other methods established in the protein–protein interactions field, is that highly diverse peptide libraries can be constructed at low cost.[56] These are some of the applications of phage display. The list is even broader and even more applications are expected to arise as the methodology evolves.[58–67]

Table 1 Engineered bacteriophages and their applications

Engineered bacteriophage	Application
M13	Phage display,[25–33] lethal delivery agent,[83] engineered protein purification,[84] nanomaterials,[87] vaccinology[109,110]
M13KE	Pathogen detection[100]
T7	Phage display,[39–43] gene therapy,[69] biofilm control[114]
Lambda (λ)	Phage display[44–47] vaccinology,[46,70] biocontrol[116]
T4	Phage display,[37,38] vaccinology[72,112]
ϕA1122	Pathogen detection[106]
A511	Pathogen detection[104]
HK620	Pathogen detection[98]
PBSPCA1	Agriculture[108]
fd	Phage display,[34–36] nanodevice fabrication and bottom-up manufacturing[85,86]

2. Applications of engineered phages

Herein, we discuss how genetically modified phages (listed in Table 1) are used in different fields (presented in Fig. 2).

2.1 Therapeutic applications

Natural phages have multiple barriers that could prevent them from being developed into viable phage therapeutic products. Issues can arise both from their entry to mammalian cells and from circumventing the immune response of the host. Modified phages can be developed to avoid inactivation by the host defense system and persist in the body, thus enhancing their therapeutic potential.

In this section, examples of recombinant phages as therapeutic agents for a variety of diseases are given. This includes phage therapy in mammalian hosts as well as phages as lethal delivery vehicles for prokaryotic hosts.

The T7 bacteriophage was genetically modified for a potential therapy of hepatitis B. Globally HCC (hepatocellular carcinoma) is the fifth most common cancer in men, and the eighth most common in women, and it was estimated that during the year 2000, more than 500 000 new cases arose.[68] The first step of hepatitis B virus (HBV) infection is the interaction between a cell surface receptor and the HBV envelope protein (specifically the PreS1 region). T7 phage was modified to display polypeptides of varying length (up to fifty amino acids) of the PreS1 region. This was achieved by fusing the peptides of interest (PreSi region variants) to the C-terminus of the capsid protein gp10B from T7. It was suggested that this system could be enhanced by displaying amino acids of the first half of the PreS1

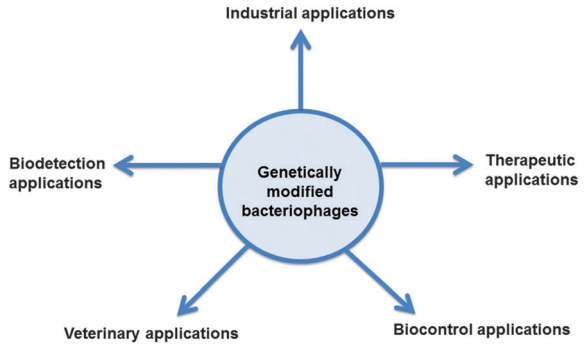


Fig. 2 A schematic representation of the synthetic bacteriophage applications discussed in this review.

region, which would increase modified phage uptake and hence the efficiency of desired gene transfer into the cell. This study introduces T7 display as a model system for gene delivery into mammalian host cells and elucidates recombinant phages expand upon natural phage therapy.[69]

Another example of therapeutic bacteriophage applications in mammalian systems is the display of a membrane glycoprotein, of H5 influenza virus, on bacteriophage lambda major capsid protein gpD.[70] Here the lambda bacteriophage is assembled without its major capsid protein first *in vivo*, and then a fusion of gpD and H5 influenza membrane glycoprotein is added *in vitro*. This approach can be used to 'decorate' gpD deficient phage with recombinant gpD fused proteins (which are heterologously expressed in insect cells). The recombinant phages bind to the receptors on red blood cells and create a lattice structure of interconnected virus and red blood cells. It is observed that two H5 influenza-specific monoclonal antibodies inhibit the binding of red blood cells to the recombinant phages. These modified phages mimic H5 influenza virus behavior by attaching to the red blood cells; this could potentially be exploited for the detection of influenza virus-specific antibodies in vaccine trials.[71]

T4 phage is another example of engineered phage use in vaccinology. The outer capsid protein of T4 phage was fused to a protective antigen from *Bacillus anthracis*, to develop a vaccine for anthrax. The fusion protein was expressed in *E. coli*, purified, and *in vitro* added to the assembled phage. The PA-T4 particles presented immunogenicity in mice in the absence of an adjuvant. This study provides a promising system for construction of customized vaccines against anthrax.[72]

Phages target bacteria more specifically than most antibiotics, and consequently have less effect on the human gut microbiome. However, lytic phages, whose concomitant cell lysis may result in the release of toxic substances (endotoxins) have encouraged the development of lysis-deficient phages. These lysis-deficient phages can be engineered by harnessing the phage machinery responsible for cell lysis. This consists of a membrane protein (normally deemed, holin) and endolysin or murein hydrolase. Holins form holes in the cell membrane, thus letting endolysin cross the membrane and degrade the peptidoglycan layer of the cell wall.[73] An example of engineering a recombinant phage that is lysis-deficient involves *Staphylococcus aureus* phage P954 where its endolysin gene was inactivated by a loss of function insertion.[74] Endolysin-deficient phages encoding lethal but non-lytic proteins are able to kill bacteria while reducing the endotoxin release.[75] Engineered non-lytic phages have been shown to be efficient in treating mice infected with *P. aeruginosa*, *E. coli* or *S. aureus* which present higher survival rates, due to the lower levels of endotoxin release.[13,74,76,77] Additionally, filamentous phages (which do not lyse the host) have been used for the specific delivery of lethal substances or genes to the site of infection.[78] Among possible genes, those encoding for modified holin,[79] lethal transcription regulator[80] and addiction toxins (which induce programmed cell death) have been reported.[81,82]

Modified phage particles can also be used as lethal delivery agents for efficient pathogen killing. A recent example is the engineering of the filamentous phage M13 to carry an integrin binding peptide and a fragment of the polymorphic membrane protein D from the sexually transmitted pathogen *Chlamydia trachomatis* (Ct), as a possible way to eliminate Ct infection. Based on this report, the engineered phage was able to significantly reduce Ct infection in both primary endocervical and HeLa cells, addressing the current lack of treatments against *Chlamydia trachomatis*.[83]

2.2 Industrial applications

High-gradient magnetic fishing (HGMF) partially purifies target products from heterogeneous bioprocess liquors. In HGMF, the target product is captured using magnetic adsorbent particles in combination with high-gradient magnetic separation equipment. HGMF binding capacity of microbeads was increased by placing an engineered M13 bacteriophage monolayer on a superparamagnetic (SPM) core of microbeads. This was achieved by genetic and chemical modification of the M13 coat proteins, pIII (minor coat) and pVIII (major coat) respectively. pIII protein was modified at its N-terminus to enable its binding onto nitrilotriacetic acid or silica coated SPMs, respectively. pVIII protein subunits of wild-type M13 were chemically cross-linked to a carboxyl-functionalized bead permitting side-on linkage of the phage to the SPMs. The phage-SPM particles, when used to fish the desired antibodies, led to $>90\%$ purified product from high protein solutions in one purification step.[84]

fd filamentous phage has been engineered to increase the affinity of gold (Au) to its protein coat. This was achieved by substituting five amino acids on the N-terminal region of p8, the fd phage's major coat protein.[85,86] These studies show how recombinant phages can be coated with a metal of interest, demonstrating the potential of recombinant phages as self-assembling templates for applications in bottom up manufacturing.

In addition, M13 filamentous bacteriophage has been used as a scaffold for the self-assembly of cobalt manganese oxide nanowires to make LiO_2 battery electrodes. Here, the phage coat protein gpVIII is modified to display peptides of negatively charged amino acids, and is able to interact with cationic metal precursors (such as cobalt and manganese) resulting in high production yield oxides. These oxides formed LiO_2 battery electrodes which were more porous and had a higher specific heat capacity in comparison to carbon electrodes.[87]

2.3 Biodetection applications

One area in which synthetic phages are proving useful is biodetection, where one of their main advantages is that they can be quickly amplified in the targeted live bacteria, compared to PCR or antibody-wash detection systems (which subsequently incur more false positives from the detection of dead bacteria).[88] Early adoptions of biodetection with synthetic phages have focused on the insertion of reporter genes into naturally isolated phages.[89–93] Most of these methods comprise of either inserting luciferase genes into the phage genome, or of fusing fluorescent proteins to the phage capsid, mimicking the engineering strategy employed in phage display.[94] It should be emphasized that these

systems are far from just academic enquires. Sample6 recently launched its DETECT/L kit on the back of luciferase-based recombinant phage technologies.[93] The kit correctly identifies 50 *Listeria* species, as well as correctly excluding 30 non-listeria species that were subjected to testing. However, converting luciferase-based assays to more optimal multiplex assays (*i.e.* assays that detect more than one bacterial species in a given sample) may not be straightforward,[95] which is an issue that could be circumvented by fluorescent-based reporter systems with compatible emission and excitation wavelengths.[96]

The precision of phage host recognition has been explored for pathogen detection from environmental samples. Water quality control is one of the major concerns for public health as well as marine environment, and rapid methods need to be developed to allow accurate pathogen identification. One such method used a phage-based fluorescent biosensor 'phagosensor' prototype for enteric bacteria detection.[97] In this system, the synthesis of the fluorescent protein only occurs after it is delivered to *E. coli* TD2158 by temperate bacteriophage HK620 carrying the fluorescent gene.[98] The recombinant phages were incubated with the sample for one hour, followed by flow cytometry, which allowed sensitive detection of environmental *E. coli* TD2158 strain in diluted samples, and in mixed co-cultures. The established template was also successfully adapted to phage P22 to detect *Salmonella enterica Typhimurium*.[99]

Colorimetry is another method for detecting pathogenic bacteria in water. In this instance, the target bacterial strain, ER2378, is trapped on a syringe filter followed by infection of a specific phage, M13KE.[100] ER2378 is a lacZα-complementing strain of *E. coli* that expresses the ω-domain of the β-gal (ωGal) and the αGal peptide is cloned in an intergenetic region of the M13KE genome.[101,102] Upon αGal peptide delivery to the bacterial strain, αGal is converted to the β-gal active form of αGal, which is detectable by colorimetric assay.[100]

Another example of bacteriophage-based pathogen detection is the high-intensity fluorophage that can detect *Mycobacterium tuberculosis*.[103] Phage A511 was modified to report cells of *Listeria* genus upon synthesis of a bacterial luciferase gene. The modified gene was placed downstream of the major capsid protein (cps) and was expressed upon phage infection of *Listeria* cells.[104]

Yersinia pestis is the etiological agent of the plague, which has seen a breakout of cases in America recently.[105] To detect *Yersinia pestis* in blood samples, a recombinant reporter phage containing bacterial luxAB reporter genes was inserted in an early-transcribed noncoding region of the plague-diagnostic lytic phage ϕA1122 by homologous recombination. Upon infection of *Yersinia pestis* with the recombinant ϕA1122, the bioluminescent phenotype was observed after 10 to 15 min.[106] In a following study, the reporter phage was assessed as a diagnostic tool for *Yersinia pestis* in samples taken directly from the blood. Even though it displayed 100% inclusivity for *Yersinia pestis*, some non-pestis *Yersinia* strains and *Enterobacteriaceae* also showed signal transduction. The reporter phage demonstrated rapid detection of antimicrobial susceptibility profiling upon antibiotic incubation of the blood samples. As a consequence these results suggest that the application of lytic reporter phages to detect bacterial pathogens in blood samples could reduce the time to diagnosis for patients afflicted with.[107]

Phage-reporter systems were also developed for agricultural settings. *Pseudomonas cannabina* pv. *alisalensis* and *Pseudomonas syringae* pv. *maculicola* are both causative agents for diseases of *Brassicaceae* family. In a recent study, phage-based diagnostic was developed to identify the cause of bacterial blight. *P. cannabina* pv. *alisalensis* or PBSPCA1 phage was modified by integration of bacterial *luxAB* genes in the place of nonessential *phoH* gene using homologous recombination upon wild-type PBSPCA1 phage infection of *P. cannabina* pv. *alisalensis* BS91 containing *luxAB* expression cassette. A successful detection of *Pseudomonas cannabina* pv. *alisalensis versus Pseudomonas syringae* pv. *maculicola* resulted in more than 100-fold increase in bioluminescence within 4 hours of tissue harvesting.[108]

2.4 Veterinary applications

Genetically engineered phages have a wide range of applications in veterinary science and medicine. In the vast majority of cases, recombinant phages deliver antigens to be used for vaccination against animal diseases.

