2018

Investigating Genetic (IN)Compatibility Between Temperate Phages and CRISPR-CAS Systems in Staphylococcus Aureus

Gregory W. Goldberg

INVESTIGATING GENETIC (IN)COMPATIBILITY BETWEEN TEMPERATE PHAGES AND CRISPR-CAS SYSTEMS IN *STAPHYLOCOCCUS AUREUS*

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Gregory W. Goldberg

June 2018

INVESTIGATING GENETIC (IN)COMPATIBILITY BETWEEN TEMPERATE PHAGES AND CRISPR-CAS SYSTEMS IN *STAPHYLOCOCCUS AUREUS*

Gregory W. Goldberg, Ph.D.

The Rockefeller University 2018

Prokaryotic organisms employ various mechanisms for defending against parasitism by viruses and other mobile genetic elements. One form of defense comprises the adaptive immune systems derived from clustered, regularly interspaced, short palindromic repeat (CRISPR) loci and CRISPR-associated (*cas*) genes. CRISPR-Cas immune systems enable the acquisition of heritable resistance to specific mobile genetic elements on the basis of nucleic acid sequence recognition, but do not necessarily discriminate between target elements which are burdensome and those which are beneficial. My thesis is concerned with the consequences of CRISPR-Cas immunity directed at a particular breed of bacterial DNA viruses, known as temperate phages, which cause both harmful (lytic) and benign (lysogenic) infections under different conditions.

Initial studies investigating prokaryotic CRISPR-Cas immunity seemed to indicate that functional, DNA-targeting systems cannot stably co-exist with their target elements *in vivo*. For example, in studies where immunity was directed at temperate phages, DNA-targeting CRISPR-Cas systems were found to prevent both lysogenic and lytic infections except when targeting was altogether abrogated via mutation or inhibition of the CRISPR-Cas system. The first part of my thesis work includes *in vivo* experiments which challenged the generality of this view, with regard to the different types of DNA-targeting CRISPR-Cas systems. Namely, I demonstrated that a staphylococcal branch of the 'type III' CRISPR-Cas systems is capable of tolerating lysogenic infections by specific temperate phages which are otherwise targeted during lytic infections. I further

established that the capacity for conditional temperate phage tolerance results from a transcription-dependent targeting modality which was not anticipated for this particular DNA-targeting type III system. In contrast, I observed only the expected genetic escape outcomes when temperate phages were targeted by a 'type II' CRISPR-Cas system with a transcription-independent (Cas9-based) DNA targeting modality. These findings laid the groundwork for subsequent studies of CRISPR-Cas immunity to phages in *Staphylococcus aureus* hosts, and guided my colleagues towards *in vitro* characterization of the type III system's transcription-dependent targeting mechanism.

CRISPR-Cas systems have been identified in about 50% of sequenced bacterial genomes, and the factors which influence this distribution are still not fully understood. My description of conditional tolerance by a staphylococcal, type III CRISPR-Cas system illustrated that, in principle, these particular systems could stably co-exist with their temperate phage target elements in lysogenic hosts while maintaining their ability to protect against lytic infections. During the second part of my thesis work, I set out to define additional phenotypic consequences for the lysogenized lineages of *S. aureus* which maintain conditional tolerance, in an effort to better understand how this phenomenon might influence the distribution and stability of type III systems among natural isolates. Notably, I found that the maintenance of certain temperate-phage-targeting systems can incur fitness costs in lysogenic populations. I showed, furthermore, that these costs are potentially greater if more than one temperate phage is targeted in populations of double lysogens, but that they can be alleviated by mutations which do not abrogate phage targeting during lytic infections. Collectively, these findings imply that long-term maintenance of type III systems in natural populations of lysogens might require additional evolutionary fine-tuning, particularly among lineages which are prone to multiple infection.

This thesis is dedicated to my family and dear friends

ACKNOWLEDGMENTS

I am deeply grateful for the opportunities, training, and resources I was provided during my time at The Rockefeller University. It was an inspiring and formative experience for me, which would not have been possible if not for a number of outstanding individuals that I had the fortune of interacting with.

First, I'd like to sincerely thank my advisor, Dr. Luciano Marraffini, for years of support and patience, and for establishing a truly vibrant and innovative research program at Rockefeller, centered on the biology of CRISPR-Cas systems in bacteria. I was entirely unfamiliar with this intriguing subject prior to reading about his newly formed lab in 2011, and ultimately drawn to the David Rockefeller Graduate program because of it. I am further grateful that, shortly after rotating with him, Dr. Marraffini warmly entertained my wishes to forego additional rotations and pursue doctoral research in his lab. It was here that I worked closely with him on many occasions, and came to know his many remarkable qualities. Dr. Marraffini is a superbly clever and devoted scientist, and I will always be inspired by his courage, wise regard for pragmatism, and ability to foster a work environment where diverse people could come together to freely share their ideas and passion for science, and yet also cultivate their own individuality.

I am also indebted to the members of my thesis committee, Drs. Charlie Rice and Vincent Fischetti, for their thoughtful suggestions, commentary, and genuine encouragement over the years—it truly was an indispensable part of my experience here. I'd further like to thank my external examiner, Dr. Bruce Levin, for graciously agreeing to fulfill this role, and for the cheerful, personal encouragement he offered me after learning of my early work in the lab.

Likewise, I am indebted to Dean Sidney Strickland, Emily Harms, Marta Delgado, Kristen Cullen, Cristian Rosario, Stephanie Fernandez, Andrea Morris, and any other members of the Dean's Office I might be missing, for maintaining a lively, warm, and welcoming place in Founder's Hall for graduate students over the years, and for running a truly superb graduate program at The Rockefeller University.

Additionally, I'd like to sincerely thank Dr. Chad Euler (formerly of the Fischetti Laboratory, now Assistant Professor at CUNY Hunter), not only for his expert handling of the mouse work during our brief collaboration, but also for years of collegial interactions at joint lab meetings and Rockefeller-related events.

Next, I'd like to thank all the members of the Marraffini Lab, past and present, for the countless scientific discussions we shared, and for humoring my often loquacious demeanor therein. In particular, I'd like to highlight those who contributed directly to this thesis at some point or another.

Dr. Wenyan Jiang, first and foremost, deserves a very special thanks for his inspiring work ethic and dedication, and for his seemingly countless contributions to the lab. Wenyan took initiative on so many occasions, and in so doing opened key research avenues for me and other members of the lab. Among his invaluable and timely contributions to this work in particular was the cloning of our first plasmid-borne CRISPR-Cas system, pWJ30β, from which all of my type III plasmids are originally derived. Wenyan was also an excellent bay-mate throughout our 5 years together in the lab, and I am extremely grateful for the technical training and feedback he provided me on numerous occasions.

I am also extremely grateful to Dr. David Bikard for his generally learned and resourceful contributions to the lab, in addition to the technical training he provided me. Dr. Bikard's lively feedback truly helped to strengthen the rigor of my early work, and I will always appreciate that.

I would like to sincerely thank Dr. Poulami Samai for inspiring me with her patient and devoted approach to both science and laboratory work, in addition to her expert knowledge of nuclease biology. Dr. Samai, I should also mention, ultimately provided a skillful *in vitro* demonstration that the type III-A CRISPR-Cas complex we study cleaves the non-template strand of target DNA in a transcription-dependent fashion, as well as its RNA transcripts. These were monumental and reassuring findings for the field, and utterly invaluable to the confident progression of our continued work on type III-A systems in the lab.

Inbal Maniv and Dr. Asma Hatoum-Aslan were instrumental in getting the lab on its feet from the very beginning, and assisted in my technical training on numerous occasions. On both accounts I am truly indebted to them. Inbal in particular had a range of designated responsibilities around the lab, and managed to keep it in order for many years before moving on to her doctoral work in Israel. I can't help but mention that she also trained me on my very first top agar experiment… Many thanks, Inbal!

Drs. Andrew Varble and Joshua W. Modell also contributed a great deal of resourcefulness and technical expertise to the lab over the years, in addition to reagents and advice that proved useful in my recent work. I am sincerely grateful to them as well.

I'd also like to thank Rahul Bhosle and Phone Ko for maintaining a highly functional and convenient workspace, ever since they assumed their various lab responsibilities (including those which were formerly taken care of by Inbal).

Finally, I'd like to thank Robert Heler, Philip M. Nussenzweig, Dr. Charlie Mo, Dr. Alexander Meeske, Pascal Maguin, Claire Kenney, and Jakob Rostol for sharing additional advice, feedback, or data on various topics pertinent to my work, and for their generally collegial presence in the Marraffini Lab.

I am also sincerely grateful for contributions that came from outside the Marraffini Lab.

Although we've never met, I am deeply grateful to Dr. Angelika Gründling for her contribution of the pCL55-iTET vector. This reagent was invaluable to my timely demonstration of transcription-dependent targeting by a type III CRISPR-Cas system.

I am grateful to The Rockefeller University Genomics Resource Center for next generation sequencing and/or library preparation services, and to Dr. Connie Zhao and Bin Zhang (as well as their assistants at the facility) in particular for their expert knowledge and reliable advice over the years.

In addition, I would like to thank Dr. Seth Darst for the general advice on transcription he provided to the Marraffini lab over the years, the Torres lab at NYU for providing us with a protocol to isolate RNA from flash-frozen samples, the Bae and Novick laboratories for providing additional strains, and Dr. Henrik Molina of The Rockefeller University Proteomics Resource Center for informative discussions and advice on a prospective collaboration project. I'd also like to thank staff members at The Rockefeller University Library, including Olga Nilova and Jeanine McSweeney, for ensuring access to articles outside their collection, and for suggested readings on copyright guidelines.

I am also grateful to Dr. John Chen of the Novick laboratory for his timely and informative response to my e-mail question regarding the construction of antibiotic resistance-marked temperate phages. We had met only briefly at the 2012 Phage Meeting in Cold Spring Harbor.

I am equally indebted to Drs. Robert (Bob) Johnston, Jr., Claude Desplan, Suse Broyde, and Patrick Eichenberger for the tremendous encouragement, training, and inspiration they provided me over the years, and in particular during my time as an undergraduate at NYU. Perhaps most relevant to this work, it was Dr. Broyde who first inspired my interest in temperate phages, almost seven years ago, with her detailed lecture on phage lambda.

I would also like to declare that the content of this thesis includes excerpts, figures, illustrations, and tables that were adapted or reprinted from two previously published works (Goldberg et al., 2014; Goldberg and Marraffini, 2015). I am grateful to Nature Publishing Group for reserving me the rights to do this, as an author of those works.

Last, but certainly not least, I would like to thank my family and dear friends (including my late grandma, Dorothy) for inspiring and supporting me in so many ways over the years. I cannot justify including within this particular document a complete account of what they mean to me, but suffice it to say that I've dedicated this thesis work to them.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS & ABBREVIATIONS

HGT/LGT…….Horizontal Gene Transfer/Lateral Gene Transfer

HDT……..Horizontal DNA Transfer

MGE…….…..Mobile Genetic Element

phage……...bacteriophage

R–M……...Restriction–Modification

CRISPR…….....Clustered Regularly Interspaced Short Palindromic Repeat

Cas……..CRISPR-associated

REase…….……Restriction Endonuclease

MTase……Methyltransferase

crRNA…….CRISPR RNA

lambdoid…….....Lambda(λ)-like

ORF…..……Open Reading Frame

PAM……..Protospacer-Adjacent Motif

^{E}LCR……..Early Lytic Control Region

^{L}LCR……Late Lytic Control Region

RecA*……...…….....activated RecA

SOS……...…….....“Save-Our-Ship” stress response

UV……Ultraviolet

MMC……...…….....Mitomycin C

bp…….base pair(s)

kbp…….kilobase pair(s)

IEC…….Immune Evasion Cluster

SaPI……...…*S. aureus* Pathogenicity Island

CFU…….Colony-Forming Unit

PFU……..Plaque-Forming Unit

SNP…...……Single Nucleotide Polymorphism

RT……..….Reverse Transcription

PCR……..……..Polymerase Chain Reaction

Cm^R…...…….Chloramphenicol-Resistant

Mup^R…...……...…...Mupirocin-Resistant

Erm^R……..……...Erythromycin-Resistant

Tet^R……...…….....Tetracycline-Resistant

ATc……...…….....Anhydrotetracycline

STU……..SaPI Transfer Unit

HEPN domain……..…… Higher Eukaryotes and Prokaryotes Nucleotide-binding domain

ssDNA……..single-stranded DNA

CHAPTER 1: INTRODUCTION

1.1 Implications of Horizontal Gene Transfer for prokaryotic evolution

The 'vertical' transfer of a genome from parent to progeny during reproduction can result in the stochastic alteration of its sequence content via mutation. Mutations are a source of genetic variation and may give rise to novel adaptive genotypes that can be selected for during evolution. However, mutations occurring at random are more likely to be deleterious than beneficial. According to Muller's ratchet mechanism (Muller, 1932, 1964), genomic degeneration occurring via random mutation can drive an organism to extinction in the absence of a recombination process to increase the chance of reconstructing fully functional genomes (Felsenstein, 1974). This fate may be averted through sexual recombination, as it is exhibited in extant eukaryotes (**Figure 1-1A**). Moreover, sexual recombination offers an additional mechanism for generating novel genotypes during reproduction, but it has evolved with reproductive barriers that help to moderate the variation produced by ensuring that genetic exchanges occur between closely related genomes (for example, within a species) (Seehausen et al., 2014). By contrast, recombination can occur between more distantly related genomes through horizontal gene transfer (HGT), which constitutes another key avenue for genome diversification (Keeling and Palmer, 2008; Thomas and Nielsen, 2005).

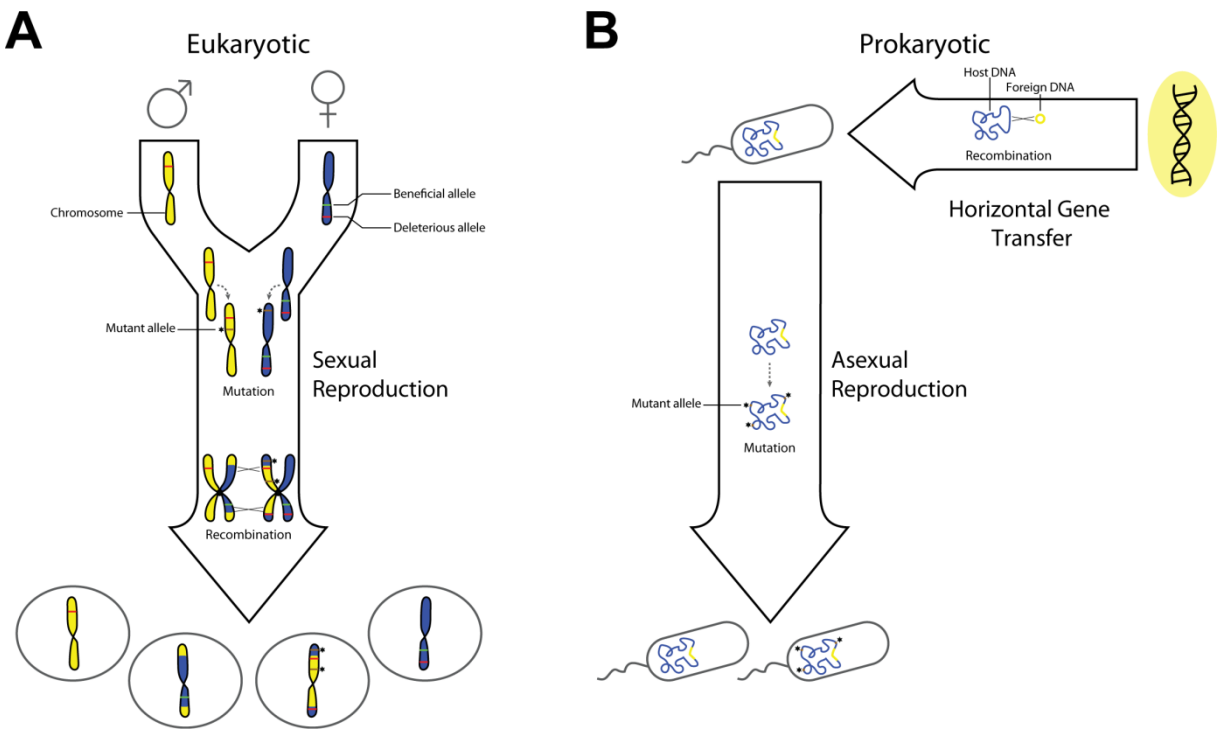


Figure 1-1. Genome evolution can result from vertical or horizontal transmission of DNA.

(A) Schematic summary of genome evolution in a unicellular and haploid eukaryote that reproduces sexually. Mutation can generate novel, mutant alleles (brown lines with stars) within chromosomes during DNA replication, which may be either deleterious or beneficial. Recombination between homologous chromosomes (yellow or blue) can occur during a tetraploid zygotic stage, before any cellular division. Sexual recombination in this manner provides opportunities for the removal of deleterious alleles (red lines) as well as for the introduction of beneficial alleles (green lines). Both processes occur vertically during reproduction. **(B)** Schematic summary of genome evolution in a haploid and unicellular prokaryote that reproduces asexually. Mutation may once again generate mutant alleles (shown as brown lines with stars) vertically during reproduction of the prokaryote's chromosome (blue); however, recombination with exogenous sources of DNA (yellow) must occur through horizontal processes that are not tied to reproduction.

Broadly speaking, HGT refers to the transmission of genetic material from one organism to another, through processes that are not tied to reproduction. Following physical transport into the recipient organism, genetic material may be incorporated into the genome through recombination with the chromosome (or chromosomes) or through autonomous replication as an episome. Evidence for HGT in eukaryotes has accumulated over the years, but most reports involve free-living protists and fungi, or endosymbionts inhabiting the cytosol of multicellular organisms (Dunning Hotopp, 2011; Keeling and Palmer, 2008). Meanwhile, the mammalian nuclear genome appears to be well insulated from HGT, and this has been suggested to result from the isolation of their germline cells in the gonads (Andersson et al., 2001). One could further speculate that this represents an evolutionary strategy that helps to preserve genomic fitness in mammalian lineages, as HGT can introduce deleterious mutations in addition to those generated vertically and might be superfluous if sufficient recombination and genetic variation is already achieved through sexual reproduction. However, for unicellular, asexually reproducing organisms such as prokaryotes, HGT provides an avenue for recombination that, in theory, can operate like sexual recombination in averting Muller's ratchet (Takeuchi et al., 2014), in addition to providing genomic variation (**Figure 1-1B**). Stricter definitions of HGT, also known as lateral gene transfer (LGT), may distinguish it from gene transfers that result in gene conversion or repair at homologous loci (Ochman et al., 2000; Vos, 2009). In other words, HGT/LGT may be

defined in certain contexts as only those transfers that introduce non-homologous (also known as heterologous) sequences to a genome. When recombinations with exogenously derived donor DNA are considered, however, both heterologous DNA insertions and gene conversions/repair at homologous loci can result from the same molecular processes (e.g., homologous recombination between related segments of DNA) (Johnston et al., 2014; Vos, 2009). Therefore, the term 'horizontal DNA transfer' (HDT) can also be employed to describe, more precisely, the transfer of DNA from one organism to another without distinguishing between outcomes (Croucher et al., 2016). In practice, because relatively little is known about horizontal RNA transfer in prokaryotes, the less stringent definition of HGT/LGT often invoked in the prokaryotic literature (and throughout this work) is essentially synonymous with HDT. Sequencing of diverse prokaryotic phyla and comparative genome analyses suggest the virtual ubiquity of HGT, as well as its potential to drive rapid genome evolution through the acquisition of novel genes en bloc (Dobrindt et al., 2004; Koonin and Wolf, 2008). Notably, HGT appears to be a common route through which evolving pathogens acquire particular traits, including resistance to antibiotics (Croucher et al., 2011). Nevertheless, in the absence of reproductive barriers, the risk of introducing deleterious information into the genome is theoretically greater with HGT than with conventional sexual reproduction. For example, HGT can introduce mobile genetic elements (MGEs) from unrelated organisms — such as transposons, viruses or plasmids — which can, in

turn, parasitize the genome (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). This matter is further complicated by the fact that MGEs are themselves often vectors of HGT in the prokaryotic domain, as I explain below.

Molecular mechanisms for prokaryotic HGT have been established using various experimental systems, beginning with the discovery of pneumococcal capsule acquisition (Griffith, 1928). Overall, three major categories have been studied extensively in bacteria: conjugation, transduction and natural transformation. Conjugation involves the direct transfer of DNA from a donor to a recipient cell during physical contact through pili (Llosa et al., 2002) or pore-like structures (Grohmann et al., 2003). It is typically orchestrated by conjugative plasmids or transposons that carry the necessary genetic functions to ensure their own transfer through this process. Transduction refers to the transfer of non-viral DNA encapsulated in viruses or virus-like particles. During lytic infections, the viruses of bacteria (known as phages) can package parts of the lysing host genome into a few of their particles, and this DNA may then be delivered to distant cells (Weinstock, 2002; Zinder and Lederberg, 1952). Natural transformation involves the uptake of free DNA from the environment and occurs after the recipient cell enters a physiological state known as 'competence' (Johnston et al., 2014). Although natural transformation of competent bacteria is a host-encoded process and usually results in unbiased DNA uptake, most known examples of conjugation and transduction are tied to the activity of MGEs that need not benefit the

host and can even be detrimental. Unlike typical genes that rely strictly on a host genome for carriage, mobile genetic elements can propagate independently of host replication. Therefore, their presence in genomes could reflect parasitism rather than natural selection for their phenotypic contributions to host survival (Orgel and Crick, 1980). For example, although conjugative plasmids and transposons can carry adaptive traits such as genes for antibiotic resistance that promote survival, horizontal dissemination of these elements can occur even when they burden their hosts with extraneous genetic cargoes that do not promote fitness or vertical transmission. Lytic phages represent an extreme form of genomic parasitism in that they immediately dispose of their host after the process of self-amplification. Hence, molecular barriers that limit the spread of MGEs can directly contribute to prokaryotic survival (Labrie et al., 2010; Makarova et al., 2013; Thomas and Nielsen, 2005). Some barriers, such as abortive infection systems, offer highly selective resistance to lytic infection by certain phages or classes of phages (Chopin et al., 2005; Depardieu et al., 2016; Labrie et al., 2010). When considering the potential benefits of HGT, there appear to be tradeoffs associated with generalized versus selective resistance mechanisms. For example, although selective barriers do not necessarily offer resistance to diverse types of elements, they are less likely to interfere with HGT outright. Among the known prokaryotic defense mechanisms, restriction–modification (R–M) and CRISPR–Cas (clustered, regularly interspaced short palindromic repeat–CRISPR-associated proteins)

systems are distinct in that they can resist diverse MGEs while still offering selectivity. Thus, both systems can safeguard against genomic parasitism without necessarily interfering with HGT. These properties, furthermore, provide a central basis for the comparisons to mammalian immunity discussed in this work.

1.2 Prokaryotic immune systems are optimized for selective resistance to MGEs

Across all domains of life, the fitness of an organism may be threatened by parasitic interactions with foreign elements. Accordingly, organisms have evolved several biological barriers that aid in the defense against parasitism. Physical barriers, such as the cellular envelope and those at the surface of mammalian skin and mucosal linings, repel foreign elements rather indiscriminately (McGuckin et al., 2011; Proksch et al., 2008). Immune systems, by contrast, offer the potential to repel particular target elements while tolerating others and can thus function as highly selective barriers. However, the balance of selectivity must be well calibrated when relying on active means of surveillance and resistance; a system that is too selective risks evasion by pathogens, whereas one that is unselective risks damaging the host organism through immunopathology. In this regard, the mammalian immune system serves as an excellent example, with innate and adaptive activities (Medzhitov, 2007) that are optimized to reliably combat parasitism and generally tolerate host constituents. Prokaryotic R–M and CRISPR–Cas systems resemble the mammalian immune system in

their ability to actively resist infectious elements with a high degree of selectivity, as well as in their capacity for tolerance (Barrangou and Marraffini, 2014; Goren et al., 2012; Tock and Dryden, 2005). In the following two subsections, I describe the core mechanisms through which these systems achieve selectivity and 'self' tolerance via nucleic acid surveillance within the cell. Analogy to features of mammalian immunity is also considered in more detail.

1.2.1 *Restriction-Modification systems*

A minimal R–M system encodes enzymes with two activities: a restriction endonuclease (REase), which cleaves double-stranded DNA upon recognition of specific target sequences, and a methyltransferase (MTase), which modifies these sites through methylation to prevent cleavage (Tock and Dryden, 2005). Depending on the type of system, these activities may be associated with separate proteins, a complex, or a single protein (for a comprehensive review on their nomenclature, see (Roberts et al., 2003)). Recognition sites for an R–M system are typically palindromic sequences of 4–8 bp in length — short enough to occur frequently in the genome of a prokaryotic organism at random. Hence, without a nuclear envelope to insulate their genetic material from the cytoplasm, prokaryotes use methylation to protect their chromosome and other native DNA elements from attack by the REase. Meanwhile, the REase can provide immunity

against invading genetic elements that have not had sufficient exposure to the MTase of the cell, including the DNA of phages and plasmids (Arber and Linn, 1969) (**Figure 1-2**).

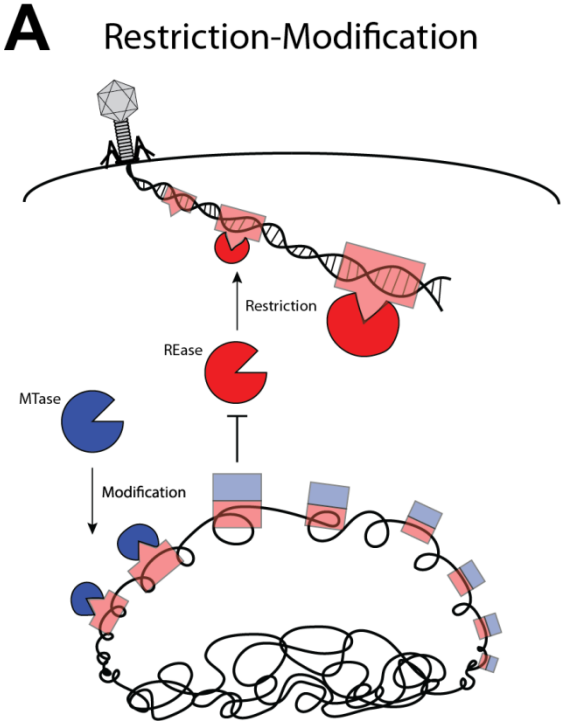


Figure 1-2. Schematic summary of antiviral defense mediated by Restriction–Modification systems.

(A) Intracellular restriction endonuclease enzymes (REases; shown in red) cleave injected viral DNA (black double helix) at short sequence motifs known as recognition sites (pink boxes). Meanwhile, methyltransferase enzymes (MTases; shown in blue) can modify DNA at the same recognition sites (appended with blue boxes) to prevent cleavage by their cognate REase. The sequence specificity of a REase is hardwired for a particular recognition site and thereby offers innate immunity to unmodified phages that harbor these sites in their DNA. Whereas unmodified invading DNA is rarely methylated fast enough to receive protection from restriction, modification is generally effective in preventing cleavage of the host chromosome (bottom) and thus allows for a rudimentary form of self and non-self discrimination.

Thus, methylation provides the R–M system with an intrinsic basis for distinguishing self from non-self, and each system is hardwired for surveillance of a particular recognition sequence, similarly to pattern recognition by mammalian pattern recognition receptors (Medzhitov, 2007). In these regards, R–M activities are akin to mammalian innate immune functions and may have an analogous role in the resistance against viruses and other MGEs (see (Vasu and Nagaraja, 2013) for a review on additional functions that have been described for some of these systems).

R–M systems have been identified in about 75% of sequenced prokaryotes, averaging roughly two systems per genome (Oliveira et al., 2014). As alluded above, the relative ubiquity of their recognition sequences in DNA dictates that the selectivity of the R–M system for foreign target elements must be primarily informed by methylation patterns (although underrepresentation of recognition sequences can also occur (Rocha et al., 2001)). However, this rudimentary mechanism for self and non-self discrimination is particularly susceptible to host mimicry evasion paradigms (Damian, 1964); even exogenous DNA may bypass restriction if it arrives pre-methylated (for example, through modification in a neighbouring cell possessing the same MTase) (Bertani and Weigle, 1953; Korona and Levin, 1993). In this scenario, the methylated foreign DNA is effectively tolerated as a self-component and recognized by the host MTase during subsequent rounds of replication. Therefore, an infectious MGE can propagate freely in spite of REase surveillance once it manages to achieve methylation. In parallel,

prokaryotic organisms may acquire resistance to these infectious elements through other defense mechanisms, such as their CRISPR–Cas systems (Dupuis et al., 2013; Hynes et al., 2014).

1.2.2 *CRISPR-Cas systems*

Among the two major classes of CRISPR-Cas systems, six distinct types (I-VI) and 19 subtypes (A-U) have now been defined (Makarova et al., 2015; Shmakov et al., 2017), but they all share two key genetic components: a CRISPR locus and a set of genes that encode Cas proteins. Cas proteins include nucleases that can degrade invading DNA or RNA targets, and they identify their targets via base pair complementarity using small RNA guides known as CRISPR RNAs (crRNAs) (Brouns et al., 2008; Gasiunas et al., 2012; Jinek et al., 2012; Zhang et al., 2012). Depending on the class of CRISPR-Cas system, crRNAs can function as guides in complex with a single effector protein or with a multiprotein effector complex (Shmakov et al., 2017; van der Oost et al., 2014). Class I CRISPR-Cas systems include the type I and type III systems that utilize a multiprotein effector complex, while Class II CRISPR-Cas systems include the type II, type V, and type VI that utilize a single protein effector. The crRNAs themselves are derived from transcription of one or more CRISPR loci, which are structured as arrays of short palindromic repeat sequences (~40 bp in length) that are intercalated with unique sequences of a similar length, known as spacers (**Figure 1-3**). After being processed into

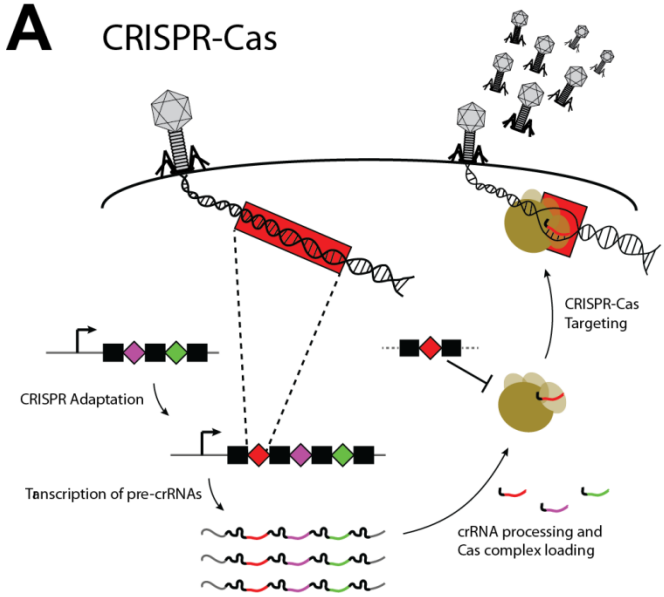


Figure 1-3. Schematic summary of antiviral defense mediated by CRISPR–Cas systems.

(A) CRISPR arrays are composed of alternating units of repeat sequences (black squares) interrupted by unique spacer sequences (colored diamonds). Newly encountered phage DNA sequences (boxed in red) can be incorporated as spacer DNA within the host CRISPR array through the process of CRISPR adaptation, which provides a genetic memory of past infection. Transcription of the CRISPR array provides primary transcripts, known as pre-CRISPR RNAs (pre-crRNAs), which are processed into short, mature species that each include a single spacer sequence. During CRISPR–Cas targeting, Cas protein complexes are guided by individual mature crRNAs to mediate the destruction of invading nucleic acids that harbor a matching target sequence. By virtue of sequences in their flanking repeat elements, spacer DNA of the CRISPR array is intrinsically spared from CRISPR–Cas targeting to prevent autoimmunity (Marraffini and Sontheimer, 2010a; Shmakov et al., 2017; Wiedenheft et al., 2012).

minimal, mature species, each crRNA contains a single spacer sequence that specifies a target for its Cas nuclease complex (Brouns et al., 2008; Carte et al., 2008; Charpentier et al., 2015; Hatoum-Aslan et al., 2011). In this way, Cas nucleases can be programmed to recognize diverse DNA or RNA targets, according to the spacer sequence of their crRNA. Importantly, the spacer content of a CRISPR locus can be actively modified through a process known as CRISPR adaptation (Barrangou et al., 2007) (reviewed in (Amitai and Sorek, 2016; Jackson et al., 2017)). During this process, spacer DNA can be acquired directly from an invading element for incorporation into the CRISPR locus (**Figure 1-3**). This provides the cell with a novel target sequence for its Cas nucleases as well as a genetic memory of the encounter to pass on to daughter cells. In this manner, CRISPR–Cas function is particularly analogous to mammalian adaptive immunity. However, it should be emphasized that memory acquired through CRISPR is fully heritable, whereas newborn mammals are only thought to receive short-lived, maternally derived passive immunity (Grindstaff et al., 2003). Furthermore, whereas the mammalian immune system relies heavily on clonal deletion and anergy of self-reactive cells to establish central tolerance (Kyewski and Klein, 2006), CRISPR–Cas systems do not require such mechanisms because spacer sequence diversity is not randomly generated but instead derived from invading genetic elements in an apparent Lamarckian evolutionary fashion (Barrangou and Marraffini, 2014; Haerter and Sneppen, 2012). Nevertheless, given that the spacer sequences in CRISPR loci are

themselves perfect matches for the crRNAs they encode, additional mechanisms are required to protect the CRISPR locus DNA from autoimmune responses (Marraffini and Sontheimer, 2010b; Shmakov et al., 2017; Wiedenheft et al., 2012). In this sense, CRISPR loci can be viewed as immune-privileged regions of the genome.

The majority of naturally occurring spacer sequences with known targets match to viral elements, although matches to plasmids and other MGEs are also observed (Brodt et al., 2011). A growing body of experimental work indicates that CRISPR–Cas systems are indeed capable of resisting a wide range of MGEs targeted by their spacers (Barrangou et al., 2007; Edgar and Qimron, 2010; Marraffini and Sontheimer, 2008; Tamulaitis et al., 2014). Unlike R–M systems, which discriminate targets primarily through methylation, CRISPR–Cas systems derive their selectivity first and foremost from an exquisitely specific sequence recognition capability based on crRNA complementarity. On the one hand, the spacer sequence of a typical mature crRNA is 20–40 nucleotides—presumably long enough to discriminate different target elements on the basis of sequence alone. Accordingly, a phage- or plasmid-derived spacer sequence is unlikely to specify targeting of the host's own chromosome (generally not exceeding ~5 Mbp for a prokaryotic genome (Koonin and Wolf, 2008)). On the other hand, it has been shown that spacers engineered to specifically target the chromosome appear to be lethally self-reactive (Bikard et al., 2014; Jiang et al., 2013a), so the observation that self-targeting spacers rarely occur in nature might be explained by the

immediate culling of cells that acquire such spacers — in a manner akin to clonal deletion. Interestingly, however, experimental evidence indicates that spacer acquisition is biased towards extra-chromosomal sequences from the outset (that is, even in the absence of CRISPR–Cas target degradation), despite the overrepresentation of chromosomal DNA by mass (Levy et al., 2015; Modell et al., 2017; Yosef et al., 2012). Evidently, this bias helps to ensure that CRISPR–Cas surveillance processes display a selective preference for foreign genetic elements, given that many types of HGT events do not involve physical linkage of foreign DNA to the chromosome. The phenomenon known as 'primed' CRISPR adaptation may also help to ensure that the adaptation machinery is selectively directed towards foreign elements, and that robust immunity is achieved in the face of recurring infections (Datsenko et al., 2012; Fineran et al., 2014; Jackson et al., 2017; Musharova et al., 2017). This phenomenon is analogous to affinity maturation and class-switching of antibodies in the mammalian immune system (Wabl and Steinberg, 1996), in the sense that it can refine the CRISPR repertoire to ensure more effective resistance during subsequent infections. Perhaps one distinction here, however, is that antibody development can be driven by antigen-presenting cells (Medzhitov, 2007) even after a particular infection is cleared, whereas primed CRISPR-Cas adaptation still requires re-exposure to the infectious element(s).

As discussed in section **1.1**, while the uptake of DNA from exogenous sources can introduce parasitic MGEs that compromise host fitness, it can also introduce

beneficial genes (Croucher et al., 2016). Four of the six (I-VI) types of CRISPR-Cas systems which have been classified to date (Makarova et al., 2015; Shmakov et al., 2017) include systems that cleave the DNA of their target elements (Jinek et al., 2012; Samai et al., 2015; Westra et al., 2012; Zetsche et al., 2015). It was therefore proposed that resistance to foreign DNA elements by CRISPR-Cas systems could jeopardize the survival of bacterial hosts which rely heavily on horizontal gene transfer (HGT) (Marraffini and Sontheimer, 2008), and in turn promote the evolution of strains that do not harbor CRISPR-Cas systems (Marraffini, 2013; Palmer and Gilmore, 2010). CRISPR-Cas systems are indeed absent from ~50% of sequenced bacterial genomes (Touchon et al., 2016), despite evidence that they can be transferred horizontally (Chakraborty et al., 2010; Godde and Bickerton, 2006; Horvath et al., 2009; Seed et al., 2013) and function in heterologous hosts (Bikard et al., 2012; Brouns et al., 2008; Heler et al., 2015). Moreover, the potential for genetic loss of CRISPR-Cas systems has been clearly demonstrated in laboratory settings. For example, population bottlenecks imposed by antibiotics were found to select for mutant or deleted CRISPR-Cas systems when a target element carrying resistance to the antibiotic was introduced prior to treatment (Edgar and Qimron, 2010; Fischer et al., 2012; Goldberg et al., 2014; Gudbergsdottir et al., 2011; Jiang et al., 2013b). In this manner, tradeoffs associated with targeting of favorable genetic elements could influence the distribution of CRISPR-Cas systems in wild populations. It should be emphasized, however, that CRISPR-Cas systems were not

found to impose a general barrier to HGT when evolutionary timescales were considered *in silico* (Gophna et al., 2015)—perhaps owing to their optimizations toward selectivity.

Fitness costs associated with the maintenance of CRISPR-Cas systems, including those that result from 'immunopathological' self-targeting, have also been proposed to influence the distribution of these systems (Goldberg and Marraffini, 2015; van Houte et al., 2016; Vercoe et al., 2013). As mentioned above, previous studies have demonstrated that chromosomal targeting by CRISPR-Cas systems in bacteria is severely detrimental to growth and potentially lethal (Jiang et al., 2013b; Vercoe et al., 2013). Therefore, high rates of CRISPR-Cas adaptation are potentially detrimental to the host, insofar as they engender a greater risk of acquiring self-targeting spacers (Heler et al., 2017). In certain cases, immunopathology may result from CRISPR-Cas immunity directed at temperate phages that integrate into their host's chromosome during lysogeny (Cady et al., 2012; Edgar and Qimron, 2010; Goldberg et al., 2014). 'Temperate' phages are distinct from strictly lytic 'virulent' phages in that they can establish an alternative infection state, known as lysogeny (Lwoff, 1953), wherein the host is spared from lysis. Integrative temperate phages, such as lambda-like (lambdoid) temperate phages, integrate into their host's chromosome as a so-called prophage upon establishment of a lysogenic infection (**Figure 1-4**). The difference between the lambdoid temperate phage's infection

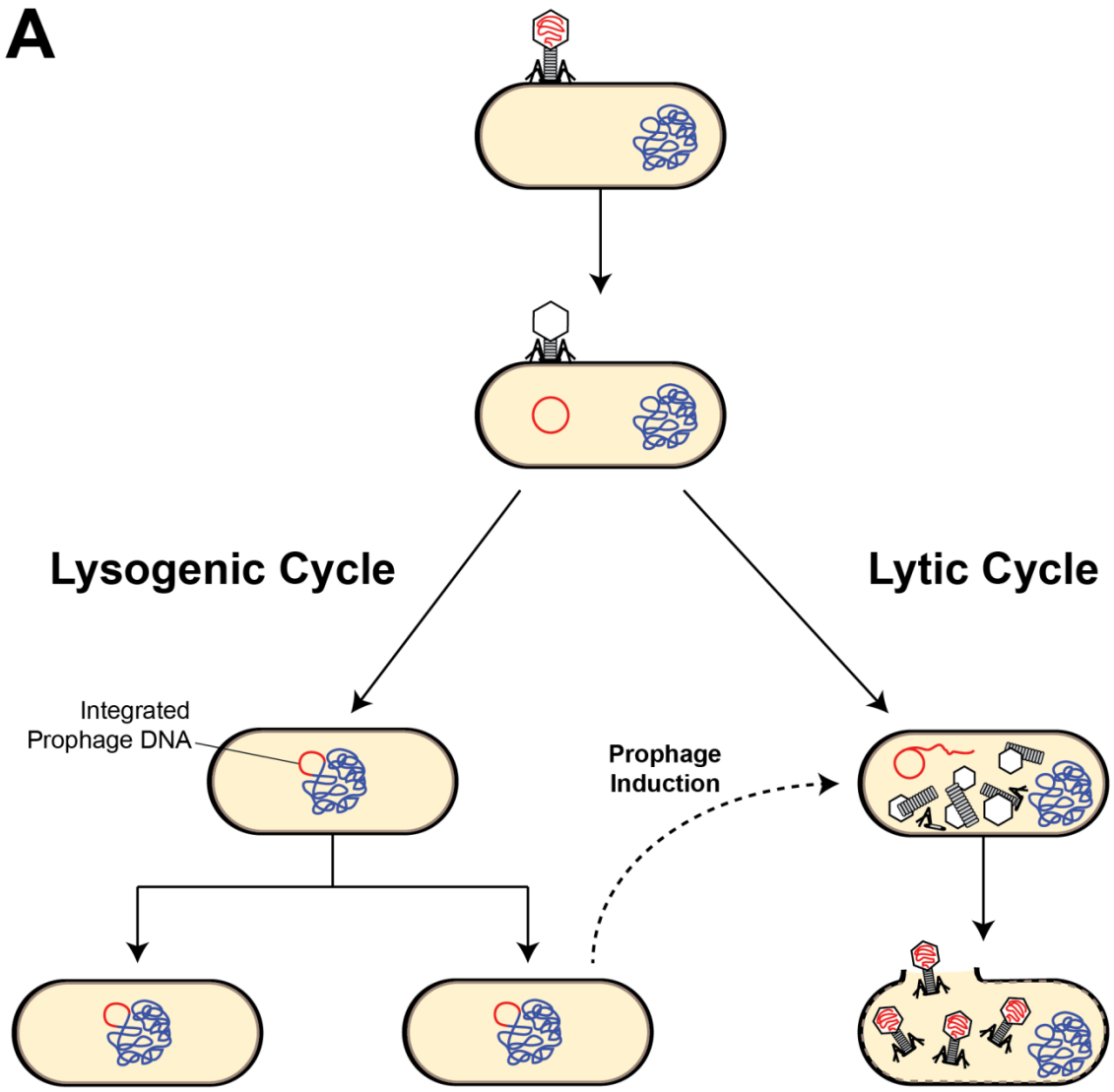


Figure 1-4. Summary of phage λ's alternate infection pathways.

(A) Following injection and circularization of the λ genome (red), infection can follow either a lysogenic or lytic pathway. During a lytic cycle, λ genomes are propagated and packaged into phage particles, which are ultimately released via lysis of the host cell. During a lysogenic cycle, the host is spared from lysis via repressor-mediated immunity, and the λ genome is integrated into the bacterial chromosome (blue) as a so-called prophage. Subsequent rounds of cell division allow passive replication of the prophage within the chromosome. During a process known as 'prophage induction', phage λ can excise from a lysogen's chromosome and re-enter the lytic cycle.

states reflects changes in its transcriptional activity within the host, rather than alterations of its DNA sequence (Johnson et al., 1981). Therefore, when CRISPR-Cas immunity is directed at an integrative temperate phage, targeting of phage DNA may occur at prophage loci in the event of lysogenization. In the next section, I briefly summarize the transcriptional regulation that determines a lambdoid phage's infection state, as well as some other relevant features of lambdoid temperate phage biology.

1.3 Lambdoid temperate phages

Typically, the term 'lambdoid' is used to describe temperate phages of *Escherichia coli* that are similar to phage λ (lambda) (Brüssow et al., 2004; Gottesman and Weisberg, 2004). The mechanisms of gene regulation (Dodd et al., 2005) and site-specific integration (Kotewicz et al., 1977) that that are central to phage lambda's lytic and lysogenic life cycles have been characterized in extensive detail. However, these life cycles are not specific to temperate phages infecting *E. coli*, and lambda's mechanisms therefore serve as more general models for understanding certain types of temperate phages. In this work, I consider a temperate phage lambdoid if its prophage integrates site-specifically, and if its life cycle decisions are principally determined by a *cI*-/*cro*-like regulatory mechanism, or 'genetic switch', akin to that of phage lambda (Johnson et al., 1981; Ptashne, 2011).

Under typical laboratory conditions, the majority of lambda infections give rise to a lytic cycle. A proportion of these infections, however, will lead to establishment of a lysogenic cycle and prophage integration (**Figure 1-4**). This occurs when lambda's *cI*-encoded repressor accumulates early during infection to sufficient levels that repress transcription of *cro*, a gene required for establishment of the lytic cycle (Casjens and Hendrix, 2015). During lysogeny, *cI* expression remains on, and its repressive action ensures that transcription of *cro* (and also as a consequence most of the phage genome) remains off (Ptashne, 2011). In cases where *cro* was not sufficiently repressed, a lytic cycle ensues. Under these conditions, the *cro* gene product—also a repressor—accumulates to sufficient levels that prevent *cI* transcription, and thereby ensures that expression of the other early-transcribed lytic genes can proceed. In turn, some of these early-transcribed lytic genes include regulators that allow transcription of late genes and subsequent completion of the lytic cycle (Casjens and Hendrix, 2015).

In order to exert their regulatory control early during infection, CI and Cro repressors compete for binding at two regions on the lambda genome, known as operators, which each possess three binding sites. The right operator (O_R) is located between the oppositely oriented *cI* and *cro* ORFs, and serves as the primary site of action for determining the decision between lytic and lysogenic life cycles (Johnson et al., 1981). Although CI and Cro can occupy each of the three O_R binding sites (O_R1, O_R2, and O_R3), they have evolved with distinct preferences that allow differential

accumulation at each site. Thus, CI and Cro can preferentially repress their partner's transcription without completely repressing their own transcription up front (**Figure 1-5**). The molecular details of transcriptional regulation by CI and Cro are discussed elsewhere in more depth (Dodd et al., 2005).

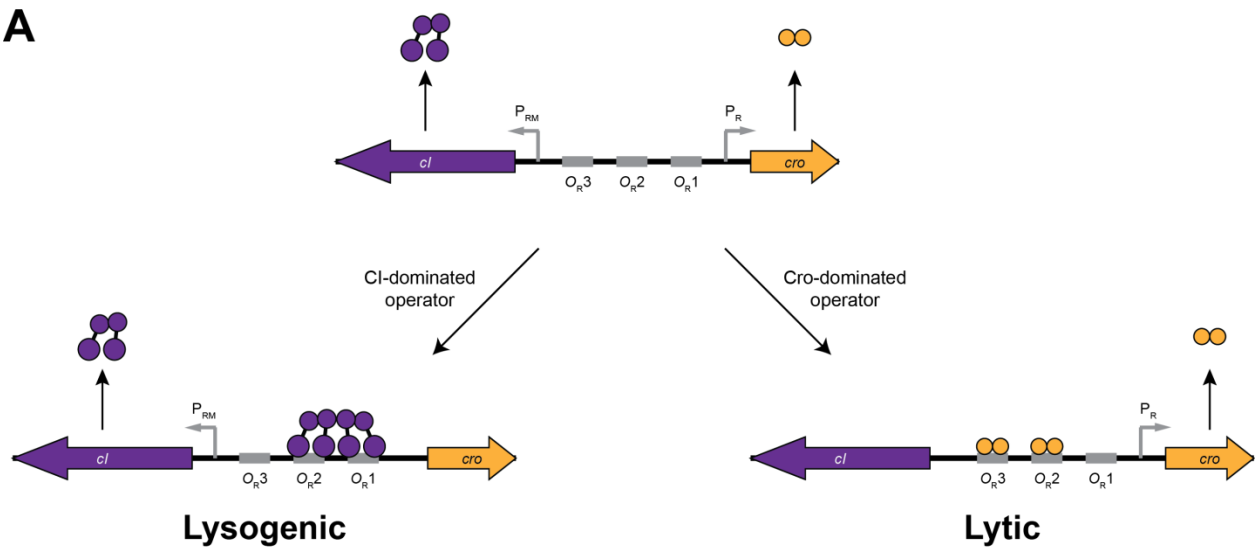


Figure 1-5. Schematic diagram summarizing transcriptional regulation that stabilizes phage λ's decision between lysogenic and lytic life cycles.

(A) During infections, the *cI* and *cro* ORFs (thick purple and orange arrows, respectively) can be transcribed from the P_{RM} and P_R promoters, respectively (grey bent arrows). Homodimers of their gene products (CI and Cro, respectively) can bind at each of the three operator binding sites (solid grey rectangles). Cro's preference for O_R3 and O_R2 over O_R1 allows it to dominate the operator such that *cI* transcription is turned off while *cro's* can initially remain on; a lytic cycle ensues (right). Meanwhile, CI's preference for O_R1 and O_R2 over O_R3 allows it to dominate the operator such that *cro* transcription is turned off while *cI*'s can remain on; a lysogenic cycle ensues (left). Features of the diagram are scaled arbitrarily.

In temperate phages with related genomic architecture, similar operators can be found between divergently transcribed *cI*- and *cro*-like ORFs. Later in the text, I'll refer to such regions more generally as early lytic control regions (^{E}LCRs) for simplicity. Another region, which I've termed the late lytic control region (^{L}LCR), refers to a conserved region that was identified within various temperate phages infecting gram positive bacteria (Ferrer et al., 2011; Quiles-Puchalt et al., 2013). As its name implies, this latter region appears to be specifically involved in regulating transcription of late lytic genes and completion of the lytic cycle, but it is not directly controlled by the CI- or Cro-like proteins.

Unlike toxic lytic infections, lysogenic infections do not necessarily reduce host fitness and can even be advantageous (Brüssow et al., 2004; Dykhuizen et al., 1978; Edlin et al., 1977). In addition, they provide immunity to lytic superinfections by the same phage, because the CI(-like) repressors expressed from a resident prophage during lysogeny can also suppress lytic development of an injected phage in *trans* (Casjens and Hendrix, 2015). However, this is not always a stable arrangement: functional lambdoid prophages retain the ability to re-initiate the lytic cycle and excise from the chromosome, both stochastically and in response to DNA damage or other signals (Bailone et al., 1979; Johnson et al., 1981; Little and Michalowski, 2010; Schubert et al., 2007). This process arises from derepression via autoproteolytic cleavage of the prophage's CI(-like) repressors, and is known as 'prophage induction' (**Figure 1-4**). CI(-

like) repressors possess an intrinsic serine protease activity, but their autoproteolysis only occurs efficiently in the presence of activated RecA (RecA*) co-factors (Erill et al., 2007; Livny and Friedman, 2004; Mo et al., 2014). Host-encoded RecA becomes activated when it associates with ssDNA, typically arising via DNA damage. Accumulation of RecA* triggers the host's 'Save-Our-Ship' (SOS) stress response, and in turn leads to upregulation of RecA precursors (Erill et al., 2007). Once a threshold of RecA* accumulates, prophage induction is also triggered. These processes can be stimulated with DNA-damaging treatments such as UV light or Mitomycin C (MMC). When lambdoid lysogens are grown in the absence of overt DNA-damaging stimuli, 'spontaneous' prophage induction can still occur at a moderate to low frequency (Nanda et al., 2015). However, even spontaneous prophage induction appears to be largely dependent on RecA*, because its frequency becomes almost undetectably low in host backgrounds that have lost their SOS-responsiveness (Fuchs et al., 1999; Little and Michalowski, 2010; Nanda et al., 2015).

Whereas superinfection of a lambdoid lysogen by identical or closely related phages is prevented if the prophage utilizes the same repressor-mediated immunity module, superinfection by divergent phages that do not share the immunity module—known as 'heteroimmune' phages—can still occur. During infection by a heteroimmune phage, a lysogenic host can become lysogenized with more than one prophage, or 'polylysogenized'. In these cases, induction of more than one prophage can still be

coordinated via the SOS response, but intracellular competition and/or cross-talk within the host may result in asymmetric propagation of one prophage over another (Refardt, 2011).

Among certain bacteria with pathogenic lifestyles, such as *Staphylococcus aureus*, polylysogeny is a common phenomenon (Brüssow et al., 2004). For example, one screen of 291 *S. aureus* clinical isolates based on site-specific integrase typing determined that most (87%) harbored one or more prophage, and about half (51%) contained two or more (Goerke et al., 2009). Similar to lambda and other lambdoid phages, the temperate phages of *S. aureus* have typically been identified as Siphoviridae with genome sizes of ~39-43 kbp (Xia and Wolz, 2014). Many of *S. aureus*'s prophages are predicted or confirmed to contribute to their host's pathogenicity (Xia and Wolz, 2014). In the next section, I elaborate on some examples relevant to this work.

1.4 Examples of temperate phages that contribute to staphylococcal pathogenicity

In cases where lysogenization is associated with a phenotype independent of repressor-mediated immunity, the phenomenon can be referred to as 'lysogenic conversion' (Bondy-Denomy and Davidson, 2014). This was first described for the ability to produce diphtheria toxin conferred by lysogenization with certain phages (Groman, 1955). Numerous other examples of phages (and in particular prophages) contributing to

bacterial pathogenicity have been documented in the literature, and are reviewed elsewhere (Brüssow et al., 2004; Canchaya et al., 2003; Casjens, 2003). Lysogenization by integrative temperate phages results in two principle genomic alterations: disruption of the integration site locus, and incorporation of a heterologous stretch of DNA—the prophage. Either alteration can result in lysogenic conversion phenotypes (Bondy-Denomy and Davidson, 2014). These phenotypes are highly homogeneous in cases where the temperate phage reliably integrates into a specific locus. Likewise, the prophage itself may encode functions that reliably confer a phenotype. For example, temperate phages can harbor seemingly extraneous segments of DNA that are not required for their lytic or lysogenic functions, known as 'morons' (Brüssow et al., 2004; Cumby et al., 2012; Juhala et al., 2000). Moron DNA may include toxin or virulence genes, and even promoter elements that allow for their efficient expression during lysogeny. In certain cases, processes which occur during prophage induction may also contribute to lysogenic conversion phenotypes (Tyler et al., 2013). For example, toxin expression can be enhanced via coordinated upregulation of phage late promoters during the lytic cycle (Wagner et al., 2002). In addition, it has been shown that lysis can allow for efficient release of toxins which accumulate in the periplasm of gram-negative bacteria (Shimizu et al., 2009). In the following two subsections, I outline some phages and phage-like elements that are proposed or confirmed to contribute to staphylococcal pathogenicity in various ways.

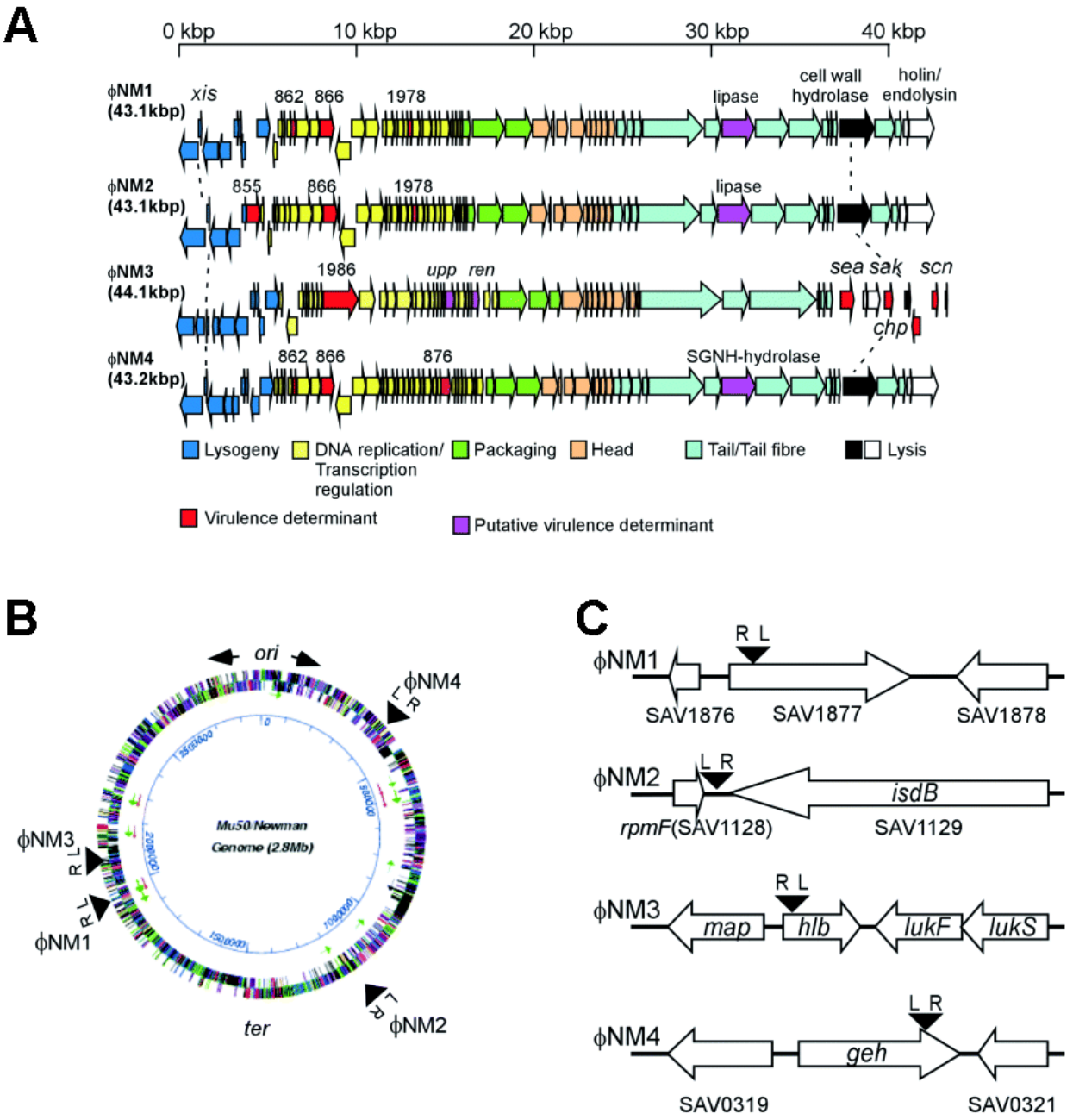
1.4.1 *β-hemolysin-converting phages*

Lysogenization with the β-hemolysin-converting phages of *Staphylococci*, also known as Sa3int phages, is found to simultaneously disrupt the β-hemolysin (*hlb*) gene harboring their integration site and also provide morons—including a cluster within the 3′ end of their genomes (Coleman et al., 1989; van Wamel et al., 2006; Verkaik et al., 2011; Xia and Wolz, 2014). This 'immune evasion cluster' (IEC) can encode for several secreted proteins: staphylococcal enterotoxin A (SEA) (Dohlsten et al., 1993), staphylokinase (SAK) (Jin et al., 2004), chemotaxis inhibitory protein of *S. aureus* (CHIPS) (de Haas et al., 2004), and staphylococcal complement inhibitor (SCIN) (Rooijakkers et al., 2005). Some of these ORFs were found to be upregulated at the transcriptional level in various strains, following prophage-inducing treatments with Mitomycin C (Cao et al., 2012; Goerke et al., 2006; Sumby and Waldor, 2003). In addition, such treatments were found to increase the extracellular concentrations of SEA released by certain strains (Cao et al., 2012). However, it remains to be clarified whether the process of prophage-mediated lysis itself can contribute to the release of functional IEC proteins (or other secreted proteins) from *S. aureus* lysogens. Given that canonical processing of signal peptides requires secretion across the plasma membrane (Schneewind and Missiakas, 2012), it's possible that rapid cell lysis via induction could even diminish the release of functional secreted proteins by disrupting the integrity of plasma membranes prematurely. Whereas the ϕ13 phage of NCTC8325 appears to be fully functional (Goerke et al., 2006;

Iandolo et al., 2002; van Wamel et al., 2006), other β-hemolysin-converting prophages appear to be at least partially defective (Bae et al., 2006; Holt et al., 2011; Sumby and Waldor, 2003; van Wamel et al., 2006). For instance, *S. aureus* Newman-derived single lysogens harboring the ϕNM3 prophage ('TB3') do not produce infective particles (Bae et al., 2006). *S. aureus* Newman has been studied as a model clinical isolate in numerous works, and its pathogenicity has also been investigated in animal models (Baba et al., 2008). In addition to its ϕNM3 prophage, Newman harbors three other prophages: ϕNM1, ϕNM2, and ϕNM4 (**Figure 1-6A**). All four prophages integrate site-specifically at distinct loci (**Figure 1-6B & C**). Unlike the ϕNM3 prophage, ϕNM1, ϕNM2, and ϕNM4 represent functional temperate phages that can form infective particles (Bae et al., 2006). They are also '*pac*' phages that utilize a headful packaging mechanism, and are therefore capable of generalized transduction (Chen et al., 2015a; Chen and Novick, 2009; Chen et al., 2015b). When a *S. aureus* Newman transposon insertion screen was used to look for virulence genes required in a nematode killing assay, one candidate ORF was identified in ϕNM3 (separate from the IEC), and five were identified among Newman's three *pac* prophages (Bae et al., 2004). The mechanistic basis for this remains unclear to me. No precise molecular function was ascribed to any of the candidate ORFs; and, all three of Newman's *pac* prophages were found to be dispensable for pathogenesis in a mouse liver abscess model (Bae et al., 2006). However, various *pac*

Figure 1-6. Summary of *S. aureus* Newman's four prophages (ϕNM1, ϕNM2, ϕNM3, and ϕNM4) and their integration sites within the host chromosome.

(A) Schematic summary of the predicted ORF (thick arrow) positions and orientations in each prophage genome. ORFs are color-coded according to tentatively assigned functions (listed at the bottom). The *sea*, *sak*, *chp*, and *scn* ORFs at the 3' end of ϕNM3 constitute its immune evasion cluster. Other virulence determinants (red) were identified in a transposon insertion screen for mutants defective in nematode killing, and are numbered according to locus tags (SAV numbers) of *S. aureus* Mu50 prophage homologues originally used as a reference (prior to the complete sequencing of Newman). Dashed lines indicate that ϕNM3's cell wall hydrolase ORF is truncated, and that its *xis* ORF is missing, although these are intact within the other prophages. **(B)** Summary of the prophage positions (black triangles) along *S. aureus* Newman's chromosome; the Mu50 genome was used as a reference. 'L' and 'R' denote the relative position of *attL* and *attR* attachment arms at each prophage (and thereby denote the relative orientation of prophage insertions at each locus). Positions of the chromosomal origin of replication (*ori*) and replication termination site (*ter*) are also indicated, with oppositely oriented black arrows used to denote bi-directional DNA synthesis. **(C)** Schematic summary as in 'B', but zoomed in to show relevant genomic regions in Newman that contain the integration sites for each of its prophages. Where a predicted function could not be assigned, ORFs were labeled according to their SAV numbers from the Mu50 annotation originally used as a reference. *hlb*, encodes β-haemolysin; *geh*, encodes glycerol ester hydrolase. Figure elements were reprinted from (Bae et al., 2006) with permission.


A
0 kbp
10 kbp
20 kbp
30 kbp
40 kbp
ϕNM1 (43.1kbp)
xis
862
866
1978
lipase
cell wall hydrolase
holin/ endolysin
ϕNM2 (43.1kbp)
855
866
1978
lipase
ϕNM3 (44.1kbp)
1986
upp
ren
sea
sak
scn
chp
ϕNM4 (43.2kbp)
862
866
876
SGNH-hydrolase
Lysogeny
DNA replication/ Transcription regulation
Packaging
Head
Tail/Tail fibre
Lysis
Virulence determinant
Putative virulence determinant
B
ori
ϕNM4
ϕNM3
ϕNM1
ϕNM2
ter
Mu50/Newman Genome (2.8Mb)
C
ϕNM1
SAV1876
SAV1877
SAV1878
ϕNM2
isdB
rpmF(SAV1128)
SAV1129
ϕNM3
map
hlb
lukF
lukS
ϕNM4
geh
SAV0319
SAV0321

phages of *S. aureus*— including Newman's—have been found to mobilize *S. aureus* Pathogenicity Islands (SaPIs) (Chen and Novick, 2009; Chen et al., 2015b; Dearborn and Dokland, 2012; Penades et al., 2015). SaPIs are a natural repository for putative and confirmed superantigen toxin genes (Penades et al., 2015; Sloane et al., 1991). Therefore, even when *pac* phages do not carry genes that contribute to pathogenesis directly, they can indirectly contribute to pathogenicity by functioning as 'helper' phages that spread SaPIs.

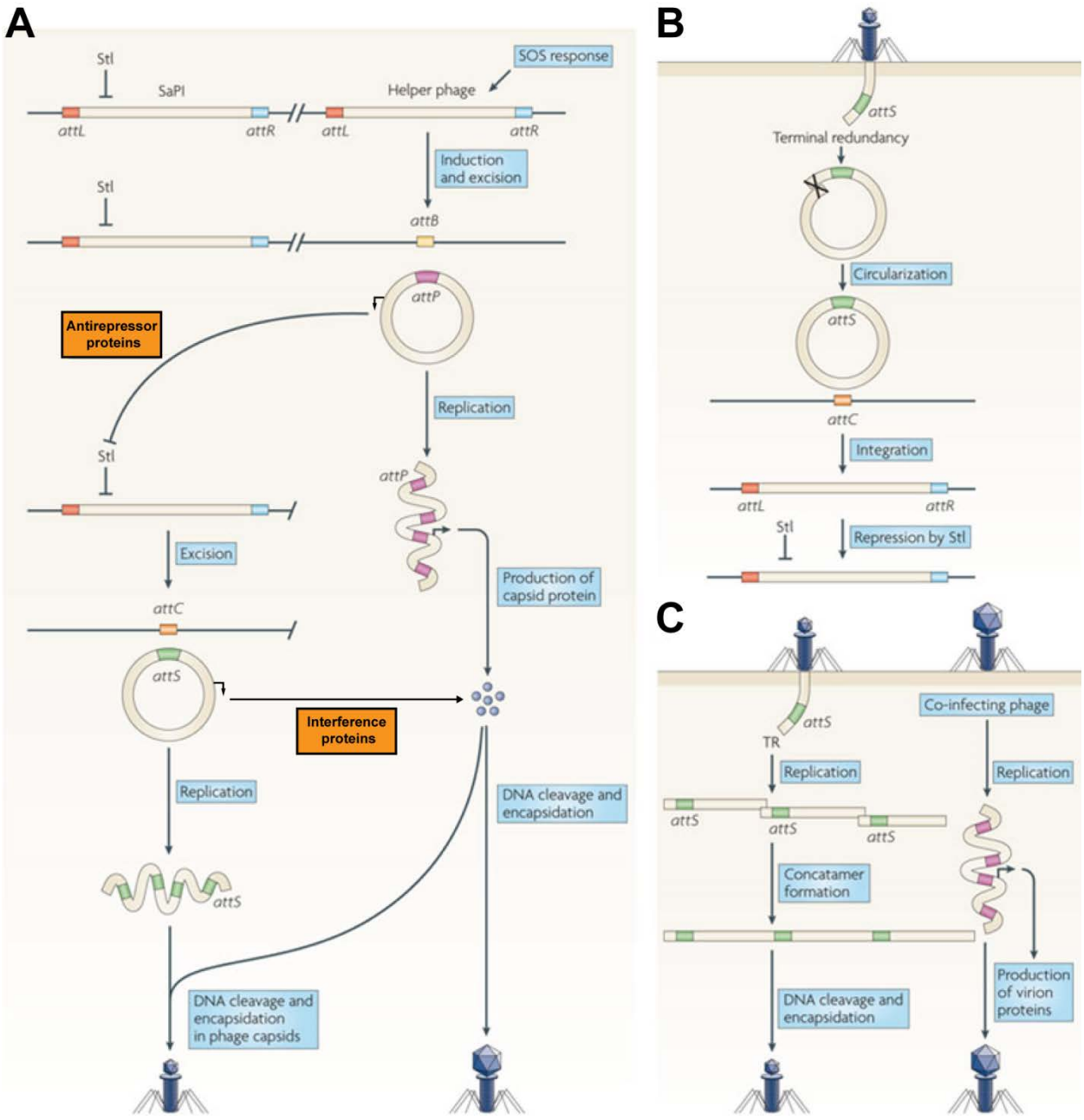
1.4.2 *SaPIs and their SaPI-mobilizing 'helper' phages*

SaPI helper phages include many of the well-studied *S. aureus* phages, such as ϕ11 and 80α (Xia and Wolz, 2014). Whereas 80α is thought to represent a laboratory recombinant (Christie et al., 2010; Lindsay et al., 1998), ϕ11 is derived from a prophage of the NCTC8325 isolate (Iandolo et al., 2002; Novick, 1967). Similar to phage lambda, both prophages encode *cI*-like repressors, and can be induced by treatment with DNA-damaging agents (Ferrer et al., 2011; Penades et al., 2015; Quiles-Puchalt et al., 2013). The ^{E}LCR of ϕ11 was also confirmed to include binding sites for its CI-like repressor *in vitro* (Ganguly et al., 2009). When a SaPI-containing strain is lysogenized with a helper phage for its SaPI, the SaPI genome can be excised, replicated, and packaged during prophage induction. In a sense, this phenomenon represents a SaPI-dependent

lysogenic conversion phenotype. Below, I provide a coarse-grained summary of how SaPIs interact with—and apparently parasitize—their helper phages.

SaPIs are small prophage-like elements (~15 kbp) found in the genomes of many *S. aureus* isolates (Novick et al., 2010; Penades et al., 2015). Their mobility is regulated by master repressor 'Stl' proteins (**Figure 1-7**), similar to how the lytic activity of lambdoid temperate phages is regulated by CI-like repressors (Johnson et al., 1981). Unlike phages, however, SaPIs do not encode the full complement of genes required for infection, and therefore rely on helper phages for packaging and transmission. Moreover, whereas the lytic cycle of many bona-fide prophages can be induced in response to host signals alone (Little and Michalowski, 2010; Nanda et al., 2015; Refardt, 2011), induction of SaPIs requires interactions with helper-phage-encoded antirepressor proteins (Penades et al., 2015; Tormo-Mas et al., 2010). Helper phages express SaPI antirepressors during their lytic cycle, such that SaPI induction only occurs during infection or prophage induction. Various SaPI antirepressor proteins have been described, such as the Dut (dUTPase) and Sri proteins found in 80α, ϕNM1, and ϕNM2 (Dearborn and Dokland, 2012; Tormo-Mas et al., 2010). Once a SaPI is de-repressed, it expresses proteins that allow it to excise from the chromosome, replicate, and hijack its helper phage's packaging machinery (Figure 1-7A). Hijacking is mediated by SaPI-encoded 'interference' proteins (Ram et al., 2012; Ram et al., 2014) which ensure

Figure 1-7. Summary of helper-phage-mediated SaPI mobilization and transfer. **(A)** Schematic summary of SaPI mobilization in lysogens during induction of a helper prophage. SaPI induction is normally repressed by its master repressor, 'Stl'. Once the lytic cycle of a cohabiting helper prophage is induced (e.g., via the host's SOS-response (Erill et al., 2007; Nanda et al., 2015)), phage-encoded antirepressor proteins are expressed, and relieve repression by Stl. SaPI excision and replication ensues, while expression of SaPI-encoded interference proteins ultimately re-directs the phage's packaging machinery to concatemeric SaPI genomes for encapsidation. Packaging of phages can still occur, albeit inefficiently. In some cases, SaPI interference involves proteins that modify phage capsids such that they are ~1/3 the size and therefore too small to package complete phage genomes (Christie and Dokland, 2012). This too can reduce the maximal yield of infective (phage) particles. **(B)** Schematic summary of SaPI transfer via modified phage particles, which are large enough to encapsidate a typical complete SaPI genome. Following injection, the SaPI can be site-specifically integrated into the host chromosome via *attS* (SaPI) and *attC* (chromosomal) attachment sites, where it is maintained by Stl-mediated repression. **(C)** Same as in 'A', except that SaPI mobilization is licensed by an infecting helper phage; note that this may occur before Stl-mediated repression of an injected SaPI was ever established. Figure adapted from (Novick et al., 2010) with permission.


A
Stl
SaPI
attL
attR
attL
attR
SOS response
Helper phage
Induction and excision
Stl
attB
attP
Antirepressor proteins
Replication
Stl
attP
Excision
Production of capsid protein
attC
attS
Interference proteins
Replication
DNA cleavage and encapsidation
attS
DNA cleavage and encapsidation in phage capsids
B
attS
Terminal redundancy
Circularization
attS
attC
Integration
attL
attR
Stl
Repression by Stl
C
attS
TR
Co-infecting phage
Replication
Replication
attS
attS
attS
Concatamer formation
DNA cleavage and encapsidation
Production of virion proteins

preferential packaging of the SaPI in phage or modified phage particles (Christie and Dokland, 2012) that are ultimately released upon completion of the helper phage's lytic cycle. These include proteins which prevent the phage's small terminase (TerSϕ) subunit from recognizing of its own genome during packaging, and thereby allow the SaPI's alternative small terminase (TerS$_{SP}$) subunit to re-direct capsids to the SaPI genome (Chen et al., 2015b; Penades et al., 2015). Upon transfer into a new host that lacks active helper phages, SaPIs may once again be established within the chromosome via site-specific integration and Stl-mediated repression (**Figure 1-7B**).

By impeding the packaging of helper phage genomes, SaPI-encoded interference proteins can substantially reduce the helper phage's infective burst size (Lindsay et al., 1998). Importantly, this effect can provide the SaPI-containing host with a survival advantage by preventing plaque formation. Functional SaPIs, therefore, offer a highly selective form of defense against (helper) phages, but operate more like abortive infection systems (Chopin et al., 2005; Depardieu et al., 2016; Labrie et al., 2010) than CRISPR-Cas systems in the sense that they do not rescue the viability of infected cells. Meanwhile, by allowing for lysis during infection or prophage induction, SaPIs permit both the transfer of host genes via generalized transduction pathways and the efficient transfer of their own genomes (Chen et al., 2015b). Accordingly, it was recently proposed that this feature of SaPIs could offer a unique advantage over CRISPR-Cas systems, and perhaps explain the relative abundance of SaPIs (and scarcity of CRISPR-

Cas systems) in *S. aureus* populations which are thought to benefit heavily from frequent horizontal gene transfer (Ram et al., 2014). This idea is compelling, but seems to disregard the contributions of helper-phage-mediated mobilization and transfer of SaPIs, which might be sufficient to explain their relative abundance in temperate-phage-rich populations of *S. aureus*. Furthermore, while type III-A CRISPR-Cas systems in *S. aureus* genomes display signatures of recent horizontal acquisition (Cao et al., 2016; Golding et al., 2010; Kinnevey et al., 2013), intrinsic mobility has yet to be demonstrated for any extant CRISPR-Cas system (Beguin et al., 2016; Krupovic et al., 2014). Hence, the relative immobility of extant CRISPR-Cas systems could also explain their scarcity among *S. aureus* genomes. A recent PCR-based screen of 636 *S. aureus* clinical isolates derived from four Chinese hospitals determined that only 6 isolates contained type III-A CRISPR-Cas systems (Cao et al., 2016). Despite this scarcity, type III-A systems are relatively well-represented among *Staphylococci* compared to other CRISPR-Cas systems, and my advisor Dr. Marraffini first noticed this while browsing the available genomes and CRISPR database (Grissa et al., 2007a, b). In the next section, I summarize some additional developments that led us to embark on our study of type III-A systems in *S. aureus* hosts, about five and a half years ago.

1.5 Rationale for the study of type III-A CRISPR-Cas systems in *S. aureus* hosts

Although the first experimental demonstration of a CRISPR-Cas system's adaptive immune functionality was carried out with *Streptococcus thermophilus* and virulent phages, prior *in silico* analyses of these systems and their CRISPR spacer content (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005) seemed to suggest that they would also be capable of defending against other classes of mobile genetic elements, such as plasmids. A seminal study led by Dr. Marraffini subsequently confirmed this prediction, in showing that a CRISPR-Cas system in *Staphylococcus epidermidis* RP62a (Gill et al., 2005) can target DNA (plasmid) elements, and thereby prevent horizontal gene transfer (Marraffini and Sontheimer, 2008). *Staphylococci* are thought to engage in frequent horizontal transfer of mobile genetic elements and accessory genes (Lindsay and Holden, 2004), and in this light the implications of Dr. Marraffini's findings were particularly intriguing. Apparently, by preventing horizontal gene transfer, CRISPR-Cas immunity could be *detrimental* to bacteria in certain circumstances. Negative selection operating on CRISPR-Cas systems that target favorable elements might therefore explain their scarcity among *S. aureus* genomes, and perhaps even explain their absence from about 50% of bacterial genomes overall (Grissa et al., 2007a, b), and Dr. Marraffini conveyed this intriguing notion to me once I'd begun work in his laboratory. Dr. Marraffini also pointed out that spacers matching to temperate phages could be identified among the CRISPR loci of type III-A systems in sequenced

Staphylococci, suggesting these systems do indeed interface with temperate phages in the wild, and presumably protect their hosts from them. For example, the second spacer in RP62a's type III-A system matches to the CNPH82 temperate phage of *S. epidermidis* (Daniel et al., 2007). Moreover, whereas plasmid or non-phage integrative elements in *S. aureus* were often found to confer antibiotic resistance, temperate phage or phage-like elements in *S. aureus* were often found to confer putative or confirmed virulence traits (Lindsay and Holden, 2006). Targeting of temperate phages by type III-A systems in *S. aureus*, therefore, might be expected to present a barrier to virulence acquisition, in the same sense that antibiotic resistance acquisition could be prevented by plasmid targeting. Consistent with this idea, CRISPR-Cas targeting of a fully functional temperate phage (lambda) had been tested in one study, and it was found to resist lysogenic infections (Edgar and Qimron, 2010). In two other studies, resistance to lytic infection by a virulent mutant of lambda (λ_{vir}) had also been shown (Brouns et al., 2008; Sapranauskas et al., 2011). The time was ripe for testing whether CRISPR-Cas targeting of temperate phages in *S. aureus* would license a similar state of genetic incompatibility.

There were practical reasons for working with type III-A systems in *S. aureus* as well. In December 2011, my colleague Wenyan showed that the type III-A system from *S. epidermidis* RP62a could license anti-conjugative-plasmid immunity in *S. aureus* hosts, once cloned on an autonomously replicating plasmid vector. This provided a proof of principle that targeting by a type III-A system could operate as expected in

heterologous staphylococcal hosts, and allowed us to readily modify the system's CRISPR array and *cas* genes via standard plasmid cloning manipulations. Modification of *S. epidermidis* RP62a and its derivatives, meanwhile, is less trivial, and we were initially lacking a phage that could infect these strains. By contrast, a variety of *S. aureus* phages were available to us, and Dr. Marraffini recommended that I begin work with the three functional temperate phages of *S. aureus* Newman: ϕNM1, ϕNM2, and ϕNM4.

CHAPTER 2: INVESTIGATION OF CRISPR-CAS IMMUNITY IN *S. AUREUS* REVEALS AN UNEXPECTED MECHANISM FOR CONDITIONAL TEMPERATE PHAGE TOLERANCE

When this work was begun, the type I-E CRISPR-Cas system's potential to resist both lytic and lysogenic infections by phage lambda or its derivatives had been demonstrated in *Escherichia coli* (Brouns et al., 2008; Edgar and Qimron, 2010; Sapranauskas et al., 2011), but little else was known about CRISPR-Cas immunity to temperate phages in bacteria. Spacers matching to temperate phages had been identified in various CRISPR-Cas systems, including type III-A systems of *Staphylococci* (Gill et al., 2005; Golding et al., 2010). However, a phage-defense function for CRISPR-Cas immunity in staphylococcal hosts had yet to be demonstrated experimentally. Enabled by my colleague Wenyan's recent cloning of the *S. epidermidis* RP62a type III-A system on a plasmid vector, I set out to characterize the genetic outcomes of temperate phage targeting by type III-A systems in *S. aureus.* The plan involved reverse-engineering of our type III-A system's upstream CRISPR locus, such that it would contain spacers matching to the functional temperate phages of *S. aureus* Newman (ϕNM1, ϕNM2, and ϕNM4). The spacer in position 1 of this CRISPR locus had already been shown to license immunity to plasmids that bore a matching *nickase* (*nes*) target sequence, so it was conceivable that temperate-phage-matching spacers, once engineered, would also be able to license anti-phage immunity. By working with three

heteroimmune temperate phages, moreover, I could presumably emulate the effects of CRISPR-Cas targeting in natural *S. aureus* populations, where infection by multiple (pro)phages is not uncommon (Brüssow et al., 2004; Goerke et al., 2009). Most of the phage- and prophage-targeting experiments described in sections **2.1**, **2.2**, and **2.3** were initially carried out with the 'TB4' host or its derivatives (Bae et al., 2006). TB4 is a prophage-cured derivative of *S. aureus* Newman that can be re-infected with ϕNM1, ϕNM2, or ϕNM4, and thereby serves as a native host background for these phages. This strain and its lysogenic derivatives lack endogenous CRISPR-Cas systems, but can be readily transformed with plasmids miniprepped from *S. aureus* RN4220—a restriction-deficient and prophage-cured laboratory derivative of NCTC8325 which also lacks an endogenous CRISPR-Cas system (Nair et al., 2011). Once it was clear that RN4220 could be used to reproduce the core findings directly, however, I focused on conducting experiments with this more genetically tractable host. The majority of the data presented in **Chapter 2**—and ultimately included for publication (Goldberg et al., 2014)—was therefore derived from experiments with RN4220.

2.1 Initial demonstration of CRISPR-Cas immunity to phages in *S. aureus*

My work on this project began with the isolation of ϕNM1, ϕNM2, and ϕNM4 phage stocks. Supernatants from overnight cultures of *S. aureus* Newman contain a mixture of spontaneously induced infective particles derived predominantly from ϕNM1 and

ϕNM2 (Bae et al., 2006). Single plaques were isolated by plating the supernatant mixture on a sensitive soft agar lawn, and then genotyping plaques with primers specific for each phage. This allowed for the direct identification of ϕNM1 and ϕNM2 plaques, which were then propagated to produce high-titer lysates. Temperate phages such as these do not produce completely clear soft agar lawns upon lysis. Rather, lawns become 'turbid' because a high proportion of cells are lysogenized during initial infections, which then survive subsequent infections via prophage repressor-mediated immunity. In turn, lysogens can be isolated by re-streaking from a region of turbid lysis. I obtained TB4::ϕNM1 and TB4::ϕNM2 single lysogens in this manner, which were genotyped using my original phage-specific primers. A TB4::ϕNM1+ϕNM2 double lysogen was subsequently constructed by infecting TB4::ϕNM1 with ϕNM2, and re-streaking from a region of turbid lysis in the same way. This double lysogen allowed me to identify ϕNM4 plaques from *S. aureus* Newman supernatants, because it counter-selects against ϕNM1 and ϕNM2 phages. Finally, a TB4::ϕNM4 single lysogen was similarly obtained during propagation of ϕNM4 on sensitive lawns.

Shortly after I began work on the phage isolations, I started working to re-engineer the CRISPR locus of Wenyan's type III-A CRISPR-Cas plasmid, which he called pWJ30β. In order to readily replace the native phage-targeting spacer in position 2 of this plasmid with engineered spacers targeting *S. aureus* Newman's phages, I

needed to create a new plasmid lacking all repeats and spacers downstream of its *nes* spacer in position 1. This new plasmid, pGG3 (**Figure 2-1**), served as the parent

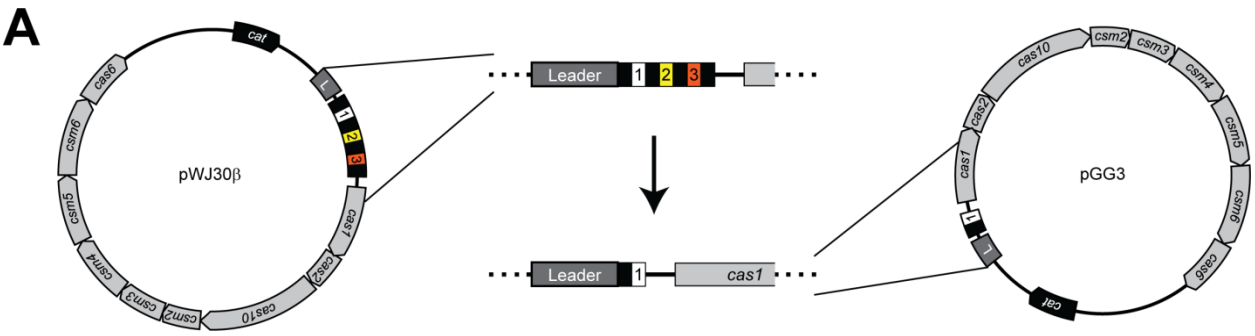


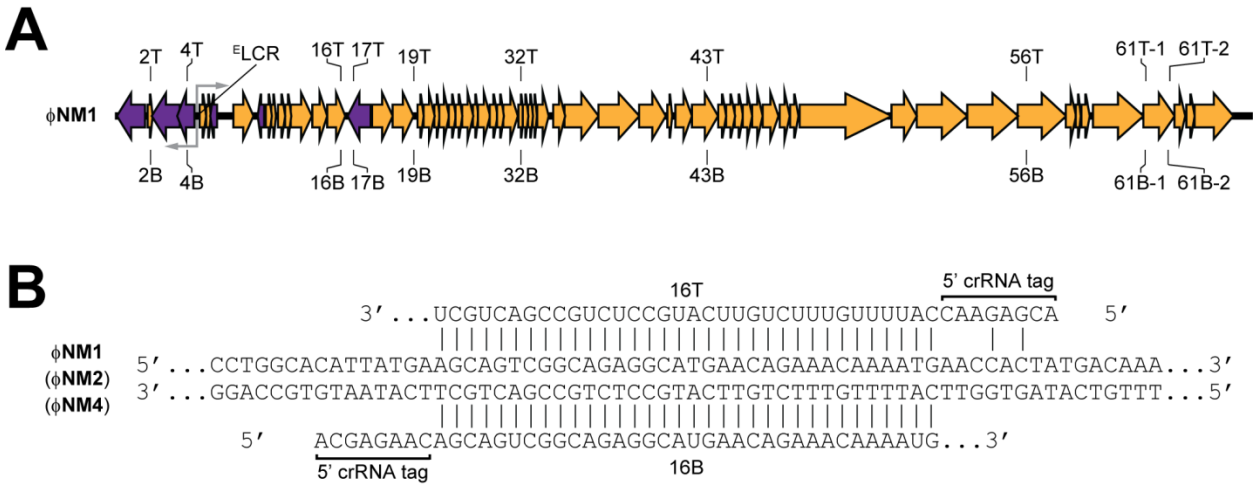
Figure 2-1. Plasmid-borne type III-A CRISPR-Cas systems for initial experiments in *S. aureus*.

(A) Schematic representations of the pWJ30β (left) and pGG3 (right) type III-A CRISPR-Cas plasmids. pWJ30β harbors the wild type system from *S. epidermidis* RP62a, including its conserved leader sequence ('L', dark grey box) immediately upstream of the CRISPR array. To construct pGG3, the CRISPR array of pWJ30β was reduced to a single repeat-spacer element (middle). pC194 backbones include a chloramphenicol acetyltransferase (*cat*) gene conferring resistance to chloramphenicol, which is also depicted alongside *cas* genes in the plasmid diagrams.

vector for my initial phage-targeting constructs. It was immediately apparent that the pGG3 plasmid in TB4 did not prevent infection by the Newman phages. In fact, TB4 derivatives harboring pGG3 were arbitrarily used for propagation during some of my initial lysogen isolations (once it was cloned), probably because this allowed me to prophylactically treat the culturing media with an antibiotic (chloramphenicol). Accordingly, the effects of phage-targeting spacers could be subsequently tested

relative to this isogenic and phage-sensitive background. PCR-based methods using our standard oligo sizes (~ 60 bp) allowed for one-step addition of repeat-spacer units to the pGG3 backbone, with the phage-targeting spacer introduced at position 2 but lacking a downstream repeat. Dr. Marraffini pointed out to me that this should not present an issue, because single repeat-spacer units were found to be sufficient for anti-plasmid immunity in conjugation assays (Hatoum-Aslan et al., 2011). In addition, Dr. Marraffini's previous work with the type III-A system had demonstrated that excessive matching between the 5' tag of the crRNA and the corresponding flanking sequence in the target would prevent immunity (Marraffini and Sontheimer, 2010b), so I designed my spacers to avoid this, as per his recommendation. Beyond this, it was not known whether there would be additional requirements for type III targeting, such as the PAMs which had been described for type I and type II systems (Deveau et al., 2008; Semenova et al., 2011). It had been pointed out to me, furthermore, that previous reports seemed to indicate CRISPR-Cas systems could provide immunity to phages and plasmids regardless of the DNA strand or genomic context targeted (Barrangou et al., 2007; Deveau et al., 2008; Garneau et al., 2010; Marraffini and Sontheimer, 2008). I naïvely doubted that I would not encounter further constraints on the type III-A (DNA) targeting mechanism, even with spacers engineered to match their targets perfectly. In any event, I reasoned that I could satisfy my own skepticism by designing a few different spacers; if there really were no other targeting constraints, I should be able to

Figure 2-2. Summary of target positions and spacer design for type III CRISPR-Cas immunity directed at ϕNM1 (or other *S. aureus* Newman phages) in Chapter 2.
(A) Schematic representation of the ϕNM1 genome's rightward-oriented (golden) and leftward-oriented (purple) ORFs, scaled in accordance with their annotated lengths (Accession number: NC_008583.1). The target positions for various type III spacers with crRNAs complementary to the top ('T') or bottom ('B') strand of the genome are also indicated. Oppositely oriented bent grey arrows represent the P_{cI} (leftward) and P_{cro} (rightward) promoters found within the early lytic control region (^{E}LCR) designated for ϕNM1. **(B)** Base-pairing potential between the spacer 16B or 16T crRNAs and the DNA of perfectly conserved target regions in ϕNM1, ϕNM2, and ϕNM4, as indicated by vertical lines between the sequences. The 5′ crRNA tag is an 8 bp sequence derived from the CRISPR repeat that must be sufficiently mismatched with the target's flanking sequence in order to license immunity. **(C)** Plaquing efficiency of ϕNM1 on lawns of RN4220 harboring pGG3- or pGG3-BsaI-derived type III CRISPR-Cas plasmids with different phage-targeting spacers. High plaquing efficiency implies that the CRISPR-Cas plasmid did not provide immunity to infection; the functionality of these plasmid backbones was verified in at least one independent CRISPR-Cas targeting assay prior to publication (Goldberg et al., 2014), such as the pG0400 conjugation assay (**Figure 2.3**). Dotted line represents the limit of detection under these assay conditions. Error bars, mean ± s.d. (n = 3, technical replicates).

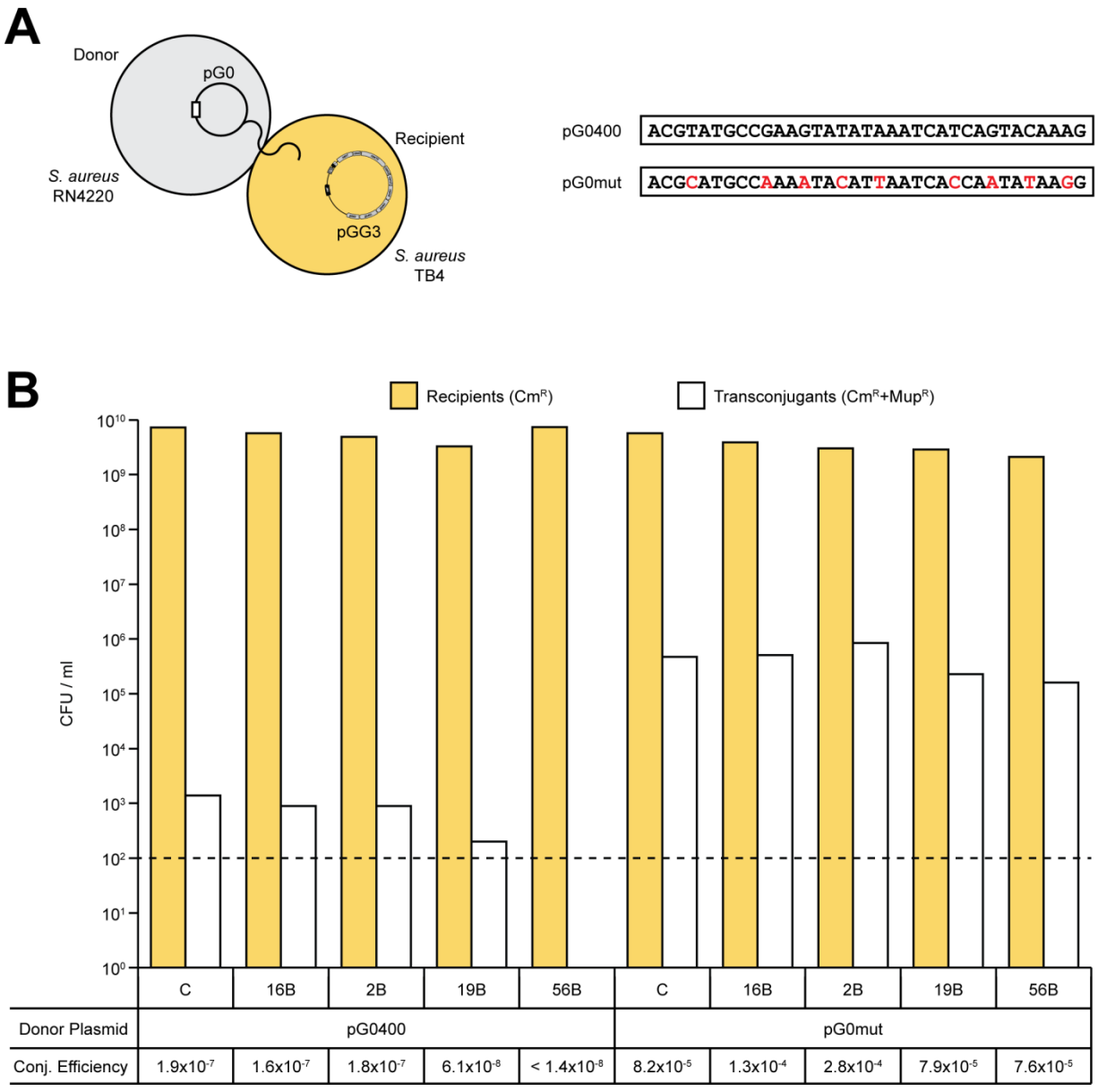

A
2T
4T
ELCR
16T
17T
19T
32T
43T
56T
61T-1
61T-2
ϕNM1
2B
4B
16B
17B
19B
32B
43B
56B
61B-1
61B-2
B
16T
5' crRNA tag
3'...UCGUCAGCCGUCUCCGUACUUGUCUUUGUUUUACCAAGAGCA 5'
ϕNM1
(ϕNM2)
(ϕNM4)
5'...CCTGGCACATTATGAAGCAGTCGGCAGAGGCATGAACAGAAACAAAATGAACCACTATGACAAA...3'
3'...GGACCGTGTAATACTTCGTCAGCCGTCTCCGTACTTGTCTTTGTTTTACTTGGTGATACTGTTT...5'
5' ACGAGAACAGCAGUCGGCAGAGGCAUGAACAGAAACAAAAUG...3'
5' crRNA tag
16B
C
Efficiency of plaquing (ϕNM1)
10-10
10-8
10-6
10-4
10-2
100
2B
2T
4T
4B
16B
16T
17B
17T
19B
19T
32B
32T
43B
43T
56B
56T
61B-1
61T-1
61B-2
61T-2

provide CRISPR-Cas immunity with virtually any spacer matching to a phage's genome. Moreover, in perhaps equally naïve fashion, I attempted to increase my odds of encountering such constraints by initially designing spacers that match to all three phages (ϕNM1, ϕNM2, and ϕNM4). Besides these considerations, the first spacer I successfully cloned, 16B, was arbitrarily designed with a sense sequence matching to one of the rightward-oriented ORFs (*gp16* in ϕNM1, numbered '866' in **Figure 1-6**) reported to encode a conserved nematode virulence determinant (Bae et al., 2006; Bae et al., 2004). When considering the rightward-oriented ORFs in these phages (e.g., those encoding various lytic genes), sense spacers produce crRNAs with complementarity to the 'bottom' strand, hence the 'B' (**Figures 2-2A & 2-2B**). The 16B spacer did not detectably reduce plaque formation by any of these phages. Representative data from experiments with ϕNM1 are shown in **Figure 2-2C**. I next designed three additional spacers (2B, 19B, and 56B) targeting different regions of the phage genomes, even though we expected that DNA targeting by CRISPR-Cas systems would provide anti-phage immunity regardless of the target's genomic context. For consistency, sense spacers were again designed to target the bottom strand of conserved sequences. However, I made a point to include a spacer targeting a gene in the lysogeny region, because I surmised that targeting of this region would be more likely to have distinct functional consequences. This was made possible by the well-conserved, rightward-oriented excisionase ORF (*gp2*) present in the lysogeny region of each phage. Although

the 19B and 56B spacers matching to lytic regions did not reduce plaque formation, the 2B spacer did. Again, representative data from experiments with RN4220 and ϕNM1 are shown in **Figure 2-2C**. Meanwhile, all three of these strains exhibited anti-plasmid immunity when I tested them as conjugative plasmid recipients in control assays (**Figure 2-3**), owing to the *nes* spacer retained at position 1 of their CRISPR. The conjugation efficiencies I observed are comparable to published values reported by our lab for pG0400 and pG0mut (Hatoum-Aslan et al., 2013; Jiang et al., 2013b; Marraffini and Sontheimer, 2008): efficiency is only reduced for the pG0400 plasmid with a wild type target for the *nes* spacer. The rare pG0400 transconjugants I observed were presumed to represent CRISPR-inactive genetic escapers, since, at this time, my colleagues Wenyan and Inbal were well under way with their characterization of similar transconjugants in *S. epidermidis* RP62a, which they found to be genetic escapers in all cases (Jiang et al., 2013b). Evidently, however, we still didn't fully understand the requirements for phage targeting by type III-A systems in plaque assays, because 3 of my first 4 spacers failed to license anti-phage immunity by otherwise functional CRISPR-Cas systems. Given that these particular phage-targeting spacers were all engineered, it initially seemed plausible to me that the non-functional spacers could be lacking in target recognition capacity, or were poorly expressed in our host background. I hadn't, for example, considered their predicted RNA-DNA hybridization potential when designing them, or assessed crRNA abundance. Moreover, a requirement for

Figure 2-3. Integrity of pGG3 backbones can be functionality validated using conjugation assays.

(A) Schematic diagram depicting the conjugative transfer of pG0(400 or mut) plasmids conferring mupirocin resistance from *S. aureus* RN4220 donors to TB4 recipients containing pGG3 or pGG3-derived CRISPR-Cas plasmids with the native *nes*-targeting spacer in position 1. The solid white rectangle on pG0 signifies the presence of a perfect or partially matching target for the *nes* spacer in pGG3. The wild type (pG0400) and mutated (pG0mut) target site sequences are shown on the right, with red lettering used for mutated bases. **(B)** Conjugation assays for validating the functionality of CRISPR-Cas plasmids conferring resistance to chloramphenicol. In this example dataset, recipients harbor type III CRISPR-Cas plasmids with one of the four indicated phage-targeting spacers, or the pGG3 parent vector (C). After filter mating with either pG0400 or pG0mut donors, recipient and transconjugant concentrations (CFU / ml) were quantified as described in the methods. Dotted line represents the limit of detection under these assay conditions. Conjugation efficiencies denote the ratio of transconjugants to recipients measured in each experiment.


A
Donor
pG0
Recipient
S. aureus
RN4220
pGG3
S. aureus
TB4
pG0400 ACGTATGCCGAAGTATATAAATCATCAGTACAAAG
pG0mut ACGCATGCCAAAATACATTAATCACCAATATAAGG
B
Recipients (CmR)
Transconjugants (CmR+MupR)
CFU / ml
10^10
10^9
10^8
10^7
10^6
10^5
10^4
10^3
10^2
10^1
10^0
C 16B 2B 19B 56B C 16B 2B 19B 56B
Donor Plasmid pG0400 pG0mut
Conj. Efficiency 1.9x10-7 1.6x10-7 1.8x10-7 6.1x10-8 < 1.4x10-8 8.2x10-5 1.3x10-4 2.8x10-4 7.9x10-5 7.6x10-5

previously unidentified target flanking motifs, such as PAMs, had still not been rigorously ruled out. In any case, it was easy for me to suspend judgment on this matter rather than attempting to pursue these ideas directly, since there were other experiments to be done. For example, Dr. Marraffini had pointed out that the rare plaques which arise on immune lawns with the 2B spacer are probably caused by escaper mutant phages (Barrangou et al., 2007; Deveau et al., 2008; Semenova et al., 2011), and he recommended that I isolate some of these phages to determine whether they had target site mutations. He also recommended that I begin by picking some big and some small plaques. Presumably, this could increase our odds of sampling different phages (i.e., with different plaque size phenotypes). These endeavors ultimately proved fruitful, as I explain below.