Recombinant M13 bacteriophage has been used to vaccinate pigs against the tapeworm *Taenia solum*, which causes cysticercosis, a disease to which humans and pigs are susceptible. KETc1, KETc12 and GK1 peptides were fused to coat protein gpVIII and a recombinant antigen KETc7 was displayed on coat protein gpIII. The pooled phages were successful in reducing the number of cysts in the murine model (990 in mice receiving M13 *vs.* 338 immunized with the pool of recombinant phage). Preliminary work demonstrated sporadic effectiveness for a small sample of pigs, but further studies were needed to corroborate this evidence.[109] Such a study was performed on a larger sample size, and significantly reduced the occurrence of cysticercosis in vaccinated pigs by 54.2%.[110]

Phages lambda and T4, as mentioned earlier in this review, have been used to vaccinate against human diseases. In this section, we briefly mention their application as possible vaccines against pathogens that target animals. Recombinant lambda phage can be used as a potential vaccine against porcine Circovirus 2,[46] a virus that causes post-weaning multisystemic wasting syndrome (PMWS), and has proved costly in the swine industry.[111]

In another study, the T4 bacteriophage was used to develop a vaccine against infectious bursal disease virus (vvIBDV), a virus that causes infectious bursal disease (IBD) in chickens. Immunization of chickens with the recombinant T4-VP2 phage resulted in no clinical death; however, some temporary bursal damages were observed.[112]

2.5 Biocontrol applications

Biocontrol is the regulation of pest/pathogen levels by biological means; a strategy that is being progressively favored by industry. One of the major problems in industrial processing is biofilm formation, and this is especially apparent in food industry.[113] To this end, T7 bacteriophage was modified to express dispersin B (DspB),[114] an enzyme produced by

Actinobacillus actinomycetemcomitans.[115] DspB acts *via* β-1,6-*N*-acetyl-D-glucosamine hydrolysis which disrupts biofilm formation and integrity. DspB was placed downstream of capsid gene 10B under the control of the strong T7 10 promoter. This allowed DspB to be expressed intracellularly so that its release would occur during cell lysis. The results showed that DspB expressing phage was significantly more effective at killing *E. coli* in comparison to wild-type T7 and wild-type T3 bacteriophages, and that this phage reduced the amount of biofilm by a factor of 2.6 in comparison to non-engineered T7 control phage.[114]

In another study lambda phage was engineered to deliver CRISPR/Cas system to sensitize bacteria with antibiotic resistance genes. Once delivered to a pathogen, the CRISPR/Cas system was transferable between bacterial hosts so that bacteria with antibiotic genes would be outcompeted. In addition, pathogens with acquired CRISPR/Cas were no longer susceptible to engineered lytic phage infection. This system showed effective reduction in infections of antibiotic-resistant pathogens and thus offering a potential biocontrol system for hospital surface treatment.[116]

3. Discussion

Natural phages may be a solution to a myriad of issues in agriculture, biocontrol, and medicine: in agriculture, they are applied against plant infections; in biocontrol, in the protection and control of crops and food products; and in medicine, a number of trials have explored their safety and efficacy.

Phage therapy has reached a critical juncture in its development. Natural phages are, according to current legislature, unpatentable, as they are no longer considered novel, having first been introduced over a century ago.[117] This prevents them from being commercially viable in industry, especially for big pharma corporations that have to undergo expensive clinical trials, and so is an impediment to the development of infrastructure that can deliver treatments to patients. The intellectual property of recombinant phages, on the other hand, can be secured, thus overcoming these issues. Yet recombinant phages do more than to provide opportunities for profit; they extend upon natural phages by creating additional functionalities (such as delivery of variety of cargos, *e.g.* depolymerase[114]) as well as overcoming some of their limitations.

One such limitation is that natural phages can induce a mammalian immune response upon their entry, a response that could be avoided by modifying the phage's coat protein. Nonetheless, in some instances, an overexcited immune response is desired, for example, when developing vaccines against viral particles. In the case of HBV, phage-displayed viral peptides can evoke immune response, and help identify the antibodies specific to these viral peptides.

Furthermore, by refactoring phage genomes,[14] phages could have their sequences rejigged to avoid host restriction systems, by removing the palindromic sequences targeted by type II restriction enzymes, or by appending peptides that inhibit the CRISPR/Cas proteins. One can also envisage the swapping or addition of exogenous endo/exonucleases to cleave host genomes faster. Further work demonstrated that rational engineering of tail fibres can change host tropism of bacteriophages[1–6] and pyocins[118,119] (phage-like particles called bacteriocins, produced by some strains of *Pseudomonas aeruginosa* for use in intra-species warfare). Employing this in combination with directed evolution, phage cocktails could be generated to target a given pathogen through separate, or a combination of, receptors, and adapted to circumvent the host-immune system. Importantly, new techniques are constantly being developed in the fields of genome engineering and gene synthesis, decreasing also the cost of gene synthesis. It is expected that also because of that, the engineering of recombinant bacteriophages will be further facilitated and will enable the corresponding broadening of their applications.

In summary, although there are still ethical, socio-economical, and experimental issues to resolve, the groundwork of phages appears promising, and will surely come to establish itself at the forefront of personalized therapeutics and diagnostics.

Conflicts of interest

The authors declare no competing financial interests.

Acknowledgements

The work is supported by the grants FP7-ICT-FET 610730 (EVOPROG), EPSRC-BBSRC BB/M017982/1 (WISB centre) and the startup allocation grant from the School of Life Sciences to A. J. A. P. S. is funded by FP7-ICT-FET 610730 (EVOPROG). A. G. is funded by a doctoral fellowship from the EPSRC & BBSRC Centre for Doctoral Training in Synthetic Biology (grant EP/L016494/1). P. M. is funded by a doctoral fellowship from the EPSRC Doctoral Training Centre *Molecular Organisation and Assembly in Cells* (Grant No. EP/F500378/1). We also wish to acknowledge Matthew Tridgett for contributing to Fig. 1.

References

1 R. Marzari, D. Sblattero, M. Righi and A. Bradbury, *Gene*, 1997, **185**, 27–33.
2 F. Pouillot, H. Blois and F. Iris, *Biosecur Bioterror*, 2010, **8**, 155–169.
3 H. Ando, S. Lemire, D. P. Pires and T. K. Lu, *Cell Syst.*, 2015, **1**, 187–196.
4 J. R. Meyer, D. T. Dobias, J. S. Weitz, J. E. Barrick, R. T. Quick and R. E. Lenski, *Science*, 2012, **335**, 428–432.
5 F. Mahichi, A. J. Synnott, K. Yamamichi, T. Osada and Y. Tanji, *FEMS Microbiol. Lett.*, 2009, **295**, 211–217.
6 M. Yoichi, M. Abe, K. Miyanaga, H. Unno and Y. Tanji, *J. Biotechnol.*, 2005, **115**, 101–107.
7 E. W. Kovacs, J. M. Hooker, D. W. Romanini, P. G. Holder, K. E. Berry and M. B. Francis, *Bioconjugate Chem.*, 2007, **18**, 1140–1147.
8 I. Yacoby, H. Bar and I. Benhar, *Antimicrob. Agents Chemother.*, 2007, **51**, 2156–2163.

9 M. L. Embleton, S. P. Nair, W. Heywood, D. C. Menon, B. D. Cookson and M. Wilson, *Antimicrob. Agents Chemother.*, 2005, **49**, 3690–3696.
10 T. K. Lu and J. J. Collins, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 4629–4634.
11 E. G. Gladstone, I. J. Molineux and J. J. Bull, *J. Biol. Eng.*, 2012, **6**, 13.
12 S. P. Denyer, N. A. Hodges and S. P. Gorman, *Hugo and Russell's pharmaceutical microbiology*, John Wiley & Sons, 8th edn, 2008.
13 V. D. Paul, S. Sundarrajan, S. S. Rajagopalan, S. Hariharan, N. Kempashanaiah, S. Padmanabhan, B. Sriram and J. Ramachandran, *BMC Microbiol.*, 2011, **11**, 195.
14 L. Y. Chan, S. Kosuri and D. Endy, *Mol. Syst. Biol.*, 2005, **1**, 0018.
15 L. J. Marinelli, M. Piuri, Z. Swigonova, A. Balachandran, L. M. Oldfield, J. C. van Kessel and G. F. Hatfull, *PLoS One*, 2008, **3**, e3957.
16 T. Feher, I. Karcagi, F. R. Blattner and G. Posfai, *J. Microb. Biotechnol.*, 2012, **5**, 466–476.
17 U. Qimron, B. Marintcheva, S. Tabor and C. C. Richardson, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 19039–19044.
18 M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna and E. Charpentier, *Science*, 2012, **337**, 816–821.
19 R. Kiro, D. Shitrit and U. Qimron, *RNA Biol*, 2014, **11**, 42–44.
20 B. Martel and S. Moineau, *Nucleic Acids Res.*, 2014, **42**, 9504–9513.
21 Y. Husimi, *Adv. Biophys.*, 1989, **25**, 1–43.
22 C. R. Merril, B. Biswas, R. Carlton, N. C. Jensen, G. J. Creed, S. Zullo and S. Adhya, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 3188–3192.
23 K. M. Esvelt, J. C. Carlson and D. R. Liu, *Nature*, 2011, **472**, 499–503.
24 J. C. Carlson, A. H. Badran, D. A. Guggiana-Nilo and D. R. Liu, *Nat. Chem. Biol.*, 2014, **10**, 216–222.
25 W. Markland, B. L. Roberts, M. J. Saxena, S. K. Guterman and R. C. Ladner, *Gene*, 1991, **109**, 13–19.
26 D. Bhardwaj, S. S. Singh, S. Abrol and V. K. Chaudhary, *J. Immunol. Methods*, 1995, **179**, 165–175.
27 B. L. Roberts, W. Markland and R. C. Ladner, *Methods Enzymol.*, 1996, **267**, 68–82.
28 J. A. Chappel, M. He and A. S. Kang, *J. Immunol. Methods*, 1998, **221**, 25–34.
29 T. Belien, K. Hertveldt, K. Van den Brande, J. Robben, S. Van Campenhout and G. Volckaert, *J. Biotechnol.*, 2005, **115**, 249–260.
30 S. S. Sidhu, B. K. Feld and G. A. Weiss, *Methods Mol. Biol.*, 2007, **352**, 205–219.
31 K. Hertveldt, T. Belien and G. Volckaert, *Methods Mol. Biol.*, 2009, **502**, 321–339.
32 D. Ghosh, Y. Lee, S. Thomas, A. G. Kohli, D. S. Yun, A. M. Belcher and K. A. Kelly, *Nat. Nanotechnol.*, 2012, **7**, 677–682.
33 Y. Ivarsson, R. Arnold, M. McLaughlin, S. Nim, R. Joshi, D. Ray, B. Liu, J. Teyra, T. Pawson, J. Moffat, S. S. Li, S. S. Sidhu and P. M. Kim, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 2542–2547.
34 M. Tornetta, S. Baker, B. Whitaker, J. Lu, Q. Chen, E. Pisors, L. Shi, J. Luo, R. Sweet and P. Tsui, *J. Immunol. Methods*, 2010, **360**, 39–46.
35 J. W. Gillespie, A. L. Gross, A. T. Puzyrev, D. Bedi and V. A. Petrenko, *Front Microbiol*, 2015, **6**, 628.
36 A. Prisco and P. De Berardinis, *Int. J. Mol. Sci.*, 2012, **13**, 5179–5194.
37 V. P. Efimov, I. V. Nepluev and V. V. Mesyanzhinov, *Virus Genes*, 1995, **10**, 173–177.
38 M. Gamkrelidze and K. Dabrowska, *Arch. Microbiol.*, 2014, **196**, 473–479.
39 H. M. Li, K. Guo, Z. Yu, R. Feng and P. Xu, *Thorac. Cancer*, 2015, **6**, 469–474.
40 C. L. Wong, C. C. Sieo and W. S. Tan, *J. Virol. Methods*, 2013, **193**, 611–619.
41 M. Tsuboyama and I. Maeda, *J. Biosci. Bioeng.*, 2013, **116**, 28–33.
42 S. Liu, Y. Sun, H. Chen, S. Song and Y. Xu, *Appl. Biochem. Biotechnol.*, 2010, **162**, 1206–1213.
43 M. Dai, J. Temirov, E. Pesavento, C. Kiss, N. Velappan, P. Pavlik, J. H. Werner and A. R. Bradbury, *Protein Eng., Des. Sel.*, 2008, **21**, 413–424.
44 N. Sternberg and R. H. Hoess, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 1609–1613.
45 E. Pavoni, P. Vaccaro, V. D'Alessio, R. De Santis and O. Minenkova, *BMC Biotechnol.*, 2013, **13**, 79.
46 L. N. Gamage, J. Ellis and S. Hayes, *Vaccine*, 2009, **27**, 6595–6604.
47 R. Levy, I. J. Molineux, B. L. Iverson and G. Georgiou, *J. Immunol. Methods*, 2007, **321**, 164–173.
48 G. N. Sundell and Y. Ivarsson, *BioMed Res. Int.*, 2014, **2014**, 176172.
49 N. Lonberg, *Curr. Opin. Immunol.*, 2008, **20**, 450–459.
50 D. J. Christensen, E. B. Gottlin, R. E. Benson and P. T. Hamilton, *Drug Discovery Today*, 2001, **6**, 721–727.
51 L. F. Wang and M. Yu, *Curr. Drug Targets*, 2004, **5**, 1–15.
52 J. M. Gershoni, A. Roitburd-Berman, D. D. Siman-Tov, N. Tarnovitski Freund and Y. Weiss, *BioDrugs*, 2007, **21**, 145–156.
53 P. Molek, B. Strukelj and T. Bratkovic, *Molecules*, 2011, **16**, 857–887.
54 M. E. Kokoszka and B. K. Kay, *Methods Mol. Biol.*, 2015, **1248**, 173–188.
55 B. Addepalli, S. Rao and A. G. Hunt, *Methods Mol. Biol.*, 2015, **1255**, 147–158.
56 C. Blikstad and Y. Ivarsson, *Cell Commun. Signaling*, 2015, **13**, 38.
57 H. J. Ren, R. D. Liu, Z. Q. Wang and J. Cui, *Parasitol. Res.*, 2013, **112**, 1857–1863.
58 E. A. Coelho, M. A. Chavez-Fumagalli, L. E. Costa, C. A. Tavares, M. Soto and L. R. Goulart, *Rev. Soc. Bras. Med. Trop.*, 2015, **48**, 370–379.
59 K. A. Henry, M. Arbabi-Ghahroudi and J. K. Scott, *Front Microbiol*, 2015, **6**, 755.

60 M. Hyvonen and P. Laakkonen, *Methods Mol. Biol.*, 2015, **1324**, 205–222.
61 H. Shim, *BMB Rep.*, 2015, **48**, 489–494.
62 K. Omidfar and M. Daneshpour, *Expert Opin. Drug Discovery*, 2015, **10**, 651–669.
63 D. Sanchez-Martin, M. D. Sorensen, S. Lykkemark, L. Sanz, P. Kristensen, E. Ruoslahti and L. Alvarez-Vallina, *Trends Biotechnol.*, 2015, **33**, 292–301.
64 G. A. Loset, G. Berntzen, T. Frigstad, S. Pollmann, K. S. Gunnarsen and I. Sandlie, *Front Oncol*, 2014, **4**, 378.
65 E. C. Roncolato, L. B. Campos, G. Pessenda, L. Costa e Silva, G. P. Furtado and J. E. Barbosa, *Toxicon*, 2015, **93**, 79–84.
66 A. Zhao, M. R. Tohidkia, D. L. Siegel, G. Coukos and Y. Omidi, *Crit. Rev. Biotechnol.*, 2015, 1–14, DOI: 10.3109/07388551.2014.958978.
67 C. E. Chan, A. P. Lim, P. A. MacAry and B. J. Hanson, *Int. Immunol.*, 2014, **26**, 649–657.
68 F. X. Bosch, J. Ribes, M. Diaz and R. Cleries, *Gastroenterology*, 2004, **127**, S5–S16.
69 K. H. Tang, K. Yusoff and W. S. Tan, *J. Virol. Methods*, 2009, **159**, 194–199.
70 J. Mattiacio, S. Walter, M. Brewer, W. Domm, A. E. Friedman and S. Dewhurst, *Vaccine*, 2011, **29**, 2637–2647.
71 W. Domm, M. Brewer, S. F. Baker, C. Feng, L. Martinez-Sobrido, J. Treanor and S. Dewhurst, *J. Virol. Methods*, 2014, **197**, 47–50.
72 S. B. Shivachandra, M. Rao, L. Janosi, T. Sathaliyawala, G. R. Matyas, C. R. Alving, S. H. Leppla and V. B. Rao, *Virology*, 2006, **345**, 190–198.
73 I. Young, I. Wang and W. D. Roof, *Trends Microbiol.*, 2000, **8**, 120–128.
74 T. Matsuda, T. A. Freeman, D. W. Hilbert, M. Duff, M. Fuortes, P. P. Stapleton and J. M. Daly, *Surgery*, 2005, **137**, 639–646.
75 R. Young and U. Blasi, *FEMS Microbiol. Rev.*, 1995, **17**, 191–205.
76 S. Hagens, A. Habel, U. von Ahsen, A. von Gabain and U. Blasi, *Antimicrob. Agents Chemother.*, 2004, **48**, 3817–3822.
77 M. J. Catalao, F. Gil, J. Moniz-Pereira, C. Sao-Jose and M. Pimentel, *FEMS Microbiol. Rev.*, 2013, **37**, 554–571.
78 M. Russel, N. A. Linderoth and A. Sali, *Gene*, 1997, **192**, 23–32.
79 S. Hagens and U. Blasi, *Lett. Appl. Microbiol.*, 2003, **37**, 318–323.
80 Z. Moradpour, Z. Sepehrizadeh, F. Rahbarizadeh, A. Ghasemian, M. T. Yazdi and A. R. Shahverdi, *FEMS Microbiol. Lett.*, 2009, **296**, 67–71.
81 C. Westwater, L. M. Kasman, D. A. Schofield, P. A. Werner, J. W. Dolan, M. G. Schmidt and J. S. Norris, *Antimicrob. Agents Chemother.*, 2003, **47**, 1301–1307.
82 T. M. Viertel, K. Ritter and H. P. Horz, *J. Antimicrob. Chemother.*, 2014, **69**, 2326–2336.
83 S. R. Bhattarai, S. Y. Yoo, S. W. Lee and D. Dean, *Biomaterials*, 2012, **33**, 5166–5174.
84 J. Muzard, M. Platt and G. U. Lee, *Small*, 2012, **8**, 2403–2411.
85 N. Korkmaz Zirpel, T. Arslan and H. Lee, *J. Colloid Interface Sci.*, 2015, **454**, 80–88.
86 N. Korkmaz, *Colloids Surf., B*, 2013, **112**, 219–228.
87 D. Oh, J. Qi, B. Han, G. Zhang, T. J. Carney, J. Ohmura, Y. Zhang, Y. Shao-Horn and A. M. Belcher, *Nano Lett.*, 2014, **14**, 4837–4845.
88 N. K. Petty, T. J. Evans, P. C. Fineran and G. P. Salmond, *Trends Biotechnol.*, 2007, **25**, 7–15.
89 S. Kim, M. Kim and S. Ryu, *Anal. Chem.*, 2014, **86**, 5858–5864.
90 M. Oda, M. Morita, H. Unno and Y. Tanji, *Appl. Environ. Microbiol.*, 2004, **70**, 527–534.
91 C. P. Kodikara, H. H. Crew and G. S. Stewart, *FEMS Microbiol. Lett.*, 1991, **67**, 261–265.
92 T. Funatsu, T. Taniyama, T. Tajima, H. Tadakuma and H. Namiki, *Microbiol. Immunol.*, 2002, **46**, 365–369.
93 M. Cappillino, R. P. Shivers, D. R. Brownell, B. Jacobson, J. King, P. Kocjan, M. Koeris, E. Tekeian, A. Tempesta, J. Bowers, E. Crowley, P. Bird, J. Benzinger and K. Fisher, *J. AOAC Int.*, 2015, **98**, 436–444.
94 G. P. Smith, *Science*, 1985, **228**, 1315–1317.
95 S. D. Alcaine, L. Tilton, M. A. Serrano, M. Wang, R. W. Vachet and S. R. Nugen, *Appl. Microbiol. Biotechnol.*, 2015, **99**, 8177–8185.
96 A. E. Smartt, T. Xu, P. Jegier, J. J. Carswell, S. A. Blount, G. S. Sayler and S. Ripp, *Anal. Bioanal. Chem.*, 2012, **402**, 3127–3146.
97 M. Vinay, N. Franche, G. Gregori, J. R. Fantino, F. Pouillot and M. Ansaldi, *PLoS One*, 2015, **10**, e0131466.
98 B. P. Cormack, R. H. Valdivia and S. Falkow, *Gene*, 1996, **173**, 33–38.
99 K. A. Datsenko and B. L. Wanner, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 6640–6645.
100 R. Derda, M. R. Lockett, S. K. Tang, R. C. Fuller, E. J. Maxwell, B. Breiten, C. A. Cuddemi, A. Ozdogan and G. M. Whitesides, *Anal. Chem.*, 2013, **85**, 7213–7220.
101 K. A. Noren and C. J. Noren, *Methods*, 2001, **23**, 169–178.
102 J. Norrander, T. Kempe and J. Messing, *Gene*, 1983, **26**, 101–106.
103 P. Jain, T. E. Hartman, N. Eisenberg, M. R. O'Donnell, J. Kriakov, K. Govender, M. Makume, D. S. Thaler, G. F. Hatfull, A. W. Sturm, M. H. Larsen, P. Moodley and W. R. Jacobs, Jr., *J. Clin. Microbiol.*, 2012, **50**, 1362–1369.
104 M. J. Loessner, M. Rudolf and S. Scherer, *Appl. Environ. Microbiol.*, 1997, **63**, 2961–2965.
105 H. A. Hill, L. D. Elam-Evans, D. Yankey, J. A. Singleton and M. Kolasa, *MMWR Morb Mortal Wkly Rep*, 2015, **64**, 889–896.
106 D. A. Schofield, I. J. Molineux and C. Westwater, *J. Clin. Microbiol.*, 2009, **47**, 3887–3894.
107 J. P. Vandamm, C. Rajanna, N. J. Sharp, I. J. Molineux and D. A. Schofield, *J. Clin. Microbiol.*, 2014, **52**, 2998–3003.
108 D. A. Schofield, C. T. Bull, I. Rubio, W. P. Wechter, C. Westwater and I. J. Molineux, *Appl. Environ. Microbiol.*, 2012, **78**, 3592–3598.
109 K. Manoutcharian, A. Diaz-Orea, G. Gevorkian, G. Fragoso, G. Acero, E. Gonzalez, A. De Aluja, N. Villalobos, E. Gomez-Conde and E. Sciutto, *Vet. Immunol. Immunopathol.*, 2004, **99**, 11–24.