2.2 Initial characterization of phages and lysogenic hosts that escape type III-A CRISPR-Cas immunity via genetic means

Plaques picked from CRISPR-immune lawns can be re-plated on the same lawns to ensure additional counter-selection against residual wild type phage particles. In this manner, I initially isolated six ϕNM1 mutants, two ϕNM2 mutants, and six ϕNM4 mutants. My colleague Wenyan had used the Greek letter, β, to distinguish his pWJ30β plasmid from its pWJ30 predecessor, so I initially adopted a similar practice when

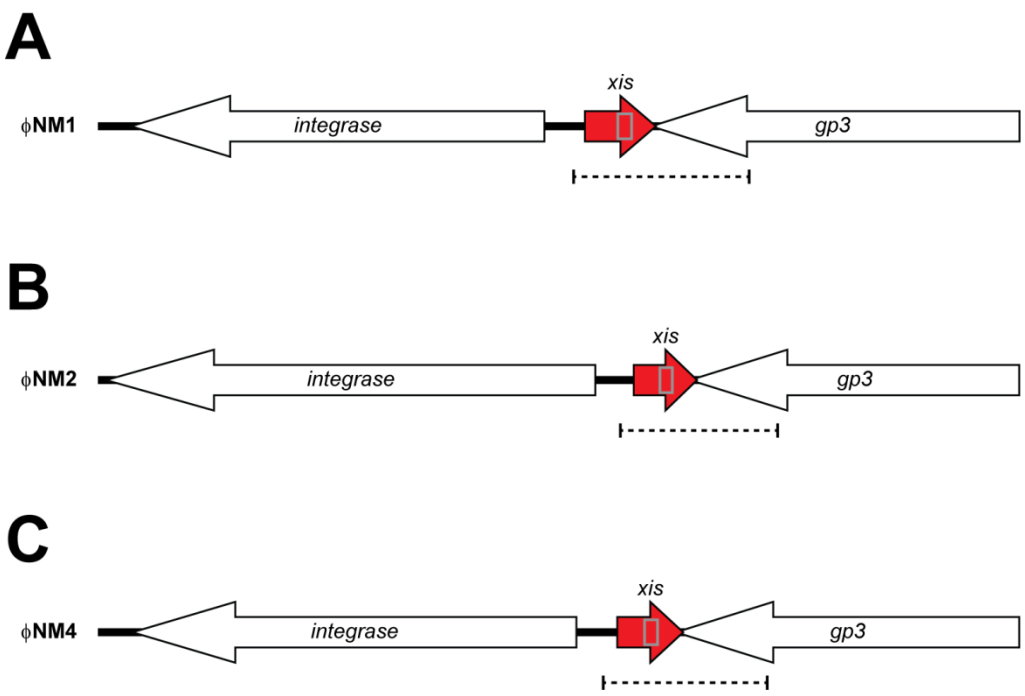


Figure 2-4. Preliminary genotyping of spacer 2B escaper phages (γ-series).
(A) Schematic diagram depicting a relevant genomic region of ϕNM1, scaled according to its reference annotation (Accession number: NC_008583.1). The *integrase* and *excisionase* (*xis*) ORFs correspond to *gp1* and *gp2*, respectively. Grey rectangle denotes the position of the *xis* target sequence. Dashed line denotes the location of target region PCRs initially attempted with primers oGG25 and oGG26. **(B)** Same as in 'A', except the schematic depicts a relevant genomic region of ϕNM2, scaled according to the ϕNM2 (NC_028913.1) reference annotation. **(C)** Same as in 'A', except the schematic depicts a relevant genomic region of ϕNM4, scaled according to the ϕNM4 (NC_028864.1) reference annotation.

naming my phage mutants. For example, the spacer 2B escapers were the first phages which I provided with a 'γ' designation, in order to help distinguish them from wild

type. To genotype these 'γ-series' mutants, I initially attempted to PCR-amplify small regions which include the conserved *xis* target for spacer 2B (**Figure 2-4**). After repeated attempts at this PCR, I was only successful at reproducibly amplifying from one of my ϕNM1 mutants, ϕNM1γ6, but Sanger sequencing confirmed that its amplicon lacked mutations (data not shown). Incidentally, this was the only one of my γ-series mutants which exhibited a small plaque phenotype when plating on spacer 2B lawns (**Figure 2-5A**). Importantly, this reduction in plaque size was not observed when control lawns were infected with ϕNM1γ6 (**Figure 2-5B**); and, a 'clear-plaque' phenotype was readily apparent on these lawns (**Figure 2-5C**). This indicated that, although its plaque size

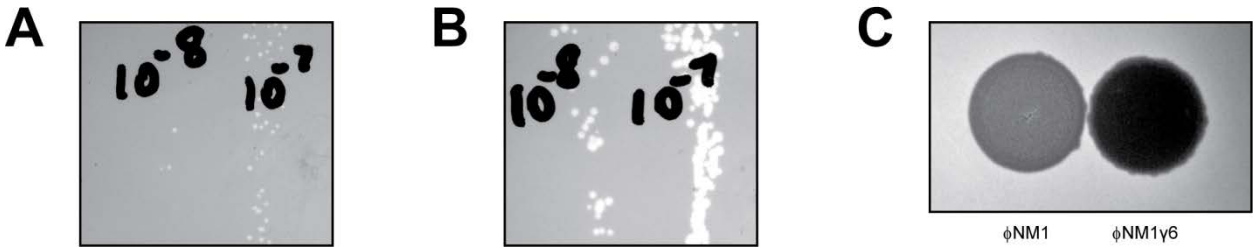


Figure 2-5. ϕNM1γ6 plaque phenotypes.
(A-B) Plaque formation by a ϕNM1γ6 lysate spotted on lawns of TB4 harboring the spacer 2B CRISPR-Cas system (A) or the pGG3 parent vector (B). Pictures are representative of at least two experiments performed under the same assay conditions. **(C)** Comparison of ϕNM1 (turbid) and ϕNM1γ6 (clear) phenotypes by spotting of high phage titers on a sensitive lawn. Picture is representative of at least four technical replicates.

phenotype was CRISPR-dependent, the ϕNM1γ6 phage did have heritable alterations—elsewhere—that could produce a clear-plaque phenotype even in the absence of

CRISPR-Cas targeting. In contrast, all of my other γ-series mutants escaped targeting with apparently wild type plaque sizes, regardless of whether they also exhibited clear-plaque phenotypes. I began to surmise that the small-plaque phenotype of ϕNM1γ6 could reflect incomplete escape of the 2B spacer, or partial targeting, enabled by the presence of its intact target sequence. The complete escape observed with the other mutants for whom an intact target sequence could not be amplified, meanwhile, might be explained by deletions at the target region which included one or more annealing site for my PCR primers. Consistent with this notion, I was generally able to PCR-amplify wild type amplicons when I used my original ϕNM1-, ϕNM2-, or ϕNM4-specific primer sets that did not encompass the target region.

At this stage, the source of the ϕNM1γ6 phenotypes was still quite puzzling. Regarding clear-plaque phenotypes, Dr. Bikard explained to me that this can arise if the phage is incapable of establishing lysogeny. Indeed, the few surviving colonies I was able to isolate after clearance of lawns with high titers of pure ϕNM1γ6 were never found to be lysogenized. These survivors were presumed to have acquired CRISPR-independent resistance mutations, since no expansion of CRISPR arrays was observed even when the infected lawn harbored a type III-A CRISPR-Cas plasmid (data not shown). Growth phenotypes were also observed in some cases while re-streaking these survivors in the absence of phage, which would not be expected for CRISPR-adapted or lysogenized lineages.

In a separate experiment, I characterized a few clones that survived infection by one of my other clear mutants, ϕNM4γ4. In this case, the infected lawn also harbored the spacer 2B CRISPR-Cas plasmid, which is fully escaped by the ϕNM4γ4 mutant. Although I was again unsuccessful at isolating type III-A CRISPR-adapted clones here, two of my isolates had apparently become lysogenized (**Figure 2-6A**), as determined by PCR-amplification with my original ϕNM4-specific primer set. One of these lysogenic survivors, ϕNM4γ4-S1, appeared to have an altered CRISPR region in its spacer 2B plasmid, given that I could not amplify its CRISPR array with my standard primers. The other survivor, ϕNM4γ4-S2, still had an intact CRISPR array amplicon. In both cases, the *xis* target region could be amplified, but the ϕNM4γ4-S2 isolate gave a shortened product (**Figure 2-6A**). Sanger sequencing revealed an 81 bp deletion, confined within the ORF, which included 23/35 base pairs of the original target sequence (**Figures 2-6B & C**). Aside from allowing CRISPR-escape, this deletion did not appear to produce a phenotype, since I could isolate plaques from lysogen supernatants, and use them to re-lysogenize TB4. Evidently, this mutant was distinct from ϕNM4γ4. Soon afterwards, I realized that the 'ϕNM4γ4' stock used for infection in this particular experiment had not been purified to homogeneity, because I was able to amplify faint, wild type products from the *xis* target region (indicative of contaminating, parental ϕNM4 phages). Therefore, these lysogenic survivors were probably derived from residual wild type phages in the lysate which ultimately

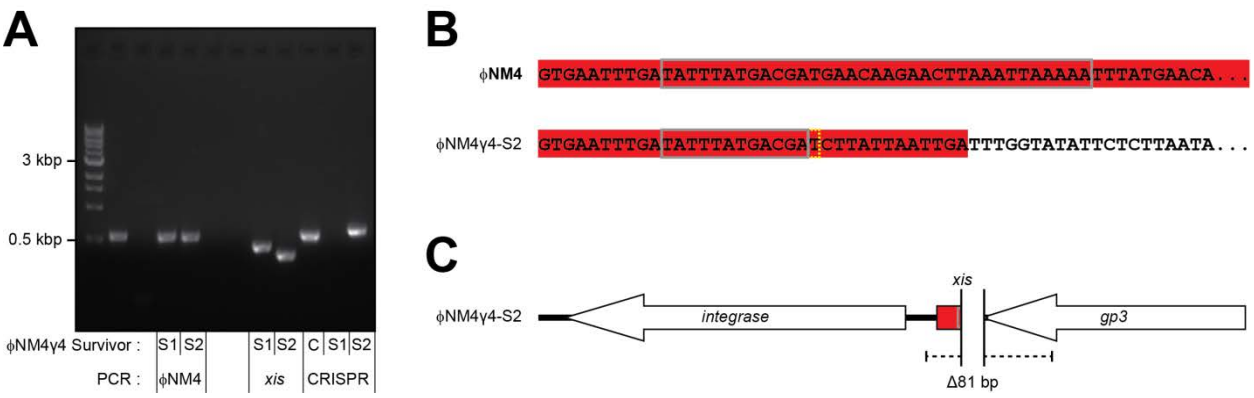


Figure 2-6. Characterization of two ϕNM4-derived lysogens that genetically escaped prophage targeting by spacer 2B.

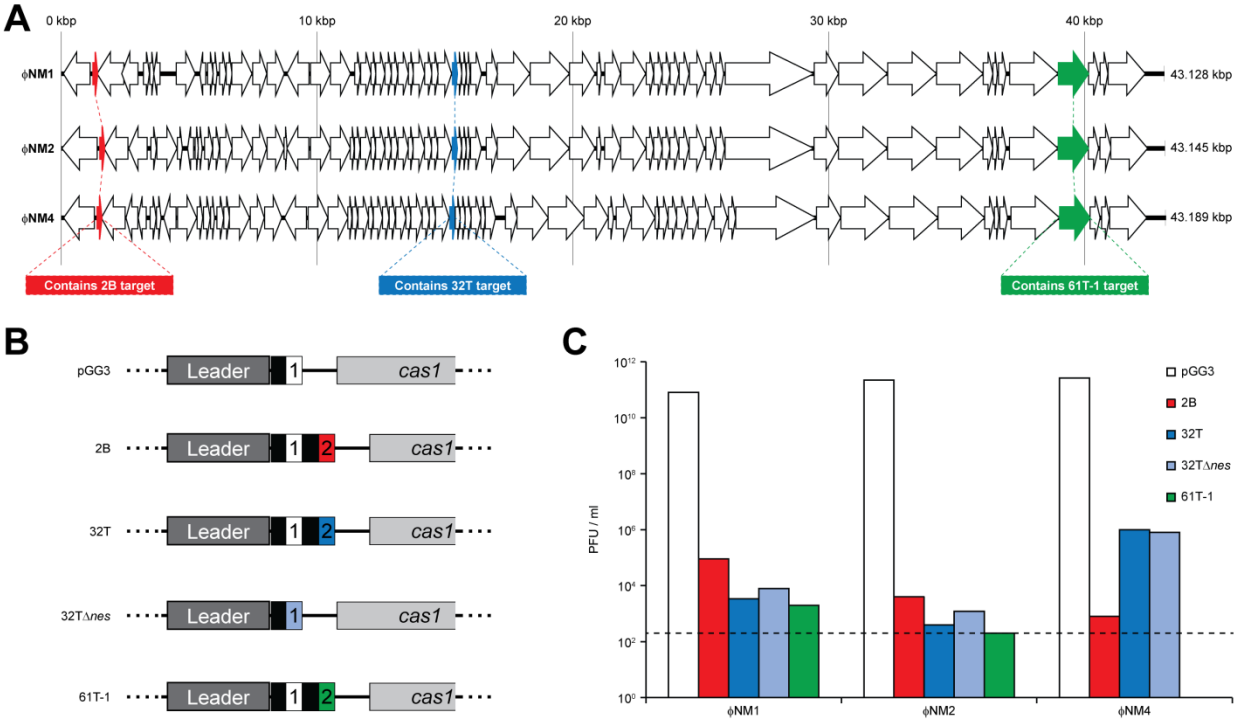
(A) PCR amplification from two clones that survived ϕNM4γ4 infection on TB4 lawns with the spacer 2B CRISPR-Cas plasmid ('S1' and 'S2'). After re-streaking, a single colony lysate from each isolate was used to template three different PCRs using primers for the *xis* target region (middle), primers for amplifying the CRISPR array (right), or primers for amplifying from ϕNM4 phages at a separate locus (left). Note the shorter size of the *xis* amplicon derived from the S2 isolate, and the lack of CRISPR array amplification from the S1 isolate. Control amplification (C) from the CRISPR array of the pGG3 parent vector lacking spacer 2B is shown in lane 10. Size markers of 3 kbp and 0.5 kbp in lane 1 are indicated. Additional replicates were not performed on the original isolates. Lanes 2, 3, 6 and 7 were used for other experiments. **(B)** Sequence of the *xis* target region in ϕNM4 and ϕNM4γ4-S2. Red highlighting denotes sequences from the *xis* ORF; target sequences matching the 2B spacer are boxed in grey. Note that while only 12 bp of the original target sequence remain in ϕNM4γ4-S2, the deletion regenerated a 'T' immediately downstream (dotted yellow box). **(C)** Schematic diagram as in **Figure 2-4C**, with the position of the ϕNM4γ4-S2 deletion indicated (Δ81 bp). PCR amplification with the oGG25/oGG26 primer set (dashed line) produces a shortened product size of 364 bp.

protected their hosts from superinfection by ϕNM4γ4 mutants. Indeed, none of my clear mutants were capable of superinfecting their respective lysogens, even when high titers were spotted on lysogen lawns around the time of initial isolation (data not shown). Notwithstanding, this result seemed to provide at least one example of how a type III-A CRISPR-Cas system could select for temperate phage mutants that completely evade immunity—even during lysogeny—via target sequence deletions.

This mutational, or genetic form of CRISPR-escape seemed to mirror what my colleagues had found in conjugation experiments (and also what had been reported previously on temperate phage targeting in *E. coli* (Edgar and Qimron, 2010)), where functional CRISPR-Cas systems do not co-exist with their DNA target elements. Consistent results were obtained in transformation assay pilot experiments where I electroporated the spacer 2B CRISPR-Cas plasmid into *S. aureus* Newman (which harbors 3 prophage targets). Although *S. aureus* Newman was comparably transformed with pGG3 and spacer 16B CRISPR-Cas plasmids (data not shown), only a single transformant was isolated from the two experiments where I attempted to introduce the spacer 2B plasmid. This transformant maintained an intact CRISPR array, but was presumed to have lost its CRISPR-Cas targeting functionality by other means (e.g., via mutation of one or more *cas* gene), in part because I was able to amplify from all four of its prophages using PCR. Accordingly, I saved the strain but proceeded with other experiments described below, under the assumption that this putative escaper would

Figure 2-7. Spacers matching to conserved sequences in ϕNM1, ϕNM2, and ϕNM4 can license type III-A CRISPR-Cas immunity to all three phages.

(A) Schematic representations of the ϕNM1, ϕNM2, and ϕNM4 genomes; ORF (thick arrow) lengths were scaled according to their respective NCBI annotations (NC_008583.1, NC_028913.1, and NC_028864.1). Colored ORFs contain a conserved or partially conserved target for spacer 2B (red), spacer 32T (blue), or spacer 61T-1 (green). **(B)** Schematic summary of the CRISPR arrays in CRISPR-Cas plasmids used for initial phage-targeting experiments in TB4 hosts. The 2B, 32T, and 61T-1 spacers are color-matched to their target ORFs shown in panel 'A'. CRISPR-Cas plasmids with a phage-targeting spacer in position 2 were derived from pGG3 as described in the methods. The spacer 32TΔ*nes* CRISPR-Cas plasmid is a derivative of the spacer 32T CRISPR-Cas plasmid, and was isolated by miniprepping from a pG0400 transconjugant that escaped mupirocin counter-selection via loss of the *nes* spacer in position 1. **(C)** Quantification of plaque-forming potential (PFU / ml) measured for lysates of ϕNM1, ϕNM2, or ϕNM4 plated on soft agar lawns of *S. aureus* TB4 harboring one of the five CRISPR-Cas plasmids summarized in 'B'. Lawns with the pGG3 parent vector plasmid lack phage-targeting spacers and thus serve as a sensitive control in this assay. Data represents single biological replicates; for each phage, the same lysate was used for plating on the five lawns. Dotted line represents the limit of detection under these assay conditions.


A
0 kbp
10 kbp
20 kbp
30 kbp
40 kbp
ϕNM1
ϕNM2
ϕNM4
43.128 kbp
43.145 kbp
43.189 kbp
Contains 2B target
Contains 32T target
Contains 61T-1 target
B
pGG3
2B
32T
32TΔnes
61T-1
Leader
1
2
cas1
C
PFU / ml
pGG3
2B
32T
32TΔnes
61T-1
ϕNM1
ϕNM2
ϕNM4

not offer particularly novel insights (about six months later, I confirmed that the strain had indeed lost CRISPR-Cas targeting functionality in a pG0400 conjugation assay; data not shown).

Meanwhile, additional phage escaper data corroborating these preliminary results was promptly procured, once I'd found two additional spacers, 32T and 61T-1, which reduced plaque formation by the three Newman phages (**Figure 2-7**). The first was the 32T spacer, which was originally identified as a naturally occurring spacer in the type III-A CRISPR-Cas locus of *S. argenteus* MSHR1132 (Holt et al., 2011). Dr. Marraffini, and also my colleague Wenyan, brought this spacer to my attention when they found that it partially matches to the Newman phages I was working with. Partial complementarity between this spacer's crRNA and the Newman phage target regions is illustrated in **Figure 2-8**. In parallel, I constructed two variants (32T*^{A} and 32T*^{B}) which perfectly matched their targets in ϕNM1/ϕNM2 and ϕNM4, respectively, because it was unclear at the time whether the naturally occurring mismatches would abrogate immunity. Ultimately, all three spacers were found to provide immunity. Having been generally advised on the pitfalls of studying engineered systems, however, I endeavored to work with the unmodified 32T spacer whenever possible. Using this spacer, I isolated three ϕNM2 and three ϕNM4 escaper plaques, which I designated as 'α-series' mutants to distinguish them from escapers of spacer 2B. Unlike the 2B spacer, which targets the bottom strand of Newman phage genomes, spacer 32T targets their

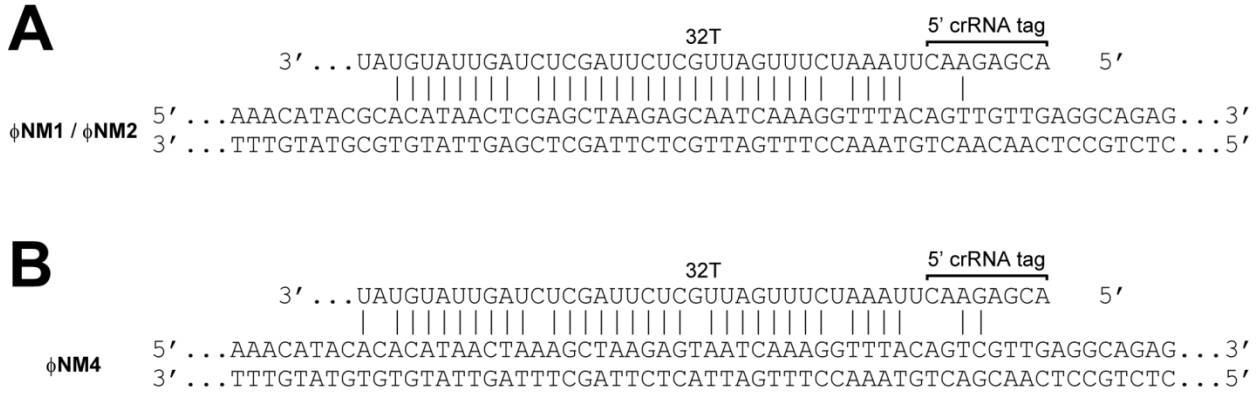


Figure 2-8. Partial complementarity between the spacer 32T crRNA and its DNA target regions in ϕNM1/ϕNM2 or ϕNM4.

(A) Vertical lines denote base-pairing potential between sequences of the 32T crRNA and the DNA of conserved ϕNM1/ϕNM2 target regions; G:U base-pairing potential was not considered. The 5′ crRNA tag is an 8 bp sequence derived from the CRISPR repeat that must be sufficiently mismatched with the target's flanking sequence in order to license immunity. **(B)** Same as in 'A', but with base-pairing potential to DNA of the ϕNM4 target region depicted instead.

'top' strands, hence the 'T'. Having now found, somewhat fortuitously, that I could obtain immunity with spacers targeting either strand, I designed four additional spacers (61B-1, 61B-2, 61T-1, and 61-T2) matching a separate region of the genome, this time with deliberate targeting of both strands. I was initially unsuccessful at cloning the 61-T2 spacer. However, among the other three spacers which I promptly cloned, the 61T-1 spacer was the only spacer which provided immunity. Using this 61T-1 spacer, I again isolated some escaper plaques, and designated them as 'τ-series' mutants. Two ϕNM1 mutants, and one ϕNM2 mutant, were isolated. None of these α- or τ-series escapers

were found to produce wild type PCR amplicons at their target regions, and in some cases I was able to map deletions which included all or part of their target sequences. Aside from their CRISPR-escape phenotypes, however, they were generally indistinguishable from wild type. One exception was the ϕNM2τ1 phage, which exhibited a small plaque phenotype (**Figure 2-9A**). Unlike the ϕNM1γ6 phenotype, however, the ϕNM2τ1 plaque phenotype was observed even on control lawns (**Figure 2-9B**), and was therefore deemed to be CRISPR-independent and not studied further.

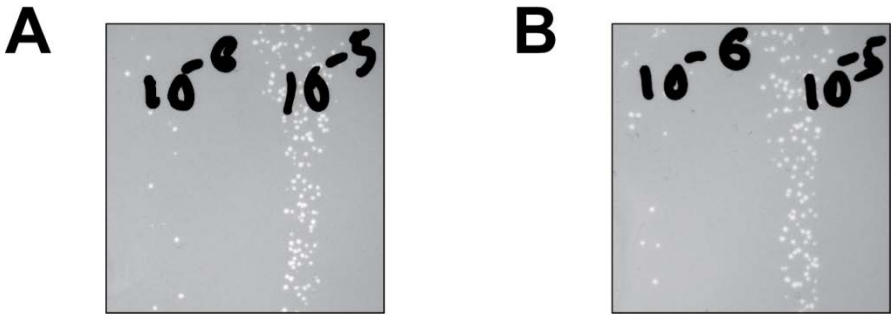
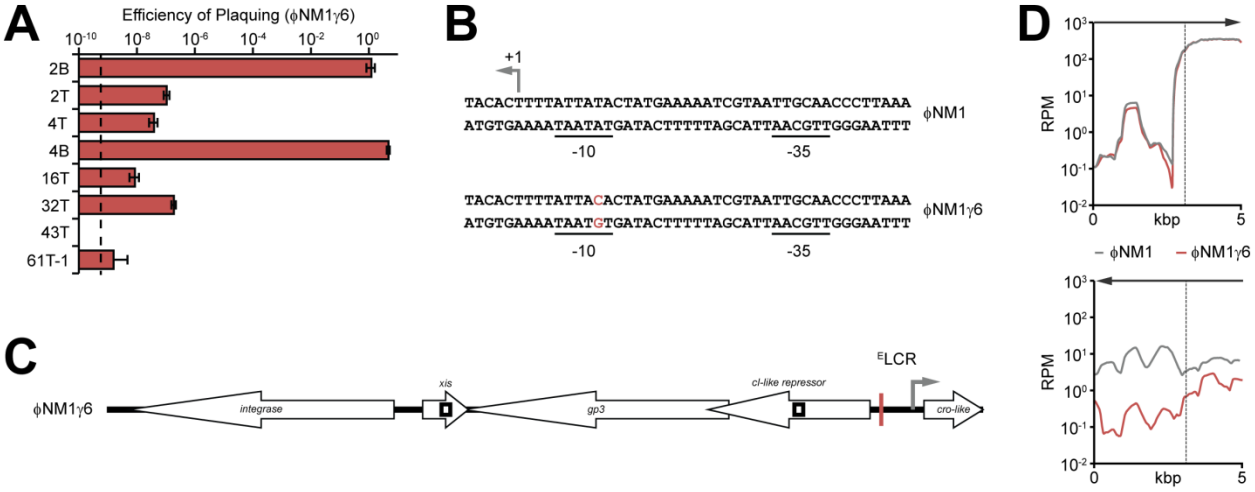


Figure 2-9. CRISPR-independent, small-plaque phenotype of ϕNM2τ1.
(A-B) Plaque formation by a ϕNM2τ1 lysate spotted on lawns of TB4 harboring the spacer 61T-1 CRISPR-Cas system (A) or the pGG3 parent vector (B). Pictures are representative of at least two experiments performed under similar assay conditions.

Given that this phage had escaped CRISPR-Cas targeting of the parental tail fiber gene, I presumed its plaque phenotype to reflect impaired phage particle functionality (such as a defect in adsorption) resulting from a deletion in that target region. Taken together, these putative and confirmed deletion escapers stood in stark contrast to the ϕNM1γ6 escaper, which did not possess target site mutations and appeared to only partially

evade targeting by spacer 2B. I initially considered the possibility that ϕNM1γ6 had in some way acquired a general CRISPR-escape phenotype, but this was promptly ruled out by plating on lawns harboring the spacer 32T CRISPR-Cas plasmid. Representative results from experiments with RN4220 are depicted in **Figure 2-10A**, and demonstrate that ϕNM1γ6's plaquing efficiency was reduced appreciably by the 32T spacer, but not the 2B spacer. Somehow, the heritable changes in ϕNM1γ6 that allowed for partial escape were still ostensibly specific to targeting by spacer 2B. Unfortunately, I could not also investigate the ϕNM1γ6 CRISPR-escape phenotype in the context of a lysogenic infection (i.e., chromosomal targeting by spacer 2B), because it apparently didn't form lysogens. It was now clear to both me and Dr. Marraffini that full genomic sequencing of this phage would likely provide additional insight into its phenotypes (which were not lost upon passaging in the absence of CRISPR, as one might expect for an epigenetic form of escape (Bertani and Weigle, 1953; Korona and Levin, 1993)). We had also discussed plans to produce a more thorough characterization of my various escaper mutations, which in many cases still hadn't been mapped by PCR, so I was already considering the prospect of sequencing entire phage genomes. Meanwhile, Dr. Bikard was isolating phages in separate experiments, and plans were conceived to pool our phages together in a multiplexed, next-generation sequencing run at the Rockefeller University Genomics Core. Both phenotypes of ϕNM1γ6 would ultimately be explained, but it was some time before this sequencing was completed. In the interim, I

Figure 2-10. ϕNM1γ6 can escape type III targeting of the bottom strand to the left of the ^{E}LCR, owing to a SNP that reduces leftward transcription from its P_{cI} promoter. **(A)** Plaquing efficiency of ϕNM1γ6 on lawns of RN4220 harboring pGG3- or pGG3-BsaI-derived type III CRISPR-Cas plasmids with different phage-targeting spacers. Dotted line represents the limit of detection under these assay conditions. Error bars, mean ± s.d. (*n* = 3, technical replicates). **(B)** Consensus sequences of the leftward-oriented P_{cI} promoter within the ^{E}LCR of ϕNM1 and ϕNM1γ6. The SNP found in ϕNM1γ6's –10 element is shown in clay-red lettering. A putative transcription start site (+1) is indicated by the bent grey arrow for ϕNM1. **(C)** Schematic diagram depicting a relevant genomic region of ϕNM1γ6 that includes the ORF2 (*xis*) and ORF4 (*cI*-like repressor) target regions, as well as the ^{E}LCR. Open black rectangles denote the approximate positions of spacer 2B/2T and 4B/4T target sequences. The location of the SNP in P_{cI} is marked by a clay-red vertical line. Rightward-oriented bent grey arrow represents the intact P_{cro} promoter. **(D)** Comparison of phage transcription profiles from cells infected with ϕNM1 (grey lines) or ϕNM1γ6 (clay-red lines), 15 min post-infection. Phage-derived transcripts are plotted in reads per million total-mapped reads (RPM) relative to their position on the genome (only the first 5 kbp are shown). To improve readability, curves were smoothened by plotting average reads per million values over a 500 bp sliding-window. Arrows up top indicate the direction of transcription plotted in each graph; the vertical dotted line marks the position of the ^{E}LCR.


A
Efficiency of Plaquing (ϕNM1γ6)
10-10
10-8
10-6
10-4
10-2
100
2B
2T
4T
4B
16T
32T
43T
61T-1
B
+1
TACACTTTTATTATACTATGAAAAATCGTAATTGCAACCCTTAAA
ATGTGAAAATAATATGATACTTTTTAGCATTAACGTTGGGAATTT
ϕNM1
-10
-35
TACACTTTTATTACACTATGAAAAATCGTAATTGCAACCCTTAAA
ATGTGAAAATAATGTGATACTTTTTAGCATTAACGTTGGGAATTT
ϕNM1γ6
-10
-35
C
ϕNM1γ6
integrase
xis
gp3
cI-like repressor
ELCR
cro-like
D
RPM
kbp
ϕNM1
ϕNM1γ6

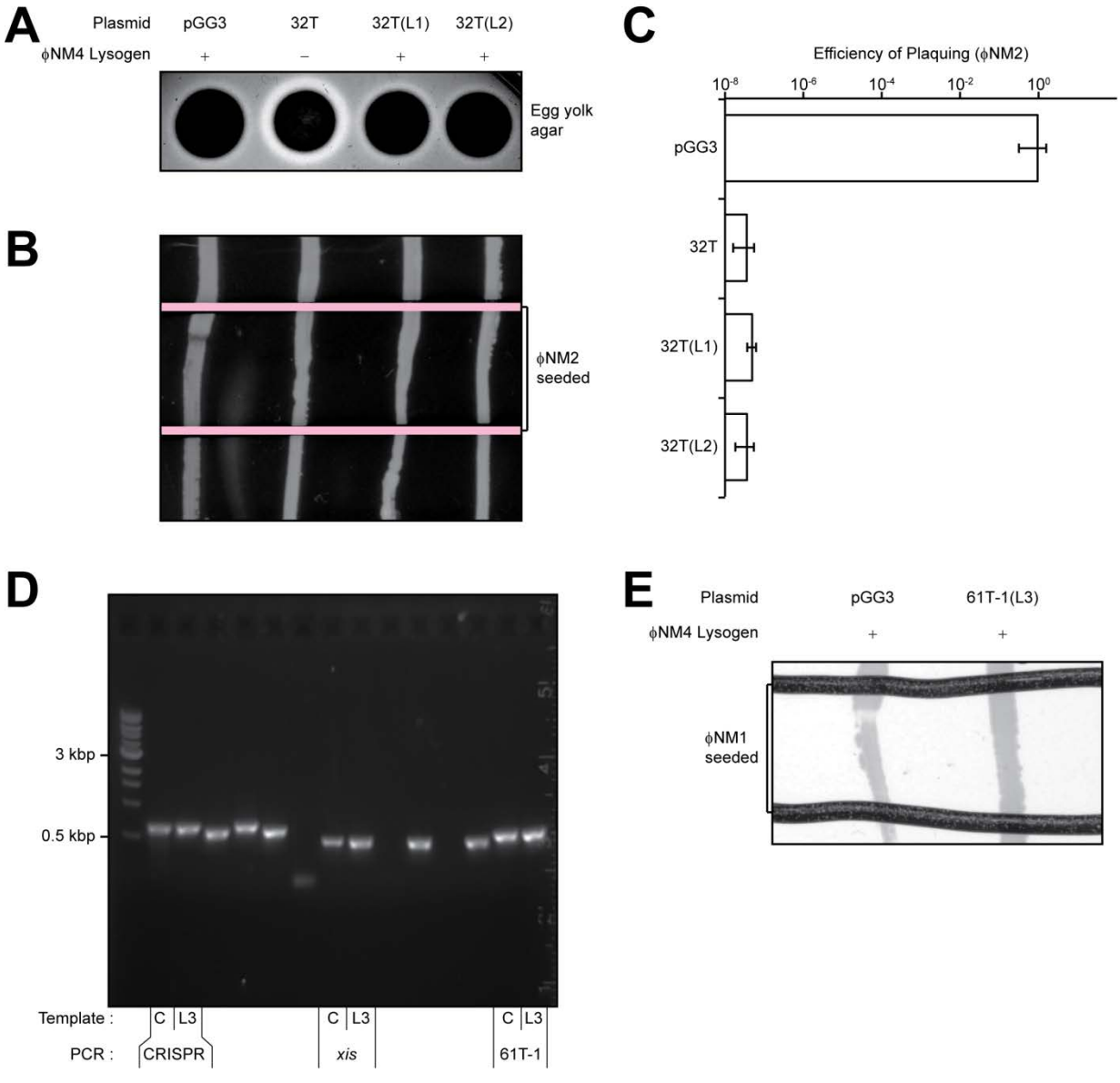
was conducting experiments with lysogens and the spacer 32T CRISPR-Cas plasmid, which also proved fruitful.

2.3 Initial evidence for co-existence between functional type III CRISPR-Cas systems and their DNA target elements in *S. aureus* lysogens

As explained in section 2.2, my analysis of two lysogenic clones which survived the infection by ϕNM4(γ4) in a spacer 2B lawn seemed to suggest that type III-A CRISPR-Cas targeting could select for genetic escape outcomes in the event of lysogenization by a lambdoid, chromosomally integrated prophage (**Figure 2-6**). In order to demonstrate this more clearly without the need for co-infecting ϕNM4 wild type and ϕNM4γ4 mutant phages, Dr. Marraffini recommended that I take advantage of the lysogenic conversion phenotype associated with ϕNM4 integration to screen for lysogenized clones, by plating cells on egg yolk-supplemented agar after a brief infection. Upon site-specific integration into TB4's glycerol ester hydrolase (*geh*) locus, the ϕNM4 prophage disrupts expression of this secreted lipase, and thereby reduces the visible breakdown of lipids caused by colonies formed on egg yolk agar plates (Bae et al., 2006). The unmodified, 32T spacer was initially tested in this 'egg-yolk screen'. Approximately 1000 colonies were plated, and two were found to be *geh*$^-$ (that is, displaying reduced lipid clearance on the egg yolk agar). A representative image of the two lysogenic isolates plated on egg yolk agar is presented in **Figure 2-11A** alongside lysogenic and

Figure 2-11. Characterization of ϕNM4 lysogens isolated from the egg yolk screen with spacer 32T or 61T-1 CRISPR-Cas plasmids.

(A) Visualization of TB4-derived strains grown on egg yolk-supplemented agar. Integration of ϕNM4 within the host's *geh* locus results in strongly reduced breakdown of lipids in the media, enabling a screen for ϕNM4 lysogenization. Right-most lanes display two lipase-negative isolates ('L1' or 'L2') from the spacer 32T lysogenization screen; left-most lanes display lipase-negative and lipase-positive control strains, respectively (according to the labeling above). Picture is representative of at least five technical replicates for each strain. **(B)** ϕNM2-sensitivity assay for the strains shown in 'A' (arranged in the same order from left to right). Single colonies were streaked through the ϕNM2-seeded region from top to bottom. The ϕNM4 lysogen harboring pGG3 and the non-lysogen harboring spacer 32T streaked in the two left-most lanes serve as sensitive and insensitive controls, respectively. Picture is representative of three technical replicates for each strain. **(C)** ϕNM2 plaquing efficiency on soft agar lawns of the strains analyzed in 'A' and 'B', calculated by comparison with plaquing on non-lysogenic lawns of TB4 harboring the pGG3 plasmid. **(D)** PCR amplification from a lipase-negative isolate ('L3') of the spacer 61T-1 lysogenization screen, or from control templates that produce wild type amplicon sizes (C). A colony lysate from the L3 isolate was used to template three different PCRs using primers for the spacer 61T-1 target region (right), primers for amplifying the CRISPR array (left), or primers for amplifying from Newman phages at a separate, *xis*, locus (left). Size markers of 3 kbp and 0.5 kbp in lane 1 are indicated. Additional replicates were not performed. Lanes 4, 5, 6, 7, 10, 11, 12, and 13 were used for other experiments. **(E)** ϕNM1 sensitivity assay for the 61T-1(L3) egg yolk isolate characterized in 'D'. The ϕNM4 lysogen harboring pGG3 in the left-most lane serves as a sensitive control. Single colonies were streaked through the ϕNM1-seeded region from top to bottom. Additional replicates were not performed.


A
Plasmid
pGG3
32T
32T(L1)
32T(L2)
ϕNM4 Lysogen
+
−
+
+
Egg yolk agar
B
ϕNM2 seeded
C
Efficiency of Plaquing (ϕNM2)
10-8
10-6
10-4
10-2
100
pGG3
32T
32T(L1)
32T(L2)
D
3 kbp
0.5 kbp
Template :
C
L3
C
L3
C
L3
PCR :
CRISPR
xis
61T-1
E
Plasmid
pGG3
61T-1(L3)
ϕNM4 Lysogen
+
+
ϕNM1 seeded

non-lysogenic controls displaying *geh*$^-$ and *geh*$^+$ phenotypes, respectively. After determining by PCR and Sanger sequencing that both isolates maintained intact CRISPR arrays, I proceeded with more functional testing. Firstly, I confirmed that each of the isolates maintained (CRISPR-dependent) immunity to ϕNM1 in plaque assays (data not shown). In other words, these isolates hadn't genetically inactivated their spacer 32T CRISPR-Cas systems. I next investigated whether the resident prophage was still wild type, by collecting supernatants with spontaneously induced PFU and re-plating them on both spacer 32T lawns and control lawns. Although some plaquing was observed on control lawns, none was observed on spacer 32T lawns (data not shown). This indicated that the phages derived from each lysogen were indeed still wild type, or at least the majority of them were. While examining these isolates again prior to publication (Goldberg et al., 2014), I confirmed that both are also resistant to ϕNM2 infection (as expected for strains with an intact spacer 32T CRISPR-Cas system). Representative experiments with ϕNM2 are depicted in **Figures 2-11B & C.** Shortly after the original egg yolk screen with spacer 32T, I repeated the screen under similar conditions using the 61T-1 or 2B spacers. Although no colonies were identified in the spacer 2B screen, in the 61T-1 screen, one *geh*$^-$ colony was again identified among the ~1000 colonies plated. I determined by PCR that this isolate had wild type amplicon sizes at both its CRISPR array and prophage target region, and subsequently confirmed its resistance to ϕNM1 by the streak-test method (**Figures 2-11D & E**). I also analyzed

the spontaneously induced PFU derived from this isolate, and found that while some plaques were formed on control lawns, none were formed on 61T-1 lawns (data not shown). Again, this finding seemed to suggest that its resident prophage was wild type.

The results of these egg yolk screens seemed to indicate that, at least with certain spacers, functional CRISPR-Cas systems could co-exist with non-mutant prophage targets within lysogenic isolates. I noticed, however, that the number of spontaneously induced PFU derived from these isolates appeared to be reduced relative to that of wild type ϕNM4 lysogens. In my mind, this phenotype was reminiscent of the ϕNM1γ6 escape phenotype: co-existence between a functional CRISPR-Cas system and intact target sequence occurred at least transiently, but not without evidence of a partial interaction (presumably targeting). Corroborating the lysogeny results, I found transformation of TB4::ϕNM4 single lysogens with spacer 32T CRISPR-Cas plasmids to be moderately efficient. In one experiment, four transformants were confirmed by PCR to have both their CRISPR array and target region amplicons intact (**Figure 2-12A**). However, seemingly conflicting results were again obtained via electroporation of spacer 2B CRISPR-Cas plasmids, this time into TB4::ϕNM1 single lysogens instead of Newman. Of the nine transformants that were found in one experiment, 8/9 had lost their CRISPR-Cas targeting functionality, as assessed by sensitivity to ϕNM4 infection in the streak-test assay (**Figure 2-12B**). The remaining ϕNM4-resistant transformant was noted to have lost its ϕNM1 prophage, although it's unclear from my notes how this

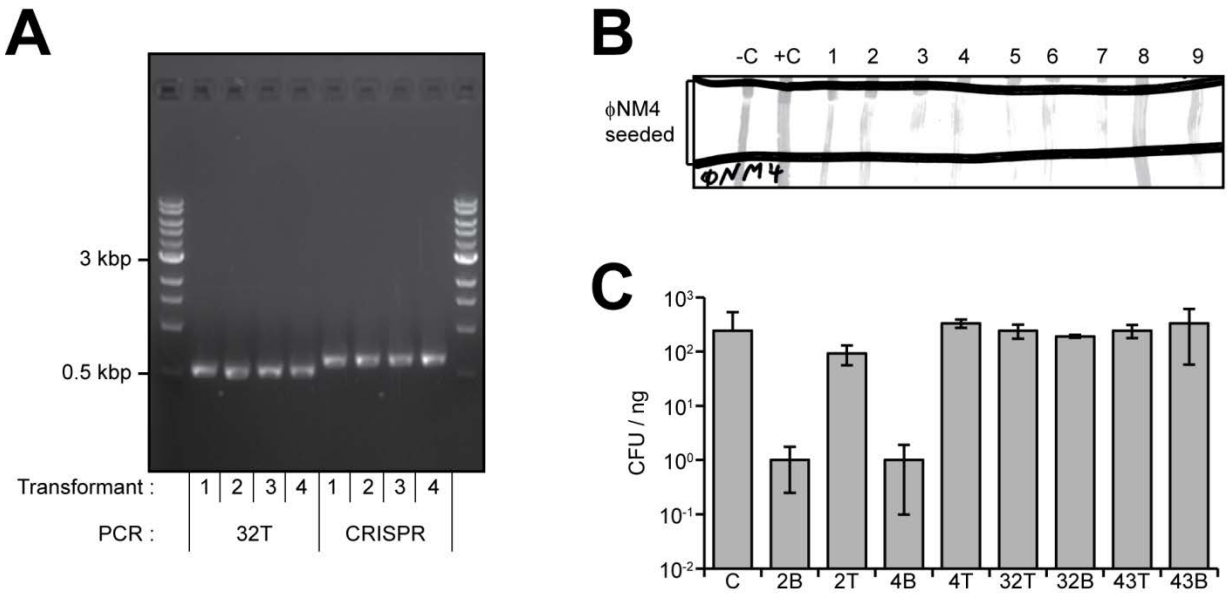


Figure 2-12. Transformation of single lysogens with type III CRISPR-Cas plasmids. **(A)** PCR amplification from four randomly selected transformants isolated via an electroporation of TB4::ϕNM4 lysogens with the spacer 32T CRISPR-Cas plasmid. After re-streaking, a single colony lysate from each transformant was used to template two different PCRs using primers for the 32T target region (left) or primers for amplifying the CRISPR array (right). Size markers of 3 kbp and 0.5 kbp in lanes 1 and 10 are indicated. Additional replicates were not performed on these isolates. **(B)** ϕNM4-sensitivity assay for the nine transformants isolated via an electroporation of TB4::ϕNM1 lysogens with the spacer 2B CRISPR-Cas plasmid. After re-streaking, single colonies were streaked through the ϕNM4-seeded region from top to bottom. A TB4::ϕNM1 lysogen harboring pGG3 and a TB4 non-lysogen harboring spacer 2B were streaked in the two left-most lanes as sensitive (-C) and insensitive (+C) controls, respectively. Additional replicates were not performed on these isolates. **(C)** Transformation of RN4220::ϕNM1 lysogens with different CRISPR-Cas plasmids as indicated. The pGG3 plasmid was transformed as a control (C). Transformation efficiencies were calculated as the number of CFU per nanogram of miniprepped DNA used for electroporation. Error bars, mean ± s.d. (n = 3, technical replicates).

was assessed at the time. Plausibly, the strain was found to be sensitive to one of my ϕNM1-derived spacer 2B escapers, such as ϕNM1γ1 or ϕNM1γ6, or that the ϕNM1 prophage simply could not be detected by colony PCR. Representative transformation efficiency data, later collected with RN4220::ϕNM1 lysogens, are shown in **Figure 2-12C**.

At this point, I was piecing together a model that could account for the various results. On the one hand, immunity to lytic infection was licensed by spacers targeting either strand of the phage genomes (2B, or 32T and 61T-1). On the other hand, the 2B spacer exhibited a distinct phenotype from the 32T and 61T-1 spacers in the context of lysogeny (**Figures 2-6 & 2-12**). Meanwhile, most of the sense-oriented spacers that had been tested so far (16B, 19B, 56B, 61B-1 and 61B-2) all failed to license immunity. The 2B spacer is distinct from other bottom-strand-targeting spacers in that its target is located within a cluster of leftward-oriented ORFs to the left of the ^{E}LCR (**Figure 2-2**). At least in part, ORF orientations are predicted to reflect differences in the directionality of transcription throughout phage genomes. Based on these predictions, I reasoned that a directional transcription requirement for type III-A targeting could explain functional consequences associated with the orientation of spacers relative to their target region(s). In other words, if type III-A immunity required that targets are transcribed in the antisense orientation relative to a spacer's crRNA, this could explain why the 2B, 32T, and 61T-1 spacers were all functional, while the others were not. An apparent

inconsistency here was that the wild type *nes* spacer is sense-oriented relative to its target ORF, but was shown to license pG0400 targeting in conjugation assays (Marraffini and Sontheimer, 2008). However, my colleague Wenyan had also found that an antisense-oriented spacer complementary to the opposite strand of this target sequence (but arbitrarily also shifted 2 bp upstream; '*nes3*') could provide immunity to pG0400 (personal communication). Therefore, the results would still be explained if this *nes* target region is bi-directionally transcribed, such that the sense-oriented wild type spacer would in fact be complementary to a non-template strand when considering transcription from downstream. Given that the downstream ORF was oriented in the opposite direction, this seemed plausible. Also note that my model assumed the directional transcription requirement to be a pre-requisite for DNA targeting, which, would not necessarily distinguish between whether the sequence was coding or non-coding, and instead distinguish directionality on the basis of whether the strand was templating or non-templating for an RNA polymerase. Thus, the rightward-oriented reading frame of *xis* would not necessarily impact targeting by 2B. Moreover, evidence was mounting that type III systems could directly cleave complementary RNAs (Hale et al., 2012; Hale et al., 2009; Zhang et al., 2012), and this did not go unnoticed by our lab. RNA transcripts, naturally, are produced from transcribed DNA, and the strand polarity of a given transcript is derived from the directionality in which it is transcribed. I reasoned, therefore, that it would not be entirely unprecedented if our type III-A

system could also in some way respond to the polarity of transcriptional events (possibly, but not necessarily, via the RNA transcripts—even if this were to ultimately result in DNA targeting). Importantly, a directional transcription requirement for type III-A targeting could account for the various results I was observing with lysogens. During lysogenic infections, lambdoid prophages repress transcription of most lytic genes, but can sustain transcription of one or more genes required for maintenance of lysogeny (Casjens and Hendrix, 2015; Ptashne, 2011). Targets of the 32T and 61T-1 spacers lie within lytic regions of the Newman phage genomes and were presumably repressed during lysogeny, whereas the spacer 2B target lies within a small intervening region of the lysogeny genes downstream of the *cI*-like repressor, just before *int*. Conceivably, the 2B target could be transcribed in lysogens by read-through from an upstream promoter, such as the leftward P_{cI} promoter found in the ^{E}LCR (**Figures 2-2**). In other words, chromosomal targeting would occur in lysogens where a target in the proper orientation falls within an actively transcribed region, but not necessarily where a repressed lytic gene is targeted. Although sustained *integrase* expression is not required for lysogenic maintenance in phage λ (Casjens and Hendrix, 2015), there are noticeable differences in the genomic organization of λ and the Newman phages, so it was reasonable to suspect that their transcriptional patterns would also differ. For example, in lambda, the *xis* ORF is oriented in the same direction as *int*, but they're oppositely oriented in the Newman phages. Furthermore, the possibility that targeting

was licensed by even low levels of transcriptional read-through in this region could not be excluded (i.e., regardless of whether the transcription was functionally relevant to *int* expression), and would perhaps explain why lytic infection was blocked by the 2B spacer despite the position of its target between lysogeny genes. Finally, I reasoned that the ϕNM1γ6 CRISPR-escape and clear-plaque phenotypes might also be explained if its mutation(s) disrupts leftward transcription of the lysogeny genes: weakened leftward transcription could both reduce the chance of DNA targeting by the spacer 2B CRISPR-Cas system during lytic infections, and also prevent the establishment of lysogeny.

A number of these points were still hypothetical of course, so I conceived some plans to test the model further. In particular, I wanted to confirm the directional transcription requirement using an inducible promoter system (ideally carried on a plasmid, for ease of manipulation) where I could insert a target sequence in both orientations relative to the promoter, and then assay for targeting upon induction. When I shared these ideas with Wenyan, he excitedly agreed that, if confirmed, the directional transcription requirement would nicely explain the phage results, and perhaps other findings as well. For example, in some of his early experiments where he designed a few chromosome-targeting spacers for the type III-A system in *S. epidermidis*, only the antisense-oriented spacers seemed to offer strong targeting. Furthermore, when we looked into some of the naturally occurring, temperate-phage-targeting spacers that had been sequenced, all of them (including the 32T spacer) appeared to be

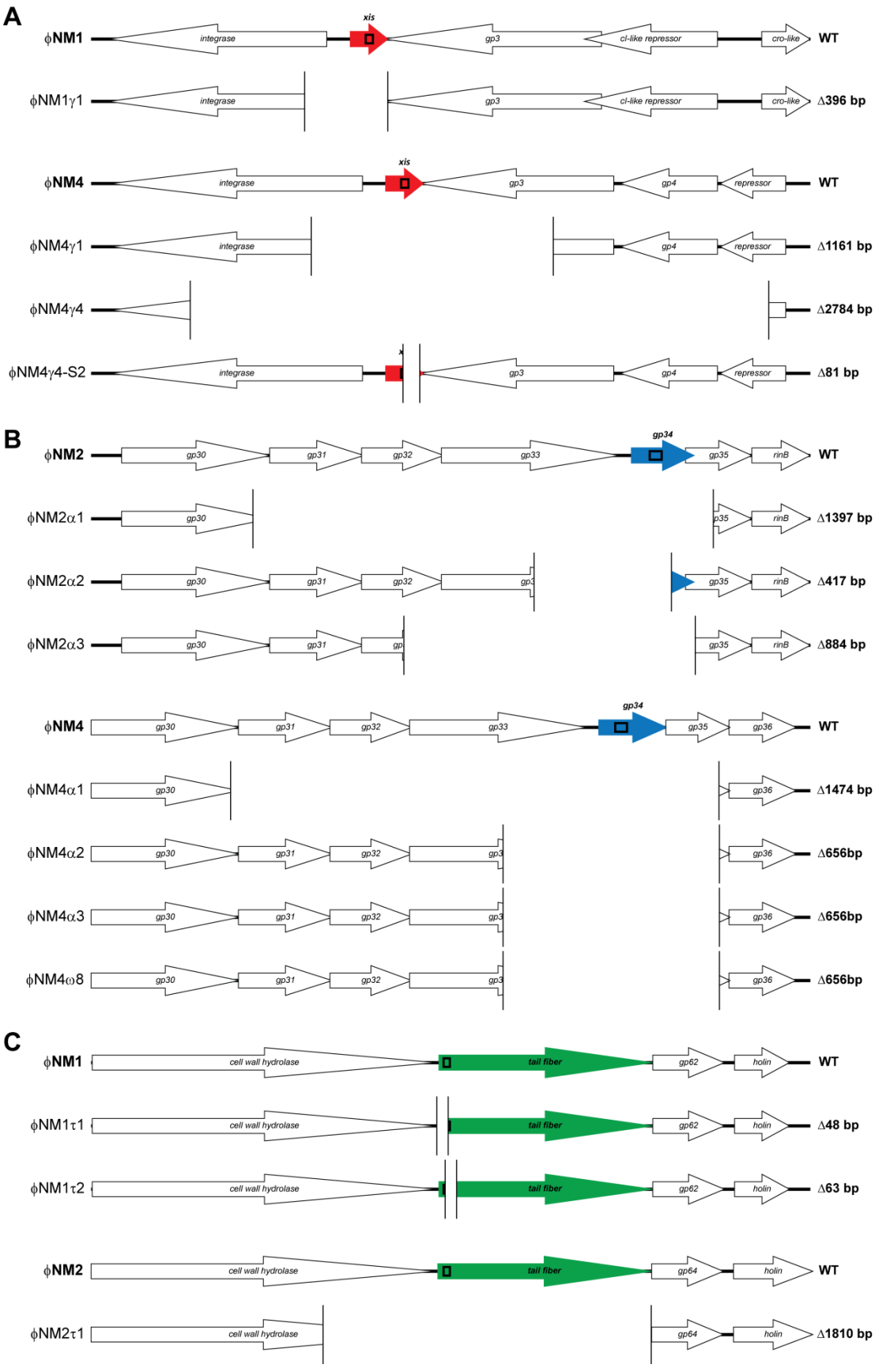
antisense-oriented relative to their target ORFs. At the time, however, no one in the lab had been working with tightly regulated inducible systems (which I assumed might be valuable in this case), so we consulted Dr. Marraffini to discuss the logistics. When I proposed the idea, he suggested that I use one of the pCL55-derived integrative vectors we had received from Angelika Gründling, which were already designed for inducible expression. These were the IPTG- and (Anhydro)tetracycline-inducible vectors referred to as pCL55-iSpac and pCL55-iTET, respectively. The pCL55-iTET plasmid, I now presume, refers to the p*itet* plasmid published previously (Grundling and Schneewind, 2007). My CRISPR-Cas plasmids contained the same chloramphenicol resistance marker, so Dr. Marraffini pointed out that I would simply need to replace the resistance marker on one of these pCL55-based vectors. Moreover, because these vectors allow plasmid manipulations to be carried out in *E. coli*, cloning with them was expected to be straightforward (and indeed, at least for the pCL55-iTET vector, generally was). Ultimately, although chromosomal targeting of these integrative vectors imposed certain experimental limitations, it provided the answers we were looking for, and nicely emulated the prophage-targeting scenario encountered during lysogeny.