110 J. Morales, J. J. Martinez, K. Manoutcharian, M. Hernandez, A. Fleury, G. Gevorkian, G. Acero, A. Blancas, A. Toledo, J. Cervantes, V. Maza, F. Quet, H. Bonnabau, A. S. de Aluja, G. Fragoso, C. Larralde and E. Sciutto, *Vaccine*, 2008, **26**, 2899–2905.

111 J. C. Harding, C. D. Baker, A. Tumber, K. A. McIntosh, S. E. Parker, D. M. Middleton, J. E. Hill, J. A. Ellis and S. Krakowka, *J. Vet. Diagn. Invest.*, 2008, **20**, 274–282.

112 Y. C. Cao, Q. C. Shi, J. Y. Ma, Q. M. Xie and Y. Z. Bi, *Acta Biochim. Biophys. Sin.*, 2005, **37**, 657–664.

113 R. Van Houdt and C. W. Michiels, *J. Appl. Microbiol.*, 2010, **109**, 1117–1131.

114 T. K. Lu and J. J. Collins, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 11197–11202.

115 Y. Itoh, X. Wang, B. J. Hinnebusch, J. F. Preston and T. Romeo, 3rd, *J. Bacteriol.*, 2005, **187**, 382–387.

116 I. Yosef, M. Manor, R. Kiro and U. Qimron, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 7267–7272.

117 K. Thiel, *Nat. Biotechnol.*, 2004, **22**, 31–36.

118 S. R. Williams, D. Gebhart, D. W. Martin and D. Scholl, *Appl. Environ. Microbiol.*, 2008, **74**, 3868–3876.

119 D. Scholl, M. Cooley, S. R. Williams, D. Gebhart, D. Martin, A. Bates and R. Mandrell, *Antimicrob. Agents Chemother.*, 2009, **53**, 3074–3080.

viruses

MDPI

Article

Comparison of CRISPR and Marker-Based Methods for the Engineering of Phage T7

Aurelija M. Grigonyte [1], Christian Harrison [2], Paul R. MacDonald [3], Ariadna Montero-Blay [4], Matthew Tridgett [1], John Duncan [1], Antonia P. Sagona [1], Chrystala Constantinidou [5], Alfonso Jaramillo [1,6,7,*] and Andrew Millard [2,*]

1 Warwick Integrative Synthetic Biology Centre (WISB) and School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK; grigonyt@ualberta.ca (A.M.G.); Matthew.Tridgett@warwick.ac.uk (M.T.); john.duncan1@fedex.com (J.D.); A.Sagona@warwick.ac.uk (A.P.S.)
2 Department Genetics and Genome Biology, University of Leicester, University Road, Leicester LE1 7RH, UK; cdh18@leicester.ac.uk
3 MOAC DTC, Senate House, University of Warwick, Coventry CV4 7AL, UK; P.R.MacDonald@warwick.ac.uk
4 EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr Aiguader 88, 08003 Barcelona, Spain; arimonbla@gmail.com
5 Warwick Medical School, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK; C.I.Constantinidou@warwick.ac.uk
6 Institute of Systems and Synthetic Biology (ISSB), CNRS, CEA, Université Paris-Saclay, 91025 Evry, France
7 Institute for Integrative Systems Biology (I2SysBio), University of Valencia-CSIC, 46980 Paterna, Spain
* Correspondence: adm39@le.ac.uk (A.M.); Alfonso.Jaramillo@synth-bio.org (A.J.); Tel.: +44-0-116-252-5743 (A.M.)

Received: 13 January 2020; Accepted: 6 February 2020; Published: 10 February 2020

Abstract: With the recent rise in interest in using lytic bacteriophages as therapeutic agents, there is an urgent requirement to understand their fundamental biology to enable the engineering of their genomes. Current methods of phage engineering rely on homologous recombination, followed by a system of selection to identify recombinant phages. For bacteriophage T7, the host genes *cmk* or *trxA* have been used as a selection mechanism along with both type I and II CRISPR systems to select against wild-type phage and enrich for the desired mutant. Here, we systematically compare all three systems; we show that the use of marker-based selection is the most efficient method and we use this to generate multiple T7 tail fibre mutants. Furthermore, we found the type II CRISPR-Cas system is easier to use and generally more efficient than a type I system in the engineering of phage T7. These results provide a foundation for the future, more efficient engineering of bacteriophage T7.

Keywords: bacteriophage; CRISPR; T7; tail fibres

1. Introduction

Bacteriophages (phages) are viruses that specifically infect bacteria. Much of the original research on bacteriophages has guided our current understanding of molecular biology. Recently, there has been a renaissance in phage research due to the emergence of multi-drug resistant bacteria and the need for viable antibiotic alternatives [1]. Thus, there has been considerable research in the area of phage therapy, to kill bacterial infections using phages [2–4]. Phages offer a number of advantages (bactericidal agents, rapid discovery, undergo replication at the site of infection) over traditional antibiotics [5–9]. However, there are several limitations to phage therapy, such as narrow host range and limited pharmacokinetics leading to insufficient build-up of phage concentration at the site of infection [10,11].

The genetic engineering of phages offers the possibility of overcoming these limitations and allowing phage therapy to be a viable alternative. For the effective engineering of phages to occur a highly efficient and reproducible system is required. From the onset of phage research a number of methods have been used to create phage mutants; this has included random mutagenesis using a variety of inducing agents including UV, as well as chemical mutants such as hydroxylamine and N-methyl-N'-nitro-N-nitrosoguanidine [12]. The use of random mutagenesis provides the ability to create a large library of mutants but requires extensive screening to identify a mutant of interest. Thus, other approaches have been developed to create targeted mutants using homologous recombination to introduce the desired mutation, followed by selection of the resulting phage [13]. The selection process has proven to be the major bottleneck in the creation of phage mutants. Unlike bacteria where an antibiotic resistance marker can be used to select for mutants, no such generic approach works so efficiently for phages, as phage lysis kills the bacterial host. Therefore, a number of other strategies have been developed for engineering and creating phage mutants.

An alternative to the antibiotic resistance markers used for positive selection of phage mutants, is the use of a bacterial host gene that is essential for phage replication but not essential for host growth [14,15]. In the infection of *E. coli* by phage T7, the proteins thioredoxin and cytidine monophosphate kinase (dCMP), encoded by *trxA* and *cmk* genes respectively, are known to be essential [14]. Therefore, these genes have previously been used as markers to positively select for phage mutants in an *E. coli* background that lack *trxA* or *cmk* [14,16]. This marker-based approach has been used to knockout a number of T7 genes including genes 0.4, 11, 12, 17 [16–18]. These knockouts were achieved by replacing the target gene with *trxA* by homologous recombination, and selection on *E. coli* BW25113 Δ*trxA* cells [16–18].

An alternative to positive selection is to reduce the number of phages that have to be screened to find the desired mutant by increasing the efficiency of homologous recombination. An example of this is the BRED system (bacteriophage recombineering with electroporated DNA) that has been used to genetically modify mycobacteriophages and coliphages [19,20]. BRED exploits the mycobacterial recombineering system, in which expression of the RecE/RecT-like proteins of the mycobacteriophage Che9c confers high levels of homologous recombination and facilitates simple allelic exchange using a linear DNA substrate. This method has been used to engineer both lytic and temperate phages [21,22].

Another method to reduce the number of phage that require screening is the use of a Yeast Artificial Chromosome (YAC) in Saccharomyces cerevisiae [19,20]. Here, a phage genome can be cloned into a YAC, either in one piece or as PCR fragments assembled through gap-repair cloning [23]. Transformation-associated transformation recombineering (TAR) systems of yeast can then be used to introduce desired mutations, with the engineered genome re-introduced into its bacterial host. TAR approach has been used to make tail fibre mutants in coliphages T7 and T3 [24].

The most recent advance in screening methods is the use of CRISPR-Cas to counter select against phages that have not been altered and enrich for the phage mutants of interest. A number of different CRISPR systems have now been used to modify the genomes of a number of phage systems, the first of which used a type I CRISPR system to target and create mutants of bacteriophage T7 [25]. A different type I CRISPR system has also been used to create mutants of the lytic phage ICP1 that infects *V. cholerae* [26]. A number of phages have been modified with type II CRISPR systems; including phage 2972 infecting *Streptococcus thermophilus*, phage P2 infecting *Lactococcus lactis*, phiKpS2 infecting *Klebsiella pneumoniae* and phages T2, T4, T7 and KF1 that infect *E. coli* respectively [25,27–31]. A type III CRISPR system has also been used to engineer phages that infect *Staphylococcus aureus* and *Staphylococcus epidermidis* [32]. Whilst CRISPR is being used for editing phage genomes, there is still much to be learned concerning what the most efficient system to use is. For example phage T4 has been engineered by three different groups, with contrasting efficiency for type II CRISPR [29,30,33].

Both Type I and II CRISPR/Cas systems have different essential components that are required for their function. Whilst both systems require a DNA repeat-spacer, each of them has additional requirements. Type II-A CRISPR-Cas needs Cas9 endonuclease and trans-activating CRISPR RNA

(*tracr*RNA), whilst the I-E CRISPR-Cas system requires Cas3 enzyme as well as a five protein complex that consists of CasA, CasB, CasC, CasD, and CasE, often referred to as cascade [34–36]. It is clear both CRISPR-Cas and marker-based selection can be used to generate phage mutants; however, no comprehensive analysis of their comparative efficiency has been carried out in a single system for a single phage.

The aim of this work was to compare efficiencies of both CRISPR-Cas systems and marker-based (*cmk* and *trx*) methods for the development of an efficient system of constructing phage T7 mutants (Figure 1). As a proof of concept, the most efficient method was then used to rapidly generate tail fibre mutants of phage T7.

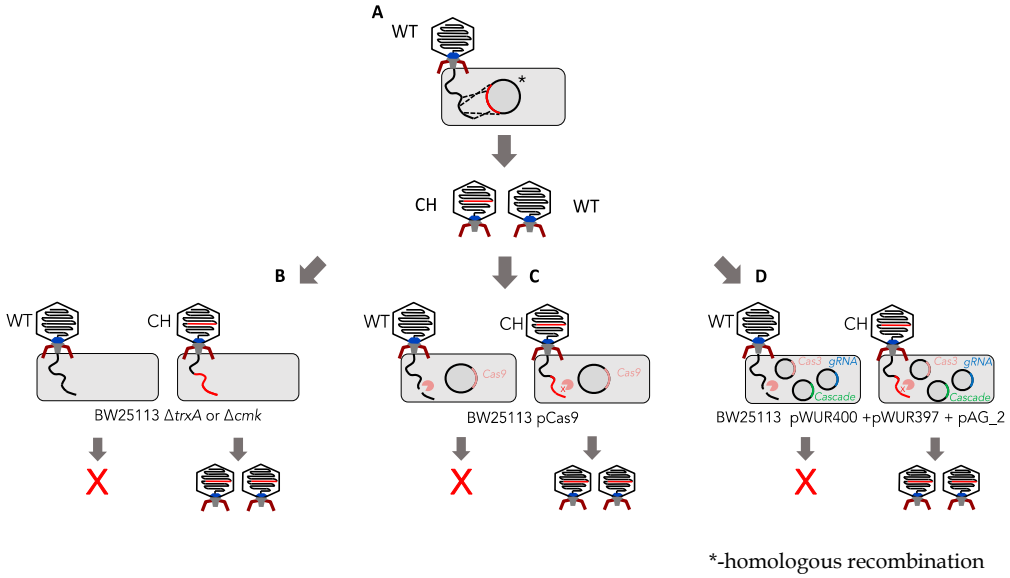


Figure 1. Summary of marker-based vs. marker-less selection methods. (**A**) The homologous recombination step allows for generation of a mixed population of chimeric/mutant (CH) and wild-type (WT) phages. (**B**) Marker-based selection requires a gene encoding an essential host factor to be incorporated into the region that will homologously recombine in step A. Mutants are then selected for on *E. coli* cells that are deficient in this host factor. For phage T7 either *E.coli* Δ*trxA or E.coli* Δ*cmk* can be used for positive selection. (**C**) Type II CRISPR system consists of one vector (pCas9: contains Cas9 as well as direct repeat regions and guide RNA). (**D**) A Type I CRISPR system consists of three vectors (pWUR400:Cas3, pWUR397:cascade system and pAG_2:guide RNA). Both CRISPR systems target wild-type phage, enriching for mutants.

2. Materials and Methods

2.1. Bacterial Strains and Phages

E. coli BW25113, *E. coli* BW25113 Δ*trxA*, *E. coli* BW25113 Δ*cmk*, BL21(DE3) and BL21(AI) strains were used as the bacterial hosts for all assays carried out in this study (Table S1). The strains were grown in LB medium (Oxoid, Basingstoke, UK) at 37 °C overnight with shaking (200 rpm). Phage T7 was propagated by inoculating *E. coli* culture at an OD 600 nm of 0.3. Phage lysates were filtered using a 0.22 µm pore size filter (Sartorius, Dublin, Ireland) and stored at 4 °C. Phage enumeration was carried out using a standard double agar overlay plaque assay or spot assays as previously described [37].

2.2. Type I CRISPR gRNA Delivery Vector Component/Vector Design

The vectors pWUR400 and pWUR397 provide the cascade and cas proteins of the type I system, respectively [36]. A third component of the system, a vector containing a gRNA to target specific sites of interest, was designed and synthesized in a pSMART Amp^r vector (Table S2). The initial vector pAG_1 contained a non-targeting gRNA control, referred to as a scrambled RNA (scr). The main

components of the construct were: (a) T7 RNA polymerase promoter at the 5′ of the construct; (b) 5′ handle; (c) gRNA for a target gene sequence; (d) BbsI restriction enzyme sites flanking the gRNA region; (e) 3′ handle; (f) T7 terminator.

2.3. Type I CRISPR gRNA Design

Type I CRISPR gRNAs were designed manually taking into consideration protospacer adjacent motif (PAM) requirements for the system. Previous research has identified that the PAM that fits most efficiently is AGG [38]. The main principle for the design was to first identify the AGG sequence; the 32 bp following AGG comprise the target sequence. All AGG motifs were identified in g17. Ten out of 37 identified potential gRNAs were chosen for gRNA design: four targeting the 5′ region and six targeting the 3′ region of the gene respectively. They were designed and synthesized as a complementary single stranded DNA fragments (IDT) and inserted into pAG_1 vector using Golden Gate cloning (Tables S2 and S3, Figure S6).

2.4. Type II CRISPR gRNA Design

Each of the type II CRISPR gRNAs was designed using DNA 2.0 (Atum, Newark, CA, USA). The following criteria were chosen when using the software: NGG was entered for the desired PAM and species off target was *E. coli*. Each of the gRNAs generated by the software was appended with BsaI enzyme restriction sites. A total of 11 gRNAs was designed and then synthesised as a complementary single stranded DNA fragments (IDT) and inserted into pCas9 vector using Golden Gate cloning (Tables S2 and S3, Figure S6).

2.5. Golden Gate Cloning

The type I CRISPR gRNAs were inserted in the vector pAG1 using a modified Golden Gate cloning method [39]. Briefly, 0.5 µL of each primer (100 mM) was mixed together with 49 µL of sterile water and incubated at 95 °C for 5 min, prior to being diluted 10 times. The 15 µL assembly mix was prepared as follows: vector of interest (100 ng), 2 µL of the diluted primers (10 mM), 1.5 µL of 10× T4 buffer (NEB, Ipswich, MA, USA), T4 Ligase (NEB), 0.15 µL of 100× BSA and made up to 15 µL with water. The assembly mix was then incubated in a thermocycler under the following conditions: 25 cycles of 3 min incubation at 37 °C followed by 4 min at 16 °C, and 1 cycle of 5 minutes at 50 °C and 5 min at 80 °C. The samples were then left overnight at room temperature and 1 µL of the assembly mix was then transformed into electro-competent *E. coli* cells using 2 mm cuvettes at 2.5 kV.

2.6. PCR

Polymerase Chain Reaction (PCR) was carried out in 0.2 mL volume PCR in a thermal cycler (T100 Thermal cycler, Bio-Rad, Hercules, CA, USA). HF Phusion Master mix (M0531S, New England Bio Labs, Hitchin, UK) and BSA (B9000S, New England Bio Labs, UK labs) were used for the reactions. Final concentrations were used as follows: primers at 0.5 µM, DNA at < 250 ng, 1× Phusion Master Mix. PCR amplification was carried out as per Protocol Phusion®High-Fidelity PCR Master Mix recommended conditions.

2.7. Gibson Assembly

A master mix was prepared of 5× ISO buffer 0.5 M Tris-HCl pH 7.5, 0.05 M $MgCl_2$, 4 mM dNTPs (1 mM of each: dGTP, dCTP, dATP, dTTP), 0.05 M DTT, 0.25% (*v*/*v*) PEG800 and 5 mM NAD. Then the following components were mixed 0.26% 5× ISO Buffer, 0.005 U T5 exonuclease, 0.03 U Phusion polymerase, 5.3 U Taq ligase. The Gibson assembly was based on previously described method [40]. Briefly, fragments required for the assembly were PCR amplified (using primers in Table S3) and gel purified. Gibson reactions were carried out in 20 µL volumes, with 15 µL of Gibson master mix (0.26% 5× ISO Buffer, 0.005 U T5 exonuclease, 0.03 U Phusion polymerase, 5.3 U Taq ligase) with the

remaining 5 µL made up of vector and inserts at the appropriate ratios. For the assembly of three or more fragments an equilmolar ratio of fragments was used; when inserting one fragment into a vector, a ratio of 1:2 (vector to insert) was used. The reaction was incubated at 50 °C for 60 min, followed by incubation at 37 °C for 60 min. The reaction volume was diluted 3-fold and 1 µL used in electroporation.

2.8. Homologous Recombination and In Trans Complementation

A strain containing a homologous recombination plasmid (HR) was grown until it reached an OD600 of 0.3–0.4, whereby phage T7 was then added at MOI ~0.01. This was then followed by incubation at 37 °C with shaking (200 rpm) for 3 hours. The lysate was then filtered through a 0.22 µm pore size filter and stored at 4 °C until further use. For the in trans selection, the lysate was plated on *E. coli* BW25113 Δ*trxA* strain containing pAG30 or pAG31 vectors that provide wild-type tail fibres.

2.9. Confirmation of Phage Mutants

Following homologous recombination and selection, plaques were picked from plates and resuspended in 1 mL of SM buffer. PCR was used to confirm the presence of phage mutants using primers to differentiate between wildtype and mutant phages (Table S3). PCR products were Sanger sequenced bidirectionally to confirm products were correct. For four phages, whole genome sequencing was carried out. Genomic DNA was extracted from 1 mL of phage lysate using a previously described method [41]. Sequencing libraries were prepared using NexteraXT (Illumina, San Diego, CA, USA) following the manufacturer's instructions. Reads were trimmed with sickle using default settings and genomes assembled using SPAdes v3.12.0 [42,43] using "-only-assembler" option and assemblies checked for errors and corrected with pilon [44]. Genomes were annotated with Prokka using phage T7 (accession: V01146) as the source of annotations. SNPs were identified using DNAdiff using default settings [45]. All data were submitted under the project accession PRJEB35760.

2.10. Type I and Type II CRISPR Screening Assays

To determine the selection efficiency of the marker-based methods (trxA and cmk), plaque assays of T7 were performed against *E. coli* BW25113, *E. coli* Δcmk and *E. coli* ΔtrxA with a starting titre of 1×10^{10} pfu/mL. The efficiency of plating on *E. coli* Δcmk and *E. coli* ΔtrxA was compared to *E. coli* BW25113. A similar approach was used to determine the efficiency of type I and type II CRISPR systems, with efficiency of plating of phage T7 on strains containing these CRISPR systems compared to *E.coli* BW25113.

3. Results

3.1. Determination of Marker-Based (cmk and trxA) Selection Efficiency against Phage T7

In order to test the efficiency of Type I and II CRISPR selection along with the selective markers *trx* and *cmk* all three methods were compared in their ability to select against *g17* of wild-type T7. To do this the same genetic background was used to try to minimise differences caused by the use of different *E. coli* strains. The strain of choice was *E. coli* BW25113 which allows the Keio collection of mutants to be used [15].

First, the extent to which the absence of host factors reduces T7 progeny was determined. To do this plaque assays were carried out on *E. coli* strains that lacked *cmk* or *trxA*. Efficiency of plating (EOP) on both strains was compared to the control strain *E. coli* BW25113 (Figure 2). In the case of *E. coli* Δ*cmk* cells, the number of progeny phage was reduced by approximately 10^5-fold, whereas with *E. coli* BW25113 Δ*trxA* cells, no progeny was detected (detection limit ~ 1), reduction by 10^{10}-fold (Figure 2).

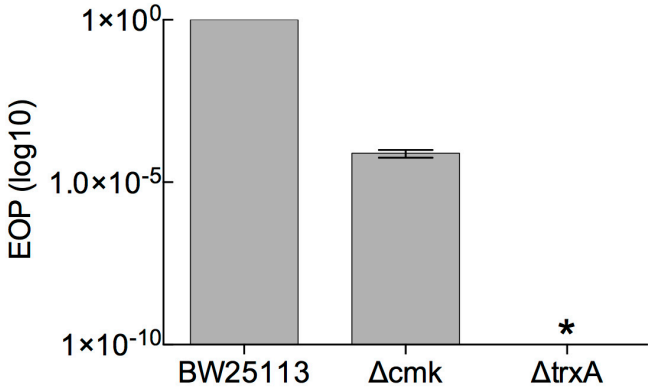


Figure 2. Efficiency of plating for T7 on *E. coli* BW25113, *E. coli* BW25113 Δ*cmk* and *E. coli* BW25113 Δ*trxA* strains. EOP was determined with respect to a reference *E. coli* strain BW25113. EOP data were log10 transformed and are presented as the mean of three independent experiments, $n = 3$. An asterisk indicates that the EOP was below the detection limit which is < 1.

3.2. Determination of the Efficiency of Type I CRISPR for Selecting against Wild-Type Phage T7

A type I CRISPR system that has previously been used to engineer phage T7 was used in this study and comprises the vectors pWUR400 and pWUR397 that carry genes encoding for the enzyme cascade and Cas3 respectively [10]. A third plasmid (pAG_1) was constructed in this study that delivers the gRNA and allows for easy swapping of gRNAs to target different regions (Figure 3). It consists of (a) T7 RNA polymerase promoter at the 5′ of the construct; (b) 5′ handle; (c) gRNA for a target gene sequence; (d) Bbs*I* restriction enzyme sites flanking the gRNA region; (e) 3′ handle; (f) T7 terminator.