Both the design and results of these inducible targeting experiments, along with additional evidence which confirmed the directional transcription requirement for type III targeting, are described in the following section (**2.4**). Before these experiments were completed, however, the data from our phage sequencing was received and proved to

be highly encouraging. I owe a great many thanks to Dr. Bikard for procuring free software to analyze this data in a timely fashion (including SAMtools (Li et al., 2009) and IGV (Robinson et al., 2011), and later also ABySS (Simpson et al., 2009) for de novo assembly). Quite excitingly, ϕNM1γ6 turned out to harbor a SNP at a crucial position of the -10 element in a leftward promoter immediately upstream of its *cI*-like repressor gene (**Figure 2-10B**). Although this SNP was located ~1700 bp away from the target sequence (**Figure 2-10C**), presumably, it would reduce leftward transcription and allow escape from spacer 2B (as well as reducing the phage's ability to lysogenize). The remaining CRISPR-escape mutants that hadn't been mapped by PCR were confirmed to harbor target site deletions. Preliminary analyses for 14 deletion mutants (**Figure 2-13A**) were presented along with the results of **Figure 2-7** at the Bacteria, Archaea, & Phages Meeting in Cold Spring Harbor (August 2012). Among these was the ϕNM4γ4 clear mutant that was ultimately adopted by my colleagues for type II adaptation experiments (Heler et al., 2015; McGinn and Marraffini, 2016; Modell et al., 2017), and which infects RN4220 somewhat more robustly than my ϕNM1γ6 clear mutant. In addition to its 2784 bp deletion in the lysogeny cluster, ϕNM4γ4 harbored a stretch of sequences (~3101 bp) in its late structural/lytic gene cluster that mapped more closely to other phages, such as bacteriophage ROSA (Accession: AY954961.1), than to ϕNM4 or any of the Newman phages. Presumably owing to the potential for recombination

Figure 2-13. Summary of γ-series, α-series, and τ-series phage mutants that escaped type III CRISPR-Cas targeting via target site deletions.

(A) Schematic diagrams depicting a relevant genomic region of ϕNM1 and ϕNM4 phages that includes the spacer 2B (*xis*) target region. **(B)** Same as in 'A', except for the spacer 32T target region escaped by three ϕNM2 and four ϕNM4 phages. The ϕNM4ω8 phage was originally isolated from a lysogen that escaped pG0400 conjugation, following ϕNM4 infection of TB4 cultures harboring the spacer 32T CRISPR-Cas plasmid. Sequencing revealed a 656 bp target site deletion that was also found in two ϕNM4 α-series escapers. **(C)** Same as in 'A', except for two ϕNM1 and one ϕNM2 escaper of the 61T-1 spacer. In each case, deleted sequences are delimited by vertical lines, and their sizes are denoted on the far right (Δ). ORFs were labeled according to their predicted gene products (where applicable), and target ORFs were color-coded as in **Figure 2-7**. Open black rectangles denote location of the 35- or 36-bp target sequence. In some cases, incomplete deletion of the target sequence was sufficient for evasion of the CRISPR-Cas system.


A
ϕNM1
integrase
xis
gp3
cI-like repressor
cro-like
WT
ϕNM1γ1
integrase
gp3
cI-like repressor
cro-like
Δ396 bp
ϕNM4
integrase
xis
gp3
gp4
repressor
WT
ϕNM4γ1
integrase
gp4
repressor
Δ1161 bp
ϕNM4γ4
Δ2784 bp
ϕNM4γ4-S2
integrase
gp3
gp4
repressor
Δ81 bp
B
ϕNM2
gp30
gp31
gp32
gp33
gp34
gp35
rinB
WT
ϕNM2α1
gp30
gp35
rinB
Δ1397 bp
ϕNM2α2
gp30
gp31
gp32
gp35
rinB
Δ417 bp
ϕNM2α3
gp30
gp31
gp35
rinB
Δ884 bp
ϕNM4
gp30
gp31
gp32
gp33
gp34
gp35
gp36
WT
ϕNM4α1
gp30
gp36
Δ1474 bp
ϕNM4α2
gp30
gp31
gp32
gp36
Δ656bp
ϕNM4α3
gp30
gp31
gp32
gp36
Δ656bp
ϕNM4ω8
gp30
gp31
gp32
gp36
Δ656bp
C
ϕNM1
cell wall hydrolase
tail fiber
gp62
holin
WT
ϕNM1τ1
cell wall hydrolase
tail fiber
gp62
holin
Δ48 bp
ϕNM1τ2
cell wall hydrolase
tail fiber
gp62
holin
Δ63 bp
ϕNM2
cell wall hydrolase
tail fiber
gp64
holin
WT
ϕNM2τ1
cell wall hydrolase
gp64
holin
Δ1810 bp

during co-infection, phage 'mosaicism' appears to be a common phenomenon (Brüssow et al., 2004; Martinsohn et al., 2008), and was assumed to explain the finding. Although we do not work with phage ROSA, it is not inconceivable that a contaminating phage from our lab environment, related to ROSA, could have produced this recombination event. However, if such a recombination did occur in our lab, it likely happened prior to the initial CRISPR selection step(s), because my other ϕNM4-derived escaper phages also harbored the same recombinant sequence at this region. Verifying the ϕNM4 prophage sequence at this region in our Newman isolate, or in our original ϕNM4 single lysogen isolate, might help to tease out this point. The ϕNM1γ1 mutant, meanwhile, was found to harbor a 396 bp deletion which included the *integrase* start codon but did not disrupt the *cI*-like repressor ORF upstream. This offered an explanation for why I was never able to isolate stable lysogens of ϕNM1γ1, despite its turbid plaques (data not shown). In other words, ablation of its *integrase* function appears to have resulted in an 'abortive lysogeny' phenotype (Gottesman and Weisberg, 2004). Regarding the escapers of spacer 32T, three out of my four ϕNM4 mutants harbored the same 656 bp deletion at the target region, whereas my three ϕNM2 mutants harbored three different deletions. The 656 bp deletions appear to have utilized 12 bp direct repeat microhomologies that flank the 32T target in ϕNM4. The downstream repeat is absent from ϕNM1 and ϕNM2, which may explain why the 656

bp deletion was not observed in ϕNM2 mutants, and perhaps also why the frequency of escape was found to be much higher for ϕNM4 than for ϕNM1 or ϕNM2 (**Figure 2-7C**).

On top of the phage sequencing results, I was further encouraged by a result that I obtained at this time using the pG0 conjugation assay, which seemed to support the notion of a directional transcription requirement for type III targeting in general. Previously, I had cloned a spacer called *nes*-γ(B) which was intended to target pG0mut on the same strand as the wild type *nes* spacer, but ~1000 bp upstream of the engineered SNPs (**Figure 2-14A**). However, this sense-oriented spacer never provided immunity. Once I began noticing the strand requirement pattern in my plaque assays, I finished cloning the inverted, antisense-oriented equivalent of this spacer (*nes*-γT) and found that it indeed prevented conjugation by pG0mut (**Figure 2-14B**). I interpreted this to mean that inverse-oriented transcription from the downstream ORF might not read all the way through to the *nes*-γB target. In hindsight, this result should be interpreted with additional caution, because the two-spacer CRISPR arrays I was using still contain the wild type *nes* spacer in position 1 (see **Figure 2-7B** for an example). Therefore, the wild type *nes* spacer might license some effect on pG0mut at the mismatched target site downstream, even though it does not reduce conjugation efficiency in the presence of these mismatches. For example, type III-A systems were eventually found to cleave complementary target RNAs

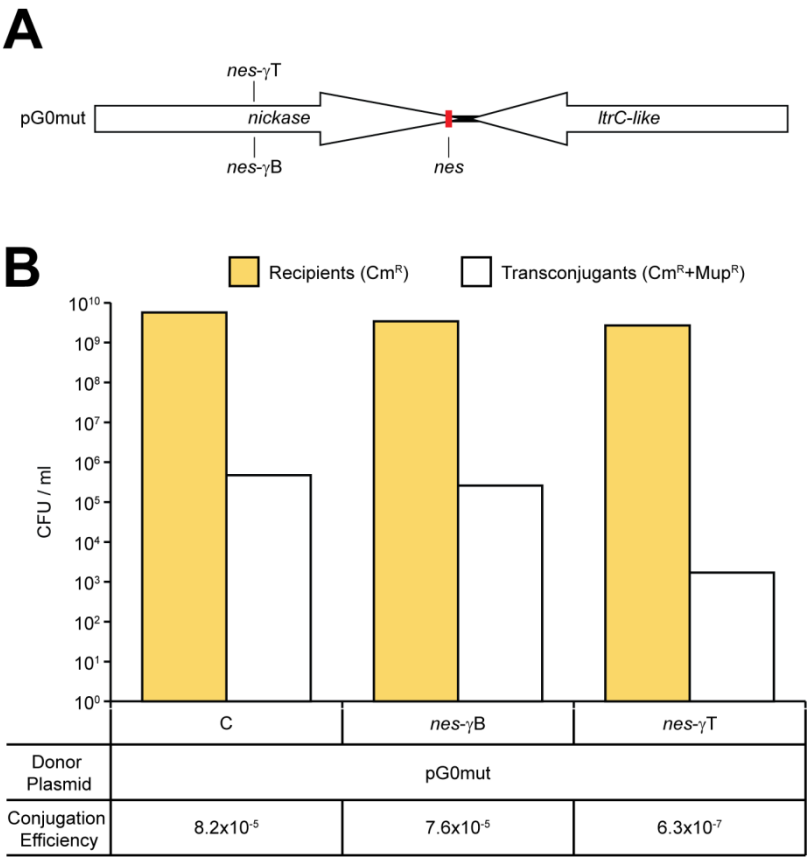


Figure 2-14. Apparent directional transcription requirement for type III targeting of pG0mut at a *nickase* sequence upstream of its mutations.

(A) Schematic diagram depicting a relevant genomic region of the pG0mut plasmid, including its *nickase* target gene and oppositely oriented downstream gene encoding an LtrC-like protein, scaled according to their ORF annotations in pGO1 (Accession number: NC_012547.1). The location of the pG0mut SNPs at the target of the native type III-A *nes* spacer in CRISPR position 1 are marked with a solid red rectangle. Along with the native *nes* spacer target, the *nes*-γB (sense) and *nes*-γT (antisense) targets ~1000 bp upstream are also indicated. **(B)** Conjugation assays performed with pG0mut as in **Figure 2-3**. In this dataset (single experiments), recipients harbor type III CRISPR-Cas plasmids with the *nes*-γB or *nes*-γT spacers in position 2, or the pGG3 parent vector harboring only the *nes* spacer in position 1 (C). Conjugation efficiencies denote the ratio of transconjugants to recipients measured in each experiment.

(Samai et al., 2015; Staals et al., 2014; Tamulaitis et al., 2014), and this even occurs in the presence of partial mismatching (Staals et al., 2014; Tamulaitis et al., 2014). Thus, the wild type *nes* spacer might potentially license cleavage of antisense transcripts originating from downstream. However, it's unclear whether this effect could result in sufficiently diminished transcriptional read-through to the *nes*-γB spacer's target, or whether some other less-expected effect(s) might arise from the mismatching.

2.4 Demonstration of a directional transcription requirement for *in vivo* targeting by a type III CRISPR-Cas system

Prior to working with the pCL55-iTET vector, I had attempted to amplify from the pCL55-iSpac vector in unrelated PCR experiments, but was unsuccessful. This time, I set out with parallel efforts to replace the antibiotic resistance marker on both of these integrative vectors. Although I was again unsuccessful at amplifying from the pCL55-iSpac vector, I managed to replace the pCL55-iTET's chloramphenicol resistance marker with a kanamycin resistance marker to create pKL55-iTET. Using this vector, I designed my first chromosomal targeting system around the 16B/16T spacer pair (**Figures 2-2 and 2-15A**), since it was the first pair of phage-targeting spacers I had designed to be the exact inverse of each other. The 61B-1/61T-1 and 61B-2/61T-2 pairs, for example, were still slightly off-set from one another (**Table 2-1**). Therefore, I could readily design targets for the 16B/16T spacers that were the exact inverse of one another, and insert

them downstream of the inducible promoter. In previously published work, chromosomal targeting by CRISPR-Cas systems was shown to be toxic, and perhaps even lethal (Edgar and Qimron, 2010). Generally, transformation experiments in the lab thus far had yielded similar outcomes (including Wenyan and Dr. Bikard's initial

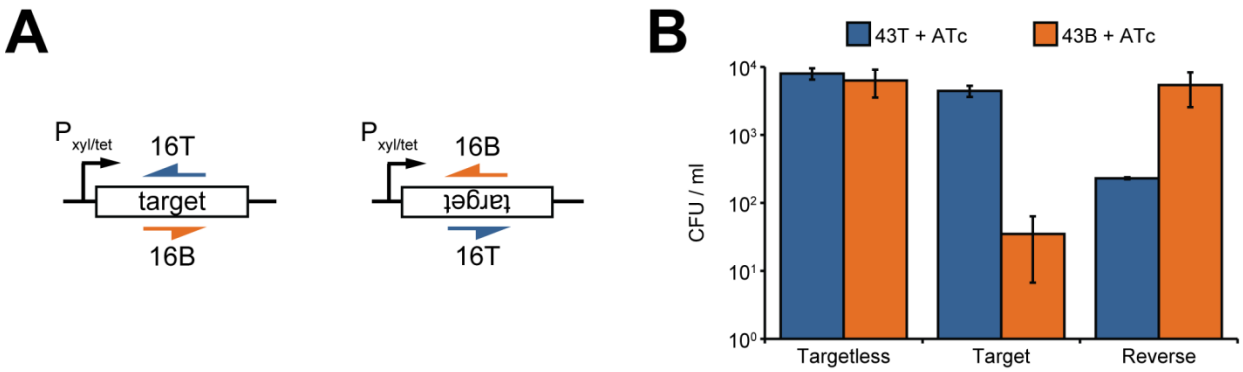


Figure 2-15. Transformation of strains containing first-generation integrative target vectors with type III CRISPR-Cas plasmids harboring spacer 16T or 16B.
(A) Schematics summarizing the orientation of spacer 16T or 16B crRNAs (orange and blue arrows) relative to a forward (left) or reversed (right) target sequence downstream of the $P_{xyl/tet}$ promoter in first-generation integrative vectors. Leftward-oriented arrows (depicted above) represent crRNAs with complementarity to the non-template strand relative to rightward $P_{xyl/tet}$ transcription; rightward-oriented arrows (depicted below) represent crRNAs with complementarity to the template strand. **(B)** Transformation of RN4220-derived strains with spacer 16T or 16B CRISPR-Cas plasmids in the presence of ATc. Electrocompetent cells possessed either the 'Targetless' parent vector insertion, the forward 'Target' vector insertion, or the 'Reverse' target vector insertion, and were transformed using the same miniprep for either CRISPR-Cas plasmid cloned in RN4220. A constant volume of each miniprep was electroporated across experiments, and the concentration of transformants (CFU / ml) recovered after electroporation was quantified by selective plating. Error bars, mean ± s.d. (n = 2, technical replicates).

genome editing experiments (Jiang et al., 2013a) with type II systems in *Streptococcus pneumoniae*, unpublished at the time). It was therefore intuited that a growth and/or transformation efficiency reduction in the presence of ATc could likewise be used as a readout for CRISPR-Cas targeting of my inducible vectors in the chromosome. Although this targeting setup provided a preliminary proof-of-principle that type III-A spacers could be either targeting or non-targeting depending on their orientation relative to a nearby promoter (**Figure 2-15B**), this particular system was not inducible (i.e., when targeting occurred, it occurred even in the absence of the ATc inducer for at least one of the two spacer-target orientations). In order to more definitively demonstrate a transcriptional requirement for type III-A targeting, I endeavored to reproduce these results with a more tightly-regulated inducible promoter system. Ideally, my spacers and targets might co-exist in the absence of the transcriptional inducer, as they could in lysogens. The $P_{xyl/tet}$ promoter architecture employed in pCL55-iTET is also found in pKOR (Bae and Schneewind, 2006), and was originally designed as an engineered promoter system for *Bacilli* (Geissendorfer and Hillen, 1990). By default, this $P_{xyl/tet}$ promoter in pWH353 had a single *tetO* operator and was known to be relatively leaky, but various modifications—including addition of a second *tetO* operator downstream—have been shown to reduce its leakiness (Geissendorfer and Hillen, 1990; Helle et al., 2011). The modified promoter system I ultimately selected was based on the pRAB12 architecture described previously (Helle et al., 2011), because it

was shown to promote virtually undetectable expression levels in the absence of inducers, while still offering moderate expression levels in their presence. Its overall regulatory architecture is similar, with divergently transcribed P_{tetR}- and $P_{xyl/tet}$-based promoters, but it includes various modifications which I introduced into pKL55-iTET in

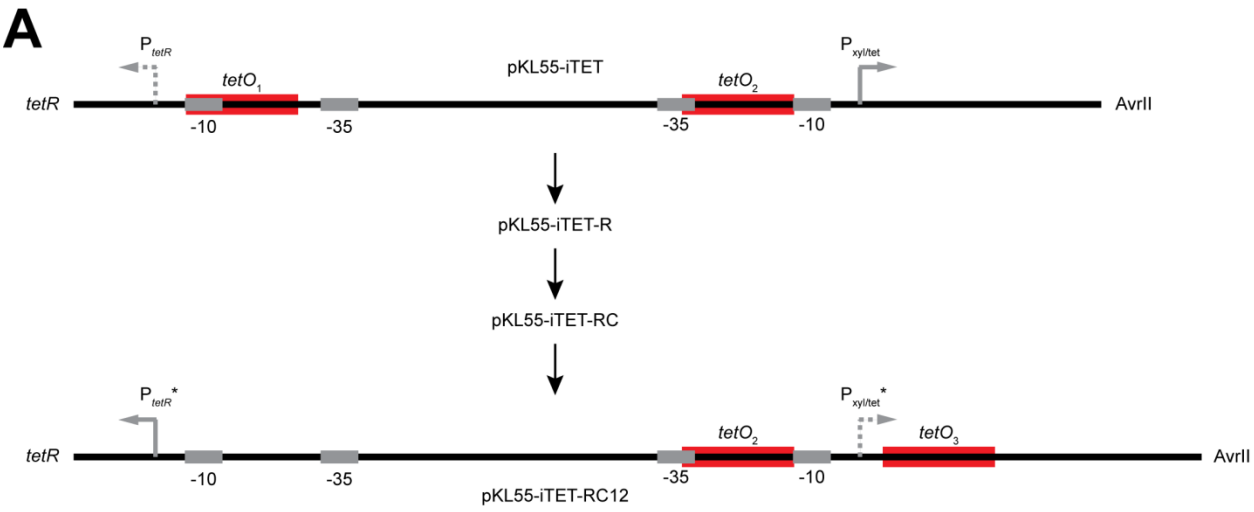


Figure 2-16. Modification of pKL55-iTET to create pKL55-iTET-RC12 second-generation target insertion vectors.

(A) Scaled representations of the ATc-inducible promoter region found within pCL55-iTET-derived vectors between the *tetR* start codon (left) and the multiple cloning site beginning with AvrII (right). The pKL55-iTET first-generation architecture is schematized at the top. Three-step modification produced the pKL55-iTET-RC12 second-generation architecture (bottom) based off pRAB12 (Helle et al., 2011). The additional 17 bp in pKL55-iTET-RC12 creates a new TetR binding site ('$tetO_3$') downstream of $P_{xyl/tet}$*. Meanwhile, the TetR binding site that overlaps the P_{tetR} promoter ('$tetO_1$') is disrupted at the modified promoter, P_{tetR}*. Bent arrows denote the approximate positions of transcriptional start sites for each promoter; dashed bent arrows signify that their promoter 's consensus is predicted to drive weaker constitutive transcription relative to its first- or second-generation counterpart.

three steps (**Figure 2-16**). Firstly, I introduced the two pRMC2 (Corrigan and Foster, 2009) SNPs which are thought to strengthen *tetR* expression by improving its promoter's –10 consensus, and perhaps also disrupting the consensus of its overlapping *tetO* operator (Stary et al., 2010). In turn, stronger expression of the *tetR* repressor gene is thought to reduce leakiness from $P_{xyl/tet}$-based promoters. In the next step, I weakened the $P_{xyl/tet}$ promoter by introducing a SNP into its –10 consensus. Finally, I introduced a second *tetO* operator downstream to strengthen the potential for repression. The resulting pKL55-iTET-RC12 vector with sequences matching to pRAB12 (Helle et al., 2011) was used in downstream manipulations.

In parallel efforts, I continued to clone spacers that would corroborate the strand-biased pattern of type III phage targeting. These included the 4B, 4T, 32B, 43B, and 43T spacers (**Figure 2-2**). The 4B/4T spacers matched to ϕNM1's *cI*-like repressor gene, while the 43B/43T spacers matched to its "head protein" gene (**Table 2-1**). By designing these additional ϕNM1-targeting spacers, I was able to further characterize the ϕNM1γ6 escape phenotype. For example, the 4B spacer was escaped by ϕNM1γ6, while the 4T, 16T, 43T and 61T-1 spacers were not (**Figure 2-10A**). Assuming a directional (i.e., non-template strand) transcription requirement for type III-A targeting, this was consistent with a reduction in leftward transcription across the spacer 2B and 4B targets that is predicted to result from the upstream promoter SNP (**Figure 2-10C**). In addition, the spacer 4B CRISPR-Cas plasmid—which provides immunity to wild type ϕNM1 in

plaque assays (**Figure 2-2C**)—was found to reduce lysogen transformation efficiency to a similar degree as the spacer 2B plasmid (**Figure 2-12C**). In this case, the reduced transformation efficiency observed solely with these 2B and 4B spacers would be explained if their targets are the only among those that I tested which are actively transcribed in the proper orientation within lysogens. Meanwhile, the other spacers target ϕNM1 within lytic regions, or within the same lysogeny region but on the template strand of transcripts arising from the leftward promoter upstream of its *cI*-like repressor. The original 2T spacer ($2T^{old}$) was also cloned at this time, but provided virtually no targeting in my plaque and transformation assays. It was later realized that the exact inverse of the 2B spacer (found in this original $2T^{old}$ construct) contained excessive 5′ handle complementarity which probably prevented type III targeting. This spacer was therefore re-cloned with a 2 nucleotide shift (**Table 2-1**) that produced the expected targeting behavior in efficiency of plaquing (**Figure 2-2C**) and transformation efficiency assays (**Figure 2-12**). However, transformants harboring this adjusted 2T CRISPR-Cas system generally exhibited a small-colony phenotype. Given what I had observed with my leaky first-generation chromosomal target vectors, where small colonies were observed in one of the orientations even in the absence of ATc, but not at all in the other (data not shown), it seemed plausible that this could reflect low levels of chromosomal targeting. It also seemed plausible that this could result, specifically, from low levels of rightward transcription across the *xis* ORF in lysogens. At this point, we

were still basing our interpretation of these results on various transcriptome *predictions*, largely inferred from the directionality of predicted ORFs within the ϕNM1 annotation. It was clear that actual transcriptome data from ϕNM1-infected cells could serve to validate these interpretations, so I discussed plans with Dr. Marraffini to obtain this using 'stranded' RNA-seq. Wenyan was also planning an RNA-seq experiment to study plasmid targeting at the time, so we ended up pooling a few samples together in what was assumed to be a mere pilot run. Ultimately, the stranded transcriptome analysis for one of the samples in this run—a logarithmically growing RN4220::ϕNM1 single lysogen—was used for publication (Goldberg et al., 2014). It confirmed, among other things, that there is indeed a slightly higher peak of rightward transcription across the spacer 2B/2T target region, relative to other target regions (**Figure 2-17**). Another sample from this run with ϕNM1-infected non-lysogens harvested 15 minutes post-infection also looked promising, and so I made plans to repeat this with four time points in succession, and also include two equivalent early time points with ϕNM1γ6.

In additional parallel efforts, I was cloning a type II CRISPR-Cas plasmid for targeting of ϕNM1 that I figured would be useful as a control for CRISPR-Cas targeting of lytic regions in *S. aureus* lysogens. In other words, I wanted to show that a presumably transcription-independent CRISPR-Cas system would prevent lysogeny in this host background even if type III systems targeting the same lytic region(s) did not. The first type II spacer I designed (16B-tII) overlapped the 16B/16T target region, and

was intended to be able to target my 'first-generation' chromosomal target insertions in either orientation, regardless of the ATc inducer's presence or absence. My colleague Wenyan had already cloned the entire *S. pyogenes* SF370 type II-A CRISPR-Cas system on pWJ40, so I made plans to modify its CRISPR array, similar to as I had done for

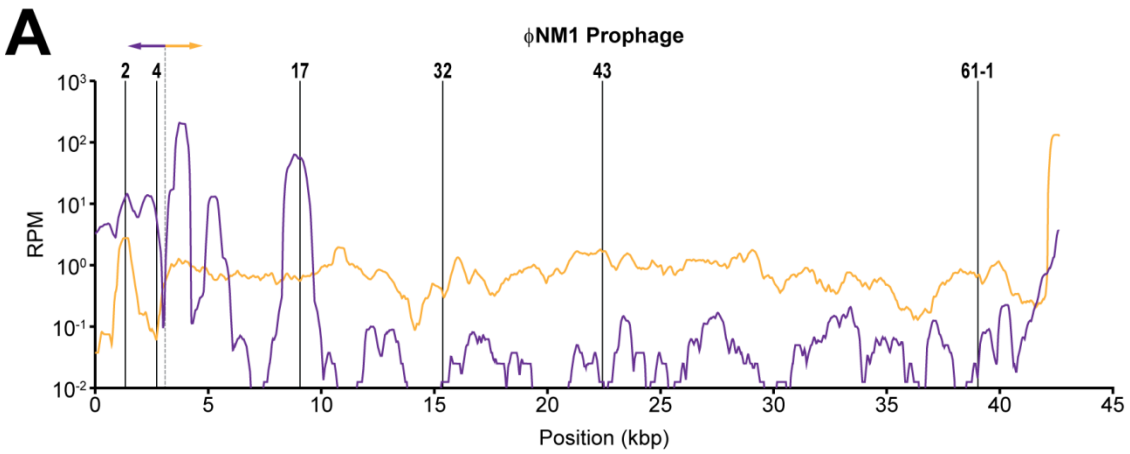


Figure 2-17. ϕNM1 prophage transcription profiles from a logarithmically growing lysogen.

(A) Rightward and leftward expression values are plotted as golden and purple lines, respectively, in reads per million (RPM). The position of relevant spacer targets are indicated with vertical solid lines. The dotted line with divergent arrowheads marks the position of the early lytic control region (^{E}LCR) with divergently oriented P_{cI} and P_{cro} promoters. To improve readability, curves were smoothened by plotting average reads per million values over a 500 bp sliding-window. Noticeable leftward expression originates from the ^{E}LCR and a few upstream regions that presumably encode functions involved in lysogenic maintenance and/or prophage morons (Brüssow et al., 2004; Cumby et al., 2012). Rightward transcription was weaker than leftward transcription as expected, but not absent, and includes a peak that overlaps the *xis* ORF targeted by the 2B/2T spacers.

pWJ30β. The first step involved reducing the CRISPR array of pWJ40 to a single repeat, creating pGG32. Type II systems were not known to function without a downstream repeat flanking their spacer(s) of interest, so I would have to construct the 16B-tII CRISPR-Cas plasmid in two steps using my blunt ligation method, without drastically increasing my primer lengths. The first step would add the spacer, and the second would add the downstream repeat. By the time this construct was cloned, I had already elected to use a different target region in my 'second-generation' chromosomal targeting vectors (explained below). Fortunately, Dr. Bikard was cloning his first type II parent vectors with a single placeholder spacer containing two oppositely oriented BsaI restriction sites. These allow for 1-step replacement of the placeholder spacer with annealed oligo pairs containing a (spacer) sequence of interest. My subsequent cloning of type II CRISPR-Cas plasmids utilized this method, along with Dr. Bikard's parent vectors. Therefore, although the pGG33 construct with spacer 16B-tII was capable of targeting ϕNM1, it was never included in my published work. However, along with the ϕNM4γ4 clear mutant (lytic) phage I had isolated, the pGG32 (single repeat) precursor ultimately proved useful for my colleague, Robert Heler, in some of his type II adaptation experiments (Heler et al., 2015). Meanwhile, after learning of the BsaI cloning method, I adopted it to construct subsequent type III CRISPR-Cas plasmids as well. This was initially accomplished with a new type III parent vector called pGG3-BsaI, which I created by introducing the same placeholder spacer from Dr. Bikard's

vectors at position two of the CRISPR array (just as I had previously introduced my phage-targeting spacers into pGG3).

When designing my second-generation chromosomal targeting system, I elected

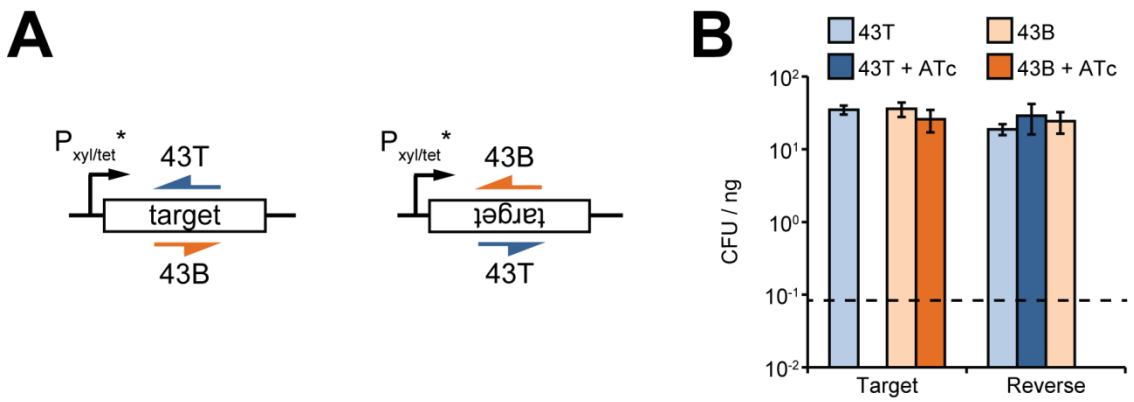


Figure 2-18. Transformation of strains containing second-generation integrative target vectors with type III CRISPR-Cas plasmids harboring spacer 43T or 43B. **(A)** Schematics summarizing the orientation of spacer 43T or 43B crRNAs (orange and blue arrows) relative to a forward (left) or reversed (right) target sequence downstream of the $P_{xyl/tet}$* promoter in second-generation integrative vectors. Leftward-oriented arrows (depicted above) represent crRNAs with complementarity to the non-template strand relative to rightward $P_{xyl/tet}$* transcription; rightward-oriented arrows (depicted below) represent crRNAs with complementarity to the template strand. **(B)** Transformation of RN4220-derived strains with spacer 43T or 43B CRISPR-Cas plasmids in the presence or absence of ATc. Electrocompetent cells possessed either the forward 'Target' or 'Reverse' target vector insertions, and were transformed using the same miniprep for either CRISPR-Cas plasmid cloned in RN4220. Transformation efficiencies were calculated as the number of transformants per ng of miniprepped DNA used for electroporation (CFU / ng). Error bars, mean ± s.d. (n = 3, technical replicates).

to use the 43B/43T spacer pair (**Figure 2-18A**). This was done in part because, as Dr. Marraffini pointed out to me, it would be easier to explain that we'd taken a phage lytic gene target if its gene product was known to be a head protein, as opposed to a 'hypothetical' protein. Furthermore, shortly after I designed my first type III spacer (16B), which was 34 bp in length, I realized that the majority of sequenced spacers our lab had identified in staphylococcal type III-A systems were actually either 35 or 36 bp, and began designing my spacers as such. Among these were the 43B and 43T spacers, which each had 36 bp. Therefore, I figured it would also be appropriate to utilize spacers exhibiting a more representative length, such as these. This time, the system was ATc-inducible; virtually no targeting was apparent in the absence of ATc for either orientation (**Figure 2-18B**).

After hearing of these results, Dr. Bikard voiced some concern that addition of the ATc inducer might be simply unmasking a strand discrepancy that arises during DNA replication (e.g., leading vs. lagging strand discrepancies), since it was always the same DNA strand being targeted in my constructs relative to the chromosome's origin of replication. I was not particularly concerned in this instance, because we had already observed type III targeting on both strands in multiple contexts. For one, my advisor originally found this while transforming with pC194-based plasmids in which he'd inserted the *S. epidermidis* RP62a *nes* spacer's target region in both orientations (Marraffini and Sontheimer, 2008). Wenyan had essentially also shown this in pG0400

conjugation assays, when comparing the wild type *nes* spacer to his inverted (and 2 bp-shifted) *nes3* spacer as described above. Now, I was finding this as well in plaque assays with ϕNM1, using my 4B/4T spacer pair (**Figure 2-2C**). The Deng *et. al.* report (Deng et al., 2013) on PAM-independent targeting by a type III-B system in *S. islandicus* REY15A was also encouraging in this regard, and it was released online shortly after I obtained preliminary replicates with my second-generation chromosomal targeting vectors. Using plasmid-targeting transformation assays, this study essentially confirmed what I had found with my first-generation chromosomal targeting vectors: type III targeting was only licensed by crRNAs with complementarity to the non-template strand of a target sequence relative to a nearby promoter. This work also found that you could complement the non-targeting, template-strand-complementary scenario by introducing an oppositely oriented promoter downstream of the target, to ensure that the spacer's crRNAs would once again bear complementarity to a non-template strand (in this case relative to the secondary promoter). Therefore, they too had demonstrated targeting on both strands of a DNA element. Presumably, these results would not have been observed if there was a strict replicative requirement for targeting of one particular strand but not the other. Based on these data, however, I could not exclude the possibility that a replicative strand discrepancy might be sufficient to license targeting on one strand in certain contexts, even if type III targeting were otherwise transcription-dependent. To rule out such possibilities, I inverted the *attP* region (including putative

cis elements adjacent to the core motif (Lee and Buranen, 1989)) on my pKL55-iTET-RC12-derived vectors such that they would integrate in the opposite orientation relative to their host's chromosome. Using PCR, I confirmed that my inversions were successful; the vector-external primers must be swapped in order to amplify across an *attL* or *attR* junction with the same vector-internal primer adjacent to *attP* (**Figure 2-19**). Essentially the same results were observed in transformation assays when I tested these inverted

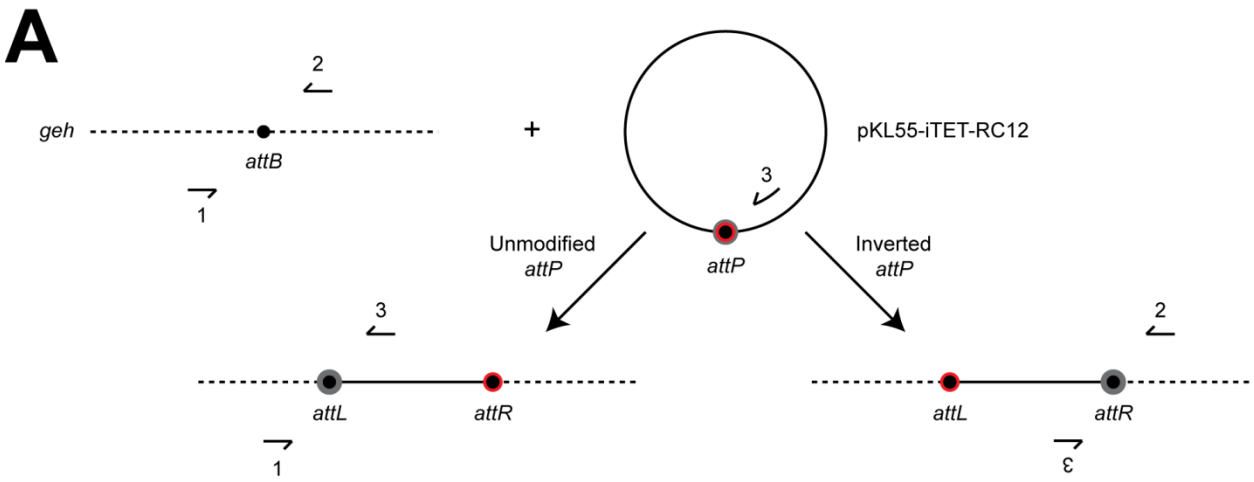
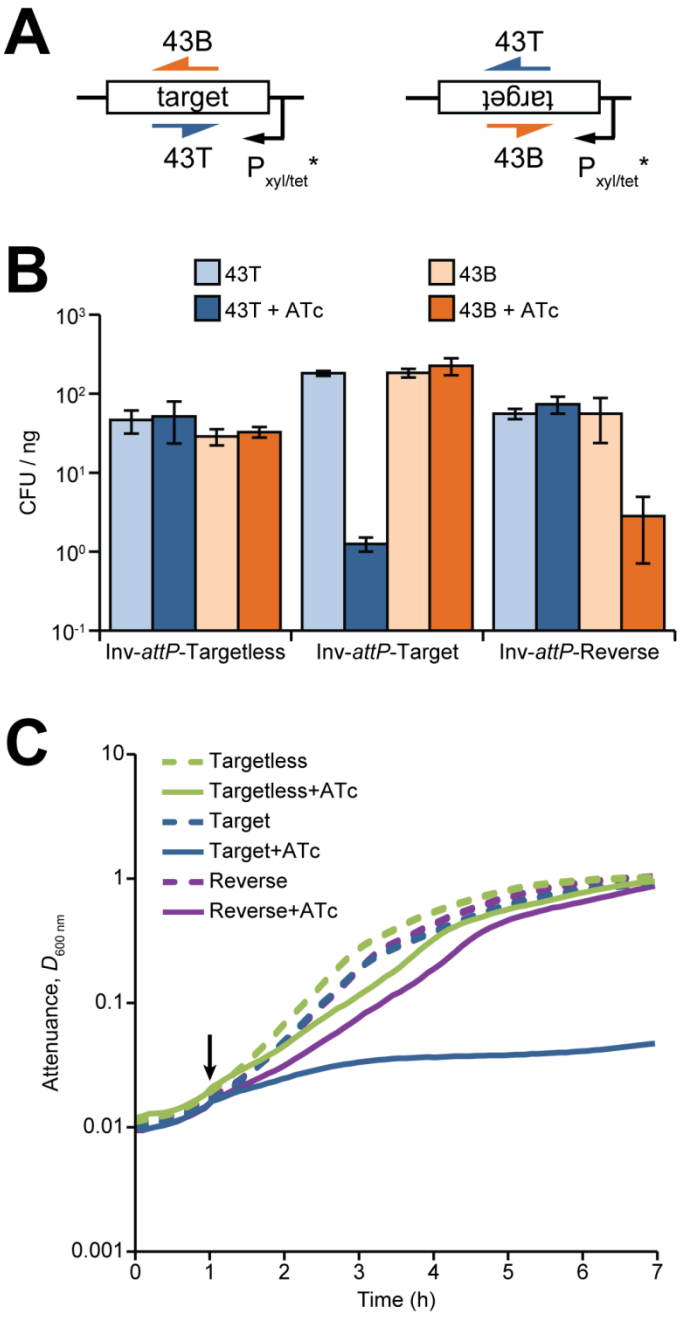


Figure 2-19. Schematic diagram summarizing site-specific integration by pKL55-iTET-RC12-derived vectors and their inverted *attP* variants.

(A) As described for the original pCL55 vectors (Lee et al., 1991), site-specific integration of pKL55-iTET-RC12-derived vectors utilizes an *attB* attachment site in the *S. aureus* RN4220 chromosomal *geh* locus, along with an *attP* motif (derived from bacteriophage L54a) present on the plasmids. Primers '1' (oGG50) and '3' (oGG96) can amplify across the *attL* junction that results from recombination with the unmodified *attP*, while primers '2' (oGG51) and 3 only amplify across the *attR* junction that results from integration of a vector with an inverted *attP* motif. Features of the diagram are scaled arbitrarily.

Figure 2-20. Testing of second-generation inducible targeting vectors integrated in the inverse orientation relative to the chromosomal origin of replication.

(A) Schematics summarizing the orientation of spacer 43T or 43B crRNAs (orange and blue arrows) relative to a forward (left) or reversed (right) target sequence downstream of the $P_{xyl/tet}$* promoter in second-generation, inverted-*attP* integrative vectors. Leftward-oriented arrows (depicted above) represent crRNAs with complementarity to the template strand relative to leftward $P_{xyl/tet}$* transcription; rightward-oriented arrows (depicted below) represent crRNAs with complementarity to the non-template strand. **(B)** Transformation of RN4220-derived strains with spacer 43T or 43B CRISPR-Cas plasmids in the presence or absence of ATc. Electrocompetent cells possessed either the 'Inv-*attP*-Targetless' vector insertion, the forward 'Inv-*attP*-Target' vector insertion, or the 'Inv-*attP*-Reverse' target vector insertion, and were transformed using the same miniprep for either CRISPR-Cas plasmid cloned in RN4220. Transformation efficiencies were calculated as the number of transformants per ng of miniprepped DNA used for electroporation (CFU / ng). Error bars, mean ± s.d. (n = 3, technical replicates). **(C)** High-resolution growth curves of Inv-*attP* strains harboring the spacer 43T CRISPR-Cas plasmid. Each of the indicated strains was grown in the presence (solid lines) or absence (dashed lines) of ATc addition at the indicated time point (black arrow), and were originally isolated as CRISPR-Cas plasmid transformants plated in the absence of ATc.


A
43B
target
43T
P xyl/tet *
43T
target
43B
P xyl/tet *
B
43T
43T + ATc
43B
43B + ATc
CFU / ng
Inv-attP-Targetless
Inv-attP-Target
Inv-attP-Reverse
C
Targetless
Targetless+ATc
Target
Target+ATc
Reverse
Reverse+ATc
Attenuance, D600 nm
Time (h)

attP variants with both forward and 'reverse' target insertions (**Figure 2-20A**), along with my inverted variant of the 'targetless' RC12 parent vector as a control (**Figure 2-20B**). The result was reassuring. At least in this region of the chromosome, unidirectional transcription appears to be sufficient to drive strand discrepancies associated with type III-A targeting. I also tested these vectors with the spacer 43T CRISPR-Cas plasmid in high-resolution growth curves (**Figure 2-20C**), and found that severe growth reductions were only observed upon addition of ATc with the forward target insertion (where spacer 43T's crRNAs are complementary to the target's non-template strand). Note that transformation efficiency data collected with the parental orientation was arbitrarily presented as 'Inv-*attP*' in the publication's extended data section (Goldberg et al., 2014), because I had not collected high-resolution growth curves with the parental orientation, and we wanted to introduce transformation and growth curve data together in a primary figure. It was decided that showing data from two different sets of strains side by side could be potentially awkward or confusing, so the main figure data utilized the inverted *attP* constructs for which I'd collected both types of data, but it was simply presented as the default orientation rather than Inv-*attP*. Note, also, that some growth reduction is apparent upon addition of ATc at this concentration (0.5 μg/ml), even for the targetless control strain (**Figure 2-20C**). This concentration was arbitrarily adopted from a modified pKOR (Bae and Schneewind, 2006) protocol circulating in the lab, which uses the $P_{xyl/tet}$ promoter for antisense *secY*

transcription. In different assays more recently, I've worked with lower ATc concentrations of 0.250 μg/ml and 0.125 μg/ml. Some general growth reduction is still observable at 0.250 μg/ml, although not at 0.125 μg/ml. This is consistent with the findings of a previous study on $P_{xyl/tet}$-based promoters in *S. aureus*, which reported that growth is not noticeably affected at ATc concentrations of 0.128 μg/ml, although strong induction could still be achieved here (Corrigan and Foster, 2009). Another study reported working with concentrations up to ~0.185 μg/ml (Helle et al., 2011). Concentrations in the range of 0.125-0.185 μg/ml are therefore probably ideal for the purpose of optimally inducing $P_{xyl/tet}$-based promoters in *S. aureus* with minimal toxicity.

At this point, I had lingering concerns of my own, regarding the absence of a repeat downstream from my phage-targeting spacers. Perhaps this CRISPR architecture was producing unforeseen effects in phage assays—or worse, allowing for the conditional tolerance phenomenon altogether! I first became suspicious when I noticed that the pGG3 plasmid has a phenotype in the conjugation assay relative to either pWJ30β or my phage-targeting derivatives, which, unlike pGG3, all have a repeat downstream of the *nes* spacer. Although pGG3 comparably reduces conjugation efficiency when colonies are scored by eye, it allows for something which I could only describe as 'pin-point puncta' to form atop the background of counter-selected ('dead') cells. I was never able to re-isolate pin-point puncta from these background spots by re-streaking; and, upon prolonged incubation, they never outgrew the background (and in

fact seemed to become part of the local lawn). In hindsight, the nature of these puncta might have been partially clarified with the use of a basic light microscope! In any case, before confirming that the downstream repeat was sufficient for alleviation of the pin-point puncta phenotype, I had decided to clone a repeat downstream of the 32T*^{B} spacer in my pGG14 plasmid (to produce the pGG22 plasmid). Testing of this 32T*^{B}(-R) (pGG22) plasmid had initially allayed my concerns because the addition of the repeat did not produce a noticeable phenotype in ϕNM4 plaque assays (data not shown). However, I subsequently confirmed that the addition of a repeat downstream of the *nes* spacer in pGG3 (creating pGG25) was sufficient to eliminate the pin-point puncta phenotype. This confirmed that, at least in certain contexts, the absence of a downstream-flanking repeat could produce a phenotype. Hence, there was still cause for concern that the absence of a repeat downstream from my phage-targeting spacers could affect my results in other assays. For example, perhaps the 'tolerance' of chromosomal targets I was finding in various assays could be an unnatural by-product of this. Note that these CRISPR array architectures are not found in natural staphylococcal type III-A systems. I therefore compared transformation with the spacer 32T*^{B} and 32T*^{B}(-R) CRISPR-Cas plasmids in lysogen transformation assays. Each plasmid produced similar numbers of TB4::ϕNM4 lysogen transformants, and also similar numbers of RN4220::ϕNM1 lysogen transformants (data not shown). This, finally, allayed my concern that the absence of a downstream repeat was not required

for tolerance in *S. aureus* per se. My colleague Wenyan would later encounter more striking phenotypes in some of his other assays if the targeting spacer lacked a downstream repeat, suggestive of partial loss-of-function (personal communication).