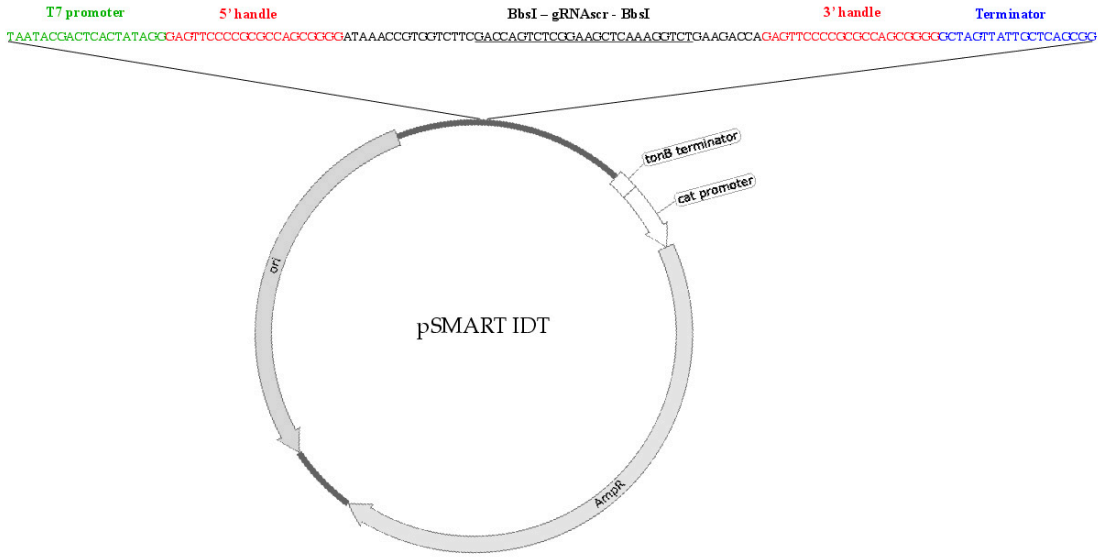


Figure 3. A schematic representation of the vector pAG_1. A 130 bp fragment consisting of a T7 promoter, two handle regions, non-targeting gRNA(scr) with two surrounding BbsI restriction sites and a T7 terminator.

It is known that different gRNAs can alter the efficacy of CRISPR selection [46,47]. Therefore, multiple gRNAS were designed and tested (Figure S6). Eleven different gRNAs, ten targeting *g17* and a scrambled gRNA, were cloned into pAG_1 by Golden Gate cloning (Tables S2 and S3). The efficiency of each gRNA to target phage DNA was determined by EOP with respect to the control gRNA (Figure 4A). All gRNAs were effective in reducing the number of T7 progeny, with gRNA7 the most effective at reducing T7 progeny by ~90-fold (Figure 4A). With gRNA2, gRNA5 and gRNA8 reducing T7 progeny by ~10-fold, whereas gRNA1, gRNA6, gRNA9 and gRNA10 showed a reduction of less than 10-fold.

The highest reduction in wild-type T7 was ~100-fold, which is significantly lower than previous reports of 10,000-fold reduction using this system [25]. No correlation between GC content and EOP was identified between the different gRNAs analysed (Figure 4B).

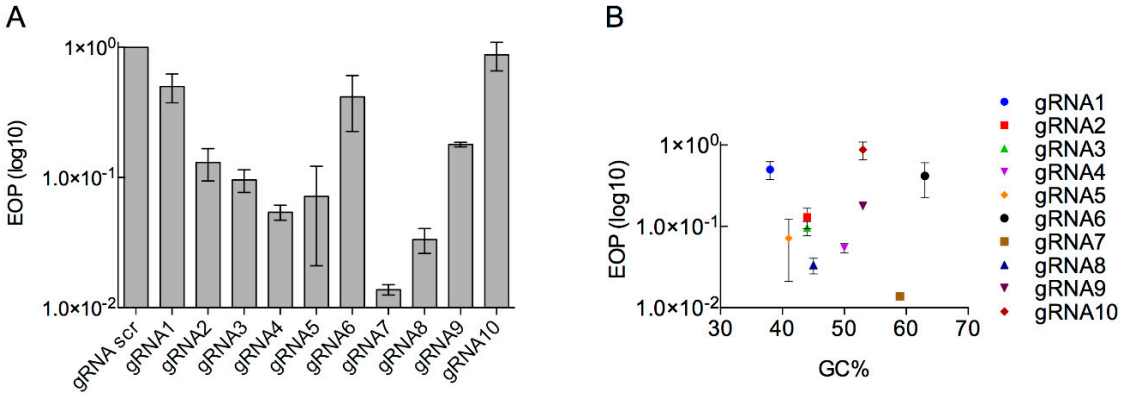


Figure 4. Efficiency of gRNA candidates as marker-less selection system in type I CRISPR-Cas system. (**A**) Efficiency of plating (EOP) for T7 against *E. coli* BW25113 containing different type I CRISPR gRNAs as well as Cas3 (pWUR400) and cascade genes (pWUR397). The (EOP) was determined with respect to a reference *E. coli* strain BW25113/pAG1. EOP data are presented as the mean of three independent experiments. (**B**) EOP data were plotted against GC% of each gRNA.

E. coli BW25113 possesses an innate type I CRISPR system under the control of the transcriptional repressor H-NS (*hns*) [48–50]. The innate CRISPR system is only active if H-NS, is inactivated (Δ*hns*) [49]. The *E. coli* BW25113 used in this study did not contain the *hns* deletion, however, it was necessary to ensure that the innate CRISPR had no effect on the type I CRISPR used here. To verify this *E. coli* BW25113 containing eight gRNAs, including the most efficient gRNA7, was evaluated for its ability to target the T7 genome and accordingly reduce viable T7 progeny. No reduction was identified, ensuring that the introduced CRISPR system used in this study was not overlapping with the innate CRISPR of *E. coli* BW25113 (Figure S1).

A type I CRISPR system has previously been used to target wild-type phage T7, with reductions of two orders of magnitude greater than we obtained [25]. It is possible the different gRNAs are responsible for observed differences in reduction of wild-type T7. However, the previous work of Kiro et al. 2014, also induced the expression of the CRISPR system prior to infection with T7. To identify if the pre-induction of a CRISPR system prior to phage addition results in increased efficiency, the same experiment was repeated in strains that contain an inducible T7 RNA polymerase: *E. coli* BL21(DE3) and *E. coli* BL21-AI (Figures S2 and S3). No difference in phage progeny reduction, between induced and un-induced systems, was identified using BL21(DE3) or BL21-AI strains (Figures S2 and S3). The carriage of all the plasmids necessary for the type I CRISPR results in substantial change in growth rate, with μ increasing from ~28 min (BL21) to > ~39 min (BL21 pWUR397 pWUR400) (Figure S4).

3.3. Determination of Type II CRISPR Efficiency as a Method for Engineered Phage T7 Selection

Eleven targeting gRNAs were designed against *g17*, that spanned the length of the gene and the efficiency of each gRNA was determined (Tables S2 and S3, Figure S6). All eleven gRNAs resulted in a reduction in T7 progeny compared to the control. The gRNA2 was the most effective, resulting in a 1000-fold reduction compared to the control (Figure 5), whereas the majority of gRNAs (gRNA5, gRNA7-gRNA11) showed a reduction in progeny of less than 10-fold, with gRNA3, gRNA4 and gRNA6 resulting in more than 10-fold but less than 100-fold reduction. In working with the type II system, we found that there was no significant change in the growth rate of cells containing Cas9 vector versus no vector containing *E. coli* BW25113 (Figure S5).

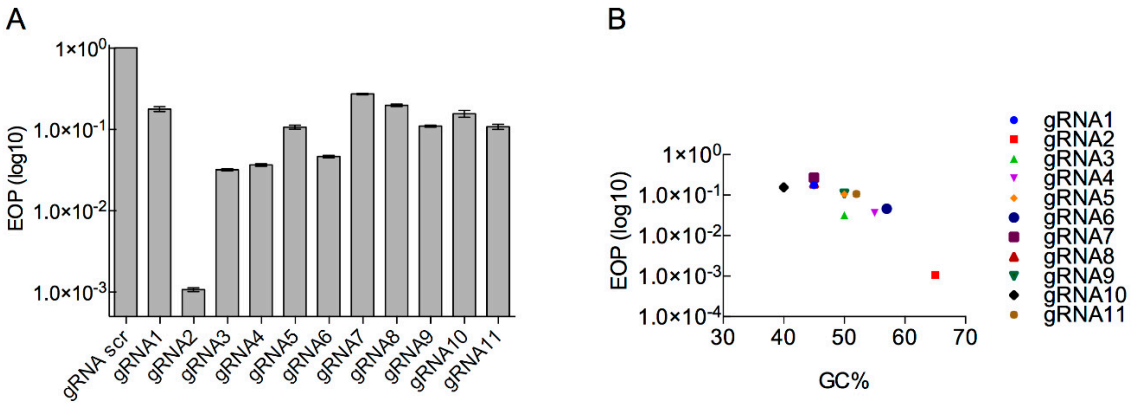


Figure 5. Efficiency of gRNA candidates as marker-less selection system in type II CRISPR-Cas system. (**A**) Efficiency of plating (EOP) for T7 against *E. coli* BW25113 containing type II CRISPR gRNAs. The EOP was determined with respect to a reference *E. coli* strain BW25113/pCas9. EOP data are presented as the mean of three independent experiments. (**B**) EOP data plotted against GC% of each gRNA.

3.4. Generation of T7 Tail Fibre Mutants

The use of marker genes was the most efficient at reducing wild-type progeny. To check that the method is also effective at selecting phage mutants, we used this method to make a number of mutants that were fusions between the tail fibre of T7 and phages BPP-1 and Yep-phi [51,52]. Phage BPP-1 is very distantly related to phage T7 [53], whereas Yep-phi is more closely related to T7 [52,54]. The tail fibre of BPP-1 contains major tropism determinant (Mtd) domain, where a single mutation allows phage BPP-1 to rapidly evolve and infect the constantly changing receptors of Bordetella pertussis [53]. Although these phages have limited similarity the fusion of the Mtd domain from phage BPP-1 with the T7 tail fibre may allow the host range of phage T7 to be expanded. The tail fibre of Yep-phi (accession: YP_009014859.1) shares 37% protein identity with the T7 homologue, with 53% protein sequence identity over the last 86 amino acids of the C-terminal [54].Thus, it may be more likely to form a functional tail fibre if fused with the T7 version. Eight T7 tail fibre mutants were designed that consisted of domains from both the T7 and BPP-1 tail fibre and one tail fibre mutant that contained domains from T7 and Yep-phi tail fibres (Table S4 and Figure S7). This was achieved by first constructing homologous recombination plasmids pAG23–pAG32 (Tables S2 and S3) with the desired chimeric tail fibre genes and *trxA* for selection. Phage T7 was then used to infect *E. coli* BW25113 Δ*trxA* containing homologous recombination plasmids.

To check that selection with *trxA* was occurring and that chimeric phages could be obtained, phage mutants were created in a two-step manner using an in trans tail fibre. Following homologous recombination, the progeny was plated on *E. coli* BW25113 Δ*trxA*, containing a plasmid expressing wild-type tail fibres (pAG_30). The resultant phage has a genome which contains a chimeric tail fibre gene, but wild-type tail fibres on the viral particle. Once confirmed as mutants, these phages were further propagated on just BW25113 Δ*trxA* to determine whether the resultant progeny is functional.

Selection efficiency was measured by PCR screening of plaques to determine the presence of mutants or wild-type phage. Between 10 and 30 plaques were screened for each mutant, all of which were positive for phage mutants (Table S5). Thus, confirming the efficiency of *trxA* marker-based selection for rapidly creating phage mutants. Only one phage mutant was able to infect *E. coli*, the fusion of T7 and Yep-phi tail fibres, with all other phages were non-infective. To ensure the lack of infectivity was due to the non-functional tail fibres and not other deleterious changes in the genome, four phage mutants were sequenced along with the initial wild-type phage T7. In addition to the desired tail fibre mutations, an additional 1-4 SNPs in phages phAG_3, phAG_4 and ph_PM (Table S6). However, these are in regions known to be non-essential for T7 replication and when a wild-type tail fibre was provided in trans the phage could replicate.