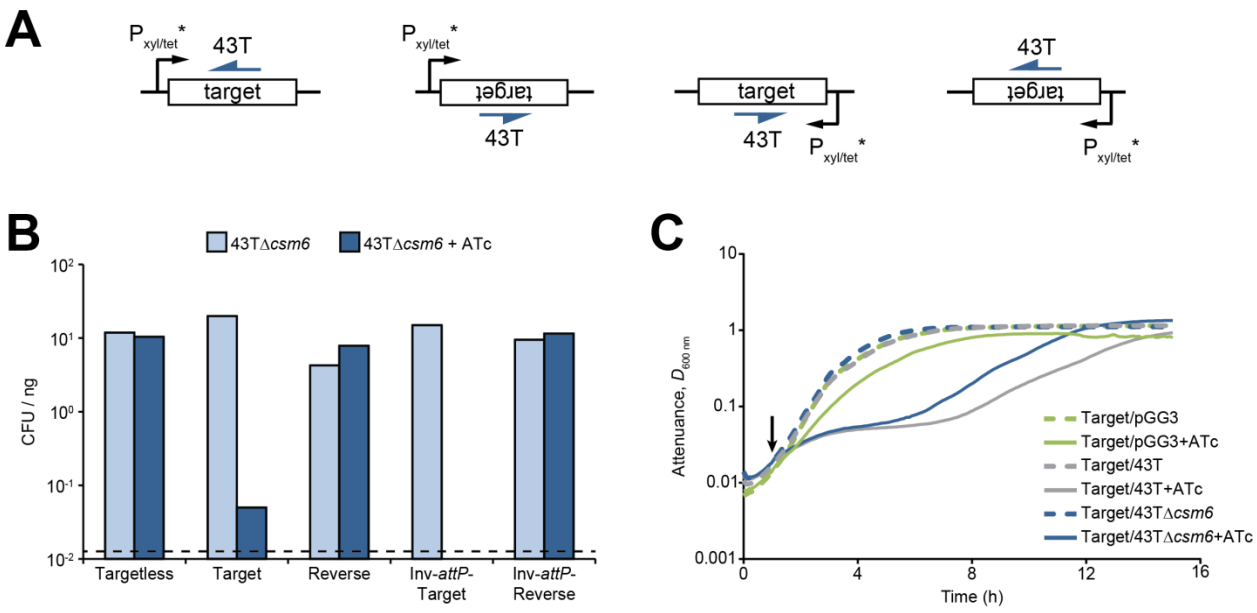
Regarding the mechanism(s) of transcription-dependent targeting I was observing in my *in vivo* experiments, some additional concerns were raised by Dr. Bikard and other members of the lab at this time. Although direct evidence for target DNA cleavage by type III systems had yet to be obtained by any groups, cleavage of RNA targets had been demonstrated (Hale et al., 2012; Hale et al., 2009; Zhang et al., 2012). On the one hand, regarding the possibility that our *S. epidermidis* RP62a system could also cleave (specific) target RNAs, I was fairly confident this would be insufficient to explain the *in vivo* results we'd observed thus far. For example, in the original work by Dr. Marraffini, knockdown of RNA at the *nickase*-derived target loci in pG0400 and pC194-based plasmids would not be expected to impair plasmid stability (Marraffini and Sontheimer, 2008). Likewise, knockdown of leftward lysogenic transcripts by the 2B or 4B spacers would not be expected to impair lytic infection by ϕNM1, and yet they licensed immunity in plaque assays (**Figure 2-2C**). I had now shown that transcription-dependent targeting of engineered chromosomal insertions is also toxic, even though these entire insertions are non-essential for host growth (**Figures 2-15, 2-18, & 2-20**). On the other hand, some compelling hypotheses had meanwhile been raised about the role of the Csm6 and Csx1 proteins found in various type III systems, including the Csm6 of

our RP62a system. Based on homology to RNase domains found in previously predicted or confirmed toxin-antitoxin systems, the HEPN RNase domains found in Csm6/Csx1 proteins were proposed to offer a similar (potentially toxic) function (Anantharaman et al., 2013; Makarova et al., 2012). For example, toxin activities can be leveraged by abortive infection systems to prevent the spread of phages, by altruistically killing infected cells before their progeny phage are released (Labrie et al., 2010). Therefore, we could not formally exclude the possibility that the type III-A targeting phenomena observed in the lab thus far had resulted from one or more toxins being activated by the presence of (transcribed) target DNA. Further adding to the complexity of the matter, Dr. Hatoum-Aslan (with help from Inbal Maniv and Dr. Samai) was finding that *csm6* is required for immunity to pG0400 in conjugation assays with the wild type *nes* target, although it does not stably associate with the complex. These findings (Hatoum-Aslan et al., 2014; Hatoum-Aslan et al., 2013), still unpublished at the time, could alternatively be explained if Csm6 were contributing to plasmid (DNA) targeting in some way. Although Csm6 itself had not been shown or predicted to harbor DNase domains (to my knowledge), it was not inconceivable that it could be recruited by the type III complex to enable or augment DNA cleavage in a manner analogous to recruitment of Cas3 nucleases by type I systems (Westra et al., 2012). In an effort to address these concerns, I constructed a Δ*csm6* knockout variant of my spacer 43T CRISPR-Cas plasmid for testing in transformation assays with my second-

generation inducible target vectors. Presumably, if transformation efficiency were still reduced in a transcription-dependent manner with this plasmid, I could rule out the possibility that Csm6/HEPN-mediated cellular suicide is strictly required for our *in vivo* targeting phenomena, and provide evidence in support of a model where the type III-A targeting complex is sufficient for DNA targeting (at least in certain contexts). Wenyan was just beginning his study on Csm6 around this time, and kindly pointed me to Dr. Marraffini's primers for deleting *csm6* from type III-A plasmids, such as pWJ30β and its derivatives. I tested the spacer 43T(Δ*csm6*) CRISPR-Cas plasmid with forward and reverse target insertions in both orientations relative to the chromosomal origin of replication, and found the same trends I'd observed with the wild type plasmid in transformation assays (**Figure 2-21B**). Inducible chromosomal targeting was also monitored in high-resolution growth curves with the 'Inv-*attP*-Target' strains, and again the toxicity licensed by the 43T(Δ*csm6*) plasmid was found to be comparable to that of wild type (**Figure 2-21C**). The recovery observed after ~6-8 hours with both wild type and Δ*csm6* targeting plasmids was assumed at the time to result from escape mutants in the population. It's also possible that the effects of ATc were simply wearing off around this time; other members of the lab have anecdotally reported that re-supplementation is required to maintain active concentrations of ATc in liquid cultures. In hindsight, plating after different durations of ATc treatments to measure viable CFUs might have helped to determine whether type III-A chromosomal targeting is lethal in

Figure 2-21. Inducible chromosomal targeting using a Δ*csm6* variant of the spacer 43T CRISPR-Cas plasmid.

(A) Schematics summarizing the orientation of spacer 43T crRNAs (blue arrows) relative to forward or reversed target sequences downstream of the $P_{xyl/tet}$* promoter in second-generation unmodified-*attP* (left two schematics) and inverted-*attP* (right two schematics) integrative vectors. Leftward-oriented arrows (depicted above) represent crRNAs with complementarity to the template strand relative to leftward $P_{xyl/tet}$* transcription; rightward-oriented arrows (depicted below) represent crRNAs with complementarity to the non-template strand. **(B)** Transformation of RN4220-derived strains with the spacer 43T(Δ*csm6*) CRISPR-Cas plasmid in the presence or absence of ATc. Electrocompetent cells possessed either the 'Targetless' vector insertion, the forward 'Target' vector insertion, the 'Reverse' target vector insertion, the 'Inv-*attP*-Target' vector insertion, or the 'Inv-*attP*-Reverse' target vector insertion, and were transformed using a single miniprep of the CRISPR-Cas plasmid cloned in RN4220. Transformation efficiencies were calculated as the number of transformants per ng of miniprepped DNA used for electroporation (CFU / ng). Dashed line represents the limit of detection under these assay conditions. Additional replicates were not performed. **(C)** High-resolution growth curves of 'Inv-*attP*-Target' strains harboring either the pGG3, spacer 43T, or spacer 43T(Δ*csm6*) CRISPR-Cas plasmid as indicated. Each of the indicated strains was grown in the presence (solid lines) or absence (dashed lines) of ATc addition at the indicated time point (black arrow), and were originally isolated as CRISPR-Cas plasmid transformants plated in the absence of ATc. Growth curves represent data from single biological replicates; additional replicates were not performed.


A
P xyl/tet *
43T
target
target
43T
target
43T
P xyl/tet *
43T
target
P xyl/tet *
B
43TΔcsm6
43TΔcsm6 + ATc
CFU / ng
10^2
10^1
10^0
10^-1
10^-2
Targetless
Target
Reverse
Inv-attP-Target
Inv-attP-Reverse
C
Attenuance, D600 nm
10
1
0.1
0.01
0.001
0
4
8
12
16
Time (h)
Target/pGG3
Target/pGG3+ATc
Target/43T
Target/43T+ATc
Target/43TΔcsm6
Target/43TΔcsm6+ATc

this assay. Nevertheless, these preliminary results did well to satisfy my curiosity, as well as the lab's. A small plaque phenotype was also noted while testing this 43T(Δ*csm6*) plasmid with ϕNM1, indicative of strongly impaired immunity. Ultimately, Wenyan would go on to show that Csm6 (and its HEPN domain's RNase activity in particular) is required for robust immunity in this context because this particular target region is transcribed late during the ϕNM1 lytic cycle (Jiang et al., 2016).

All of these findings were exciting and encouraging to me, but their general novelty was now somewhat compromised by a publication showing evidence in support of a transcription-dependent DNA targeting mechanism for an archaeal type III system (Deng et al., 2013). Accordingly, I made an effort to shift our study's focus onto the more novel findings concerning temperate phages: namely, that transcription-dependent immunity to lytic infection provides a mechanism for averting chromosomal targeting during lysogeny. The strategy was to distinguish type II systems and other transcription-independent CRISPR-Cas systems from type III systems, in that the former would provide indiscriminate immunity both to lytic and lysogenic infections by temperate phages, while the latter would provide immunity only under certain conditions (which were generally only met during lytic infections). Hence, I initially took to distinguishing the type III targeting mechanism in terms of its capacity for 'conditional targeting' of temperate phages. It was later pointed out to me by Dr. Marraffini that the non-targeting scenario we observe during lysogeny—when spacers

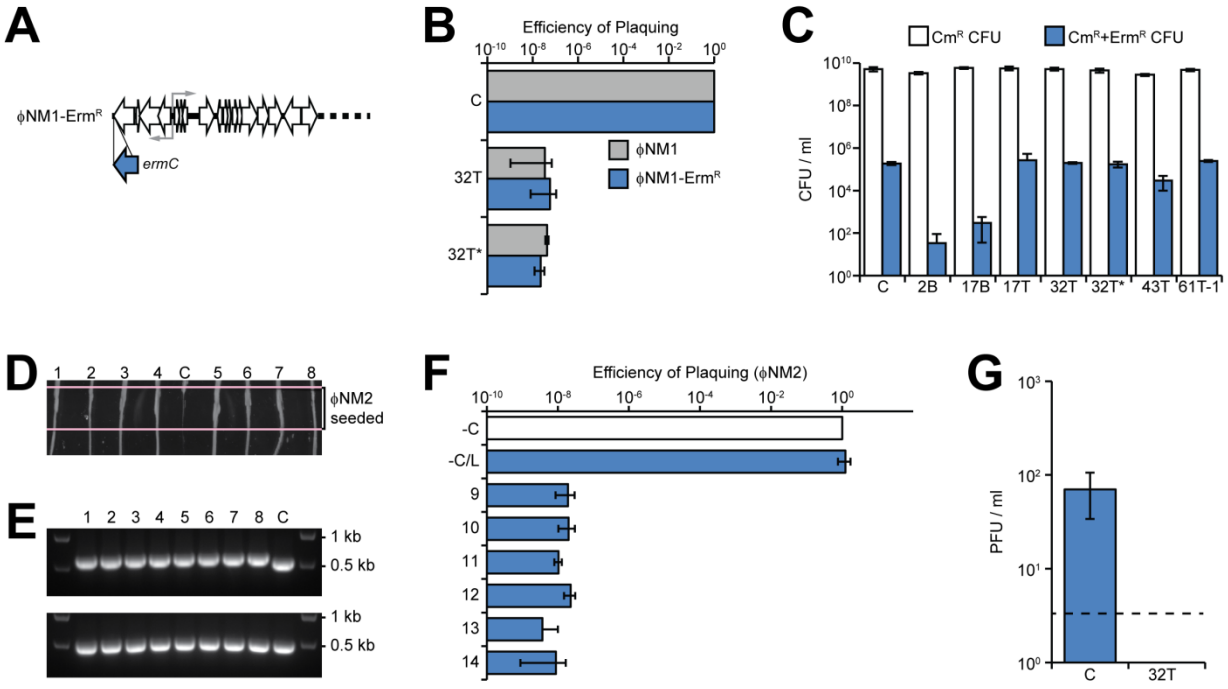
match to lytic genes—could be considered a form of 'tolerance'. Focusing on the type III-A system's capacity for tolerance, in turn, allowed us to expand on the analogy between CRISPR-Cas systems and mammalian immune systems (Goldberg and Marraffini, 2015). I elaborate on this point in **Chapter 4**. However, considering that the type III system maintains its capacity for targeting during lytic infections, and that tolerance of prophages is not absolute (e.g., during prophage induction), we ultimately decided that the phenomenon could best be summarized as 'conditional tolerance'. Experiments described in the next section helped us to convey this point more clearly in the publication (Goldberg et al., 2014).

2.5 Transcription-dependent targeting by a type III-A CRISPR-Cas system allows conditional tolerance of temperate phages

In stark contrast to what was reported for *E. coli* with a type I CRISPR-Cas system targeting temperate phage λ (Edgar and Qimron, 2010), I was finding that *S. aureus* strains with various type III-A CRISPR-Cas plasmids could tolerate lysogenic infections by the same Newman temperate phages which are resisted during lytic infections. Moreover, this conditional tolerance phenomenon was found to be contingent upon immunity directed at lytic cycle genomic regions which are sufficiently repressed during lysogeny (**Figures 2-11, 2-12, & 2-17**). However, my core findings thus far had been gleaned from results with lysogenic transformants and egg yolk screen isolates. In

order to corroborate these findings, and also demonstrate more quantitatively that tolerance allows lysogeny to be established efficiently during infection, I adopted a lysogenization efficiency assay akin to what had been tested with the type I system (Edgar and Qimron, 2010). For sensitive hosts, lysogenization frequency can be calculated by using survival as a direct readout for lysogenization. For CRISPR-resistant hosts, the number of cells which survive infection is not necessarily correlated with lysogenization frequency, because CRISPR immunity can allow cells to survive infection even in cases where that infection was not lysogenic. To overcome this, the type I study utilized a recombinant λ phage carrying a tellurite resistance marker that is expressed from the prophage in lysogens. Thus, by virtue of the tellurite-resistance-converting prophages they've acquired, lysogenized CFU can be quantified after brief infections (Edgar and Qimron, 2010). Accordingly, I endeavored to engineer a similar recombinant of ϕNM1. Unsure of where to insert a resistance marker, however, I asked Dr. John Chen of the Novick laboratory via e-mail. His reply was prompt and informative. He explained, among other things, that marker insertions tend to be functional on either side of the phage attachment site, and that erythromycin resistance would be preferable to kanamycin resistance for *S. aureus*. While it was too late for my pKL55-iTET vectors conferring kanamycin resistance, I heeded his advice regarding erythromycin resistance going forward. I designed a pKOR-based allelic exchange plasmid (Bae and Schneewind, 2006) for inserting an erythromycin resistance marker,

Figure 2-22. Conditional tolerance in RN4220 hosts infected with ϕNM1-Erm^R. **(A)** Schematic diagram depicting a relevant genomic region of ϕNM1-Erm^R that includes the *ermC* insertion (blue thick arrow) immediately downstream of *int*. ORFs were scaled by length in accordance with the ϕNM1 (NC_008583.1) or pE194 (NC_005908.1) annotations. Grey bent arrows signify the presence of divergently oriented promoters in the ^{E}LCR of ϕNM1. **(B)** Plaquing efficiency of ϕNM1 and ϕNM1-Erm^R on lawns of RN4220 harboring the 32T or 32T* CRISPR-Cas plasmids, or the pGG3 control plasmid (C). Error bars, mean ± s.d. (*n* = 3, technical replicates). **(C)** Lysogenization of RN4220 derivatives harboring CRISPR-Cas plasmids with the indicated targeting spacers, or the pGG3 control plasmid (C), quantified as the concentration (CFU / ml) of chloramphenicol- and erythromycin-resistant colonies obtained upon plating after brief infection with ϕNM1-Erm^R. Prior to infection, cultures were plated in the presence of chloramphenicol alone to estimate the concentration (CFU / ml) of total recipients harboring CRISPR-Cas plasmids. Error bars, mean ± s.d. (*n* = 3, biological replicates). **(D)** ϕNM2-sensitivity assay for eight (1-8) randomly-selected clones isolated from ϕNM1-Erm^R lysogenization experiments with the spacer 32T CRISPR-Cas plasmid. A ϕNM1-Erm^R lysogen harboring the pGG3 plasmid was used as the sensitive control (C). After re-streak isolation, single colonies were streaked through the ϕNM2-seeded region from top to bottom. **(E)** PCR amplification of the CRISPR array (upper panel) and spacer 32T target region (lower panel) from the isolates tested in 'D'. Size markers of 1 kb and 0.5 kb are indicated. All eight target region PCR products were sequenced by Sanger and no mutations were found (data not shown). **(F)** ϕNM2 plaquing efficiency on soft agar lawns of an additional six randomly selected clones (9–14) isolated from ϕNM1-Erm^R lysogenization experiments with the spacer 32T CRISPR-Cas plasmid. The ϕNM1-Erm^R lysogen harboring the pGG3 control plasmid was also tested (–C/L), and plaquing efficiency with the non-lysogenic indicator strain harboring pGG3 is shown for comparison (–C). Error bars, mean ± s.d. (*n* = 3, technical replicates). **(G)** Plaque-forming potential of filtered overnight culture supernatants from a ϕNM1-Erm^R-lysogenic isolate harboring the spacer 32T CRISPR-Cas plasmid. PFU concentrations (PFU / ml) were measured by plating on soft agar lawns of RN4220 harboring either the pGG3 control (C) or spacer 32T CRISPR plasmid. Dashed line represents the limit of detection under these assay conditions. Error bars, mean ± s.d. (*n* = 3, biological replicates). Pictures in 'D' and 'E' were derived from single experiments.


A
ϕNM1-ErmR
ermC
B
Efficiency of Plaquing
10-10 10-8 10-6 10-4 10-2 100
C
32T
32T*
ϕNM1
ϕNM1-ErmR
C
CmR CFU
CmR+ErmR CFU
CFU / ml
10^10
10^8
10^6
10^4
10^2
10^0
C 2B 17B 17T 32T 32T* 43T 61T-1
D
1 2 3 4 C 5 6 7 8
ϕNM2
seeded
E
1 2 3 4 5 6 7 8 C
1 kb
0.5 kb
1 kb
0.5 kb
F
Efficiency of Plaquing (ϕNM2)
10-10 10-8 10-6 10-4 10-2 100
-C
-C/L
9
10
11
12
13
14
G
PFU / ml
10^3
10^2
10^1
10^0
C
32T

ermC, immediately downstream of ϕNM1's *integrase* (**Figure 2-22A**). Recombination with the prophage allowed me to isolate a modified single lysogen, which in turn produced erythromycin-resistance-converting phage particles via spontaneous induction. Using this ϕNM1-Erm^R phage, which seemed to behave no differently than ϕNM1 in plaque assays (**Figure 2-22B**), I confirmed that various type III-A CRISPR-Cas plasmids do not reduce lysogenization efficiency in RN4220 (**Figure 2-22C**). These included some of the same 'conditionally tolerant' plasmids that could efficiently transform pre-established RN4220::ϕNM1 lysogens (**Figure 2-12C**). In addition, I corroborated my results from the egg yolk screens by characterizing some Erm^R isolates harboring the conditionally tolerant spacer 32T CRISPR-Cas plasmid (**Figures 2-22D, 2-22E, & 2-22F**). For one of these isolates, I confirmed that the spontaneously induced PFU recovered from supernatants of overnight cultures were not spacer 32T escaper mutants (**Figure 2-22G**). Note that the number of PFUs derived from conditionally tolerant lysogens was again found to be reduced relative to that of non-targeting control lysogens. In turn, this was corroborated with wild type ϕNM1 by comparing the number of spontaneously induced PFUs derived from three lysogenic transformation isolates harboring either the spacer 32T, spacer 43T, or pGG3 (non-targeting) CRISPR-Cas plasmids (**Figure 2-23A**). In order to visualize the effect of CRISPR-Cas targeting on prophage induction in conditionally tolerant lysogens, I monitored some of my strains in high-resolution growth curves following treatment with the DNA-damaging agent,

Mitomycin C (MMC). MMC treatments are known to be SOS-inducing for various bacteria, including *S. aureus* (Anderson et al., 2006). SOS-induction, in turn, can trigger

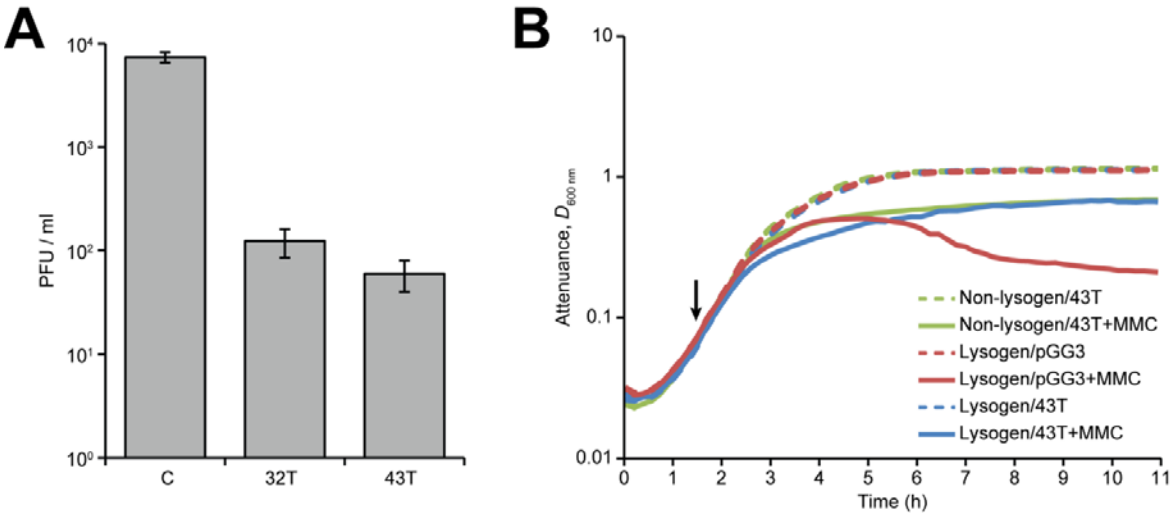


Figure 2-23. Conditionally tolerant type III-A CRISPR-Cas systems limit prophage induction in lysogenic hosts.

(A) Plaque-forming potential of filtered overnight culture supernatants from RN4220::ϕNM1 lysogens transformed with either the pGG3 (C), 32T, or 43T CRISPR-Cas plasmids. PFU concentrations (PFU / ml) were measured by plating on soft agar lawns of RN4220 non-lysogens harboring the pGG3 control plasmid. Dotted line represents the limit of detection under these assay conditions. Error bars, mean ± s.d. (*n* = 3, biological replicates). **(B)** Growth of ϕNM1-Erm^R lysogens harboring the pGG3 (red) or spacer 43T (blue) CRISPR-Cas plasmids as indicated, monitored at high-resolution in the presence (solid lines) or absence (dashed lines) of Mitomycin C added at the indicated time point (black arrow). An RN4220 non-lysogen harboring the 43T CRISPR-Cas plasmid control plasmid was also tested (green).

prophage induction by stimulating autocatalytic cleavage of lambdoid prophage CI-like repressors (Erill et al., 2007; Livny and Friedman, 2004). Whereas lysogenic cultures

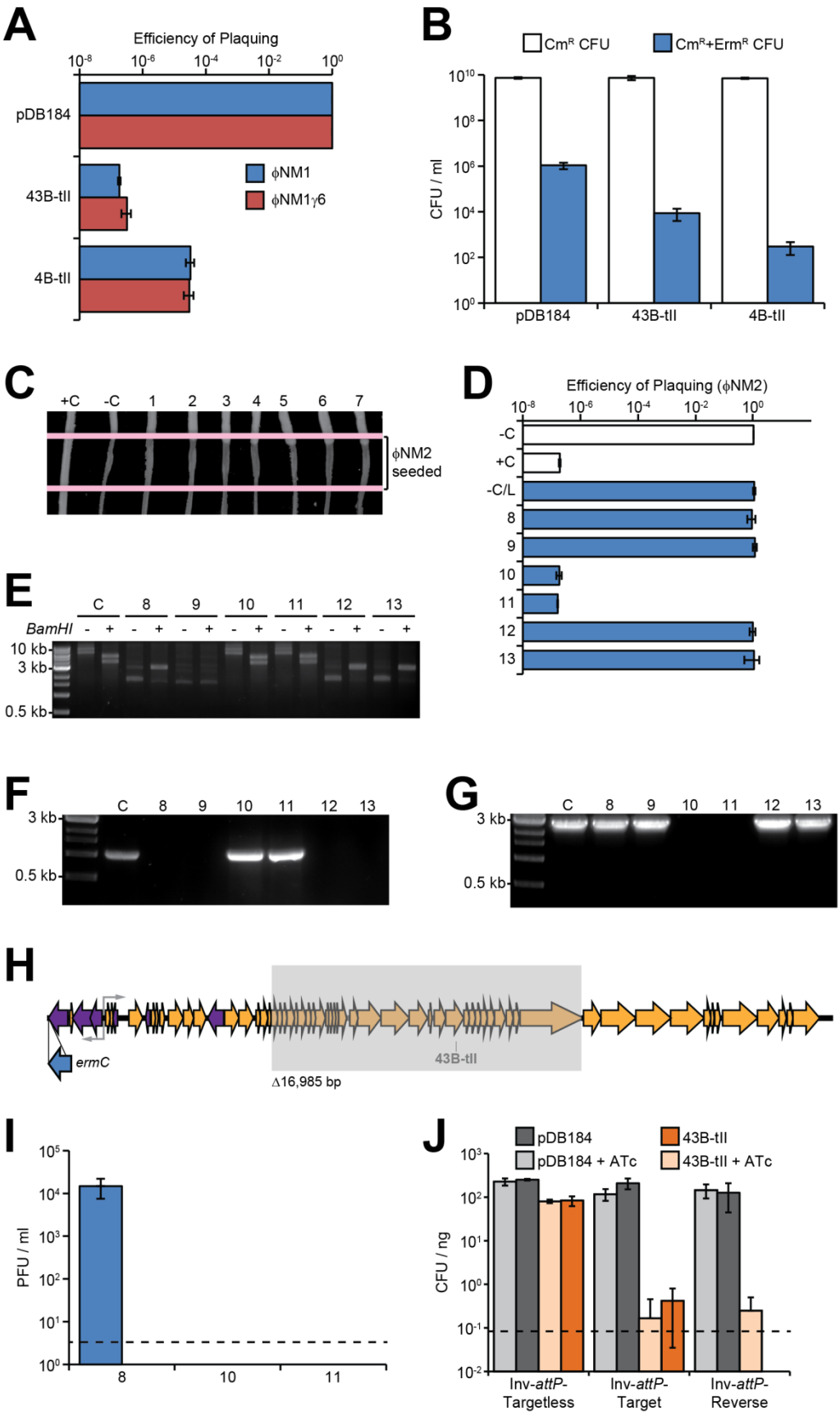
with the non-targeting pGG3 plasmid lysed after MMC treatment, the conditionally tolerant lysogens with spacer 43T did not (**Figure 2-23B**). Whether or not this discrepancy reflects differences in viability was never rigorously determined, but my preliminary experiments with short (~45 min) MMC treatments of RN4220 and its derivatives in HIB media seemed to suggest that viability was diminished to a similar degree for both non-targeting and conditionally tolerant lysogens, even when using a subinhibitory concentration (0.5 μg/ml) of MMC such as this (Goerke et al., 2006). Therefore, while the conditionally tolerant lysogens were found to reach the same attenuance ($D_{600\ nm}$) readings as the non-lysogens harboring spacer 43T, this does not necessarily imply that they would display equivalent viabilities upon plating, because inviable cells that do not lyse may still contribute to attenuance measurements. Conditionally tolerant plasmids aside, the spacer 2B type III CRISPR-Cas plasmid was found to be 'intolerant' of lysogenization, in that it reduced the number of Erm^R CFUs that result from infection (**Figure 2-22C**). Additional characterizations of the rare lysogenized and transformed clones isolated from spacer 2B experiments were presented to a skeptical reviewer during the first revision (Goldberg et al., 2014), and in all cases the isolates were shown to represent genetic escapers that had either inactivated their CRISPR-Cas system or lost their *xis* target region amplicon (data not shown).

Soon after, I demonstrated that (transcription-independent) targeting by type II CRISPR-Cas plasmids can likewise reduce lysogenization efficiency—in addition to preventing lytic infection by both ϕNM1 and ϕNM1γ6 (**Figures 2-24A & B**). Note that lytic infection by ϕNM1γ6 was prevented even when its bottom strand was targeted in a lysogeny gene (using spacer 4B-tII), presumably because type II targeting is largely indifferent to the reduced transcription in this region. Note also that the observed lysogenization efficiency was considerably higher for one of my type II CRISPR-Cas plasmids (harboring spacer 43B-tII), so I characterized a few Erm^R isolates derived from those infections to confirm that they were in fact genetic escapers (**Figure 2-24**). A definitive explanation for the discrepancy between my two type II spacers in the lysogenization assay is still lacking, but some ideas were offered in response to a reviewer's comments during the review process (although they were not explicitly discussed within the publication). Firstly, the discrepancy might result in part from differences in the frequency of target site mutation. Two escapers of the 43B-tII spacer were found to harbor identical large deletions within the prophage lytic region. These deletions (**Figure 2-24H**) appear to have utilized a 14 bp direct-repeat microhomology, suggesting that target site recombinogenicity is a contributing factor (e.g., via 'hot-spots'). Meanwhile, the location of the 4B-tII target within the *cI*-like repressor gene that is required for lysogenization, along with the proximity of this target to the *ermC* insertion, should reduce the likelihood of detecting mutations in this region when we

Figure 2-24. Transcription-independent targeting by a type II CRISPR-Cas system in *S. aureus* precludes detectable tolerance of DNA target elements.

(A) Plaquing efficiency of ϕNM1 and ϕNM1γ6 on lawns of RN4220 harboring the indicated type II CRISPR-Cas plasmids. The pDB184 parent vector was used as the non-targeting control. Error bars, mean ± s.d. (*n* = 3, technical replicates). **(B)** Lysogenization of RN4220 derivatives harboring the type II CRISPR-Cas plasmids tested in 'A', quantified as the concentration (CFU / ml) of chloramphenicol- and erythromycin-resistant colonies obtained upon plating after brief infection with ϕNM1-Erm^R. Prior to infection, cultures were plated in the presence of chloramphenicol alone to estimate the concentration (CFU / ml) of total recipients harboring CRISPR-Cas plasmids. Error bars, mean ± s.d. (*n* = 3, biological replicates). **(C)** ϕNM2-sensitivity assay for seven (1-7) randomly-selected clones isolated from ϕNM1-Erm^R lysogenization experiments with the spacer 43B-tII CRISPR-Cas plasmid. For comparison, a resistant non-lysogen harboring the spacer 43B-tII plasmid and a sensitive lysogen harboring the pDB184 plasmid were included as controls (+C and –C, respectively). After re-streak isolation, single colonies were streaked through the ϕNM2-seeded region from top to bottom. **(D)** ϕNM2 plaquing efficiency on soft agar lawns of an additional six randomly selected clones (8–13) isolated from ϕNM1-Erm^R lysogenization experiments with the spacer 43B-tII CRISPR-Cas plasmid, measured alongside a ϕNM1-Erm^R lysogen harboring the pDB184 plasmid (–C/L). For comparison, plaquing efficiency of ϕNM2 on the non-lysogenic indicator strain harboring pDB184 or the spacer 43B-tII targeting plasmid is also shown (–C and +C, respectively). **(E)** Agarose gel electrophoresis of plasmid DNA purified from isolates 8–13 from panel 'D', and the parental strain harboring the spacer 43B-tII CRISPR-Cas plasmid (C). The symbols + or – indicate the presence or absence of treatment with the BamHI restriction enzyme, which produces two bands for the wild-type spacer 43B-tII plasmid: 5367 bp and 3972 bp. Size markers correspond to 10 kb, 3 kb and 0.5 kb bands of the 1 kb DNA ladder from NEB. **(F)** Colony PCR spanning the

type II CRISPR array for isolates 8–13. Spacer 43B-tII plasmid DNA was used as a template for the control (C). Size markers of 3 kb and 0.5 kb are indicated. **(G)** Colony PCR spanning the target region for isolates 8–13 and a ϕNM1-Erm^R lysogen harboring the pDB184 control plasmid (C). Isolates 10 and 11 harbor identical deletions within the prophage that remove the target region (see below). Size markers of 3 kb and 0.5 kb are indicated. The presence of *attL* and *attR* prophage integration arms was also verified independently for each isolate using PCR (data not shown). **(H)** Location of the 16,985 bp deletion identified within the prophage of isolates 10 and 11 (shaded grey box). The location and orientation of the *ermC* insertion cassette is also shown (thick blue arrow). Deletion was mapped by primer walking. An ~9.1 kb product spanning the deletion was ultimately amplified using primers oGG6 and oGG241, and the deletion junction was sequenced by Sanger using oGG245. A perfect 14 bp direct repeat microhomology flanks the deletion. **(I)** Plaque forming potential of overnight culture supernatants from isolates 8, 10 and 11. Supernatants were also plated with spacer 43B-tII targeting lawns, yielding no detectable PFUs. No PFUs were detected from the supernatants of isolates 10 and 11 whatsoever, presumably resulting from their deletion of genes essential for prophage induction, including the ORF 43 head protein. Dotted line represents the limit of detection for this assay. Error bars, mean ± s.d. (n = 3, biological replicates). **(J)** Transformation of RN4220-derived strains with spacer 43B-tII or pDB184 CRISPR-Cas plasmids in the presence or absence of ATc. Electrocompetent cells possessed either the 'Inv-*attP*-Targetless' vector insertion, the forward 'Inv-*attP*-Target' vector insertion, or the 'Inv-*attP*-Reverse' target vector insertion, and were transformed using the same miniprep for either CRISPR-Cas plasmid cloned in RN4220. Transformation efficiencies were calculated as the number of transformants per ng of miniprepped DNA used for electroporation (CFU / ng). Error bars, mean ± s.d. (n = 3, technical replicates). Pictures in 'C', 'E', 'F', and 'G' are derived from single experiments.


A
Efficiency of Plaquing
10-8
10-6
10-4
10-2
100
pDB184
43B-tII
4B-tII
ϕNM1
ϕNM1γ6
B
CmR CFU
CmR+ErmR CFU
CFU / ml
1010
108
106
104
102
100
pDB184
43B-tII
4B-tII
C
+C
-C
1
2
3
4
5
6
7
ϕNM2
seeded
D
Efficiency of Plaquing (ϕNM2)
10-8
10-6
10-4
10-2
100
-C
+C
-C/L
8
9
10
11
12
13
E
C
8
9
10
11
12
13
BamHI
- + - + - + - + - + - + - +
10 kb
3 kb
0.5 kb
F
C
8
9
10
11
12
13
3 kb
0.5 kb
G
C
8
9
10
11
12
13
3 kb
0.5 kb
H
ermC
43B-tII
Δ16,985 bp
I
PFU / ml
105
104
103
102
101
100
8
10
11
J
pDB184
pDB184 + ATc
43B-tII
43B-tII + ATc
CFU / ng
103
102
101
100
10-1
10-2
Inv-attP-
Targetless
Inv-attP-
Target
Inv-attP-
Reverse

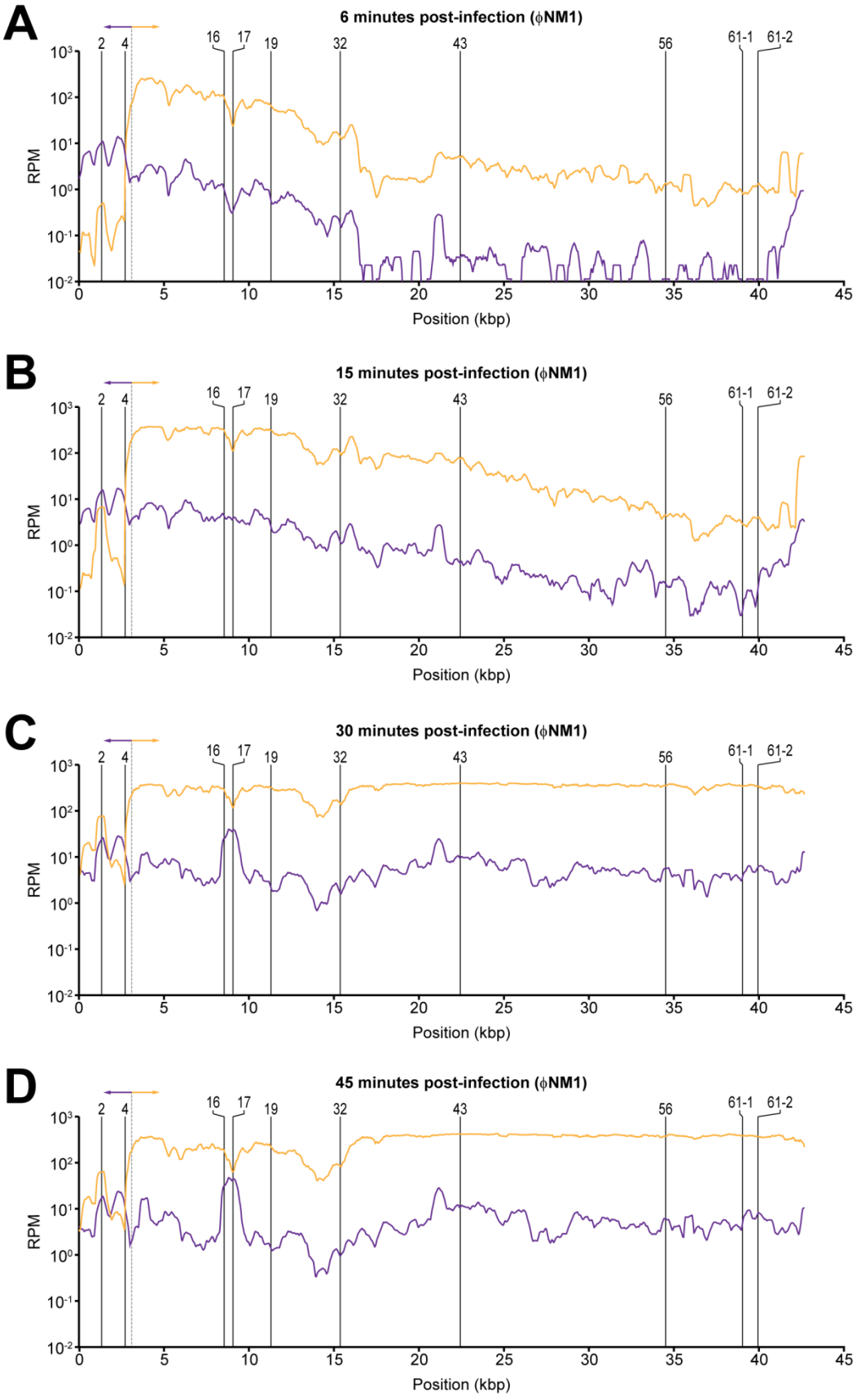
select for Erm^R CFU. In other words, the 4B-tII spacer might be more difficult to escape because it targets lysogenization functions that are required for integration of the *ermC* cassette associated with prophages. Furthermore, it has been shown that point mutations in the target which abrogate Cas9-mediated target cleavage do not necessarily abrogate Cas9-mediated binding altogether, and might therefore still allow repression via the dCas9-like mechanism (Bikard et al., 2013). In this scenario, knockdown of the *cI*-like repressor gene should also be selected against. However, these factors may be insufficient to explain the discrepancy, given that the dominant form of escape I observed with the 43B-tII spacer was due to CRISPR-Cas plasmid inactivation. On this note, it's possible that the 43B-tII spacer promoted higher frequencies of plasmid rearrangement in the overnight cultures used for lysogenization, either due to the spacer sequence's recombinogenicity itself, or due to stress associated with off-target effects. In support of this latter idea, partial matching to the RN4220 chromosome appears to be greater for the 43B-tII spacer than for the 4B-tII spacer. To determine this, I took advantage of a script that Dr. Bikard had designed to look for partial matching adjacent to 'NGG' PAMs. When the parameters called for a minimum of 8 consecutive bp matched at the seed sequence adjacent to the PAM, no partial matches were found for the 4B-tII spacer, while eight were found for the 43B-tII spacer. Among these 43B-tII partial matches was a 15 bp match to an ORF encoding a putative pyrimidine nucleoside phosphorylase. In turn, partial matching might allow for occasional cleavage

of the chromosome, or dCas9-like repression of target gene(s) that are potentially essential for normal growth. The 43B-tII CRISPR-Cas plasmid was also used in transformation assays to target my second-generation chromosomal insertion vectors harboring the inverted *attP* motif. As expected, the 43B-tII transformation efficiency was reduced for both forward and reverse targets, regardless of the presence or absence of ATc (**Figure 2-24J**).

Around this time, I also obtained the results from my second RNA-seq experiment. Dr. Bikard, again demonstrating his resourcefulness, provided me with a script for normalizing and plotting the reads. The results corroborated my previous results 15 minutes post-infection with ϕNM1, and expanded it to include additional time points. Perhaps as expected, strong rightward transcription of the early lytic genes immediately to the right of the ^{E}LCR was observed even at the early time points, whereas late genes further downstream were not strongly transcribed until later time points (**Figure 2-25**). Only moderate leftward transcription was observed to the left of the ^{E}LCR, but it remained essentially constant across time points. A moderate peak of rightward transcription across the *xis* ORF to the left of the ^{E}LCR was also discernible, and thereby confirmed our prediction that this region is bi-directionally transcribed during lytic infections. In a separate endpoint RT-PCR experiment (**Figure 2-26**), my results tested positive for bi-directional transcription across the target region of Dr. Marraffini's original pC194-derived target plasmids harboring *nes* target insertions in

Figure 2-25. Visualization of ϕNM1 transcription profiles at various time points after infection (MOI 20).

(A) RNA was harvested 6 minutes post-infection of a logarithmically growing culture of RN4220 cells harboring the non-targeting pGG3 plasmid. Rightward and leftward expression values are plotted as golden and purple lines, respectively, in reads per million (RPM). The position of relevant spacer targets are indicated with vertical solid lines. The dotted line with divergent arrowheads marks the position of the early lytic control region (^{E}LCR) with divergently oriented P_{cI} and P_{cro} promoters. To improve readability, curves were smoothened by plotting average reads per million values over a 500 bp sliding-window. **(B)** Same as in 'A', except RNA was harvested 15 minutes post-infection. **(C)** Same as in 'A', except RNA was harvested 30 minutes post-infection. To the left of the ^{E}LCR, rightward expression is comparable to leftward expression by 30 minutes post-infection, consistent with the strand-independent targeting observed for this region. **(D)** Same as in 'A', except RNA was harvested 45 minutes post-infection.


A
6 minutes post-infection (ϕNM1)
B
15 minutes post-infection (ϕNM1)
C
30 minutes post-infection (ϕNM1)
D
45 minutes post-infection (ϕNM1)
2
4
16
17
19
32
43
56
61-1
61-2
RPM
Position (kbp)

either orientation (Marraffini and Sontheimer, 2008), which may also explain why he observed targeting in both cases. Importantly, my RNA-seq experiment indicated that

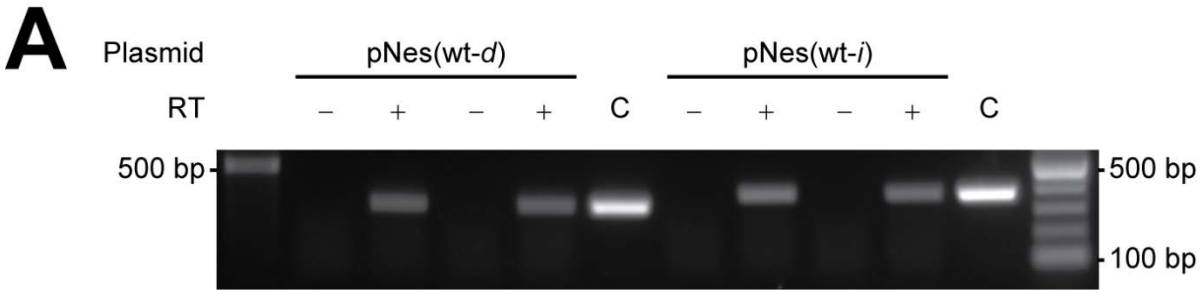


Figure 2-26. Detection of transcription across target insertions for the pNes(wt-*d*) and pNes(wt-*i*) plasmids.

(A) Reverse transcription was performed in both directions with DNase-treated total RNA from RN4220 cells harboring the indicated plasmids, using either forward or reverse primers for cDNA synthesis in two separate reactions for each target plasmid (Marraffini and Sontheimer, 2008). PCR was performed on cDNA products, or plasmid DNA templates for control (C) lanes. The symbols + or – indicate the presence or absence of reverse transcriptase enzyme in the reverse transcription reaction mixture used to template the PCR. Size markers of 500 bp and 100 bp are indicated. Picture is derived from a single experiment; reaction components were not tested in additional technical replicates.

leftward transcription from the ^{E}LCR was reduced for the ϕNM1γ6 mutant relative to ϕNM1 at equivalent time points, while rightward transcription of ϕNM1γ6 was essentially unchanged (**Figures 2-10D & 2-27**). In particular, when I calculated the fold change for the sum of leftward expression values across the region bounded by the start of the phage genome and the ^{E}LCR, as well as for the sum of leftward expression values

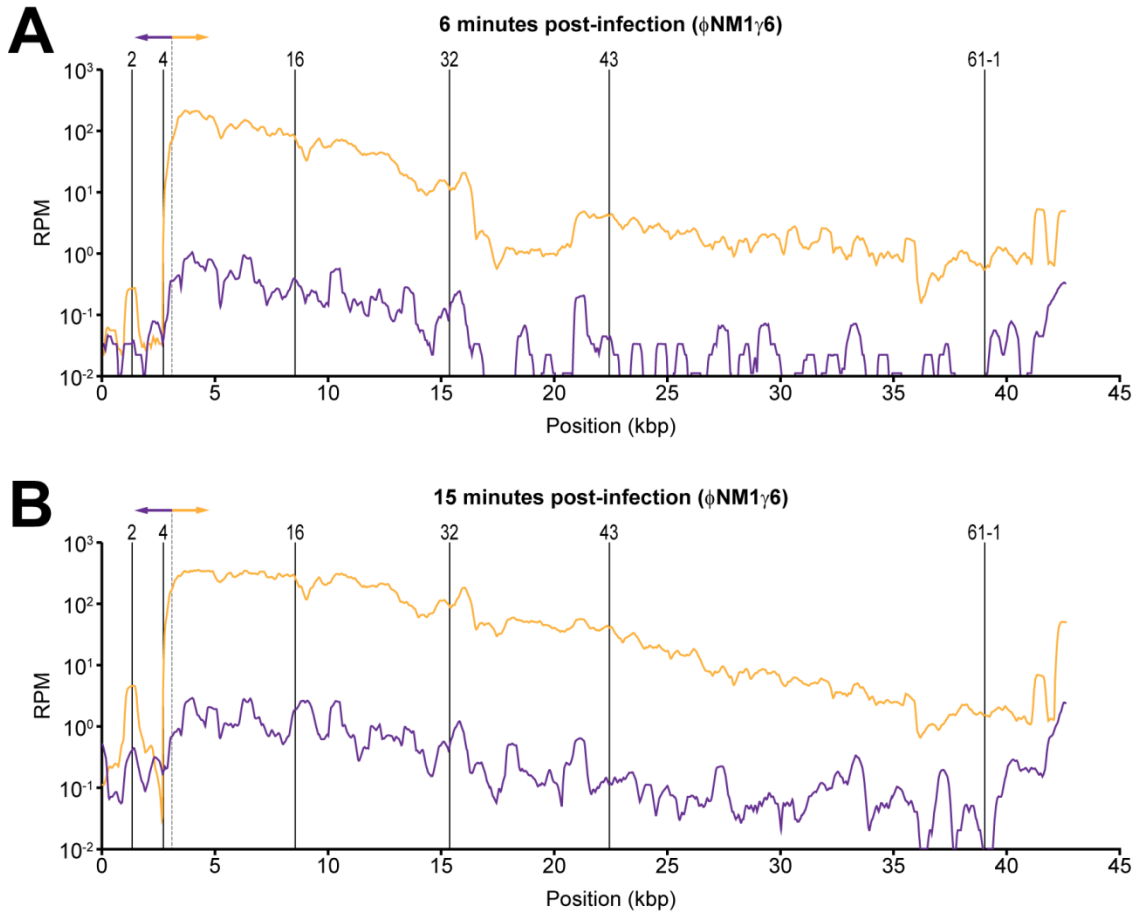


Figure 2-27. Visualization of ϕNM1γ6 transcription profiles at two time points post-infection (MOI 20).

(A) RNA was harvested 6 minutes post-infection of a logarithmically growing culture of RN4220 cells harboring the non-targeting pGG3 plasmid. Rightward and leftward expression values are plotted as golden and purple lines, respectively, in reads per million (RPM). The position of relevant spacer targets are indicated with vertical solid lines. The dotted line with divergent arrowheads marks the position of the early lytic control region (^{E}LCR) with divergently oriented P_{cI} and P_{cro} promoters. To improve readability, curves were smoothened by plotting average reads per million values over a 500 bp sliding-window. **(B)** Same as in 'A', except RNA was harvested 15 minutes post-infection.

across the entire phage genome, I found approximately 32-fold and 4-fold reductions, respectively, for ϕNM1γ6 at the 15 minute time point (suggesting a net 8-fold reduction in leftward expression across this region).

During the initial review process, some concern was raised by an anonymous reviewer regarding our evidence for *in vivo* DNA targeting. Our efficiency of plaquing and chromosomal targeting results were difficult to explain without invoking a DNA targeting argument, but the evidence was still arguably indirect. Similar to Dr. Bikard's initial concerns regarding Csm6, the reviewer seemed to be concerned that a toxic activity associated with targeting—perhaps of RNA alone—could explain our results. Toxicity associated with specific targeting of RNA by the complex *in vivo* still seemed highly unlikely. However, the reviewer may have been privy to knowledge about Csm6, or at least the hypotheses surrounding its non-specific, toxin-like RNase function (Anantharaman et al., 2013; Makarova et al., 2012). Given the mounting evidence in our lab that *csm6* was not strictly required for targeting in these assays, we were not particularly concerned (but to be fair, no mention of Csm6 or its mutagenesis was included in the manuscript). Nevertheless, we provided some experiments that helped to address the reviewer's concern. In particular, I provided high-resolution growth curves of strains harboring different CRISPR-Cas plasmids, following infection with ϕNM1 at MOI ~10 or MOI ~100. Growth was largely unperturbed at MOI ~10 for all the spacers tested, relative to uninfected cultures (**Figure 2-28A**). This suggested that cells

were not strictly arresting upon infection (i.e., that the immunity observed in plaque assays was not due to an abortive-infection-like mechanism), because presumably all or most of the cells are injected at MOI ~10. In other words, infected cells appeared to be clearing the phage, or at least surviving infection. At higher MOIs (~100), growth was more noticeably perturbed (**Figure 2-28B**). This was assumed to reflect the CRISPR-Cas

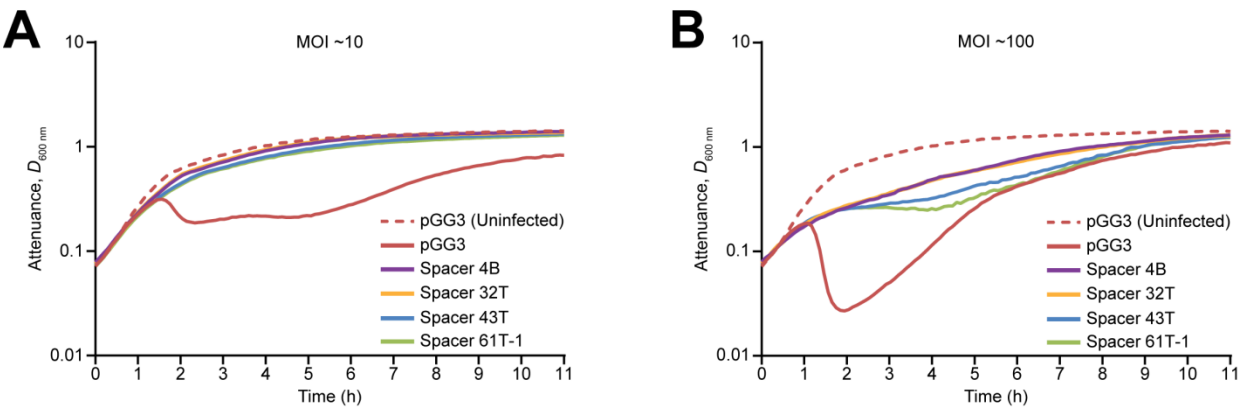


Figure 2-28. Infection with ϕNM1 in liquid culture.