4. Discussion

Comparison of selective markers and Type I and II CRISPR systems revealed that selection with *trxA* was the most efficient method for reducing phage numbers. The use of marker-based selection produced an approximately 10^{10}-fold drop in phage progeny, compared to a ~100× and ~1000× decrease with the most efficient gRNAs using Type I and II CRISPR systems respectively. Previously, the use of *E. coli* Δ*trxA* and *cmk* mutants has been reported to produce an EOP of 10^{-2} and 0.05 respectively [25], whereas, within this study, EOPs of 10^{-10} and 10^{-4} were observed. The differences in the observed EOP of plating may simply be due to the growth state of the cells used during a plaque assay. Within this study exponentially growing cells were used, whereas the original work of Kiro et al. 2014 used cells that had been grown overnight and likely to be in stationary phase of growth [25]. Whilst T7 can infect cells in stationary phase, more productive infections occur in the exponential phase of growth [55]. Thus, when using *E. coli* Δ*trxA* or Δ*cmk* for positive selection we would suggest infecting *E. coli* in the exponential phase of growth.

The maximum reduction in wild-type T7 progeny, and hence the potential T7 mutant selection efficiency, was found to be 100-fold for the type I CRISPR, when targeting *g17*. This 100-fold reduction was 100 times less than had been achieved by Kiro et al. when targeting *g1.7* using the same CRISPR system [25]. Although the same type of CRISPR system was used, there were differences in the protocol used as we did not pre-induce the CRISPR system, unlike Kiro et al. [25]. However, when repeating the experiment using *E. coli* BL21(DE3) and BL21-AI, allowing for induction of the CRISPR system prior to phage addition, we found no difference in Type I CRISPR efficiency with or without induction. Therefore, the difference in observed EOP may simply be down to the selection of the most optimal gRNA. The EOP obtained in this study with different gRNAs varied from 1 to 1×10^{-4}, which is within the range reported for other studies where the EOP obtained with different gRNAs varies from 1 to 1×10^{-6} [29,30].

Using a type II CRISPR system we obtained a 1000-fold reduction when targeting *g17* with gRNA2, which is higher than any gRNA used for the type I system in this study. In targeting *g17* with a similar number of gRNAs distributed across the region of interest, it is clear to see that that type II CRISPR system was generally more efficient than the type I system. The 1000-fold reduction we obtained with a type II CRISPR system is within the range of 10^{-1} and 10^{6} -fold that has previously been reported when using a type II CRISPR system to target phage T4 [29,30]. What gives rise to the large variation in efficiency of different gRNAs is not clear. For non-bacteriophage systems where type II CRISPR has been widely used, there have been numerous studies to predict the efficiency of different gRNAs (see Wilson et al. 2018 for a review) with large databases of validated gRNAs for different organisms [56,57]. Very high or low GC content of gRNAs has been shown to be an important factor in the decreasing efficiency of gRNAs when targeting human cell lines [58], whereas, we found that the efficiency of gRNAs increased with GC content, albeit on small sample. These differences are probably not unexpected given the vastly different backgrounds when using a type II CRISPR system in human cell lines and bacteria.

Given the number of CRISPR systems available to engineer phages, it is often difficult to compare across systems. For instance, the Cas9 used in this study originated from *S. pyogenes*, whereas other studies have used Cas9 that originates from *S. thermophiles*, potentially adding to differences in efficiency, combined with the choice of gRNA that is known to be critical [59–61]. Within this study, we directly compared a marker-based system with both type I and II CRISPR systems. Based on the results obtained here, it is clear the most efficient gRNAs for the type II CRISPR system had a higher GC content, an important factor to consider when designing new gRNAs. In contrast with type I system, there was no clear and obvious pattern that may lead to the design of gRNAs that are more effective.

In addition to assessing the effectiveness of a type I and II CRISPR systems and marker-based selection for the ability to target *g17*, we observed differences in the ease of use of these systems. In this study we designed a new vector to be used with type I systems that allows gRNAs to be quickly inserted. Whilst this allows the rapid changing of different gRNAs, the system still requires the use

of three plasmids. The growth of cells that contain this system was severely inhibited in our hands, making the system difficult and cumbersome to work with compared to a type II system. The choice of method for creating phage mutants will depend on a number of factors and the use of CRISPR allows the creation of clean deletions that cannot be achieved using marker-based methods, whilst also allowing the creation of multiple mutations, whereas in comparison, marker-based methods are limited to the number of markers that can be selected for. However, the use of marker-based methods has the advantage of being extremely efficient, in particular *trxA*, for the selection of phage mutants and does not require multiple gRNAs to be designed and tested. Within this study we were able to rapidly generate a number of tail fibre mutants (the majority of which ultimately proved to be non-functional in their ability to infect *E. coli)* using *trxA*-based selection. The functional T7–Yep–phi tail fibre hybrid adds to the growing evidence that functional synthetic phages can be engineered [16,24,25].

5. Conclusions

The use of the *trxA* marker-based positive selection proved the most efficient at selecting against wild-type phage T7. We were able to use this system to rapidly engineer eight T7 tail fibre mutants. Although only one of these mutants was ultimately found to be functional. The comparison of type I and II CRISPR systems demonstrated that the type II system was the most efficient at targeting wild-type phage T7. This method was also easier to use and would be method of choice if choosing a CRISPR based selection approach. Furthermore, we highlighted that when using CRISPR selection to engineer phage T7 the GC content of the spacer can impact the efficiency. The direct comparison of these three methods provides a basis for further T7 engineering, and aid others who wish to engineer phages in choosing the most appropriate method.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1999-4915/12/2/193/s1.

Author Contributions: The work was conceptualized by A.M.G. and A.J. The methodology was carried out by A.M.G., C.H., P.R.M., A.M.-B., M.T., J.D., and A.P.S. Analysis was carried out by A.M.G., C.H., and A.M. Supervision was carried out by A.M., A.J. and C.C. An initial draft was written by A.M.G. All authors reviewed the final manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: A.M.G. is funded by a doctoral fellowship from the EPSRC & BBSRC Centre for Doctoral Training in Synthetic Biology (grant EP/L016494/1). P.R.M is funded by a doctoral fellowship from the EPSRC Doctoral Training Centre Molecular Organisation and Assembly in Cells (Grant No. EP/F500378/1). C.H. is funded by a PhD scholarship from the Dept Genetics and Genome Biology, University of Leicester. A.M.B, A.P.S. were funded by the European Commission grant (FP7-ICT-610730, EVOPROG) to A.J. J.D. was funded by the European Commission grant (FP7-KBBE-613745, PROMYS) to A.J. A.J. is funded by the European Commission grants (FP7-ICT-610730, EVOPROG; FP7-KBBE-613745, PROMYS; H2020-MSC-642738, MetaRNA), the BBSRC grant EVO-ENGINE BB/P020615/1, FP7-KBBE (no 613745, PROMYS), and EPSRC-BBSRC (BB/M017982/1, WISB center). A.M. is funded by MRC (MR/L015080/1) and NERC (NE/N019881/1).

Acknowledgments: We thank Luciano Marraffini for plasmid pCas9, Udi Qimron for plasmids pWUR400 and pWUR397 and Mark J van Raaij for his support with tail fibre fusion design.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Czaplewski, L.; Bax, R.; Clokie, M.; Dawson, M.; Fairhead, H.; Fischetti, V.A.; Foster, S.; Gilmore, B.F.; Hancock, R.E.W.; Harper, D.; et al. Alternatives to antibiotics-a pipeline portfolio review. *Lancet Infect. Dis.* **2016**, *16*, 239–251. [CrossRef]
2. Sulakvelidze, A.; Alavidze, Z.; Morris, J.G. Bacteriophage Therapy. *Antimicrob. Agents Chemother.* **2001**, *45*, 649–659. [CrossRef]
3. Kęsik-Szeloch, A.; Drulis-Kawa, Z.; Weber-Dąbrowska, B.; Kassner, J.; Majkowska-Skrobek, G.; Augustyniak, D.; Łusiak-Szelachowska, M.; Zaczek, M.; Górski, A.; Kropinski, A.M. Characterising the biology of novel lytic bacteriophages infecting multidrug resistant *Klebsiella pneumoniae*. *Virol. J.* **2013**, *10*, 100. [CrossRef] [PubMed]

4. Oliveira, H.; Pinto, G.; Hendrix, H.; Noben, J.P.; Gawor, J.; Kropinski, A.M.; Łobocka, M.; Lavigne, R.; Azeredo, J. A lytic Providencia rettgeri virus of potential therapeutic value is a deepbranching member of the T5virus genus. *Appl. Environ. Microbiol.* **2017**, *83*. [CrossRef] [PubMed]
5. Stratton, C.W. Dead bugs don't mutate: Susceptibility issues in the emergence of bacterial resistance. *Emerg. Infect. Dis.* **2003**, *9*, 10–16. [CrossRef]
6. Erez, Z.; Steinberger-Levy, I.; Shamir, M.; Doron, S.; Stokar-Avihail, A.; Peleg, Y.; Melamed, S.; Leavitt, A.; Savidor, A.; Albeck, S.; et al. Communication between viruses guides lysis-lysogeny decisions. *Nature* **2017**, *541*, 488–493. [CrossRef] [PubMed]
7. Burns, N.; James, C.E.; Harrison, E. Polylysogeny magnifies competitiveness of a bacterial pathogen in vivo. *Evol. Appl.* **2015**, *8*, 346–351. [CrossRef]
8. Clokie, M.R.J.; Kropinski, A.M. *Bacteriophages: Methods and Protocols Volume 1: Isolation, Characterization, and Interactions*; Springer: Berlin/Heidelberg, Germany, 2009; ISBN 978-1-60327-164-6.
9. Krylov, V.N. Phage Therapy in Terms of Bacteriophage Genetics: Hopes, Prospects, Safety, Limitations. *Russ. J. Genet.* **2001**, *37*, 715–730. [CrossRef]
10. Qadir, M.I.; Mobeen, T.; Masood, A. Phage therapy: Progress in pharmacokinetics. *Braz. J. Pharm. Sci.* **2018**. [CrossRef]
11. Abedon, S.; Thomas-Abedon, C. Phage Therapy Pharmacology. *Curr. Pharm. Biotechnol.* **2010**, *11*, 28–47. [CrossRef]
12. Lucchesi, P.; Carraway, M.; Marinus, M.G. Analysis of forward mutations induced by N-methyl-N'-nitrosoguanidine in the bacteriophage P22 mnt repressor gene. *J. Bacteriol.* **1986**, *166*, 34–37. [CrossRef]
13. Court, D.L.; Sawitzke, J.A.; Thomason, L.C. Genetic Engineering Using Homologous Recombination. *Annu. Rev. Genet.* **2002**, *36*, 361–388. [CrossRef] [PubMed]
14. Qimron, U.; Marintcheva, B.; Tabor, S.; Richardson, C.C. Genomewide screens for *Escherichia coli* genes affecting growth of T7 bacteriophage. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 19039–19044. [CrossRef] [PubMed]
15. Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K.A.; Tomita, M.; Wanner, B.L.; Mori, H. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol. Syst. Biol.* **2006**, *2*. [CrossRef] [PubMed]
16. Yosef, I.; Goren, M.G.; Globus, R.; Molshanski-Mor, S.; Qimron, U. Extending the Host Range of Bacteriophage Particles for DNA Transduction. *Mol. Cell* **2017**, *66*, 721–728. [CrossRef] [PubMed]
17. Kiro, R.; Molshanski-Mor, S.; Yosef, I.; Milam, S.L.; Erickson, H.P.; Qimron, U. Gene product 0.4 increases bacteriophage T7 competitiveness by inhibiting host cell division. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 19549–19554. [CrossRef] [PubMed]
18. Auster, O.; Globus, R.; Yosef, I.; Qimron, U. Optimizing DNA transduction by selection of mutations that evade bacterial defense systems. *RNA Biol.* **2019**, *16*, 595–599. [CrossRef]
19. Balachandran, A.; Oldfield, L.M.; Julia, C.; Marinelli, L.J.; Piuri, M.; Swigon, Z.; Kessel, V.; Hatfull, G.F. BRED: A Simple and Powerful Tool for Constructing Mutant and Recombinant Bacteriophage Genomes. *PLoS ONE* **2008**, *3*, 3. [CrossRef]
20. Fehér, T.; Karcagi, I.; Blattner, F.R.; Pósfai, G. Bacteriophage recombineering in the lytic state using the lambda red recombinases. *Microb. Biotechnol.* **2012**, *5*, 466–476. [CrossRef]
21. Van Kessel, J.C.; Hatfull, G.F. Mycobacterial recombineering. *Methods Mol. Biol.* **2008**, *435*, 203–215.
22. Marinelli, L.J.; Hatfull, G.F.; Piuri, M. Recombineering: A powerful tool for modification of bacteriophage genomes. *Bacteriophage* **2012**, *2*, 5–14. [CrossRef] [PubMed]
23. Jaschke, P.R.; Lieberman, E.K.; Rodriguez, J.; Sierra, A.; Endy, D. A fully decompressed synthetic bacteriophage øX174 genome assembled and archived in yeast. *Virology* **2012**, *434*, 278–284. [CrossRef] [PubMed]
24. Ando, H.; Lemire, S.; Pires, D.P.; Lu, T.K. Engineering Modular Viral Scaffolds for Targeted Bacterial Population Editing. *Cell Syst.* **2015**, *1*, 187–196. [CrossRef] [PubMed]
25. Kiro, R.; Shitrit, D.; Qimron, U. Efficient engineering of a bacteriophage genome using the type I-E CRISPR-Cas system. *RNA Biol.* **2014**, *11*, 42–44. [CrossRef] [PubMed]
26. Box, A.M.; McGuffie, M.J.; O'Hara, B.J.; Seed, K.D. Functional analysis of bacteriophage immunity through a Type I-E CRISPR-Cas system in Vibrio cholerae and its application in bacteriophage genome engineering. *J. Bacteriol.* **2016**, *198*, 578–590. [CrossRef] [PubMed]