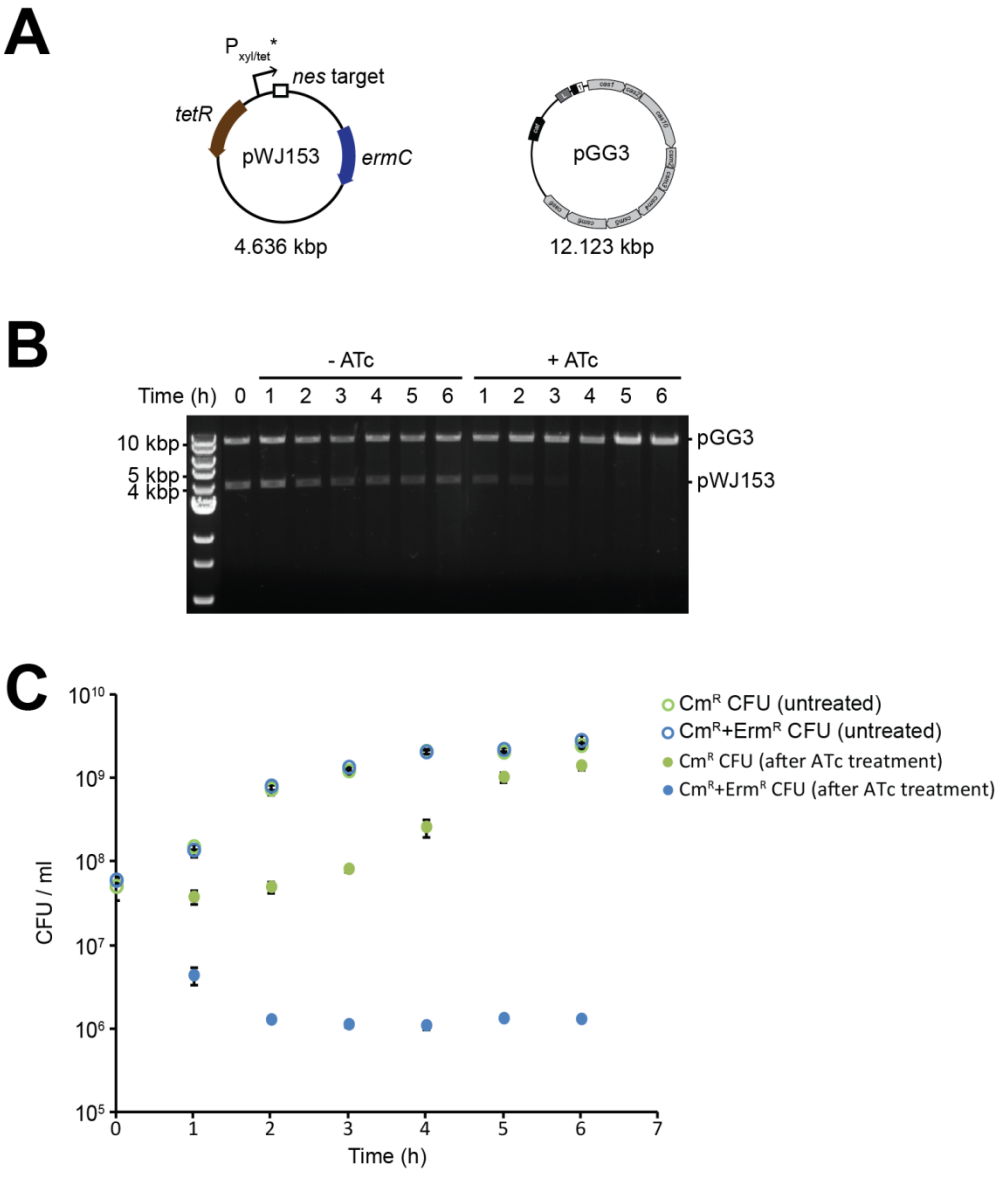
(A) Growth of RN4220 hosts harboring various CRISPR-Cas plasmids (as indicated) was monitored at high-resolution following infection with ϕNM1 (MOI ~10) at time zero (solid lines). Growth of the uninfected RN4220/pGG3 host is also shown for comparison (dashed clay-red lines). **(B)** Same as in 'A', except cultures were infected with ϕNM1 at MOI ~100.

system being overwhelmed with an unnaturally high amount of phage. Furthermore, spacers targeting late-transcribed genes seemed to perform even more poorly. This could conceivably still be explained in terms of a transcription-dependent DNA targeting model, if these discrepancies were assumed to result from differences in the

onset of DNA targeting. In other words, delays in DNA targeting for late-transcribed targets might simply give the phage a head start, and generally overwhelm the CRISPR-Cas system more readily. Wenyan and Dr. Samai (with help from others in the lab) would later go on to show that the robustness of type III-A immunity at normal MOIs (< ~10) is also due in part to the contributions of its Csm3 and Csm6 RNase activities (Jiang et al., 2016; Samai et al., 2015). However, prior to establishing a function for Csm6 in phage defense, Wenyan had also been working to engineer an inducible system for targeting (and perhaps curing) of plasmids with the *nes* protospacer, such as pG0400. His initial attempts utilized Lac-inducible promoters for expression of the type III-A system, either in its entirety or in parts. Despite these efforts, the systems were generally still leaky, and targeting would occur in the absence of induction. Furthermore, curing of pG0400 could not be achieved using wild type CRISPR-Cas systems. Once I had demonstrated that the modified ATc-inducible promoter architecture based off pRAB12 (**Figure 2-16**) allowed sufficient tightness to prevent severe chromosomal targeting by our constitutively-expressed type III-A systems (**Figures 2-18 & 2-20**), and thereby allowed for conditional targeting of DNA elements by such CRISPR-Cas systems, it became clear that this type of setup might offer a solution in plasmid-targeting assays. Wenyan's new approach was therefore to use this modified ATc-inducible architecture and clone the *nes* protospacer downstream

Figure 2-29. Inducible curing of a target plasmid.

(A) Schematic representation of the plasmid components used in this plasmid-curing experiment. The pGG3 CRISPR-Cas plasmid harbors a single spacer ('1') matching to the *nes* target sequence inserted downstream of the $P_{xyl/tet}$*-inducible promoter in pWJ153. The size of each plasmid is indicated below in kilobase pairs (kbp). Also depicted on pWJ153 is the *ermC* cassette conferring resistance to erythromycin, as well as the *tetR* gene encoding the inducible promoter's repressor. **(B)** Agarose gel electrophoresis of linearized plasmid DNA purified both from 0.250 μg/ml ATc-treated (+ATc) and untreated (–ATc) cultures at the indicated time points. Size markers of 10 kb, 5 kb and 4 kb are indicated. Picture is representative of a single technical replicate. **(C)** Concentration (CFU / ml) of colonies recovered upon plating of the cultures analyzed in 'B' at each time point. Cultures were plated with selection for either Cm^R CFUs (green) or Cm^R+Erm^R CFUs (blue) as indicated. Targeting of the pWJ153 plasmid via induction with ATc (filled circles) is accompanied by a severe drop in Erm^R CFUs relative to untreated cultures (open circles). Error bars, mean ± s.d. (n = 3, technical replicates).


A
P xyl/tet *
nes target
tetR
pWJ153
ermC
4.636 kbp
pGG3
12.123 kbp
B
- ATc
+ ATc
Time (h) 0 1 2 3 4 5 6 1 2 3 4 5 6
10 kbp
5 kbp
4 kbp
pGG3
pWJ153
C
CmR CFU (untreated)
CmR+ErmR CFU (untreated)
CmR CFU (after ATc treatment)
CmR+ErmR CFU (after ATc treatment)
CFU / ml
Time (h)

of its $P_{xyl/tet}$* promoter onto a pE194-based plasmid, creating pWJ153. This proved fruitful; the target plasmid was indeed stable in the absence of ATc, and could be cured by pGG3 upon induction (**Figure 2-29**). Although this was only demonstrated for one target orientation, it provided a final confirmation of *in vivo* DNA targeting by a type III-A system in *S. aureus*. Whereas growth was essentially halted upon induction in my chromosomal targeting assays (**Figure 2-20**), Wenyan found that growth could continue during plasmid targeting, and thereby allow cells to lose their target plasmid's DNA (**Figure 2-29B**) and its associated resistance (**Figure 2-29C**) over time. Growth was not completely unperturbed, however; increases in the number of Cm^R CFUs were delayed upon ATc treatment. This is probably due, at least in part, to the generalized toxicity associated with ATc at this concentration (0.250 μg/ml). However, the possibility that this was also the result of side effects or toxicity associated with targeting should not be ruled out, in light of recent evidence concerning the potential for collateral RNA degradation by HEPN-containing Cas nucleases (Shmakov et al., 2017). I elaborate on this point in **Chapter 3**.

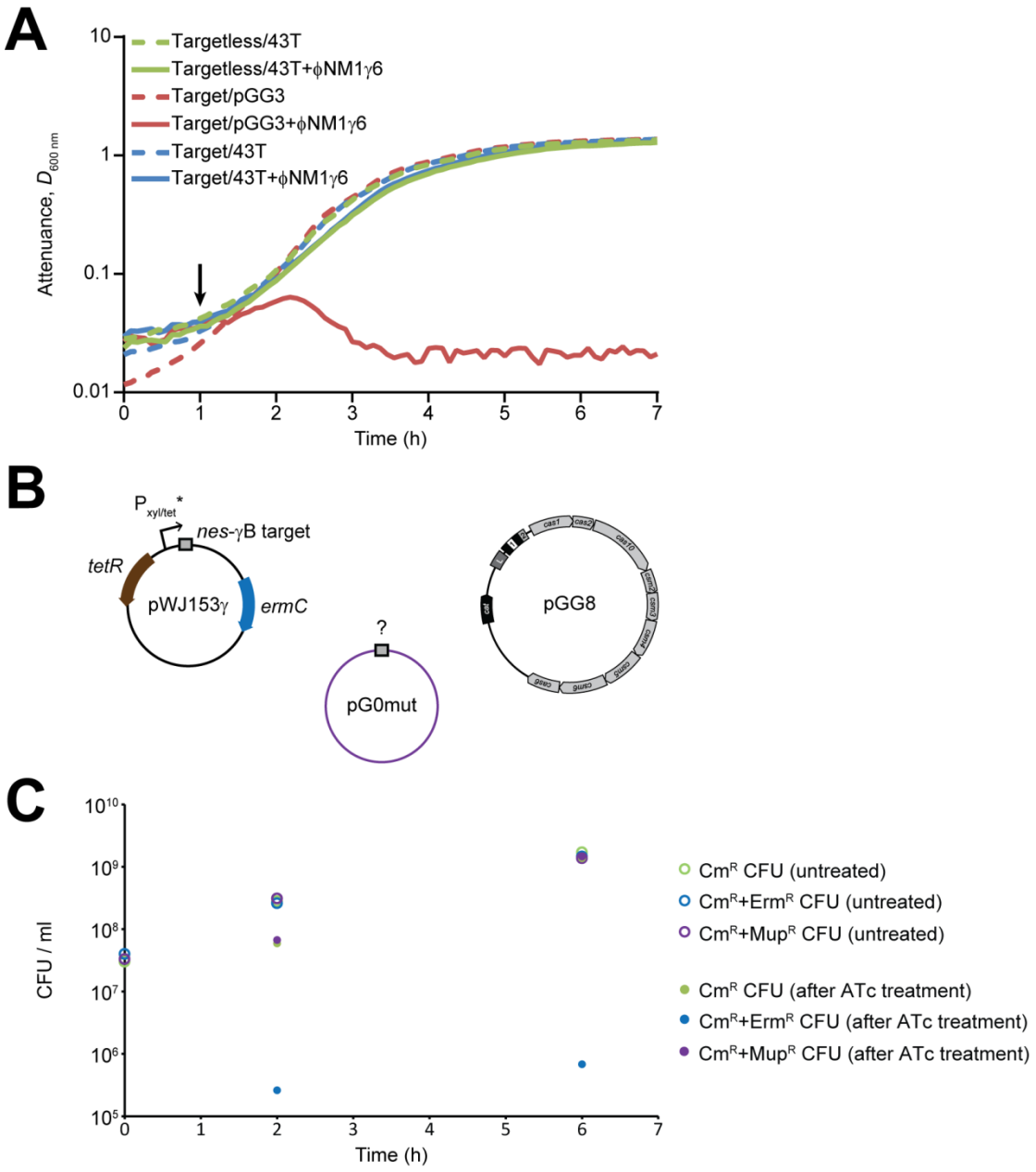
2.6 Proposed implications of transcription-dependent targeting and conditional tolerance by type III-A CRISPR-Cas systems

The work described in **Chapter 2** illustrates how transcription-dependent targeting by a type III-A CRISPR-Cas system offers a mechanism for conditional tolerance of genetic

elements in *S. aureus* hosts. In particular, my experiments with temperate phages indicated that conditional tolerance engenders several consequences for the type III-A systems that target temperate phages, as well as their host population(s). For one, tolerance helps ensure the genetic stability of these CRISPR-Cas systems during temperate phage infection, given that pressure to integrate prophages in the presence of intolerant spacers was found to select for genetic CRISPR-Cas inactivation (**Figures 2-4, 2-12, & 2-24**)—similar to what has been reported in the context of plasmid uptake (Fischer et al., 2012; Gudbergsdottir et al., 2011; Jiang et al., 2013b). In other words, conditionally tolerant spacers ensure that type III-A systems can provide CRISPR-Cas immunity to lytic infections without compromising their own stability in the event of lysogenization. An important consequence of this is that lysogenized lineages can maintain type III-A immunity to lytic infection by heteroimmune phages if the heteroimmune phages also contain a matching target for the conditionally tolerant spacer (**Figures 2-11 & 2-22**). This implies, moreover, that type III-A immunity operates via transcription-dependent targeting of transcribed elements in *cis* (because lysogenic cells survive infection by the heteroimmune phage despite the presence of their prophage with a matching target). I corroborated this in two assays. Firstly, I monitored growth of ϕNM1γ6-infected strains harboring my second-generation, inverted *attP* integrative vectors at high-resolution in the absence of ATc. Growth was essentially unperturbed relative to uninfected cultures for the strains that harbored the spacer 43T

Figure 2-30. Type III-A immunity to a transcribed target element does not result in detectable targeting of repressed DNA target elements in *trans*.

(A) Growth of the indicated 'Inv-*attP*' chromosomal insertion strains from **Figure 2-20** harboring either spacer 43T or pGG3 CRISPR plasmids, monitored at high-resolution in the absence (dotted lines) or presence (solid lines) of ϕNM1γ6 addition at an MOI of ~10. Black arrow denotes the time of phage addition; no ATc was used in this assay. The presence of a chromosomal target for spacer 43T has no discernable effect on culture growth during spacer 43T-mediated immunity to ϕNM1γ6 (compare green and blue solid lines). The strain harboring the pGG3 plasmid serves as a sensitive control (clay-red lines). **(B)** Schematic representation of the plasmid components used in the plasmid-curing experiment of panel 'C'. The pGG8 CRISPR-Cas plasmid harbors a spacer in position 2 that matches to the *nes*-γB target sequence inserted downstream of the $P_{xyl/tet}$*-inducible promoter of pWJ153γ. In pG0mut, the *nes*-γB target is thought to be lacking antisense transcription, because it is not targeted by pGG8 in conjugation assays. Induction of the target in pWJ153γ produces the same RNA that would result from antisense transcription of the *nes*-γB target in pG0mut. Also depicted on pWJ153 is the *ermC* cassette conferring resistance to erythromycin, as well as the *tetR* gene encoding the inducible promoter's repressor. The pG0mut plasmid confers resistance to mupirocin. **(C)** Concentration (CFU / ml) of colonies recovered upon plating of cultures either treated (filled circles) or untreated (open circles) with ATc at 0.125 μg/ml. Cultures were plated with selection for either Cm^R CFUs (green), Cm^R+Erm^R CFUs (blue), or Cm^R+Mup^R CFUs (purple) as indicated. Relative to untreated cultures (open circles), targeting of the pWJ153γ plasmid via induction with ATc (filled circles) is accompanied by a severe drop in the number of Cm^R+Erm^R CFUs, but not Cm^R+Mup^R CFUs. After plating of untreated cultures at time zero, only the 2-hour and 6-hour time points were plated. Additional replicates were not performed under these assay conditions.


A
Targetless/43T
Targetless/43T+ϕNM1γ6
Target/pGG3
Target/pGG3+ϕNM1γ6
Target/43T
Target/43T+ϕNM1γ6
Attenuance, $D_{600\ nm}$
Time (h)
B
$P_{xyl/tet}$*
nes-γB target
tetR
pWJ153γ
ermC
?
pG0mut
pGG8
C
CFU / ml
Time (h)
Cm^R CFU (untreated)
Cm^R+Erm^R CFU (untreated)
Cm^R+Mup^R CFU (untreated)
Cm^R CFU (after ATc treatment)
Cm^R+Erm^R CFU (after ATc treatment)
Cm^R+Mup^R CFU (after ATc treatment)

CRISPR-Cas plasmid, regardless of whether they bore a matching target in the chromosome as well (**Figure 2-30A**). In another assay, I attempted to detect loss of mupirocin resistance associated with pG0mut while curing a modified version of pWJ153 (pWJ153γ) with a target for the *nes*-γB spacer (**Figure 2-30B**). In pWJ153γ, the *nes* target of pWJ153 was replaced with the *nes*-γB sequence which is not normally targeted by pGG8 in pG0mut (**Figure 2-14**). Again, immunity appeared to be specifically directed at the transcribed target element upon induction of antisense transcription (in *cis*), because I still did not detect targeting of pG0mut by the *nes*-γB spacer in *trans* while curing the pWJ153γ plasmid (**Figure 2-30C**).

If confirmed for other type III systems, the capacity for conditional tolerance might be particularly vital to the stability of type III systems in general, since type III targeting in archaea has been shown to provide immunity despite up to 15 spacer-target mismatches (Manica et al., 2013). Evidence for type III immunity occurring within staphylococcal hosts despite the presence of spacer-target mismatches was provided in this work (**Figures 2-8, 2-11, & 2-22**), and also in three subsequent works (Cao et al., 2016; Jiang et al., 2016; Maniv et al., 2016). Therefore, without the potential for phages to readily evade targeting via point mutation that is seen for type I and type II systems (Deveau et al., 2008; Semenova et al., 2011), transcription-dependent type III-A targeting offers an alternative route to lysogenization that need not provide selection for mutants. In addition, the ability of the type III-A system to provide immunity in the presence of

spacer-target mismatches appears to result in more stringent selection against phage mutants during lytic infections. For example, ϕNM1 escapers were not observed with our spacer 43T type III CRISPR-Cas plasmid (**Figure 2-2**), while they were observed with the 43B-tII type II CRISPR-Cas plasmid (**Figure 2-24**)—despite the fact that its 3'-most 12 bp (and 'AGG' PAM) are contained within the 36 bp of the type III target. However, with other type III spacers where the target sequences were apparently non-essential for lytic infection, escapers were observed. For three such spacers, preliminary characterization (section **2.2**) indicated that target site deletions are the dominant route of escape in these cases (**Figures 2-7 & 2-13**). Where microhomologies (< 20 bp) could be identified flanking the deletions, I speculate that deletion may have occurred through a *recA*-independent pathway (Bzymek and Lovett, 2001; Chayot et al., 2010), since the minimum homology for efficient *recA*-dependent recombination is thought to be closer to ~30 bp (Martinsohn et al., 2008; Shen and Huang, 1986). It's also possible that our lambdoid Newman phages encode lambda *red*-like homologs that can facilitate these deletions (Datta et al., 2008). For example, the Newman phage annotations include ORFs encoding SSB-like proteins which might provide a recombinase function. However, the efficiency of *red*-mediated recombination was also shown to drop drastically for homologies shorter than ~30 bp (Yu et al., 2000). Whether the activity of Cas nucleases during immunity stimulates one or more of these pathways remains to be determined. On this note, the type of CRISPR-Cas system used for targeting could

potentially influence the sequence or frequency of deletions that arise, even when the same site is targeted (e.g., if different Cas nuclease lesions produce different substrates that favor deletion via one pathway over another). Evidently, certain type III spacers can also be escaped by mutants that reduce transcription of their target sequences during lytic infection (**Figure 2-10**). However, this was presumably enabled by the fact that those spacers were engineered to target lysogenization genes. Targeting of lytic genes should not be readily escaped in this manner, since reduced transcription of lytic genes would simultaneously compromise phage viability.

Although a few spacers targeting putative or confirmed lysogenization genes did not provide tolerance (**Figures 2-6, 2-12, & 2-22**), it is important to note that these genes only constitute a small portion of the phage genome. Hence, spacers targeting these regions should be acquired less frequently, even if spacer acquisition occurs randomly without an additional mechanism for distinguishing tolerant from intolerant spacers during acquisition. Interestingly, our survey of sequenced staphylococcal type III spacers showed that naturally acquired spacers with known target sequences produced crRNAs complementary to the non-template strand of predicted ORFs in nine out of ten cases (**Figure 2-31**). This bias suggests negative selection on non-functional spacers targeting template strands. Alternatively, type III systems may utilize an unknown mechanism to discriminate template and non-template strands during spacer

A

Organism Target element	Spacer sequence Target sequence	Target strand
Staphylococcus aureus MSHR1132 Staphylococcus phage vB_SauM_Remus	GTTTTTCATAGTTAATCAATCCCTTTTCTTTTTT ATTAAAGTATCAAATAGTTAGGGAAAAGAAAATA	Non-template
Staphylococcus aureus MSHR1132 Staphylococcus aureus phage phiNM1	TATGTATTGATCTCGATTCTCGTTAGTTTCTAAATT GCACATAACTCGAGCTAAGAGCAATCAAAGGTTTAC	Non-template
Staphylococcus aureus MSHR1132 Staphylococcus phage SP6	CACGCTGTAGTGAAGTATAGAAACGGCATGAGTACAAT GTGCGACATCACTTCATATCTTTGCCGTACTCATGTTA	Non-template
Staphylococcus epidermidis RP62a Staphylococcus aureus pGO1 plasmid	ACGTATGCCGAAGTATATAAATCATCAGTACAAAG TGCATACGGCTTCATATATTTAGTAGTCATGTTTC	Template
Staphylococcus epidermidis RP62a Staphylococcus phage CNPH82	TAGTAATAATTGTCATTTGCATACGTTACATCGAT ATCATTATTAACAGTAAACGTATACAATGTAGCTA	Non-template
Staphylococcus aureus 08BA02176 Staphylococcus phage S25-3 DNA	TAGAATGTTATTATCTAAGTGGTCGATGTATTCC ATCCTACAATAATAGATACACCAGCTACATAAGG	Non-template
Staphylococcus aureus 08BA02176 Staphylococcus phage S13' DNA	AAGTTAACGGCATTACCTAATAAAAATATTTTAGG TTCAACTGGCGTAATGGATTATTTTTGTAAAATCC	Non-template
Staphylococcus aureus 08BA02176 Staphylococcus sp. CDC3 plasmid SAP020A	TCATCTTTCATGTCACTGATTAATTCATTTGTA AGTAGAAAGTATAGTAACTAATTAAGTAAACAT	Non-template
Staphylococcus intermedius NCTC 11048 Staphylococcus phage K	CCAAACCATTTAGCACGATATTTATTAAAACCATA GGTTTGGTAAACCGTGCTATAAATAATTTAGGCAT	Non-template
Staphylococcus intermedius NCTC 11048 Staphylococcus phage S25-3 DNA	TATTTTTCTCCTTTAGCAATCATTCTGTCTAGTAC ATAACAAGAGGAGATCGTTAGTAAGACAGATTATG	Non-template

Figure 2-31. Survey of unique spacers associated with staphylococcal type III-A systems possessing known targets with five or fewer mismatches.

(A) Spacer-containing organisms and target elements are listed in the left column; the sequence of each spacer and its complementary target are provided in the middle column. Right column indicates whether the target strand is a template or non-template strand, inferred from annotated open reading frames. In cases where additional target elements were identifiable for a given spacer, a target element with the most available matches was chosen arbitrarily.

acquisition. Moreover, the temperate-phage-targeting spacers that we identified were found to match to lytic regions. A recent study corroborated our *in silico* findings, and expanded the analysis to include sequences from six additional *S. aureus* isolates with type III-A systems which were previously unavailable (Cao et al., 2016).

Finally, conditional tolerance by type III-A systems ensures that a host population can efficiently sample phenotypes that result from prophage integration (i.e., lysogenic conversion phenotypes)—and again whilst maintaining immunity to lytic infection. However, prophage induction was found to be disrupted by type III-A targeting in lysogenized lineages that maintain conditional tolerance (**Figure 2-23**). Therefore, it is worth noting here that lysogenic conversion phenotypes which depend on processes which occur via prophage induction, including rapid lysis of induced cells, might also be curbed by conditional tolerance in lysogenized lineages. As discussed in section ***1.4.2***, the ability of a helper prophage to mobilize a resident SaPI via prophage induction could be considered a lysogenic conversion phenotype. In **Chapter 3**, I present my preliminary evidence indicating that this too can be disrupted in lysogens where the helper prophage is targeted by a type III-A CRISPR-Cas system.

2.7 Tables associated with Chapter 2

Table 2-1. Spacers tested in Chapter 2.

Spacer[(a)]	Sequence (5'-3')[(b)]	Coordinates[(c)]	Gene product[(d)]	Plasmid Name
2T^{old}	TTTTTAATTTAAGTTCTTGTTCATCGTCATAAATA	1317-1351	Excisionase	pGG34
2T	AATTTTTAATTTAAGTTCTTGTTCATCGTCATAAA	1319-1353	Excisionase	pGG52
2B	TATTTATGACGATGAACAAGAACTTAAATTAAAAA	1317-1351	Excisionase	pGG9
4T	AATTTCGAGGAAGTTGCAATTGATAATGAAAAATT	2706-2740	Repressor	pGG31
4B	AATTTTTCATTATCAATTGCAACTTCCTCGAAATT	2706-2740	Repressor	pGG30
4B-tII	GAAATTTCCAGCAGAAACTTTACCGAAATA	2735-2764	Repressor	pGG51
16T	CATTTTGTTTCTGTTCATGCCTCTGCCGACTGCT	8527-8560	Hyp. protein	pGG19
16B	AGCAGTCGGCAGAGGCATGAACAGAAACAAAATG	8527-8560	Hyp. protein	pGG4
16B-tII	AGACCTGGCACATTATGAAGCAGTCGGCAG	8509-8538	Hyp. protein	pGG33
17T	TAATAAGTTTTATGCTCCTCAGTTTTTAAATCACTT	9045-9080	Hyp. protein	pGG60
17B	AAGTGATTTAAAAACTGAGGAGCATAAAACTTATTA	9045-9080	Hyp. protein	pGG59
19T	TTTTTAAAAATTCTTTGGTTACCATGCATCTCGCT	11293-11327	Replication	pGG53
19B	AGCGAGATGCATGGTAACCAAAGAATTTTTAAAAA	11293-11327	Replication	pGG10
32T	TTAAATCTTTGATTGCTCTTAGCTCTAGTTATGTAT	15352-15387	Hyp. protein	pGG12
32T*^{A}	GTAAACCTTTGATTGCTCTTAGCTCGAGTTATGTGC	15352-15387	Hyp. protein	pGG13
32B	ATACATAACTAGAGCTAAGAGCAATCAAAGATTTAA	15352-15387	Hyp. protein	pGG36
43T	ATTCGTCATCTTCAAGTAATGCCTCTAAATCAATAA	22411-22446	Head protein	pGG41
43B	TTATTGATTTAGAGGCATTACTTGAAGATGACGAAT	22411-22446	Head protein	pGG40
43B-tII	ACTTCACACAAGATAACATTATTGATTTAG	22393-22422	Head protein	pGG37-full
56T	GCATGCACCTTGCCTGAATGTTTTAAAAATTCATT	34512-34546	Hyp. protein	pGG54
56B	AATGAATTTTTAAAACATTCAGGCAAGGTGCATGC	34512-34546	Hyp. protein	pGG11
61T-1	ATGTCACCTAAGTCAACACCATCATTTTTTATTCT	39013-39047	Tail fiber	pGG17
61B-1	CTTAGGTGACATTGGCTGTCGATTTTACACTGAAG	39036-39070	Tail fiber	pGG15
61T-2	TTATGATTTTTTGGAGCATATAAATCATTTAGTGT	39949-39983	Tail fiber	pGG18
61B-2	CAGAAAGTGTATTGCAACAGATTGGCTCAAAAGTT	39884-39918	Tail fiber	pGG16
nes-γB	ATAAATATGTATAAGGAAAATGAAAGACTATATGA	9232-9266	Nickase	pGG8
nes-γT	TCATATAGTCTTTCATTTTCCTTATACATATTTAT	9232-9266	Nickase	pGG24
nes(-R)	ACGTATGCCGAAGTATATAAATCATCAGTACAAAG	8174-8208	Nickase	pGG25
32T*^{B}	GTAAACCTTTGATTACTCTTAGCTTTAGTTATGTGT	15256-15291	Hyp. protein	pGG14
32T*^{B}(-R)	GTAAACCTTTGATTACTCTTAGCTTTAGTTATGTGT	15256-15291	Hyp. protein	pGG22

Table 2-1. Spacers tested in Chapter 2 (continued).

(a) Numbers refer to the nearest ORF designation for the closest ϕNM1 target match, where applicable; crRNA complementarity to the 'Top' or 'Bottom' strand is denoted with a 'T' or 'B'. Type II spacers are specified with a '-tII' suffix.
(b) Type III spacers were chosen to avoid homology between the 5' crRNA tag and the target flanking sequences, which was shown to prevent immunity in *S. epidermidis* (Marraffini and Sontheimer, 2010b). Type II spacers match the nearest sequence with a 'NGG' PAM motif flanking the target's 3' end, required for Cas9-mediated immunity (Deveau et al., 2008; Jiang et al., 2013a).
(c) Numbers reflect the coordinates of target regions for each spacer in the ϕNM1 (NC_008583.1), pGO1 (NC_012547.1), or ϕNM4 (NC_028864.1) genomes.
(d) Predicted gene product of the nearest ORF for each target, according to the NCBI annotations for ϕNM1, pGO1, or ϕNM4.

Table 2-2. Oligos used in Chapter 2.

Name	Sequence	Primary Purpose
oGG6	TACCCTAGTTAACGTCTCTTG	Specific identification of ϕNM1
oGG7	GATATCAACTTGTAGTGCATCG	Specific identification of ϕNM1
oGG8	TTGTAGCAAATCAAGAATTGACAG	Specific identification of ϕNM2
oGG9	CTAATTCATCTTGATTAGACATCTC	Specific identification of ϕNM2
oGG10	GCTGACTTACAAGAAGGTGGAC	Specific identification of ϕNM4
oGG11	GTTGTAATTGGATTAAATTCAGTC	Specific identification of ϕNM4
A10	CTTTGTACTGATGATTTATATACTTCGGCATACG	Construction of pGG3 control/parent vector
L55	TAAATCTAACAACACTCTAA	Construction of pGG3 control/parent vector
L6	AAAGGTACCAAATTTAATGCTATTTTCCTTCGC	Type III CRISPR array verification
L50	AAAAGATCTAATAATGTATTTACGCTGGGGC	Type III CRISPR array verification
oGG12	GTTCTCGTCCCCTTTTCTTCGGGGTGGGTATCGATCCTTTGTACTGATGATTTATATACTTC	Common primer for spacer cloning via PCR
oGG13	AGCAGTCGGCAGAGGCATGAACAGAAACAAAATGTAAATCTAACAACACTCTAAAAAATTG	Construction of pGG4
oGG18	ATAAATATGTATAAGGAAAATGAAAGACTATATGATAAATCTAACAACACTCTAAAAAATTG	Construction of pGG8
oGG19	TATTTATGACGATGAACAAGAACTTAAATTAAAAATAAATCTAACAACACTCTAAAAAATTG	Construction of pGG9
oGG20	AGCGAGATGCATGGTAACCAAAGAATTTTTAAAAATAAATCTAACAACACTCTAAAAAATTG	Construction of pGG10
oGG21	AATGAATTTTTAAAACATTCAGGCAAGGTGCATGCTAAATCTAACAACACTCTAAAAAATTG	Construction of pGG11
oGG22	TTAAATCTTTGATTGCTCTTAGCTCTAGTTATGTATTAAATCTAACAACACTCTAAAAAATTG	Construction of pGG12
oGG23	GTAAACCTTTGATTGCTCTTAGCTCGAGTTATGTGCTAAATCTAACAACACTCTAAAAAATTG	Construction of pGG13
oGG24	GTAAACCTTTGATTACTCTTAGCTTTAGTTATGTGTTAAATCTAACAACACTCTAAAAAATTG	Construction of pGG14
oGG46	CTTAGGTGACATTGGCTGTCGATTTTACACTGAAGTAAATCTAACAACACTCTAAAAAATTG	Construction of pGG15
oGG47	CAGAAAGTGTATTGCAACAGATTGGCTCAAAAGTTTAAATCTAACAACACTCTAAAAAATTG	Construction of pGG16
oGG48	ATGTCACCTAAGTCAACACCATCATTTTTTATTCTTAAATCTAACAACACTCTAAAAAATTG	Construction of pGG17
oGG49	TTATGATTTTTTGGAGCATATAAATCATTTAGTGTTAAATCTAACAACACTCTAAAAAATTG	Construction of pGG18
oGG56	CATTTTGTTTCTGTTCATGCCTCTGCCGACTGCTTAAATCTAACAACACTCTAAAAAATTG	Construction of pGG19
oGG72	TCATATAGTCTTTCATTTTCCTTATACATATTTATTAAATCTAACAACACTCTAAAAAATTG	Construction of pGG24
oGG84	AATTTTTCATTATCAATTGCAACTTCCTCGAAATTTAAATCTAACAACACTCTAAAAAATTG	Construction of pGG30
oGG85	AATTTCGAGGAAGTTGCAATTGATAATGAAAAATTTAAATCTAACAACACTCTAAAAAATTG	Construction of pGG31
oGG91	TTTTTAATTTAAGTTCTTGTTCATCGTCATAAATATAAATCTAACAACACTCTAAAAAATTG	Construction of pGG34
oGG101	ATACATAACTAGAGCTAAGAGCAATCAAAGATTTAATAAATCTAACAACACTCTAAAAAATTG	Construction of pGG36
oGG115	ATTCGTCATCTTCAAGTAATGCCTCTAAATCAATAATAAATCTAACAACACTCTAAAAAATTG	Construction of pGG41
oGG121	TGAGACCAGTCTCGGAAGCTCAAAGGTCTCTTAAATCTAACAACACTCTAAAAAATTG	Construction of pGG3-BsaI
oGG122	GAACTTATTGATTTAGAGGCATTACTTGAAGATGACGAAT	Construction of pGG40
oGG123	TTTAATTCGTCATCTTCAAGTAATGCCTCTAAATCAATAA	Construction of pGG40
oGG175	GAACAATTTTTAATTTAAGTTCTTGTTCATCGTCATAAA	Construction of pGG52
oGG176	TTTATTTATGACGATGAACAAGAACTTAAATTAAAAATT	Construction of pGG52
oGG177	GAACTTTTTAAAAATTCTTTGGTTACCATGCATCTCGCT	Construction of pGG53
oGG178	TTTAAGCGAGATGCATGGTAACCAAAGAATTTTTAAAAA	Construction of pGG53
oGG179	GAACGCATGCACCTTGCCTGAATGTTTTAAAAATTCATT	Construction of pGG54
oGG180	TTTAAATGAATTTTTAAAACATTCAGGCAAGGTGCATGC	Construction of pGG54
oGG225	GAACAAGTGATTTAAAAACTGAGGAGCATAAAACTTATTA	Construction of pGG59
oGG226	TTTATAATAAGTTTTATGCTCCTCAGTTTTTAAATCACTT	Construction of pGG59
oGG227	GAACTAATAAGTTTTATGCTCCTCAGTTTTTAAATCACTT	Construction of pGG60
oGG228	TTTAAAGTGATTTAAAAACTGAGGAGCATAAAACTTATTA	Construction of pGG60
oGG58	GAAGAAAAGGGGACGAGAACTAAATCTAACAACACTCTAAAAAATTG	Construction of pGG22/pGG25
oGG61	GGGGTGGGTATCGATCACACATAACTAAAGCTAAGAGTAATC	Construction of pGG22
oGG74	GGGGTGGGTATCGATCCTTTGTACTGATGATTTATATACTTC	Construction of pGG25
L342	AAAGGGCCCAAATAATATTTTCATTATAGCACCTC	Construction of pGG41Δ*csm6*
L343	AAACGGCCGGAAAAAAATAAGGAATTTAAAGAGC	Construction of pGG41Δ*csm6*
W18	TATTCTGAAAAGGTCAATCAAGG	pGG41Δ*csm6* ligation junction verification
W122	AGGTGTTTGTATATTAAGACGTAC	pGG41Δ*csm6* ligation junction verification
L362	AAACTCGTGGATTCTGTGATTTGGATCCTTCC	Construction of pWJ40
W278	AAAAAGATCTTATGACTGTTATGTGGTTATCG	Construction of pWJ40
W270	AAAAAGATCTTGCATAATTCACGCTGACCTC	Construction of pWJ40
W282	AAAACACGAGCGTTTGTTGAACTAATGGGTGC	Construction of pWJ40
oGG82	AACATTGCCGATGATAACTTGAG	Construction of pGG32
oGG83	GTTTTGGGACCATTCAAAACAGCATAGCTCTAAAACCTCGTAG	Construction of pGG32
oGG86	TGAAGCAGTCGGCAGAACATTGCCGATGATAACTTGAGAAAG	Construction of pGG33s
oGG87	TAATGTGCCAGGTCTGTTTTGGGACCATTCAAAACAGCATAG	Construction of pGG33s
oGG89	TTTGAATGGTCCCAAAACAACATTGCCGATGATAACTTGAGAAAG	Construction of pGG33
oGG90	ACAGCATAGCTCTAAAACCTGCCGACTGCTTCATAATGTGC	Construction of pGG33
B220	AAAAAGCGCAAGAAGAAATCAACCAGCGCACTCGTAGACTATTTTTGTCTAAA	Construction of pDB184
B334	ACACTGAGACTTGTTGAGTTCAAACGAAAATTGGATAAAGTGGG	Construction of pDB184
L448	ATTATTTCTTAATAACTAAAAATATGG	Construction of pDB184
B333	CTTTATCCAATTTTCGTTTGAACTCAACAAGTCTCAGTGTGCTG	Construction of pDB184
oGG148	AAACACTTCACACAAGATAACATTATTGATTTAGG	Construction of pGG37-full
oGG149	AAAACCTAAATCAATAATGTTATCTTGTGTGAAGT	Construction of pGG37-full
oGG166	AAACGAAATTTCCAGCAGAAACTTTACCGAAATAG	Construction of pGG51
oGG167	AAAACTATTTCGGTAAAGTTTCTGCTGGAAATTTC	Construction of pGG51
oGG191	GAAGCTTTAGCTTTGCAGTGG	ϕNM1(-Erm) *attL* junction verification
W277	CTGTAATAGACATCGTTCGCAG	ϕNM1(-Erm) *attL* junction verification
oGG206	GCTACATTAATTATAGGGAATCTTAC	ϕNM1(-Erm) *attR* junction verification
W276	TCCTAACAGAAATTGCGTTAAAG	ϕNM1(-Erm) *attR* junction verification
L8	TTTTATACAATACTATTTATAAGTGC	RT-PCR
L86	CATATAGTTTTATGCCTAAAAACC	RT-PCR

Table 2-2. Oligos used in Chapter 2 (continued).

L87	ATATATTTATTTGGCTCATATTTGC	RT-PCR
L484	AAACTCGAGCGCGCAAGCTGGGGATCCG	Construction of pKL55-iTET-B
L485	AAACTCGAGTAGGTACTAAAACAATTCATCCAG	Construction of pKL55-iTET-B
L482	AAACTCGAGCTGAGAGTGCACCATATGCGG	Construction of pKL55-iTET-B
L483	AAACTCGAGCTTAATAGCTCACGCTATGCCG	Construction of pKL55-iTET-B
oGG108	TAATTCCTCCTTTTTGTTGACATTATATCATTGATAGAGTTATTTG	Construction of pKL55-iTET-RC12
oGG109	ACTCTATCAATGATATAATGTCAACAAAAAGGAGGAATTAATGATG	Construction of pKL55-iTET-RC12
oGG110	TGACACTCTATCATTGATAGAGCATAATTAAAATAAGCTTGATATC	Construction of pKL55-iTET-RC12
oGG111	AAGCTTATTTTAATTATGCTCTATCAATGATAGAGTGTCAATATTT	Construction of pKL55-iTET-RC12
oGG112	TTGATAGAGTGATATCGAATTCGGAGGCATATC	Construction of pKL55-iTET-RC12
oGG113	TGATAGAGAGCTTATTTTAATTATGCTCTATC	Construction of pKL55-iTET-RC12
oGG124	GATCTCAAGATAACATTATTGATTTAGAGGCATTACTTGAAGATGACGAATTAGAAGCAAACCGC	'Forward' target insertion
oGG125	GGTTTGCTTCTAATTCGTCATCTTCAAGTAATGCCTCTAAATCAATAATGTTATCTTGA	'Forward' target insertion
oGG126	GATCTTTTGCTTCTAATTCGTCATCTTCAAGTAATGCCTCTAAATCAATAATGTTATCTTGCCGC	'Reverse' target insertion
oGG127	GGCAAGATAACATTATTGATTTAGAGGCATTACTTGAAGATGACGAATTAGAAGCAAAA	'Reverse' target insertion
oGG64	TCTTATTCAAGACAACACTTACAC	Verification of inducible target insertions
oGG88	ATCTAACATCTCAATGGCTAAGG	Verification of inducible target insertions
oGG102	CCACATACCTATATCTGCCCTTTTTCTGCCCTTTTTATTTTTAAAG	*attP* inversion
oGG103	GTGTACTAAAAGGTAATCGATACGGTTATATTTATTCCC	*attP* inversion
oGG104	GGCAGAAAAAGGGCAGATATAGGTATGTGGTTTTGTATTGG	*attP* inversion
oGG105	TATAACCGTATCGATTACCTTTTAGTACACAAGTTTTTC	*attP* inversion
oGG50	GTTAATGTTACGAATGATGAACC	ϕNM4 and pKL55-iTET *attL* junction
oGG51	TTGGCAAGTTCTGCACCTTTAC	ϕNM4 and pKL55-iTET *attR* junction
oGG96	AAGATGCAACAATGGGAACCAAG	ϕNM4 and pKL55-iTET *attL/R* junctions
oGG25	CTAAATGTGATATAATAAAATAAAAAG	Spacer 2B (*xis*) target verification
oGG26	ATAAAGACACCGATTCAACTATG	Spacer 2B (*xis*) target verification
oGG38	AAGATAAAGAATTTGCTCAAGACG	Spacer 32T target verification
oGG39	TTCATCAGCTGACATTAACTCAC	Spacer 32T target verification (ϕNM1/ϕNM2)
oGG40	ACCATTAAAACTCGTCATTCTTTC	Spacer 32T target verification (ϕNM4)
oGG52	GGTGCTAGCTTCGTAAGAAAGG	Spacer 61T-1 target verification
oGG53	CAGCTTACAACGAACATAACCAG	Spacer 61T-1 target verification
oGG233	GCAAGAGAGTTAAAAGGTATACG	Spacer 43B-tII target amplification
oGG234	CTGTATATCCTTGTATCAACTATC	Spacer 43B-tII target amplification
W176	CCTATCTGACAATTCCTGAATAG	Type II CRISPR array amplification
W234	GCTTATTAACGATTCATTATAACC	Partial sequence verification of pWJ153(γ)
W235	TCATAAAGTCTAACACACTAGAC	Partial sequence verification of pWJ153(γ)
oGG229	TCCTTATACATATTTATAGGTGTTACATGTTCATATTTATCAGAGCTCGTG	Construction of pWJ153γ
oGG230	AAATGAAAGACTATATGATTTAAAACAAGATCTCCTAGGTCATTTGATATGC	Construction of pWJ153γ
oGG241	CGTTTCGGTACTTATTTCAACAC	43B-tII target deletion escaper mapping
oGG245	GTTAATTCTATGTCCATTTGTAACC	Deletion junction sequencing
oGG181	GGGGACAAGTTTGTACAAAAAAGCAGGCTATTCGAAATTGTACCTGTTTCATCTC	ϕNM1-Erm construction
oGG182	CGAAAAAAGAGTGTCTTGTGATGGTATCATATCGGTATCAAATAAC	ϕNM1-Erm construction
oGG183	TGATACCGATATGATACCATCACAAGACACTCTTTTTTCGCACC	ϕNM1-Erm construction
oGG184	CTATGAACATATTTGATTAACGTATATAGATTTCATAAAGTCTAAC	ϕNM1-Erm construction
oGG185	CTTTATGAAATCTATATACGTTAATCAAATATGTTCATAGCTTGATG	ϕNM1-Erm construction
oGG186	GGGGACCACTTTGTACAAGAAAGCTGGGTCATTAGATATAAAGATGTATACGG	ϕNM1-Erm construction
L29	TACGACTCACTATAGGGG	ϕNM1-Erm sequencing
oGG192	TCTACTTAATCTGATAAGTGAGC	ϕNM1-Erm sequencing
L325	AAACCCGGGACGCAAACCGCCTCTCCCC	ϕNM1-Erm sequencing

2.8 Materials and methods used in Chapter 2

Bacterial strains and growth conditions.

Cultivation of *S. aureus* RN4220 (ref. (Nair et al., 2011)), TB4 (ref. (Bae et al., 2006)) and derivative strains was done in TSB media (BD) at 37 °C, except when phage infections were performed, or when otherwise noted (see below). Whenever applicable, media were supplemented with chloramphenicol at 10 μg ml^{-1} to ensure CRISPR plasmid maintenance. Selection for Mup^R strains possessing pG0400 or pG0mut was carried out with mupirocin at 5 μg ml^{-1}. RN4220 strains harboring pCL55-derived insertion vectors were grown similarly, but kanamycin was provided at 25 μg ml^{-1} except during re-culture for competent cell preparation. *E. coli* DH5a was grown in LB Broth (BD) supplemented with kanamycin at 25–50 μg ml^{-1} to maintain pCL55-derived plasmids. Selection for ϕNM1-Erm^R lysogens with resistance to erythromycin (10 μg ml^{-1}) was only applied during the lysogenization protocol as described below, and, where applicable, during the subsequent ϕNM2 sensitivity assays. Soft agar lawns were composed of 50% HIB-agar (BD) supplemented with 5 mM $CaCl_2$ (and chloramphenicol at 10 μg ml^{-1} where applicable) and 100 μl of an overnight culture, and poured over similarly supplemented 100% HIB-agar plates. High-titer phage lysates were harvested from infected soft agar lawns that were sufficiently lysed via overnight growth at 37 °C. Within 72 hours of storage at 4 °C, soft agar was hydrated with 600 μl fresh HIB, scraped and decanted into a 50 ml conical-bottom Falcon tube, supplemented with an

additional 600 µl fresh HIB, and then centrifuged at 4303*g* for 8-10 min. Filtered supernatants containing the phage particles were stored in autoclaved 1.5 ml tubes (Eppendorf) at 4 °C.

Estimation of phage lysate titers.

Serial dilutions were prepared in triplicate and plated on soft agar lawns of RN4220 (technical replicates). Plates were incubated at 37 °C for 16–24 h after drying at room temperature (25 °C).

DNA preparation and cloning.

Plasmid DNA was purified from 2 to 6 ml of *E. coli* DH5a or *S. aureus* RN4220 overnight cultures. For preparation from *S. aureus* cultures, cells were pelleted, re-suspended in 100 µl TSM buffer (50 mM Tris-HCl pH7.5, 10 mM $MgCl_2$, 0.5 M sucrose) then treated with 5 µl lysostaphin (2 mg ml^{-1}) at 37 °C for 1.5 h before treatment with plasmid miniprep reagents from Qiagen. Purification used Qiagen or EconoSpin columns. Restriction enzymes were obtained from New England Biolabs (NEB).

Cloning in *S. aureus* used RN4220 electrocompetent cells unless otherwise stated. For most type III CRISPR plasmids, scarless addition of repeat-spacer units to the pGG3 parent vector was accomplished by 'round-the-horn PCR (ref. (Moore and Prevelige, 2002)) followed by blunt ligation, using common primer oGG12 and spacer-specific

oligonucleotides listed in **Table 2-2**. The pGG3 vector was itself constructed by 'round-the horn PCR using primers L55 and A10 to remove extraneous repeat-spacer elements from the pWJ30β CRISPR array. Construction of pWJ30β was described previously: "p*crispr*" (Hatoum-Aslan et al., 2013). To construct pGG22, the pGG14 plasmid was used as a template for 'round-the-horn PCR using oligos oGG58 and oGG61. The pGG25 plasmid was constructed similarly, except that the oGG74 primer was used with oGG58, and the pGG3 plasmid was used as a template. For construction of the remaining type III CRISPR-Cas plasmids, a modified parent vector (pGG3-BsaI) was created by introducing a placeholder spacer harboring two BsaI restriction sites, to facilitate scarless cloning of spacers by replacement with annealed oligonucleotide pairs possessing BsaI-compatible overhangs. Type III-A CRISPR arrays were amplified with primers L50/L6 and sequenced by Sanger using either forward or reverse primers. The pGG41Δ*csm6* plasmid was obtained via intramolecular ligation of doubly digested PCR products amplified from pGG41 by 'round the horn PCR with the primer pair L342/L343. PspOMI/EagI restriction enzymes were used for digestion, and clones were verified using primers W18/W122 to PCR-amplify and Sanger sequence the ligation junction. The BsaI cloning method was also used to construct type II CRISPR plasmids from the pDB184 parent vector, a modified version of pWJ40 with only the single placeholder spacer. Type II CRISPR arrays were amplified with primers L448 and W176, and sequenced by Sanger using L448. After the cloning of each spacer, plasmid

sizes were verified by restriction digest with BssSI for type III plasmids or BtgI for type II plasmids. The pC194-derived pWJ40 vector contains the full *S. pyogenes* SF370 type II CRISPR-Cas system and was constructed by amplifying *S. pneumoniae* LAM226 genomic DNA with oligonucleotides L362/W278 and pC194 (ref. (Horinouchi and Weisblum, 1982a)) with oligonucleotides W270/W282, followed by digestion of the PCR products with BglII and BssSI and a subsequent ligation. pDB184 was created via Gibson assembly (Gibson et al., 2009) of two PCR fragments: a pWJ40 backbone amplified using primers B220/B334, and a CRISPR array amplified from pCas9 (ref. (Jiang et al., 2013a)) using primers L448/B333.

For construction of pCL55-iTET-derived inducible target vectors, cloning used chemically competent DH5a cells. Briefly, the chloramphenicol resistance cassette was first replaced with a kanamycin resistance cassette amplified from *S. pneumoniae* LAM202-3 using primers L484/L485. This was accomplished by 'round-the-horn PCR on the pCL55-iTET parent vector using primers L482/L483, followed by blunt ligation with the PCR-amplified resistance cassette to create the new pKL55-iTET(-B) parent vector (first-generation). Directionality of the insertion was verified afterwards with an analytical restriction digest using BtgI. Modification of the $P_{xyl/tet}$ and P_{tetR} promoters in accordance with pRAB12 (ref. (Helle et al., 2011)) architecture was achieved via three steps. First, two consecutive overlap PCR steps were used to introduce point mutations with oligonucleotide pairs oGG108/oGG109 and oGG110/oGG111, creating the pKL55-

iTET-R and pKL55-iTET-RC intermediates, respectively. The third step involved 'round-the-horn PCR using oligonucleotides oGG112 and oGG113, followed by blunt ligation, to introduce the downstream operator sequence. The resulting pKL55-iTET-RC12 parent vector (second-generation) harboring the $P_{xyl/tet}$* modifications was used for downstream manipulations. For forward and reverse target insertions, annealed oligonucleotide pairs (oGG124/oGG125 and oGG126/oGG127, respectively) with appropriate overhangs were ligated into the multiple cloning site after digesting the vector with BglII and SacII restriction enzymes. Target insertions were verified by PCR amplification and Sanger sequencing using the primers oGG64 andoGG88. Inversion of the *attP* motif for both forward and reverse target vectors, and for the targetless (pKL55-iTET-RC12) parent vector, was achieved by Gibson assembly of two PCR fragments, using oligonucleotides oGG102/oGG103 for the *attP* motif and oGG104/oGG105 for the backbones. Directional integration into the RN4220 chromosome was verified by amplification of either the *attL* or *attR* junctions using primer pairs oGG50/oGG96 and oGG51/oGG96, respectively. The pWJ153-inducible target vector is a pKL55-iTET-RC12- and pE194- (ref. (Horinouchi and Weisblum, 1982b)) derived plasmid constructed via multiple steps of either 'round-the-horn PCR followed by blunt ligation or Gibson assembly. A region that contains the $P_{xyl/tet}$* inducible promoter components and the *nes* target was sequenced by Sanger using forward (W234) and reverse (W235) primers that flank the region. The expected sequence of pWJ153, including this Sanger-verified

region, was included as a supplementary sequence in the published work (Goldberg et al., 2014). The pWJ153γ plasmid is a derivative of pWJ153 that was constructed by replacing the *nes* protospacer with the *nes*-γB target via 'round-the-horn PCR followed by blunt ligation using primers oGG229 and oGG230. Clones were checked by PCR-amplifying the ligation junction using the W234/W235 primers, and Sanger sequencing with W234.

Construction of the ϕNM1-Erm^R lysogen was achieved via pKOR allelic exchange (Bae and Schneewind, 2006). Homology arms (~1 kb) were amplified from the chromosome of *S. aureus* RN4220::ϕNM1 using primer pairs oGG181/oGG182 and oGG185/oGG186, while the ~1.25 kb *ermC* resistance cassette was amplified from a pE194 plasmid preparation using primers oGG183 and oGG184. An ~3.25 kb fragment was assembled by SOEing PCR (Horton, 1993) using external primers oGG181 and oGG186 with clonase (QuikChange) *attB* adapters that allowed directional integration into the pKOR vector (Bae et al., 2006). Sequence integrity of the ~3.25 kb insertion was verified by Sanger using primers L29, oGG191, oGG192, W277 and L325.

Preparation of electrocompetent *S. aureus* cells.

S. aureus RN4220, TB4, or derivative strains were grown overnight in TSB medium, diluted 1:100 in fresh medium without antibiotics, then allowed to grow to an attenuance ($D_{600\ nm}$) reading of 0.8–1.0 for RN4220 or 0.7–0.9 for TB4. Measurements

were taken using a NanoDrop 2000c Spectrophotometer (Thermo Scientific) and disposable polystyrene cuvettes. After re-culture, cells were pelleted at 4 °C, and two or three washes were performed using chilled, sterile dH_2O or 10% glycerol. Cells were ultimately re-suspended in 1/100th volume of chilled, sterile 10% glyercol and 50 ml aliquots were distributed for storage at –80 °C.

Conjugation assays.

Conjugation assays were carried out by the filter mating method essentially as described previously (Marraffini and Sontheimer, 2008), except that TSB and TSA media were used throughout, overnight cultures were mixed directly in 5 ml fresh broth at approximately 1:1 attenuances ($D_{600\ nm}$), and growth was carried out at 30 °C for approximately 24 hours following the serial dilution and plating step. Recipients and transconjugants were quantified as the number of Cm^R CFUs or Cm^R+Mup^R CFUs, respectively, scored by eye. Conjugation efficiencies were subsequently calculated as the ratio of transconjugants to recipients.

Efficiency of plaquing assays.

High-titer lysates (~10^{11} PFU ml^{-1}) of either ϕNM1, ϕNM1γ6 or ϕNM2 were serially diluted in triplicate and applied to RN4220 or TB4 soft agar lawns harboring different CRISPR-Cas plasmids, including pGG3 or pDB184 non-targeting control lawns infected

in parallel (technical replicates). Plates were incubated at 37 °C for 18 h. After incubation, plates were monitored at bench top for up to 24 h to facilitate quantification of plaque-forming units. Efficiency ratios were calculated as the number of plaques formed on a lawn of interest divided by the number of plaques formed on the non-targeting control lawn infected in parallel.

Isolation of type III CRISPR-escape mutant phages.

Soft agar lawns of TB4 harboring a CRISPR-Cas plasmid of interest were infected with high-titer lysates (~10^{11} PFU ml^{-1}) of ϕNM1, ϕNM2, or ϕNM4 either by spotting on the lawn or mixing with cells prior to pouring. Plaques that escape targeting were picked, resuspended in a small volume of HIB broth, and subjected to at least one additional round of counter-selection on lawns with the same CRISPR-Cas plasmid. Subsequent propagation on non-targeting lawns initially used TB4 lawns harboring the pGG3 plasmid. High-titer phage lysates were ultimately harvested from sufficiently lysed soft agar lawns, as described in the "Bacterial strains and growth conditions" section above.

Quantification of erythromycin-resistant lysogens.

Overnight cultures of RN4220 with respective CRISPR-Cas plasmids were inoculated in triplicate from single colonies in HIB medium supplemented with chloramphenicol (biological replicates). After chilling at 4 °C, 1:10 dilutions were prepared in 1 ml fresh

HIB supplemented with chloramphenicol and 5mM $CaCl_2$. Diluted cultures were infected with ϕNM1-Erm^R at ~MOI 10 and incubated on ice for 30 min. After incubation on ice, cultures were transferred to a 37 °C incubator for 30 min with shaking. Serial dilutions from each culture were then applied to HIB-agar plates supplemented with chloramphenicol, erythromycin and 5 mM $CaCl_2$ for quantification of lysogenic colony-forming units. In selected cases, type III-A CRISPR locus and target sequence integrity was verified by colony PCR after re-streaking single colonies using primer pairs L6/L50 (CRISPR array) and oGG25/oGG26 (ORF 2) or oGG38/oGG39 (ORF 32). Where applicable, Sanger sequencing of PCR products was also performed using these primers. When verifying type II lysogenization isolates, the spacer 43B-tII target region was amplified using primers oGG233 and oGG234, and the type II CRISPR array was amplified using L448 and W176. The presence of integrated ϕNM1 or ϕNM1-Erm^R prophages was confirmed by colony PCR using primer pairs oGG191/ W277 and oGG206/W276 to amplify the *attL* and *attR* junctions, respectively. To estimate the total number of recipient cells, serial dilutions of untreated overnight cultures were plated on TSB- or HIB-agar supplemented with chloramphenicol.

Streak-test (phage-sensitivity) assays.

High-titer phage lysates (~10^{11} PFU ml^{-1}) were applied to the surface of a pre-dried HIB-agar plate supplemented with 5 mM $CaCl_2$ and appropriate antibiotics, then allowed to

dry for an additional ~30 min at room temperature (25 °C). Single colonies isolated from egg yolk screens, CRISPR-Cas plasmid transformations, or ϕNM1-Erm^R lysogeny experiments were streaked through the phage-seeded region using a sterile plastic loop, then incubated for ~12-16 h at 37 °C.

Quantification of PFU in lysogenic culture supernatants or other lysates.

Overnight cultures of either RN4220::ϕNM1-Erm^R or RN4220::ϕNM1 lysogens harboring targeting or non-targeting CRISPR-Cas plasmids were inoculated in triplicate from single colonies in HIB media supplemented with chloramphenicol (biological replicates). After overnight growth, cells were transferred to 4 °C then pelleted by centrifugation at 4696g for 5 min. Supernatants were filtered, and 100 μl from each lysate was mixed with 100 μl of either an indicator strain or targeting strain overnight culture for plating by the soft agar method. After drying at room temperature (25 °C), plates were incubated for 18 h at 37 °C. When quantifying plaque-forming potential (PFU / ml) from high-titer lysates of ϕNM1, ϕNM2, or ϕNM4 raised on lawns of TB4 harboring the pGG3 plasmid, 5 ul of diluted or undiluted lysate was spotted and dripped on an uninfected lawn of interest after drying at room temperature (25 °C), and then incubated for 18 h at 37 °C.

Screens for lipase-negative ϕNM4 lysogens.

An overnight culture of *S. aureus* TB4 harboring the spacer 32T or spacer 61T-1 CRISPR-Cas plasmid was re-cultured to log phase growth in HIB medium supplemented with 5 mM $CaCl_2$. Re-cultures were treated with ϕNM4 at ~MOI 50 after measurement of attenuance ($D_{600\ nm}$) to approximate cell densities. After incubation with phage for 1 h, cells were plated on TSA supplemented with 5% egg yolk emulsion. After ~24 h incubation at 37 °C, approximately 1,000 colonies were inspected for lipase secretion. Lipase-negative candidates were re-streaked to single colonies for PCR analysis and functional testing.

Phage DNA isolation and deep sequencing.

Samples of high-titer phage lysates (~10^{11} PFU ml^{-1}) were treated with DNase and RNase to a final volume of 150 μl for 1 h at 37 °C. Samples were treated with EDTA (pH 8.0) to a final concentration of 20mM, followed by treatment with SDS to a final concentration of 0.5% and 2 μl proteinase K. Samples were incubated for 1 h at 65 °C, then subjected to a PCR purification protocol (Qiagen). Paired-end library preparation was performed on purified phage DNA using a Nextera Tagmentation protocol (Illumina), and samples were pooled for multiplexed sequencing on a MiSeq (Illumina). *De novo* assembly of phage genomes used ABySS (Simpson et al., 2009).

RNA preparation for reverse transcription PCR and RNA sequencing.

For reverse transcription PCR (RT-PCR), overnight cultures were diluted 1:20 in 25 ml fresh media and grown for 2.5 h at 37 °C with shaking. After re-culture, cells were pelleted and washed twice in 1 ml ice cold TSM buffer, then treated with 3 μl lysostaphin (2 mg ml^{-1}) for 20 min at 37 °C in 500 μl TSM buffer. Treated cells were pelleted then re-suspended in 750 μl cold TRIzol Reagent (Life Technologies) after discarding the supernatant. The following chloroform extraction and precipitation was performed according to the manufacturer's protocol. After resuspension in dH_2O, samples were treated with Qiagen DNase I for 45 min at 30 °C, then re-purified using RNeasy Cleanup columns (Qiagen). In some cases, it was necessary to repeat this step a second time to ensure the complete removal of DNA. After cleanup, all samples were again treated with DNase I (Sigma-Aldrich) for 30–45min at 30 °C, before use in the reverse transcription reaction.

For RNA sequencing, overnight cultures were diluted 1:100 in fresh HIB supplemented with chloramphenicol and 500 μM $CaCl_2$, and grown for 1.5 h (approximately mid-log phase) at 37 °C with shaking. Cultures were removed, infected at MOI ~20 then split into 10 ml portions for an additional 6, 15, 30 or 45 min of growth. Immediately following incubation, samples were mixed with 10 ml of a 1:1 acetone/ethanol solution and transferred to –80 °C. The ϕNM1 lysogen was grown similarly, except without antibiotics, and harvested immediately after the 1.5 h re-

culture at 37 °C. After at least one overnight at –80 °C, samples were thawed on ice and pelleted by centrifugation at 4696g for 10 min. After two washes of 1 ml TE buffer, cells were re-suspended in 1 ml RLT buffer (Qiagen) supplemented with BME, and transferred to 2-ml tubes pre-loaded with ~0.5–1 cm^3 of 0.1 mm glass beads (BioSpec). Samples were processed in a Mini-Beadbeater instrument (BioSpec) three times for 10 s at 4,200 oscillations per minute, with 40 s of chilling on ice between runs. After beadbeating, samples were spun down for 2 min at 16,100g in a refrigerated microcentrifuge. Supernatant (750 μl) was transferred to a clean tube for mixing with 500 μl of 100% ethanol, and the following RNeasy purification was done according to the manufacturer's protocol (Qiagen). After elution, samples were treated with either Qiagen or Sigma-Aldrich DNase I for 30–45 min at 30 °C, then re-purified using RNeasy cleanup columns. In some cases, it was necessary to repeat this step a second time to ensure the complete removal of DNA. rRNA-depleted samples were subsequently generated using the RiboZero Magnetic Kit for bacteria (Epicentre), according to the manufacturer's protocol.

RNA preparation for reverse transcription PCR and RNA sequencing.

Reverse transcription used M-Mulv Reverse Transcriptase (NEB), with DNA-free total RNA isolated from RN4220 cultures harboring either the pNes(wt-*d*) or pNes(wt-*i*) plasmids as templates for cDNA synthesis. For pNes(wt-*d*), reverse transcription used

either the L8 or L86 primers in two separate 30 ml reactions, alongside mock reactions (–RT enzyme). For pNes(wt-*i*), the same was done using primers L8 or L87. After incubation, 1 μl of each reaction was used as a template for PCR, with respective primer pairs for each sample.

Phage transcriptome analysis and visualization.

Reads were aligned to reference genomes using Bowtie and sorted using Samtools. Using a custom script, sorted reads were accessed via Pysam, normalized as reads per million values, and plotted in log scale as the average over consecutive windows of 500 base pairs using matplotlib tools for IPython.

Transformation assays.

S. aureus RN4220 plasmid preparations were dialyzed on 0.025 μm nitrocellulose filters (Millipore) then quantified using a NanoDrop 2000c Spectrophotometer (Thermo Scientific). Aliquots (50 μl) of electrocompetent cells were transformed in triplicate with 80 ng dialyzed miniprep DNA per transformation (or as otherwise noted in figure legends) using a Gene Pulser Xcell (BioRad) with the following parameters: 2900 V, 25 μF, 100 Ω, 2 mm (technical replicates). After electroporation, cells were immediately re-suspended in TSB to a final volume of 200 μl and recovered at 30 °C for 2 h with shaking. Serial dilutions were then prepared for plating with appropriate antibiotics.

For targeting of iTET insertion vectors, additional plating in the presence of ATc at a final concentration of 0.5 μg ml^{-1} was performed in parallel using the same dilutions. Plates were incubated at 37 °C for 18–24 h.

Plate reader high-resolution growth curves.

For ATc induction experiments, overnight cultures were launched from single colonies in triplicate (or as otherwise noted in the figure legend) and diluted 1:200 in TSB broth (biological replicates). After 1 h of growth, ATc was added at a final concentration of 0.5 μg ml^{-1} where applicable. Measurements were taken every 5 min. For mitomycin C induction experiments, overnight cultures were launched from single colonies in duplicate and diluted 1:100 in HIB broth (biological replicates). After 1.5 h of growth, mitomycin C was added at a final concentration of 0.5 μg ml^{-1} where applicable. Measurements were taken every 10 min. For ϕNM1 infections, overnight cultures were launched from single colonies in triplicate and diluted 1:100 in HIB broth supplemented with $CaCl_2$ 5 mM (biological replicates). After 1 h 25 min of growth, attenuance ($D_{600\,nm}$) was measured for three representative cultures to estimate MOI. Aliquots were then loaded into 96-well plates along with ϕNM1 at the appropriate MOI (10 or 100), where applicable. Measurements were taken every 5 min. For ϕNM1γ6 infections, overnight cultures were launched from single colonies in triplicate and diluted 1:200 in HIB supplemented with $CaCl_2$ 5 mM (biological replicates). An average attenuance ($D_{600\,nm}$)

measurement was taken after 1 h of growth, and ϕNM1γ6 was added at an MOI of 10 on the basis of this value, where applicable. Measurements were taken every 5 min.

Plasmid-curing assays.

RN4220 strains harboring the CRISPR-Cas plasmid (pGG3 or pGG8) and the target plasmid (pWJ153 or pWJ153γ) were cultured in TSB supplemented with chloramphenicol (10 μg ml^{-1}) to an attenuance ($D_{600\ nm}$) of 0.45. Strains harboring the pGG8 and pWJ153γ combination also harbored the pG0mut plasmid as a bystander plasmid to assay for targeting in *trans*. After splitting the cultures in two, transcription across the target was induced for one of the cultures via the addition of ATc to a final concentration of 0.250 μg ml^{-1} or 0.125 μg ml^{-1} as noted in the figure legends. Where applicable, aliquots of cells from each culture were harvested before (0) and after (1, 2, 3, 4, 5 and 6 h) the time of induction for miniprepping of plasmids. In parallel, serial dilutions of both cultures were prepared in triplicate for each time point and plated on TSA plates supplemented with chloramphenicol and erythromycin or chloramphenicol alone, for quantification of antibiotic-resistant CFU (technical replicates). In experiments with pG0mut, serial dilutions for each time point were also plated on TSA plates supplemented with chloramphenicol and mupirocin. After miniprepping, plasmids were linearized with the common single cutter BamHI and subjected to agarose gel electrophoresis.

CHAPTER 3: INVESTIGATION OF PHENOTYPIC CONSEQUENCES ASSOCIATED WITH MAINTENANCE OF CONDITIONAL TOLERANCE IN LYSOGENIZED *S. AUREUS* LINEAGES

The work described in **Chapter 2** established key insights into the targeting mechanism of a type III-A CRISPR-Cas system, and thereby provided a foothold for understanding the biology of these systems in the context of temperate phage infections. Prior to that work, it was unclear whether stable co-existences between active CRISPR-Cas systems and their temperate phage targets could occur, because CRISPR-Cas targeting was shown to resist lysogenic infections by phage lambda (Edgar and Qimron, 2010). This finding, however, made use of a particular transcription-independent type I-E CRISPR-Cas system in *E. coli*. While I observed similar results with a transcription-independent (type II-A) CRISPR-Cas system in *S. aureus*, the type III-A system I tested, in contrast, was found to prevent temperate phage propagation when targets are transcribed during lytic infection or prophage induction, and tolerate prophages in cases where targets are sufficiently repressed in the chromosome (Goldberg et al., 2014). Thus, this work provided the first demonstration that a temperate phage which was otherwise targeted by a CRISPR-Cas system could be acquired via lysogeny without mutating its target, and even in the absence of CRISPR-Cas inactivation via inhibition (Bondy-Denomy et al., 2013) or mutation. Unlike inactivation of transcription-independent

CRISPR-Cas systems by temperate-phage-encoded targeting inhibitors (Bondy-Denomy et al., 2013; Pawluk et al., 2014), conditional tolerance via transcription-dependent targeting ensures that the host can maintain type III-A resistance to subsequent lytic infections by heteroimmune phages that bear a matching target. Another consequence of transcription-dependent targeting is that tolerance breaks down during prophage induction, when the lytic cycle reinitiates and target sequences are transcribed. However, additional consequences for the lysogenic hosts that maintain conditional tolerance have not been described.

In preliminary attempts to determine the effect of prophage induction on RN4220::ϕNM1 lysogens harboring the spacer 43T CRISPR-Cas plasmid, I found no evidence that these conditionally tolerant lysogens survived better than non-targeting lysogens following treatment with MMC, even at a sub-inhibitory concentration of 0.5 μg/ml (data not shown). Therefore, an *a priori* reason to suspect that conditional tolerance could provide a direct advantage to lysogenized lineages was, in my view, still lacking. In fact, the scarcity of type III-A systems observed among staphylococcal isolates seemed to argue the opposite: that these systems might be too detrimental—or at least insufficiently advantageous—to be maintained in those populations. Furthermore, given that conditional tolerance contributes to the genetic stability of the type III-A CRISPR-Cas system itself, it could be viewed as a paradigm of genetic 'selfishness' (Doolittle and Sapienza, 1980; Orgel and Crick, 1980) first and foremost—

and perhaps *should* be. The stabilizing effects of conditional tolerance, for example, might be sufficient to explain why this phenomenon has also evolved as an emergent property of other phage defense mechanisms, such as SaPI interference (Ram et al., 2014) and abortive infection (Depardieu et al., 2016). Viewing conditional tolerance in a selfish light makes it easier to envision a scenario where negative side-effects associated with its maintenance over the long term could provide pressure to abandon the phage-defense system responsible—despite the short-term stabilizing effects it can afford. For example, considering that type III-A targeting was found to occur during prophage induction (which, can occur spontaneously even in growing populations), it seemed plausible to me that maintenance of conditionally tolerant type III-A systems in lysogenized lineages could be accompanied by unforeseen phenotypic consequences. The work described in **Chapter 3** was designed to investigate this possibility. An understanding of such phenotypes, in turn, could offer valuable insight into factors that influence the stability and distribution of type III-A systems among natural, lysogenization-prone populations of *S. aureus* (Aanensen et al., 2016; Cao et al., 2016; Goerke et al., 2009).

In an effort to emulate natural systems more closely, I once again began using the Newman phages' native host background, TB4, as a model host for investigating the biology of type III-A temperate phage targeting in clinically-relevant strains of *S. aureus*. Various mutations had been identified previously (Nair et al., 2011) in the genome of *S.*

aureus RN4220, including mutations in DNA repair and housekeeping genes, which could potentially have an impact during CRISPR-Cas targeting (although not necessarily). Moreover, I resolved to exclusively work with wild type CRISPR array architectures where spacers are flanked by repeats on both sides. To facilitate the cloning of such CRISPR-Cas systems, I constructed a new set of BsaI-based parent vectors. The pGG-BsaI-R parent vector was constructed by adding a repeat downstream of the BsaI placeholder spacer in pGG-BsaI (Samai et al., 2015), and is suitable for constructing single-spacer arrays with the 'R-S-R' architecture. However, after a slew of inefficient cloning attempts with this vector, I endeavored to improve the cloning efficiency by eliminating a base pair from the 3' end of the BsaI placeholder spacer to produce different downstream overhangs upon restriction digest. Aside from this change, the resulting pGG78 plasmid is otherwise identical to pGG-BsaI-R. The pGG79 parent vector, also constructed at this time, is a variant of pGG78 with the *nes* spacer still present in position 1, and was derived from pGG3-BsaI (**Figure 3-1**). Whether or not the placeholder spacer modification was effective in improving the cloning efficiency was not rigorously determined, and remains unclear to me. In any event, I began using the pGG79 vector for downstream manipulations, because I decided that it might still on occasion be useful to assay whether the backbone is functional via the *nes* spacer in position 1 (independently of the variable spacers I introduce at position 2). Such

functional testing became even easier after my colleague, Dr. Varble, engineered a variant of ϕNM4γ4 with the 32T spacer's target replaced by the *nes* protospacer. Indeed,

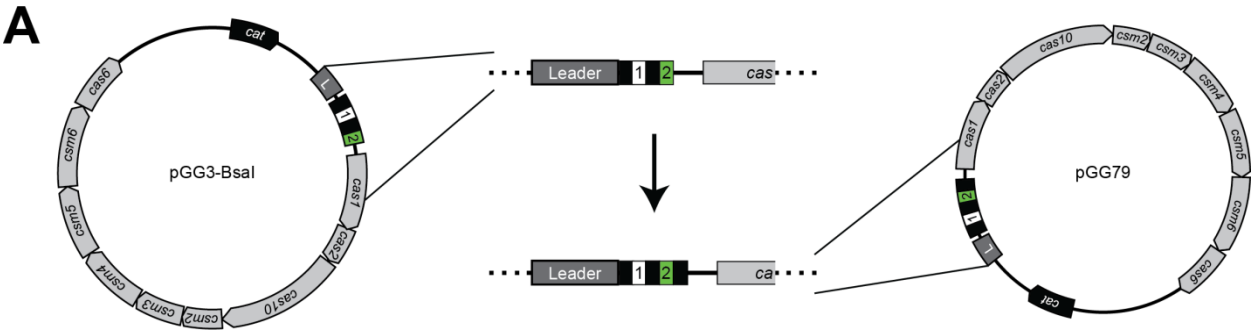


Figure 3-1. Generation of a modified BsaI parent vector for cloning of type III-A CRISPR-Cas plasmids with wild type CRISPR architecture.

(A) Schematic representations of the pGG3-BsaI (left) and pGG79 (right) type III-A CRISPR-Cas plasmids. pGG3-BsaI harbors the type III-A system from *S. epidermidis* RP62a, including its conserved leader sequence ('L', dark grey box), but with a modified CRISPR array containing only two spacers. The wild type *nes* spacer is located at position 1, while its 30 bp placeholder spacer (green) containing BsaI restriction sites to facilitate oligo cloning is located at position 2. To construct pGG79, an additional CRISPR repeat was added downstream of the BsaI placeholder spacer in pGG3-BsaI (middle). pC194 backbones include a chloramphenicol acetyltransferase (*cat*) gene conferring resistance to chloramphenicol, which is also depicted alongside *cas* genes in the plasmid diagrams.

CRISPR-Cas-mediated resistance to this phage, which we called ϕNM4γ4α2, could be assayed with a simple streak-test if needed.

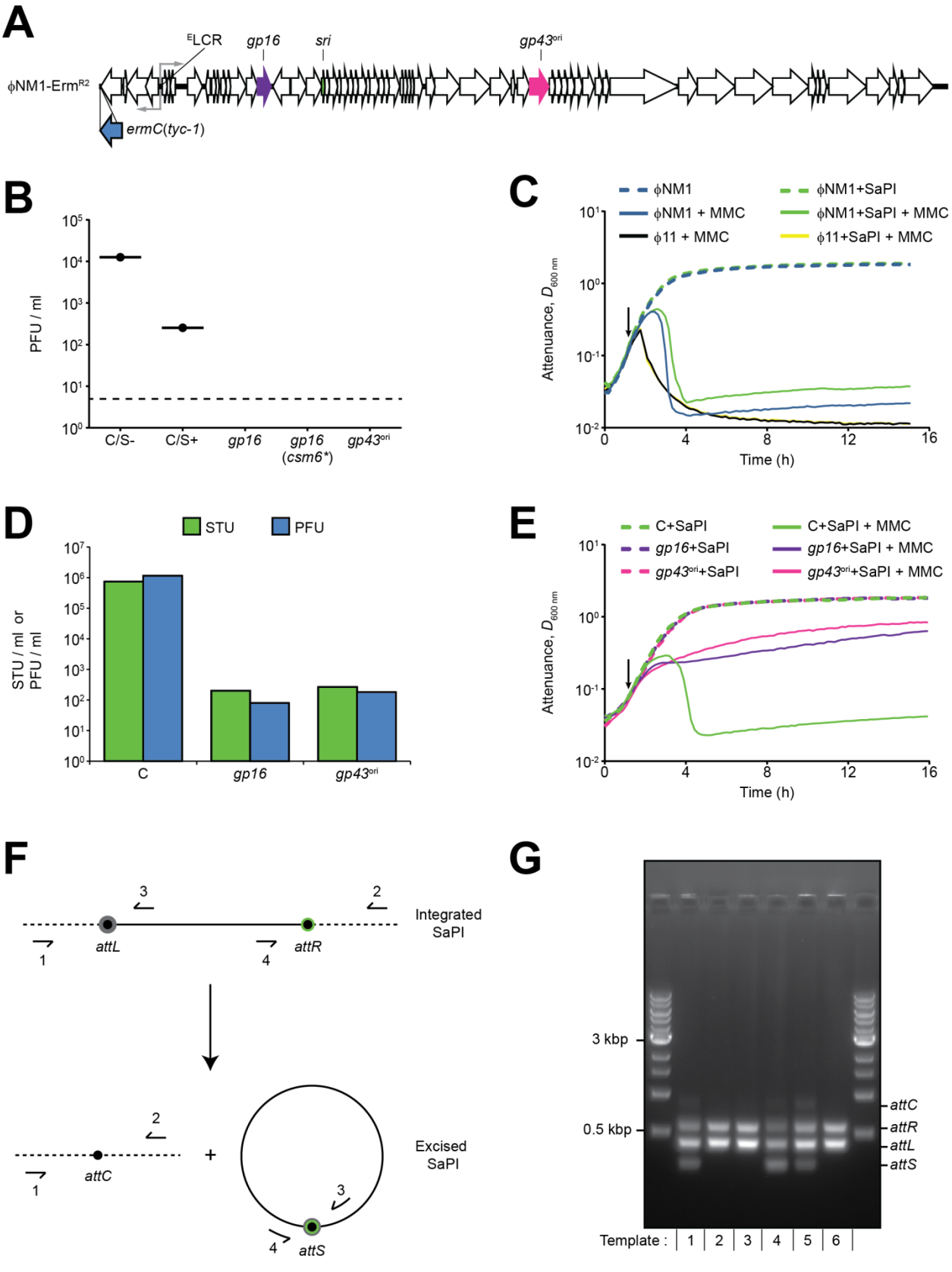
Now that I was testing various lysogenic derivatives of TB4, it became more efficient to introduce plasmids via generalized transduction rather than transformation.

Transformation required that I would have to prepare different batches of competent cells for each strain, and then transform them with each (miniprepped) plasmid. Transduction only required that a single transducing lysate was prepared per plasmid, which could then be used to deliver the plasmid to each strain directly during log-phase growth. One potential obstacle here is that preparation of the transducing lysate requires efficient lysis of the CRISPR-Cas plasmid's host by a phage; i.e., the phage should not harbor a target for the CRISPR-Cas system. However, I now had various clear-plaque phages at my disposal, so I could readily transduce each CRISPR-Cas plasmid by simply lysing its host with a suitable phage that lacks a target. In addition, I had at my disposal a *tetM*-marked SaPI—SaPI1*tst*::*tetM* (Lindsay et al., 1998) —which our lab had received from the Novick group. This SaPI confers resistance to tetracycline upon delivery and integration into a suitable host chromosome, and its transfer can therefore be readily monitored using standard transduction protocols. In the first section of this chapter, I describe some experiments illustrating how the mobilization of this SaPI in lysogenic hosts can be impeded by type III-A targeting of its helper phage(s) during prophage induction.

3.1 Conditionally tolerant type III-A CRISPR-Cas systems that target helper prophages can impede SaPI mobilization in lysogenic hosts

The SaPI1(*tst*::*tetM*) element can be mobilized by various helper phages, including 80α and ϕNM1 (Dearborn and Dokland, 2012; Lindsay et al., 1998). SaPI1 induction by 80α requires the Sri antirepressor protein (Tormo-Mas et al., 2010), and induction by ϕNM1 is thought to be licensed by its homolog in ϕNM1 (Dearborn and Dokland, 2012). The ORF encoding this *sri* homolog is missing from the original ϕNM1 annotation (NC_008583.1), but is located roughly between ORFs 19 and 20. To facilitate lysogen strain construction, I used an Erm^R-marked derivative of ϕNM1 (ϕNM1-Erm^{R2}) as my helper phage of interest (such that I could readily select for lysogens on the basis of erythromycin resistance). The ϕNM1-Erm^{R2} phage (**Figure 3-2A**) is a derivative of ϕNM1-Erm^R (Goldberg et al., 2014) that I constructed in an effort to improve plating efficiency during lysogenization assays. Its *ermC* insertion was trimmed upstream of the 5′ UTR and contains the *tyc-1* allele SNP that was shown to improve constitutive expression of this cassette in *B. subtilis* (Gryczan et al., 1980). After lysogenizing TB4 (or derivatives of TB4 harboring CRISPR-Cas plasmids), the SaPI1*tst*::*tetM* element was introduced into lysogens via transduction. I first confirmed that the presence of the SaPI alone did not appreciably alter lysis profiles mediated by ϕNM1-Erm^{R2} during MMC-stimulated prophage induction: as expected, lysis was still observed (**Figure 3-2B**). For comparison, I tested TB4 derivatives lysogenized with ϕ11, which is not a helper phage

Figure 3-2. Effects of helper phage targeting in conditionally tolerant, SaPI-containing lysogens.
(A) Schematic representation of the ϕNM1-Erm^{R2} prophage genome, with ORFs (thick arrows) scaled by length in accordance with the ϕNM1 annotation (NC_008583.1). The *ermC*(*tyc-1*) insertion (blue) is located immediately downstream of the ϕNM1 *integrase*. The *sri* antirepressor ORF (green) is located between ORFs 19 and 20 in the original ϕNM1 annotation. Precise locations of the target sites for the *gp16* and *gp43*^{ori} spacers tested in this chapter are also shown above their respective color-coded ORFs. Divergently oriented grey bent arrows signify the presence of P_{cI} and P_{cro} promoters in the ^{E}LCR. **(B)** Quantification of PFU concentrations (PFU / ml) in filtered supernatants from overnight cultures of TB4::ϕNM1-Erm^{R2} lysogens with the indicated CRISPR-Cas plasmid, grown in TSB supplemented with chloramphenicol for 14 hours. Lysogens with the control CRISPR-Cas plasmid are either lacking (C/S-) or harboring (C/S+) the SaPI1*tst*::*tetM* element. Horizontal bars emphasize the measurement from a single experiment, plotted as a black dot in each column. Dashed line represents the limit of detection under these assay conditions. **(C)** High-resolution growth curves of TB4::ϕNM1-Erm^{R2} or TB4::ϕ11 lysogens either lacking or possessing the SaPI1 element as indicated, with MMC added at the indicated time point (black arrow). Solid lines represent attenuance ($D_{600\ nm}$) measurements from single biological replicates in the presence of MMC; for comparison, attenuance measurements from the same biological replicate grown in the absence of MMC is plotted for two strains (dashed lines). **(D)** Quantification of STU and PFU concentrations (STU /ml or PFU / ml) in filtered supernatants from subcultures of TB4::ϕNM1-Erm^{R2} lysogens harboring the SaPI and a CRISPR-Cas plasmid as indicated, four hours post-treatment with MMC at 2.0 μg/ml. After 95 minutes of growth and immediately prior to treatment, subculture attenuances ($D_{600\ nm}$) were normalized to ~0.5. The pGG115 plasmid was used as the non-targeting control (C) in this assay. Additional biological replicates were not performed. **(E)** Same as in 'C', except with TB4::ϕNM1-Erm^{R2} lysogens that harbor both the SaPI and a CRISPR-Cas plasmid as indicated, and all three untreated curves were plotted. The pGG115 plasmid was also used as the non-targeting control (C) in this assay. **(F)** Schematic diagram summarizing the multiplex PCR reaction that allows for a coarse-grained assessment of integrated and excised SaPI1 genome relative abundances. The *attC* and *attS* junctions are only present when the SaPI is excised by a helper phage. Primers '1' through '4' correspond to oGG338 through oGG341. Features are scaled arbitrarily. **(G)** Multiplexed PCR amplification with the oGG338-oGG341 primer set and one of 6 colony lysate templates. Templates 1 through 4 were from TB4::ϕNM1-Erm^{R2} lysogens harboring the pGG79, *gp16*, *gp16*(*csm6**), or *gp43*^{ori} CRISPR-Cas plasmids (in that order). Templates 5 and 6, respectively, were from TB4::ϕNM1 and TB4::ϕ11 lysogens lacking CRISPR-Cas plasmids. All lysogens harbored the SaPI. Size markers of 1 kb and 0.5 kb are indicated, along with the expected banding position for the intact *attC*, *attR*, *attL*, and *attS* junction amplicons. Additional technical replicates were not performed on these templates, but similar banding patterns were observed for the *gp16* and *gp43*^{ori} strains in a separate biological replicate (not shown).


A
ELCR
gp16
sri
gp43ori
ϕNM1-ErmR2
ermC(tyc-1)
B
PFU / ml
C/S-
C/S+
gp16
gp16 (csm6*)
gp43ori
C
ϕNM1
ϕNM1 + MMC
ϕ11 + MMC
ϕNM1+SaPI
ϕNM1+SaPI + MMC
ϕ11+SaPI + MMC
Attenuance, D600 nm
Time (h)
D
STU
PFU
STU / ml or PFU / ml
C
gp16
gp43ori
E
C+SaPI
gp16+SaPI
gp43ori+SaPI
C+SaPI + MMC
gp16+SaPI + MMC
gp43ori+SaPI + MMC
Attenuance, D600 nm
Time (h)
F
attL
attR
Integrated SaPI
attC
attS
Excised SaPI
G
3 kbp
0.5 kbp
attC
attR
attL
attS
Template : 1 2 3 4 5 6

for SaPI1*tst*::*tetM* (Lindsay et al., 1998). Meanwhile, the presence of the SaPI was sufficient to reduce spontaneously induced PFU from ϕNM1-Erm^{R2} lysogens harboring a non-targeting CRISPR-Cas plasmid (**Figure 3-2C**), presumably as a result of SaPI-mediated interference. For comparison, spontaneous induction was measured under the same assay conditions for TB4::ϕNM1-Erm^{R2} lysogens lacking the SaPI but harboring conditionally tolerant type III-A CRISPR-Cas plasmids with either the *gp16* or *gp43*^{ori} spacers. The *gp43*^{ori} spacer harbors the same 36 bp sequence as the 43T spacer published previously (Goldberg et al., 2014), and is identical to the 35 bp variant published as '*gp43*' (Jiang et al., 2016) aside from the one bp at its 3' end. For the next experiments, I introduced the *gp16* or *gp43*^{ori} CRISPR-Cas plasmids into TB4::ϕNM1-Erm^{R2} lysogens harboring the SaPI. The presence of these conditionally tolerant type III-A systems prevented lysis when prophage induction was stimulated with MMC (**Figure 3-2D**). To demonstrate that type III-A targeting during prophage induction could impede SaPI transfer, I next measured the number of SaPI transfer units (STUs) in supernatants of lysogenic cultures after a four-hour treatment with MMC. For comparison, PFU were also measured. By interfering with lysis during prophage induction, type III-A plasmids are expected to prevent release of STU and PFU altogether. Indeed, the concentration of STUs and PFUs in culture supernatants were both reduced appreciably for strains harboring conditionally tolerant type III-A plasmids, four hours post-induction (**Figure 3-2E**). To determine whether intracellular mobilization of the SaPI was also reduced by

either of these spacers, I designed a multiplex PCR reaction for amplifying across excised and integrated SaPI attachment junctions (**Figure 3-2F**). The presence of an *attS* and/or *attC* amplicon indicates that the SaPI genome is excised (and possibly already replicated or packaged). Meanwhile, only the *attL* and *attR* junctions can be amplified if the SaPI is stably integrated. PCR was performed on single colony lysates to probe the effects of helper prophage targeting during spontaneous induction. Whereas excised SaPIs were detectable in TB4::ϕNM1-Erm^{R2} lysogens harboring the non-targeting parent vector, this was not the case for lysogens harboring the *gp16* CRISPR-Cas plasmid (compare to control amplification from TB4::ϕ11 lysogens harboring a non-helper prophage, **Figure 3-2G**). Interestingly, excised SaPIs were still detectable with the *gp43*^{ori} CRISPR-Cas plasmid. In fact, the *attC* amplicon appeared to be slightly enriched relative to non-targeting lysogens (**Figure 3-2G**). I speculate that this enrichment results from the accumulation of replicated SaPI genomes in cells which cannot lyse via prophage induction, due to targeting by the *gp43*^{ori} spacer. Previously published work (Jiang et al., 2016) from my colleagues showed that immunity licensed by the *gp43* spacer allows for intracellular accumulation of phage DNA because its target is transcribed late during the lytic cycle, when replication is already under way. In contrast, phage DNA accumulation was not observed when an early-transcribed ORF (*gp14*) was targeted, presumably because the phage could be cleared before replication (Jiang et al., 2016). Given that the *sri* antirepressor ORF falls within an early transcribed region upstream of

gp43^{ori}, it could presumably be expressed before targeting by *gp43*^{ori} occurs. Similarly, the effect on intracellular SaPI mobilization observed with the *gp16* spacer—which also targets an early transcribed ORF—might be explained by the location of the *gp16* target upstream of the *sri* ORF (**Figure 3-2A**). In other words, type III-A targeting at the *gp16* ORF during prophage induction might prevent intracellular SaPI mobilization by interfering with expression of the downstream *sri* ORF. I also tested a variant of the *gp16* CRISPR-Cas plasmid with catalytic site mutations in *csm6* that abolish its HEPN domain's non-specific RNase activity (Jiang et al., 2016), in order to rule out the possibility that the effects observed with *gp16* were due to this activity operating on phage, SaPI, or host transcripts at other loci during targeting. As expected for a spacer matching to an early-transcribed region of ϕNM1, the RNase activity of Csm6 appeared to be dispensable for targeting effects (**Figures 3-2C, D, E, & G**). Whether or not the *gp43*^{ori} spacer allows intracellular accumulation of intact SaPI particles was not assessed, and it's possible that targeting of the head protein disrupts intracellular packaging in addition to preventing completion of the lytic cycle. Extraction of intracellular particles via mechanical or chemical lysis could help to clarify these points. Interactions between a phage-encoded type I-F CRISPR-Cas system and an ~18-kb element, resembling a SaPI-like phage-inducible chromosomal island (PICI), were previously described in *Vibrio cholerae* (Seed et al., 2013). In that work, however, the CRISPR-Cas system directly targeted the PICI-like element, and therefore did not interfere with completion

of the helper phage's lytic cycle. Type III-A systems that directly target SaPIs might be similarly expected to allow a helper phage to complete its lytic cycle, by neutralizing the SaPI, but it is unclear if such systems could evolve naturally.

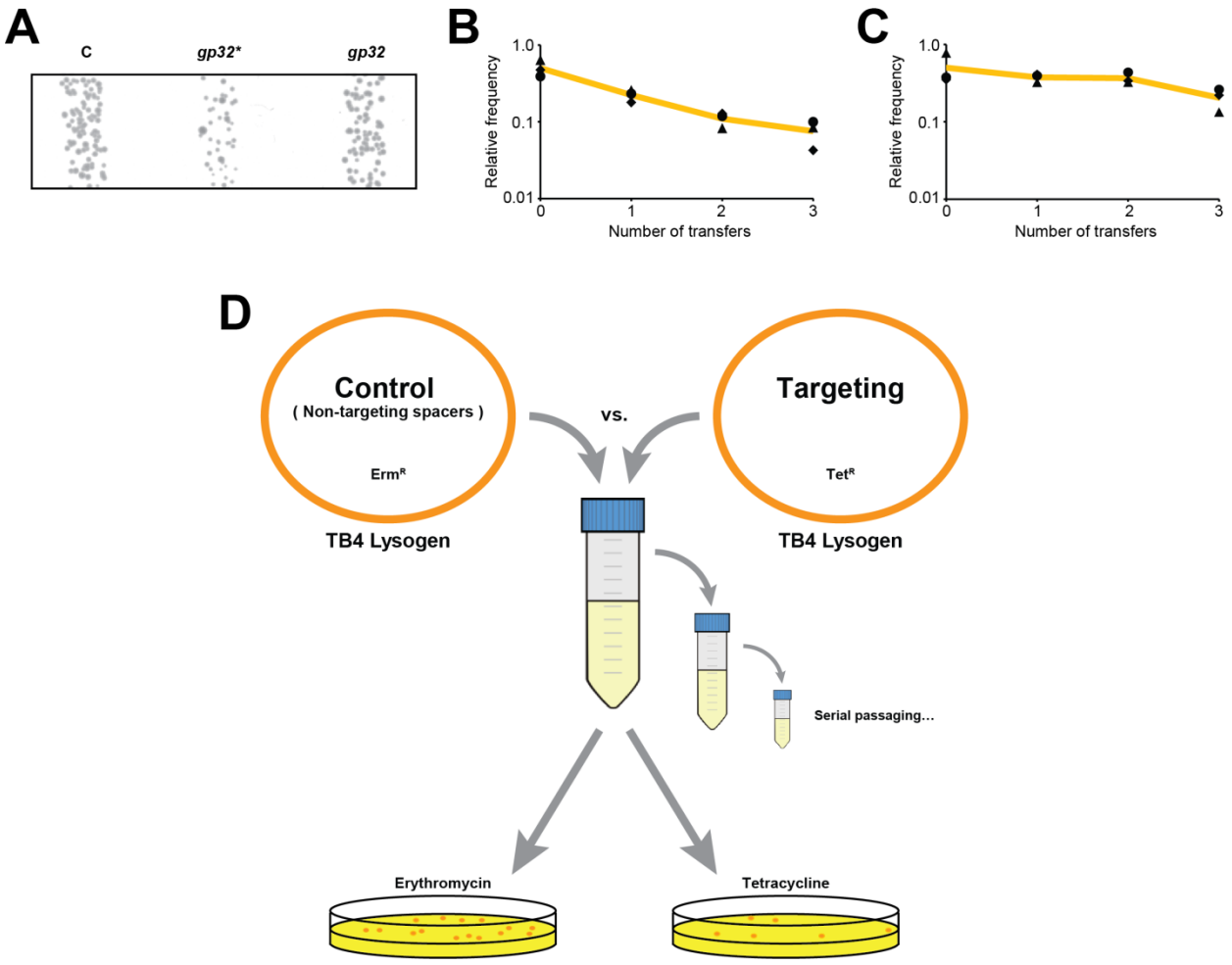
The results described above indicate that targeting of helper prophages by conditionally tolerant type III-A CRISPR-Cas systems can impede the transfer of SaPI particles from lysogens via prophage induction. Similar outcomes would be expected during lytic infections by heteroimmune helper phages bearing a target for the type III-A system, although this was not tested. By preventing the transfer of genes carried directly on the SaPI genome, and perhaps also by limiting the transfer of host genes via island-mediated (or by extension, phage-mediated) generalized transduction (Chen et al., 2015b), phage targeting by type III-A systems could be detrimental in natural populations of *S. aureus* that thrive off such forms of HGT. In turn, the loss of type III-A CRISPR-Cas systems from *S. aureus* genomes might help to ensure that such populations remain adaptive in rapidly changing environments, and/or that they can readily resist Muller's ratchet mechanism (Muller, 1932, 1964) over the long term. However, the possibility that maintenance of conditionally tolerant type III-A systems in lysogenic populations could impact host fitness more directly has not been ruled out. In the remainder of this chapter, I provide evidence that conditionally tolerant type III-A systems can impose constitutive fitness costs on lysogenic hosts which lack SaPIs altogether.

3.2 Detection of incomplete prophage tolerance by type III-A CRISPR-Cas systems in lysogenic hosts

In **Chapter 2**, the CRISPR-Cas system of *S. epidermidis* RP62a was re-engineered to contain spacers targeting the temperate phages of *S. aureus* Newman, and introduced into heterologous RN4220 (Nair et al., 2011) or TB4 hosts (Bae et al., 2006) on pC194-based plasmids (Horinouchi and Weisblum, 1982a). The spacers were all introduced into the CRISPR loci of my original parent vectors (pGG3 or pGG3-BsaI), and many of them were found to license conditional tolerance. However, a small-colony phenotype was clearly apparent when the spacer 2T CRISPR-Cas plasmid was introduced into RN4220::ϕNM1 lysogens. As explained in section **2.4**, this was assumed to result from slightly elevated rightward transcription across the prophage's *xis* target ORF, relative to other target regions (**Figure 2-17**). In other words, this result seemed to suggest that leaky prophage transcription could license an incomplete tolerance phenotype. Once I began working with pGG79-derived CRISPR-Cas plasmids, and also primarily TB4 lysogens, it became clear that this incomplete tolerance phenotype was more common than previously expected. For example, a small-colony phenotype was clearly apparent when the pGG79-derived *gp32** CRISPR-Cas plasmid was introduced into TB4::ϕNM1 lysogens (**Figure 3-3A**). A colony size reduction, albeit less pronounced, was also observed with the mismatched *gp32* variant of this spacer. Hypothesizing that these phenotypes reflect a general growth rate reduction relative to non-targeting lysogens, I

Figure 3-3. Maintenance of conditionally tolerant type III-A CRISPR-Cas systems in lysogenic hosts can incur fitness costs.

(A) Colony size comparison of Tet^R-marked TB4::ϕNM1 single lysogens harboring CRISPR-Cas plasmids containing either non-targeting spacers alone (C), the perfectly matched *gp32* spacer (*gp32**), or the originally identified *gp32* spacer with 5 mismatches (*gp32*). Picture is representative of a biological replicate for each strain plated in the presence of chloramphenicol to select for the CRISPR-Cas plasmids. **(B)** Pairwise competition experiments with the *gp32** spacer that licenses conditional tolerance of ϕNM1. An Erm^R-marked control lysogen harboring the parent vector with non-targeting spacers was competed against the Tet^R-marked conditionally tolerant lysogen. Relative frequencies (y-axis) are plotted against the number of transfers (x-axis), with one transfer per day. Individual values from each biological replicate are depicted in black as a triangle, circle, or rhombus. Solid lines represent the average change in relative frequency across the three replicates, and share color coding with the target ORF indicated in **Figure 3-4A**. **(C)** Same as in 'B', except that the Tet^R-marked conditionally tolerant lysogen used for competitions harbored the *gp32* (partially mismatched) spacer. **(D)** Schematic summary of pairwise competition assays used in this work to assess the relative fitness of antibiotic resistance-marked TB4 lysogens. Lysogens harboring CRISPR-Cas plasmids with either non-targeting or targeting spacers were mixed ~1:1 and passaged daily, with selective plating at each interval to calculate relative abundances as the ratio of Tet^R CFUs over total (Tet^R + Erm^R) CFUs. Chloramphenicol was also maintained in all media and plates to ensure selection for CRISPR-Cas plasmids. Falcon tube graphic adapted from an online source: https://clipartfest.com/download/c8aa2089bcf055d864c0317e14ff369ded0b3d10

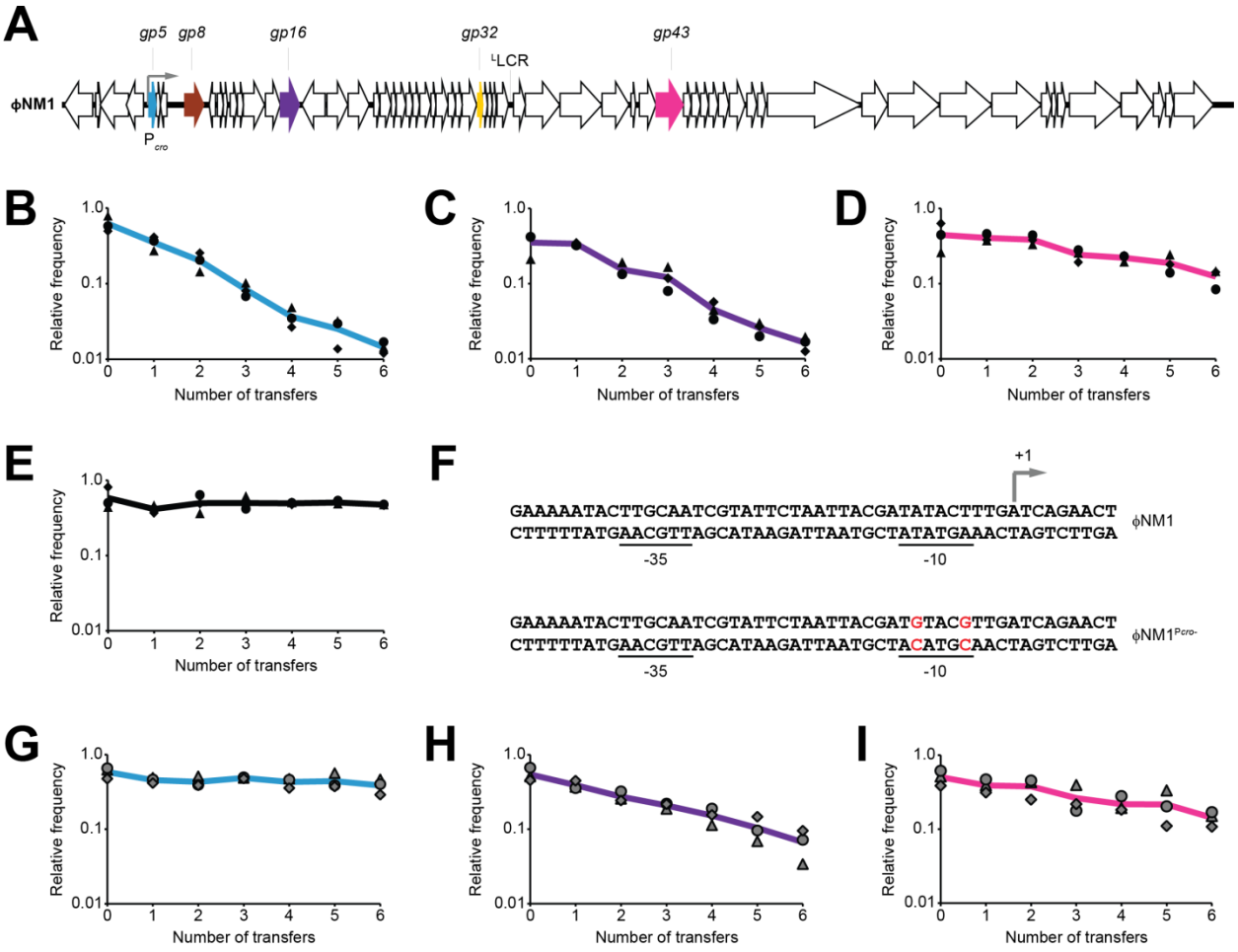

A
C
gp32*
gp32
B
Relative frequency
1.0
0.1
0.01
0
1
2
3
Number of transfers
C
Relative frequency
1.0
0.1
0.01
0
1
2
3
Number of transfers
D
Control
(Non-targeting spacers)
Erm^R
TB4 Lysogen
vs.
Targeting
Tet^R
TB4 Lysogen
Serial passaging...
Erythromycin
Tetracycline

sought to quantify them more precisely in terms of a fitness cost. This was accomplished by marking the TB4::ϕNM1 lysogens with small plasmids (Horinouchi and Weisblum, 1982b; Khan and Novick, 1983) conferring resistance to either tetracycline (Tet) or erythromycin (Erm), and measuring their relative abundances over time in pair-wise competition co-cultures (**Figure 3-3D)**. As anticipated, the relative abundance of lysogens with the perfectly matching *gp32** spacer declined more rapidly than lysogens with the mismatched spacer, when competed against control lysogens containing the parent vector with non-targeting spacers (**Figures 3-3B & C**). Fitness costs were quantified in terms of a selection coefficient, *s*, based on instruction from Robert Heler on how to calculate these values for a set of competition data points, using formulas in Microsoft Excel. **Table 3-1** lists the calculated values for all competition experiments performed in this work. To probe the generality of spacer-dependent fitness costs, I tested four additional spacers with perfect identity to sequences within *gp5*, *gp8*, *gp16*, and *gp43* of ϕNM1 (**Figure 3-4A**), and assayed them for relative fitness using the same setup. In three cases, a decline in relative abundance was clearly apparent (**Figures 3-4B, C, & D**), whereas no change in relative abundance was found in control co-cultures where competing lysogens harbored the same non-targeting parent vector (**Figure 3-4E**).