27. Lemay, M.L.; Tremblay, D.M.; Moineau, S. Genome Engineering of Virulent Lactococcal Phages Using CRISPR-Cas9. *ACS Synth. Biol.* **2017**, *6*, 1351–1358. [CrossRef]
28. Shen, J.; Zhou, J.; Chen, G.-Q.; Xiu, Z.-L. Efficient genome engineering of a virulent Klebsiella bacteriophage using CRISPR-Cas9. *J. Virol.* **2018**, *92*. [CrossRef]
29. Tao, P.; Wu, X.; Tang, W.C.; Zhu, J.; Rao, V. Engineering of Bacteriophage T4 Genome Using CRISPR-Cas9. *ACS Synth. Biol.* **2017**, *6*, 1952–1961. [CrossRef]
30. Yaung, S.J.; Esvelt, K.M.; Church, G.M. CRISPR/Cas9-mediated phage resistance is not impeded by the DNA modifications of phage T4. *PLoS ONE* **2014**, *9*, 3–9. [CrossRef]
31. Hoshiga, F.; Yoshizaki, K.; Takao, N.; Miyanaga, K.; Tanji, Y. Modification of T2 phage infectivity toward *Escherichia coli* O157:H7 via using CRISPR/Cas9. *FEMS Microbiol. Lett.* **2019**. [CrossRef]
32. Bari, S.M.N.; Walker, F.C.; Cater, K.; Aslan, B.; Hatoum-Aslan, A. Strategies for Editing Virulent Staphylococcal Phages Using CRISPR-Cas10. *ACS Synth. Biol.* **2017**, *6*, 2316–2325. [CrossRef] [PubMed]
33. Bryson, A.L.; Hwang, Y.; Sherrill-Mix, S.; Wu, G.D.; Lewis, J.D.; Black, L.; Clark, T.A.; Bushman, F.D. Covalent modification of bacteriophage T4 DNA inhibits CRISPRCas9. *MBio* **2015**, *6*. [CrossRef] [PubMed]
34. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J.A.; Charpentier, E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **2012**, *337*, 816–821. [CrossRef] [PubMed]
35. Makarova, K.S.; Grishin, N.V.; Shabalina, S.A.; Wolf, Y.I.; Koonin, E.V. A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* **2006**, *1*, 7. [CrossRef] [PubMed]
36. Brouns, S.J.J.; Jore, M.M.; Lundgren, M.; Westra, E.R.; Slijkhuis, R.J.H.; Snijders, A.P.L.; Dickman, M.J.; Makarova, K.S.; Koonin, E.V.; van der Oost, J. Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science* **2008**, *321*, 960–964. [CrossRef] [PubMed]
37. Michniewski, S.; Redgwell, T.; Grigonyte, A.; Rihtman, B.; Aguilo-Ferretjans, M.; Christie-Oleza, J.; Jameson, E.; Scanlan, D.J.; Millard, A.D. Riding the wave of genomics to investigate aquatic coliphage diversity and activity. *Environ. Microbiol.* **2019**, *21*, 2112–2128. [CrossRef] [PubMed]
38. Shah, S.A.; Erdmann, S.; Mojica, F.J.M.; Garrett, R.A. Protospacer recognition motifs: Mixed identities and functional diversity. *RNA Biol.* **2013**, *10*, 891–899. [CrossRef]
39. Engler, C.; Kandzia, R.; Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* **2008**, *3*, e3647. [CrossRef]
40. Gibson, D.G.; Young, L.; Chuang, R.Y.; Venter, J.C.; Hutchison, C.A.; Smith, H.O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **2009**, *6*, 343–345. [CrossRef]
41. Rihtman, B.; Meaden, S.; Clokie, M.R.; Koskella, B.; Millard, A.D. Assessing Illumina technology for the high-throughput sequencing of bacteriophage genomes. *PeerJ* **2016**, *4*, e2055. [CrossRef]
42. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [CrossRef] [PubMed]
43. Joshi, N.A.; Fass, J.N. Sickle: A Sliding-Window, Adaptive, Quality-Based Trimming Tool for FastQ Files (Version 1.33) [Software]. Available online: https://github.com/najoshi/sickle (accessed on 12 January 2020).
44. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **2014**, *9*, e112963. [CrossRef] [PubMed]
45. Kurtz, S.; Phillippy, A.; Delcher, A.L.; Smoot, M.; Shumway, M.; Antonescu, C.; Salzberg, S.L. Versatile and open software for comparing large genomes. *Genome Biol.* **2004**, *5*. [CrossRef] [PubMed]
46. Labuhn, M.; Adams, F.F.; Ng, M.; Knoess, S.; Schambach, A.; Charpentier, E.M.; Schwarzer, A.; Mateo, J.L.; Klusmann, J.H.; Heckl, D. Refined sgRNA efficacy prediction improves largeand small-scale CRISPR-Cas9 applications. *Nucleic Acids Res.* **2018**, *46*, 1375–1385. [CrossRef]
47. Bin Moon, S.; Lee, J.M.; Kang, J.G.; Lee, N.E.; Ha, D.I.; Kim, D.Y.; Kim, S.H.; Yoo, K.; Kim, D.; Ko, J.H.; et al. Highly efficient genome editing by CRISPR-Cpf1 using CRISPR RNA with a uridinylate-rich 3′-overhang. *Nat. Commun.* **2018**, *9*, 3651. [CrossRef]
48. Luo, M.L.; Mullis, A.S.; Leenay, R.T.; Beisel, C.L. Repurposing endogenous type i CRISPR-Cas systems for programmable gene repression. *Nucleic Acids Res.* **2015**, *43*, 674–681. [CrossRef]

49. Majsec, K.; Bolt, E.L.; Ivančić-Baće, I. Cas3 is a limiting factor for CRISPR-Cas immunity in *Escherichia coli* cells lacking H-NS. *BMC Microbiol.* **2016**, *16*. [CrossRef]
50. Ueguchi, C.; Kakeda, M.; Mizuno, T. Autoregulatory expression of the *Escherichia coli* hns gene encoding a nucleoid protein: H-NS functions as a repressor of its own transcription. *MGG Mol. Gen. Genet.* **1993**, *236*, 171–178. [CrossRef]
51. Liu, M.; Deora, R.; Simons, R.W.; Doulatov, S.; Hodes, A.; Dai, L.; Mandhana, N.; Zimmerly, S.; Miller, J.F.; Liu, M.; et al. Tropism switching in Bordetella bacteriophage defines a family of diversity-generating retroelements. *Nature* **2004**, *431*, 476–481.
52. Zhao, X.; Wu, W.; Qi, Z.; Cui, Y.; Yan, F.; Guo, Z.; Wang, Z.; Wang, H.; Deng, H.; Xue, Y.; et al. The complete genome sequence and proteomics of Yersinia pestis phage Yep-phi. *J. Gen. Virol.* **2011**, *92*, 216–221. [CrossRef]
53. Liu, M.; Gingery, M.; Doulatov, S.R.; Liu, Y.; Hodes, A.; Baker, S.; Davis, P.; Simmonds, M.; Churcher, C.; Mungall, K.; et al. Genomic and Genetic Analysis of Bordetella Bacteriophages Encoding Reverse Transcriptase-Mediated Tropism-Switching Cassettes. *J. Bacteriol.* **2004**, *186*, 1503–1517. [CrossRef] [PubMed]
54. Zhao, X.; Cui, Y.; Yan, Y.; Du, Z.; Tan, Y.; Yang, H.; Bi, Y.; Zhang, P.; Zhou, L.; Zhou, D.; et al. Outer Membrane Proteins Ail and OmpF of Yersinia pestis Are Involved in the Adsorption of T7-Related Bacteriophage Yep-phi. *J. Virol.* **2013**, *87*, 12260–12269. [CrossRef] [PubMed]
55. You, L.; Suthers, P.F.; Yin, J. Effects of *Escherichia coli* physiology on growth of phage T7 in vivo and in silico. *J. Bacteriol.* **2002**, *184*, 1888–1894. [CrossRef]
56. Wilson, L.O.W.; O'Brien, A.R.; Bauer, D.C. The current state and future of CRISPR-Cas9 gRNA design tools. *Front. Pharmacol.* **2018**, *9*, 749. [CrossRef]
57. Park, J.; Kim, J.S.; Bae, S. Cas-Database: Web-based genome-wide guide RNA library design for gene knockout screens using CRISPR-Cas9. *Bioinformatics* **2016**, *32*, 2017–2023. [CrossRef]
58. Wang, T.; Wei, J.J.; Sabatini, D.M.; Lander, E.S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **2014**, *343*, 80–84. [CrossRef] [PubMed]
59. Martel, B.; Moineau, S. CRISPR-Cas: An efficient tool for genome engineering of virulent bacteriophages. *Nucleic Acids Res.* **2014**, *42*, 9504–9513. [CrossRef]
60. Marraffini, L.A. *The CRISPR-Cas System of Streptococcus Pyogenes: Function and Applications*; University of Oklahoma Health Sciences Center: Oklahoma, OK, USA, 2016.
61. Jiang, W.; Bikard, D.; Cox, D.; Zhang, F.; Marraffini, L.A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **2013**, *31*, 233–239. [CrossRef]