Previous studies have demonstrated that chromosomal targeting by CRISPR-Cas systems in bacteria is severely detrimental to growth and potentially lethal

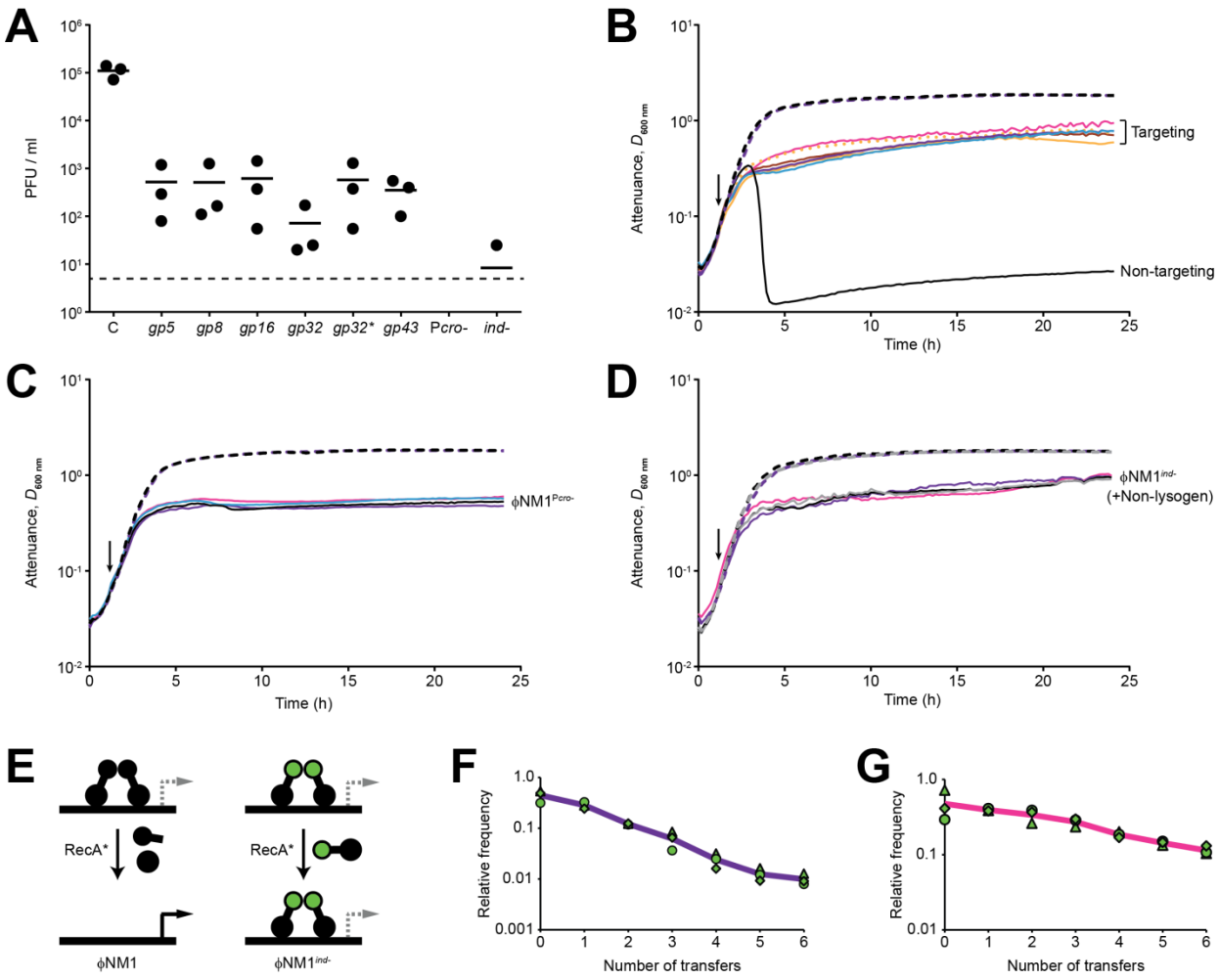
Figure 3-4. Incomplete prophage tolerance by type III-A systems can be licensed by prophage-internal promoter activity.

(A) Schematic representation of the integrated ϕNM1 prophage genome, with ORFs (thick arrows) scaled by length according to the annotation (NC_008583.1). The *gp5*, *gp8*, *gp16*, *gp32*, and *gp43* ORFs are colored to denote inclusion of a target sequence for one or more spacers tested in this work. Grey bent arrow indicates the position of ϕNM1's early lytic promoter (P_{cro}). ^{L}LCR denotes the location of an intergenic region in ϕNM1 with homology to regulatory sequences reported to control late gene expression for related phages (Ferrer et al., 2011; Quiles-Puchalt et al., 2013). **(B)** Pairwise competition experiments for testing the perfectly matching *gp5* spacer in marked single lysogens of TB4::ϕNM1, performed as described in **Figure 3-3**. **(C)** Same as in 'B', except the perfectly matching *gp16* spacer was tested. **(D)** Same as in 'B', except the perfectly matching *gp43* spacer was tested. **(E)** Control co-culture competition experiment; same as in 'B', except that both Erm^{R}- and Tet^{R}-marked TB4::ϕNM1 lysogens harbored the non-targeting pGG79 parent vector. **(F)** Sequences of the P_{cro} promoter in wild type ϕNM1 or its inactive variant in ϕNM1$^{P_{cro}-}$, with point mutations (red lettering) in the -10 element. The location of P_{cro}, along with the approximate position of its transcriptional start site (+1), was inferred from RNA-sequencing analysis of ϕNM1 during lytic infections (Goldberg et al., 2014). This designation is further supported by research on *S. aureus* temperate phages with related lambdoid regulatory architecture (Ganguly et al., 2009; Iandolo et al., 2002; Quiles-Puchalt et al., 2013). **(G)** Pairwise competition experiments for testing the perfectly matching *gp5* spacer in marked single lysogens of TB4::ϕNM1$^{P_{cro}-}$, performed as described in **Figure 3-3**. Individual values from each biological replicate are depicted in grey with black outlines as a triangle, circle, or rhombus. **(H)** Same as in 'G', except the perfectly matching *gp16* spacer was tested. **(I)** Same as in 'G', except the perfectly matching *gp43* spacer was tested.


A
gp5
gp8
gp16
gp32
gp43
LCR
ϕNM1
P cro
B
C
D
E
F
G
H
I
Relative frequency
Number of transfers
1.0
0.1
0.01
0 1 2 3 4 5 6
+1
GAAAAATACTTGCAATCGTATTCTAATTACGATATACTTTGATCAGAACT
CTTTTTATGAACGTTAGCATAAGATTAATGCTATATGAAACTAGTCTTGA
ϕNM1
-35
-10
GAAAAATACTTGCAATCGTATTCTAATTACGATGTACGTTGATCAGAACT
CTTTTTATGAACGTTAGCATAAGATTAATGCTACATGCAACTAGTCTTGA
ϕNM1Pcro-
-35
-10

(Jiang et al., 2013b; Vercoe et al., 2013), even in cases where prophage loci are targeted (Cady et al., 2012; Edgar and Qimron, 2010; Goldberg et al., 2014). In light of these findings, I hypothesized further that the spacer-dependent fitness costs result from growth defects associated with incomplete prophage tolerance, licensed by leaky transcription across targets from one or more promoter within the prophage. To explore this possibility, I introduced inactivating mutations into the early lytic promoter upstream of ϕNM1's *cro*-like regulator, encoded by *gp5* (**Figure 3-4F**). Mutation of this promoter (ϕNM1$^{P_{cro-}}$) virtually abolished the fitness cost associated with the *gp5* spacer (**Figure 3-4G**). Interestingly, the cost associated with the *gp16* spacer was only partially reduced (**Figure 3-4H**), suggesting that downstream promoters might license different growth defects in the absence of a dominant promoter upstream. In support of this notion, I observed a clear discrepancy in the magnitude of the costs associated with the *gp5* or *gp16* spacers compared to that of the *gp43* spacer in wild type lysogens (**Figures 3-4B, C, & D**), despite their similar capacities to disrupt phage lytic propagation via spontaneous or MMC-stimulated prophage induction (**Figures 3-5A & B**). The *gp43* target, unlike the *gp5* and *gp16* targets, is located within a gene cluster that is transcribed late during the lytic cycle of ϕNM1 (Goldberg et al., 2014). In staphylococcal phages with related genomic architecture, this cluster lies downstream of a regulatory region that was shown to contain a conserved transcriptional terminator (Ferrer et al., 2011), and I identified such a region (^{L}LCR) in ϕNM1 between *gp36* and *gp37* (**Figure 3-**

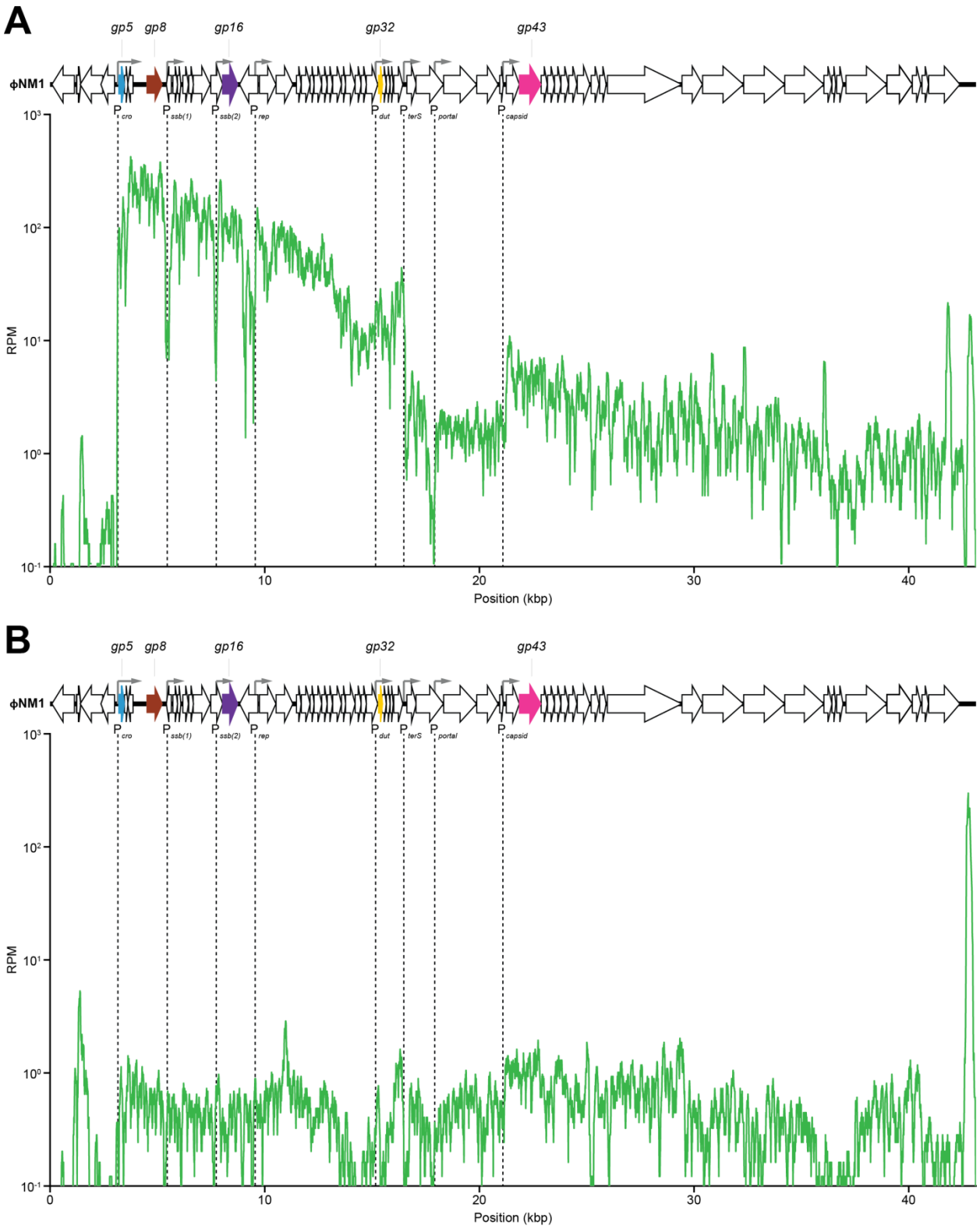
Figure 3-5. Prophage inducibility is not a prerequisite for incomplete tolerance by type III-A systems.
(A) Quantification of PFU concentrations (PFU / ml) in filtered supernatants from overnight cultures of Tet^R-marked TB4::ϕNM1 single lysogens harboring CRISPR-Cas plasmids with one of several targeting spacers, or the control parent vector with non-targeting spacers (C). Horizontal bars indicate the mean value of three biological replicates, plotted as black dots in each column. Tet^R-marked lysogens harboring the control parent vector and mutant prophages ϕNM1^{Pcro-} or ϕNM1^{ind-} were also tested, and plaques were only detected in one of the replicates for ϕNM1^{ind-}. Dotted line represents the limit of detection under these assay conditions. **(B)** High-resolution growth curves of Tet^R-marked TB4::ϕNM1 single lysogens harboring the CRISPR-Cas plasmids tested in panel 'A'. Solid lines represent the average attenuance ($D_{600\ nm}$) measured from three biological replicates with MMC added at the indicated time point (black arrow), and are colored to match the target ORFs highlighted in **Figure 3-4A** or black for the control parent vector. A dotted line is used for the *gp32* (mismatched) spacer, but note that individual dots do not correspond to measurements taken every 10 minutes. For comparison, average attenuance measured from three biological replicates in the absence of MMC is plotted for two strains (dashed lines). **(C)** Same as in 'B', except using Tet^R-marked TB4::ϕNM1^{Pcro-} lysogens with select CRISPR-Cas plasmids. **(D)** Same as in 'B', except using Tet^R-marked TB4::ϕNM1^{ind-} lysogens with select CRISPR-Cas plasmids; a Tet^R-marked TB4 non-lysogen harboring the pGG79 parent vector was also tested (grey lines). **(E)** Schematic diagram illustrating different outcomes expected for the wild type ϕNM1 prophage (left) or the ϕNM1^{ind-} prophage mutant (right) in response to a canonical SOS-inducing signal. Monomers of CI-like repressor are depicted as black bar bells, with green spots for the ϕNM1^{ind-} mutant indicating the presence of a S124A substitution in the C-terminal serine protease domain. When intact repressors are present, only leaky transcription originates from the regulated promoter (dashed grey bent arrows). Under inducing conditions, autoproteolysis of wild type repressor is stimulated by the presence of activated RecA (RecA*) and allows transcription required for the lytic cycle (solid black bent arrow), while ϕNM1^{ind-} mutant repressor remains intact and continues to allow only leaky transcription. **(F)** Pairwise competition experiments for testing the perfectly matching *gp16* spacer in marked single lysogens of TB4::ϕNM1^{ind-}, performed as described in **Figure 3-3**. **(G)** Same as in 'F', except the perfectly matching *gp43* spacer was tested.


A
PFU / ml
C
gp5
gp8
gp16
gp32
gp32*
gp43
Pcro-
ind-
B
Attenuance, $D_{600\,nm}$
Targeting
Non-targeting
Time (h)
C
ϕNM1^{Pcro-}
D
ϕNM1^{ind-}
(+Non-lysogen)
E
RecA*
ϕNM1
ϕNM1^{ind-}
F
Relative frequency
Number of transfers
G

4A). Conceivably, the magnitude of the *gp43* spacer's fitness cost is determined by one or more late promoters that are insulated from upstream transcription during the non-inducing conditions of lysogeny, perhaps in part due to this regulatory region. In order to better assess the plausibility of this surmise, I used an online promoter prediction tool to identify putative promoters upstream of my targets (**Table 3-2**). Candidates were selected from among the initial hits by cross-referencing with the position of relative transcriptional peaks identified along the traces of non-averaged plots from a stranded RNA-seq dataset collected previously (Goldberg et al., 2014). Particularly when considering the relative peaks observed 6 minutes post-infection with ϕNM1 (**Figure 3-6A**), it is apparent that there are multiple regions that display a dip in coverage followed by a peak, which could thus indicate the presence of a promoter. However, although this stranded RNA-seq dataset distinguished between the directionality of transcription, it did not distinguish transcriptional start sites from processed or cleaved RNA ends (which could also manifest as a local dip in coverage). Additional analyses would therefore be required to confirm the presence of bona-fide promoters at these sites. Note that the magnitudes of rightward transcription observed in the RN4220::ϕNM1 lysogen dataset (**Figure 3-6B**) should be interpreted with caution, because rightward transcription may be derived from a subset of the cells in the population undergoing spontaneous prophage induction, in addition to the stably lysogenic cells. Note, also, that both of these expression profiles were measured with

Figure 3-6. Location of ϕNM1 promoter candidates inferred from promoter predictions and stranded RNA-seq profiles.

(A) Schematic representation of the ϕNM1 genome as presented in **Figure 3-4A**, aligned with ϕNM1's rightward transcription profile (solid green line), obtained six minutes after infection of sensitive RN4220/pGG3 cells. Values are plotted in reads per million (RPM), but were not averaged over a 500 bp sliding-window as displayed previously (Goldberg et al., 2014). Vertical dashed lines below grey bent arrows represent the position of candidate promoters as indicated. Their sequences are listed in **Table 3-2**. **(B)** Same as in 'A', except that data was collected from a logarithmically growing RN4220::ϕNM1 lysogen.


A
gp5
gp8
gp16
gp32
gp43
ϕNM1
P cro
P ssb(1)
P ssb(2)
P rep
P dut
P terS
P portal
P capsid
RPM
Position (kbp)
B
gp5
gp8
gp16
gp32
gp43
ϕNM1
P cro
P ssb(1)
P ssb(2)
P rep
P dut
P terS
P portal
P capsid
RPM
Position (kbp)

the RN4220 host background. Nevertheless, I speculate that leaky transcription from at least some of these candidates—in addition to P_{cro}—could license incomplete prophage tolerance and a concomitant fitness reduction for the lysogen. Consistent with this notion, the ϕNM1^{Pcro-} mutation had no discernible effect on the cost associated with the *gp43* spacer (**Figure 3-4I**). Furthermore, this mutation was found to completely abolish the lysogen's capacity for detectable prophage induction in two assays (**Figures 3-5A & C**). This suggests that the costs I've observed do not require transcription derived from lytic propagation. In further support of this idea, fitness costs associated with the *gp32** spacer in the ϕNM1^{Pcro-} background were comparable to that observed in the wild type background prior to day 3, and in fact somewhat elevated (**Table 3-1**). However, a fitness equilibrium plateau was not reached at this point when P_{cro} was mutated in the ϕNM1^{Pcro-} background (**Figure 3-7**). Evidently, the pronounced curing of the wild type ϕNM1 prophage detectable by day 3 with the *gp32** spacer (data not shown) requires that the prophage's lytic capabilities are intact, although the fitness costs *per se* do not. In fact, by eliminating the target sequence, prophage curing apparently has the effect of rapidly extinguishing the fitness cost associated with maintenance of this CRISPR-Cas system. Meanwhile, prophage curing was not observed with the *gp5*, *gp16*, and *gp43* spacers in wild type ϕNM1 lysogens, or the *gp32** spacer in ϕNM1^{Pcro-} lysogens, where genetic escape via CRISPR-Cas plasmid inactivation was instead apparent by day 6 (data not shown). The reason for this discrepancy between the *gp32** spacer and the

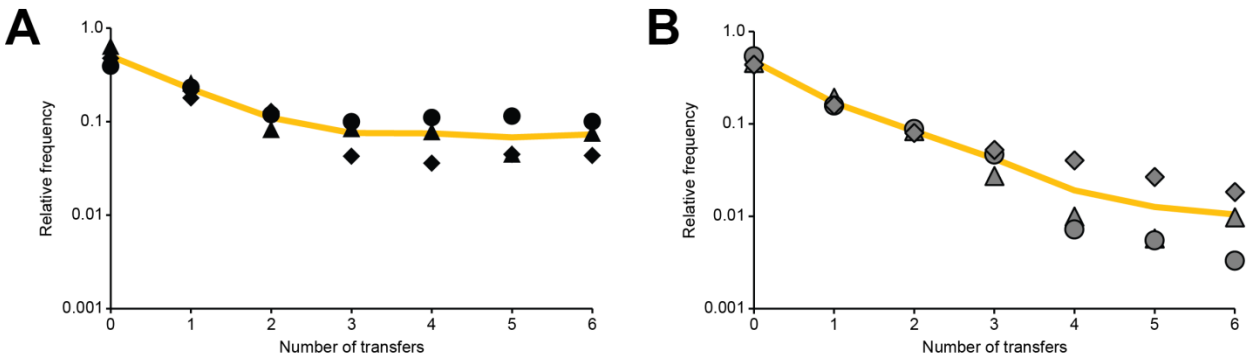


Figure 3-7. An intact ϕNM1 prophage is required for high-frequency genetic escape of *gp32-dependent fitness costs observed in TB4::ϕNM1 lysogens.**

(A) Pairwise competition experiments for testing the perfectly matching *gp32** spacer in marked single lysogens of TB4::ϕNM1, performed as described in **Figure 3-3**. **(B)** Same as in 'B', except that the perfectly matching *gp32** spacer was tested in marked single lysogens of TB4::ϕNM1^{Pcro-}, and individual values from each biological replicate are depicted in grey with black outlines as a triangle, circle, or rhombus.

other perfectly matching spacers in wild type lysogens is still unclear to me. I speculate, however, that the unique position of its target within the prophage genome, and the dispensability of the *gp32* ORF (based on homology to ϕNM2 and ϕNM4; see **Chapter 2**) for lytic infections, might be responsible. In temperate phages of *Pseudomonas aeruginosa*, accessory genes have been identified at an analogous region upstream of late-transcribed phage structural genes (Bondy-Denomy et al., 2013; Pawluk et al., 2014), and they are thought carry their own promoters (Pawluk et al., 2014). It is possible that this region in ϕNM1 also encodes accessory genes with their own promoters, and that the regulation of these promoters is relatively uncoordinated with

the phage's other promoters. In turn, such unregulated transcription might allow for bursts of targeting at opportune times during prophage induction, and targeting of this region would not necessarily disrupt the early lytic functions required for curing (because they are located upstream). It is also interesting to note that the fitness plateau was reached within a mixed population where non-targeting (Erm^R-marked) lysogens release normal levels of infective phage particles. CRISPR-Cas immunity to lytic infection licensed by the *gp32** spacer therefore appears to be sufficient to maintain the fitness of Tet^R-marked non-lysogens alongside control lysogens, at least under the growth conditions tested.

In order to further demonstrate that these spacer-dependent fitness costs arise independently of prophage induction, I disrupted ϕNM1's SOS-inducibility without altering its promoters by introducing a serine to alanine missense mutation in the catalytic domain of its *cI*-like repressor (**Figures 3-5D & E**). The CI-like repressors of lambdoid phages possess a serine protease activity that is required for prophage induction via RecA*-stimulated autoproteolysis and de-repression (Erill et al., 2007; Livny and Friedman, 2004; Mo et al., 2014). Spontaneous prophage induction was almost undetectable for this ϕNM1^{ind-} mutant as well (**Figure 3-5A**). A previous report with phage lambda showed that when the primary *recA*-dependent pathway of prophage induction is prevented, rare spontaneously induced plaques can be found with compensatory mutations in either *cI* or its operator region that destabilize the

repression circuit (Little and Michalowski, 2010). I therefore used PCR and Sanger sequencing to check this region in the five plaques I detected, and confirmed that the plaques were indeed created by mutants of ϕNM1^{ind-} that harbor additional SNPs (**Figure 3-8**). I next tested the marked TB4::ϕNM1^{ind-} lysogens in pairwise competition assays. For reasons unclear, I was not able to isolate stable transductants of TB4::ϕNM1^{ind-} harboring the *gp5* CRISPR-Cas plasmid. I speculate that, in the absence of RecA*-stimulated repressor cleavage, the stability of the repression circuit in ϕNM1^{ind-} may be more prone to fluctuations that increase transcription of its *gp5* (*cro*-like) ORF. In lambda, CI represses its own promoter at higher concentrations in a negative feedback loop (Ptashne, 2011). Therefore, if CI-like repressors accumulate to high concentrations more readily in the ϕNM1^{ind-} background, they may in turn repress their own promoter more frequently than in wild type (and thus allow for more frequent bursts of *gp5* transcription). Notwithstanding, when the *gp16* and *gp43* spacers were tested with this prophage mutant in competition assays, no reduction in their fitness costs was observed (**Figures 3-5F & G**). Taken together, these results demonstrate that conditionally tolerant type III-A systems can impose spacer-dependent fitness costs even if their target prophages are not inducible, and thus suggest that the costs arise from transcription of prophage targets within otherwise stably lysogenic cells.

Supporting this view, I did not find the relative fitness of conditionally tolerant lysogens at steady state to be correlated with their ability to survive brief MMC

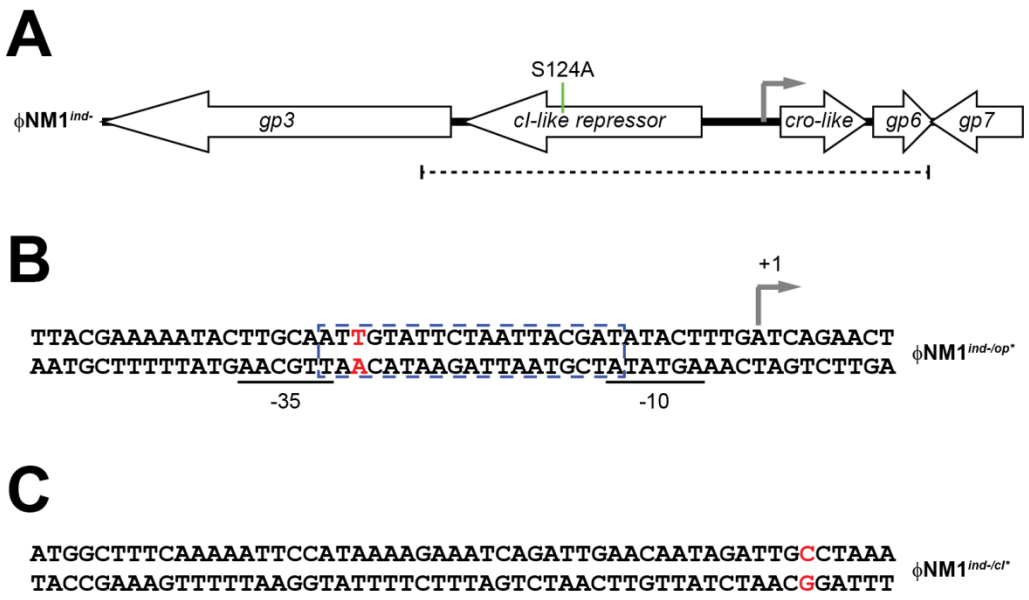


Figure 3-8. Characterization of five ϕNM1^{ind-} spontaneous induction escapers.
(A) Schematic diagram depicting a relevant genomic region of the ϕNM1^{ind-} prophage delimited by *gp3* and *gp7*, according to the ϕNM1 annotation (Accession: NC_008583.1). Green vertical line denotes location of the engineered missense mutation in ϕNM1^{ind-} that causes a Serine→Alanine substitution at residue 124 of its CI-like repressor gene product. Dashed line denotes the location of Sanger-sequenced PCR products (1355 bp) amplified from the five spontaneously induced PFUs using primers oGG31 and oGG33. Grey bent arrow indicates the position of the inferred P_{cro} transcriptional start site. **(B)** Sequence of the P_{cro} promoter region in ϕNM1$^{ind-/op*}$. Whereas the putative P_{cro} -35 and -10 elements (underlined) are intact, a point mutation (red lettering) was found within a putative operator binding site that partially overlaps these elements (dashed blue box). This point mutation appears to diminish the palindromy of this sequence (consult **Table 3-2** for the wild type sequence), and was found in 2/5 plaques. **(C)** Sequence of the first 18 codons in the mutant *cI*-like repressor gene of ϕNM1$^{ind-/cI*}$. The T→C missense mutation (red lettering) within codon 17 is expected to produce a Serine→Proline substitution. This mutation was found in 3/5 plaques.

treatments at a subinhibitory concentration (**Figure 3-9A**). For example, although fitness costs were undetectable with conditionally tolerant lysogens harboring the *gp8* spacer (**Figure 3-9B**), they survived MMC treatment worse than conditionally tolerant lysogens with the *gp43* spacer for which a fitness cost was observed (**Figure 3-4D**). Given that completion of the lytic cycle is not required for a decline in viability associated with lambdoid prophage induction (Livny and Friedman, 2004), and that CRISPR-Cas targeting of induced prophages could potentially occur prior to excision from the

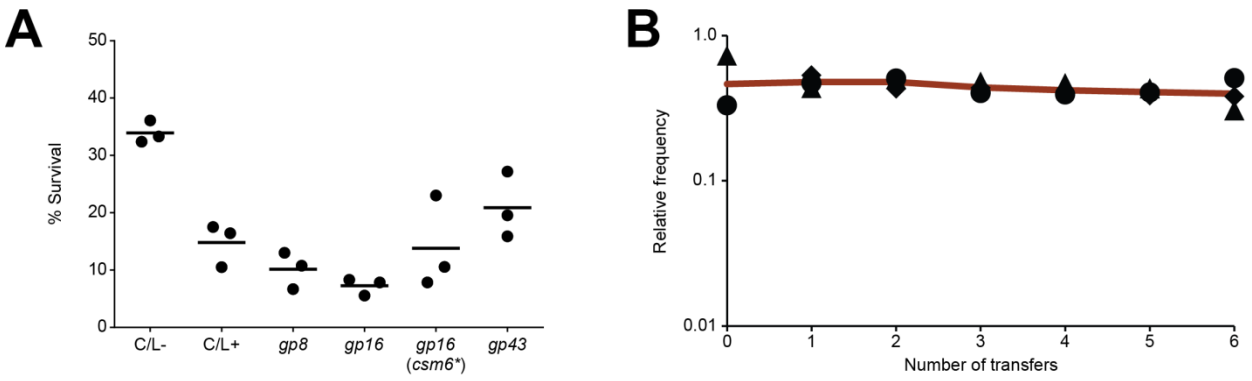


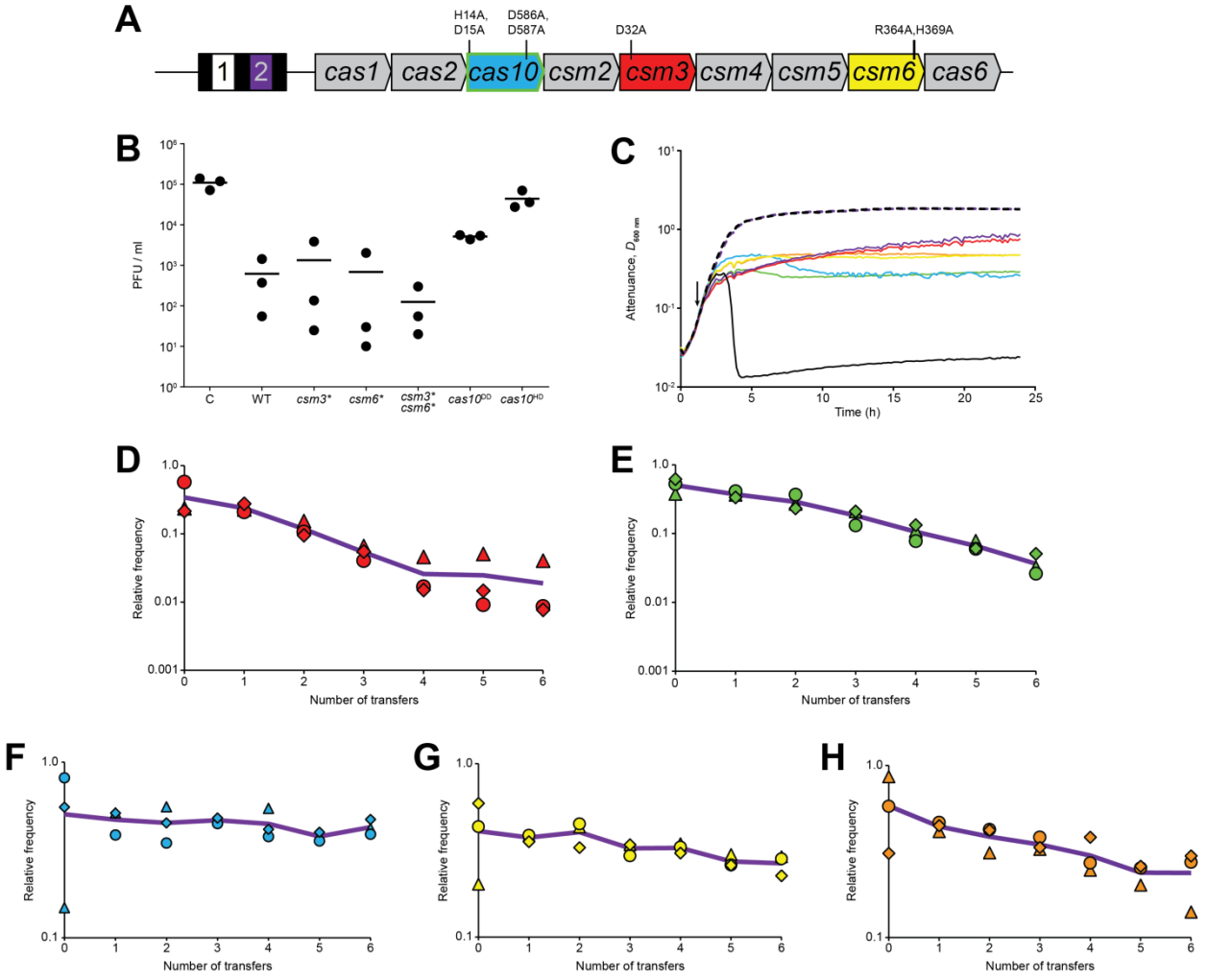
Figure 3-9. Survival of conditionally tolerant lysogens following MMC-stimulated prophage induction is not a predictor of relative fitness in untreated cultures.
(A) Percent survival of Tet^R-marked TB4::ϕNM1 lysogens harboring the indicated conditionally tolerant CRISPR-Cas plasmid or the pGG79 parent vector (C/L+), following a 45-minute treatment with MMC at a subinhibitory concentration (0.5 μg/ml). Survival of a Tet^R-marked TB4 non-lysogen harboring the pGG79 parent vector is also shown (C/L-). **(B)** Pairwise competition experiments for testing the perfectly matching *gp8* spacer in marked single lysogens of TB4::ϕNM1, performed as described in **Figure 3-3**.

chromosome, it is expected that prophage induction might still have a lethal outcome even when targeting spacers are present.

3.3 Contributions of type III-A targeting nucleases

Having demonstrated that lysogenic hosts which maintain type III-A systems can incur spacer-dependent fitness costs, I sought to clarify whether this phenomenon reflects CRISPR-Cas targeting. To this end, I tested my *gp16* spacer with mutations in Cas nuclease active sites previously implicated in targeting by the *S. epidermidis* type III-A system (**Figure 3-10A**). Mutation of Csm3 (D32A) was previously shown to abolish the targeting complex's capacity for specific cleavage of complementary RNAs, while mutation of Cas10's palm polymerase domain (D586A,D587A) was shown to abolish the complex's capacity for transcription-dependent cleavage of DNA on the non-template strand (Samai et al., 2015). Furthermore, mutations in the HEPN domain of Csm6 (R364A,H369A) were shown to abolish its RNase activity and eliminate its contributions to phage defense (Jiang et al., 2016). I first assessed the targeting capacity of mutant CRISPR-Cas plasmids during prophage induction. When *csm3*, *csm6*, or both RNases were mutated, spontaneous release of plaque forming units (PFUs) from overnight cultures was reduced to a similar degree as wild type, with individual replicates exhibiting as much as 3.2- to 4-log reductions relative to the average measured for lysogens harboring the non-targeting parent vector (**Figure 3-10B**). In

Figure 3-10. Effect of targeting nuclease active site mutations on conditional tolerance and spacer-dependent fitness costs associated with type III-A systems.
(A) Schematic diagram summarizing the arrangement of CRISPR-Cas loci used throughout the work, with position 2 of the CRISPR array occupied by the gp16 spacer (purple rectangle) in this example. The naturally occurring *nes* spacer (white rectangle) from *S. epidermidis* RP62a is maintained at position 1 in all cases. ORFs encoding Cas nucleases implicated in targeting are highlighted in blue with green outline (*cas10*), red (*csm3*), or yellow (*csm6*), and the positions of their active site mutations are labeled to scale within each ORF. Other features of the diagram are not drawn to scale. **(B)** Quantification of PFU concentrations (PFU / ml) in filtered supernatants from overnight cultures of Tet^R-marked TB4::ϕNM1 single lysogens harboring mutant or wild type CRISPR-Cas plasmids with the *gp16* spacer, or the wild type pGG79 parent vector with non-targeting spacers (C). Horizontal bars indicate the mean value of three biological replicates, plotted as black dots in each column. **(C)** High-resolution growth curves of Tet^R-marked TB4::ϕNM1 single lysogens harboring the mutant or wild type CRISPR-Cas plasmids tested in panel 'B'. Solid lines represent the average attenuance ($D_{600\ nm}$) measured from three biological replicates with MMC added at the indicated time point (black arrow), and are colored to match the *cas* gene ORFs highlighted in panel 'A' for each mutant, or otherwise green for the $cas10^{HD}$ mutant, blue for the $cas10^{DD}$ palm domain mutant, orange for the *csm3**/*csm6** double mutant plasmid, purple for the wild type *gp16* plasmid, or black for the non-targeting pGG79 plasmid. For comparison, average attenuance measured from three biological replicates in the absence of MMC is plotted for two strains (dashed lines). **(D)** Pairwise competition experiments with marked TB4::ϕNM1 single lysogens, performed as described in **Figure 3-3**, except that the targeting strain harbored the *gp16*(*csm3**) mutant CRISPR-Cas plasmid. **(E)** Same as in 'D', except with the *gp16*($cas10^{HD}$) mutant plasmid. **(F)** Same as in 'D', except with the *gp16*($cas10^{DD}$) palm domain mutant plasmid. **(G)** Same as in 'D', except with the *gp16*(*csm6**) mutant plasmid. **(H)** Same as in 'D', except with the *gp16*(*csm3**/*csm6**) double mutant plasmid.


A
H14A, D15A
D586A, D587A
D32A
R364A,H369A
1
2
cas1
cas2
cas10
csm2
csm3
csm4
csm5
csm6
cas6
B
PFU / ml
C
WT
csm3*
csm6*
csm3* csm6*
cas10DD
cas10HD
C
Attenuance, D600 nm
Time (h)
D
Relative frequency
Number of transfers
E
Relative frequency
Number of transfers
F
Relative frequency
Number of transfers
G
Relative frequency
Number of transfers
H
Relative frequency
Number of transfers

contrast, spontaneously induced PFU were never reduced more than ~1.3 logs among replicates with the *cas10*^{DD} mutant, indicating that its targeting is partially impaired. Similar results were obtained in high-resolution growth curves following treatment with MMC, where partial lysis could be detected with the *cas10*^{DD} mutant, but not with either or both of the RNAses mutated (**Figure 3-10C**). These results corroborate previous findings which demonstrated that both the Csm3 and Csm6 RNAse activities are dispensable for immunity during the lytic cycle when the target falls within an early-transcribed region of the phage genome such as this (Jiang et al., 2016).

Recent reports on type III systems from other groups have shown that the HD domain of Cas10 proteins is required for RNA-activated cleavage of ssDNA *in vitro* (Elmore et al., 2016; Estrella et al., 2016; Han et al., 2017; Kazlauskiene et al., 2016; Park et al., 2017), although a requirement for cleavage of the non-template DNA strand in a transcription-coupled assay has yet to be demonstrated. Moreover, while mutation of the HD domain was not found to abolish anti-plasmid immunity by the type III-A system of *S. epidermidis* RP62a in a pG0400 conjugation assay (Hatoum-Aslan et al., 2014), the effect of this mutation on type III-A targeting in lysogens has not been explored. I first tested the *cas10*^{HD} mutant in prophage induction assays. Interestingly, while this mutant still appeared to be largely capable of preventing lysis in high-resolution growth curves following MMC-stimulated prophage induction, almost no reduction in spontaneously released PFUs was observed in this background. The reason

for this discrepancy is currently unclear, but I speculate that the HD domain's DNase activity might be required to prevent phage DNA from being packaged efficiently, and yet not per se required to prevent completion of the lytic cycle. Additional insight could perhaps be obtained by quantifying the number of PFUs released (or produced intracellularly) at different time points after induction with MMC in this mutant background. For example, it's possible that a residual targeting activity—potentially derived from one or both of the type III-A system's RNases—could be delaying lysis without degrading phage DNA or sufficiently preventing its packaging. I further speculate that the residual targeting observed in high-resolution growth curves reflects the same activity or activities that allow for immunity to pG0400 conjugation via the native *nes* spacer in this background (Hatoum-Aslan et al., 2014), because the immunity to pG0400 licensed by this spacer was, incidentally, found to depend on *csm6* (Hatoum-Aslan et al., 2014). Testing of double or triple mutants with the HD domain and one or both RNase functions mutated would likely clarify this point.

I next measured the relative fitness of lysogens harboring mutant CRISPR-Cas plasmids. The fitness cost associated with the *gp16* spacer was not reduced by the *csm3** mutation (**Figure 3-10D**), and perhaps only slightly reduced by the *cas10*^{HD} mutation (**Figure 3-10E**). Meanwhile, the fitness cost appeared to be completely abolished by the *cas10*^{DD} palm domain mutation (**Figure 3-10F**). Interestingly, both the *csm6** and *csm3**/*csm6** mutants also displayed a strongly reduced fitness cost (**Figures 3-10G &**

H). This finding implies that the RNase activity of Csm6 can contribute to growth defects associated with chromosomal targeting, although the mechanism for this remains to be elucidated. An important caveat of these nuclease mutant data is that isogenic plasmid backbones were not used when measuring relative fitness in the competition experiments; as such, the observed reduction in *gp16*-associated fitness costs could in fact reflect reduced general toxicity associated with expression of the intact nucleases, rather than reduced targeting at prophage loci. To rule this out, mutant CRISPR-Cas plasmids with the *gp16* spacer should be tested against their respective non-targeting mutant parent vectors. Notwithstanding, my results suggest that costs associated with incomplete prophage tolerance by type III-A systems may depend on the activity of at least one Cas nuclease implicated in transcription-dependent targeting.

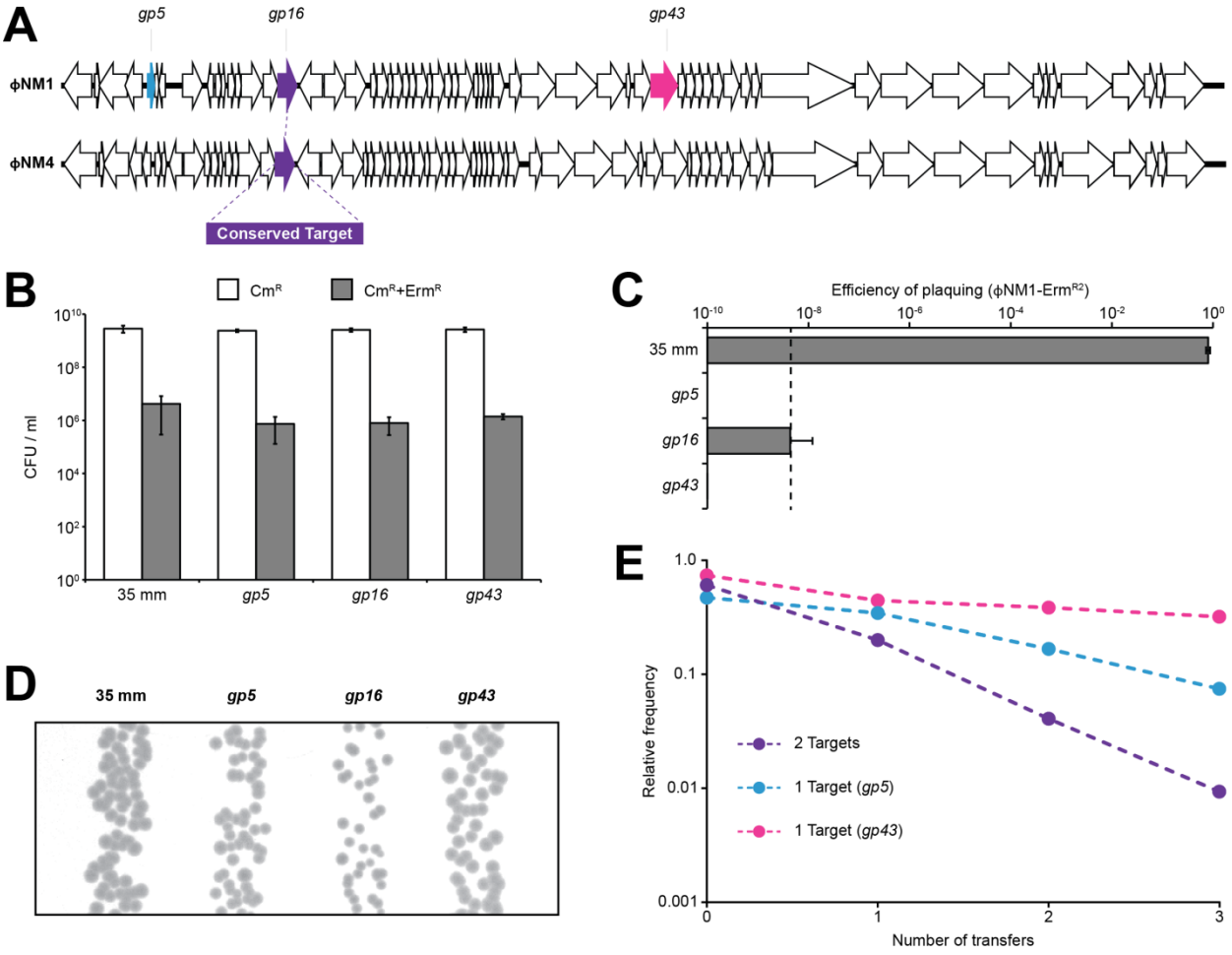
3.4 Consequences of conditionally tolerant type III-A systems in polylysogens

Clinical isolates of *S. aureus* are seldom found to possess CRISPR-Cas systems (Cao et al., 2016), and in many cases are polylysogens harboring more than one prophage (Aanensen et al., 2016; Goerke et al., 2009). The clinical isolate *S. aureus* Newman, for example, does not harbor an endogenous CRISPR-Cas system but harbors four heteroimmune lambdoid prophages (Bae et al., 2006). However, among the staphylococcal type III-A CRISPR-Cas systems which have been identified so far, some spacers have been found to match conserved temperate phage lytic genes, at least

partially (Cao et al., 2016; Goldberg et al., 2014). These observations prompted us to investigate the consequences of maintaining type III-A systems in a polylysogenic scenario where more than one prophage can be targeted. This was accomplished using ϕNM1 and ϕNM4, which integrate into distinct chromosomal loci but share regions of sequence homology, including a perfectly conserved target for the *gp16* spacer (**Figure 3-11A**). Provided that each target sequence falls within lytic gene clusters which are sufficiently repressed during lysogeny, I anticipated that polylysogeny would be tolerated even in the presence of perfectly matching spacers. To confirm this, TB4::ϕNM4 lysogens harboring CRISPR-Cas plasmids with spacers matching to ϕNM1's *gp5*, *gp16*, or *gp43* targets were infected with an erythromycin resistance-conferring derivative of ϕNM1 (ϕNM1-Erm^{R2}) that allows quantification of newly formed lysogens. Lysogens with singly (*gp5*, *gp43*) and doubly (*gp16*) targeting spacers tolerated secondary lysogenization by ϕNM1-Erm^{R2} at a comparable frequency as control lysogens harboring a fully mismatched (35 mm) spacer (**Figure 3-11B**), while plaque formation on the same hosts was reduced only in the presence of targeting spacers (**Figure 3-11C**). However, differences in colony size phenotype were clearly discernible among the resulting double lysogens (**Figure 3-11D**). Relative to double lysogens with the control spacer, double lysogens with the *gp5* or *gp43* spacers which do not target ϕNM4 had a less severe colony size reduction than lysogens with the *gp16* spacer targeting both prophages. This trend was likewise observed with unmarked

Figure 3-11. Type III-A systems which target multiple temperate phages can impose greater fitness costs in polylysogenic hosts.

(A) Schematic representation of the integrated ϕNM1 (NC_008583.1) and ϕNM4 (NC_028864.1) prophage genomes. ORFs (thick arrows) were scaled in accordance with their annotated lengths. Colored ORFs denote the presence of a target sequence for one or more spacer tested in section **3.4**. ϕNM1's *gp16* ORF includes a target that is found with 100% conservation in *gp17* of ϕNM4, as indicated by color-matched ORFs (purple). **(B)** Secondary lysogenization of TB4::ϕNM4 with ϕNM1-Erm^{R2} in the presence of the 35 mm, *gp5*, *gp16*, or *gp43* CRISPR-Cas plasmids as indicated. The '35 mm' spacer is a fully mismatched variant of the *gp16* spacer that serves as a non-targeting control. Lysogenization was quantified as the concentration (CFU / ml) of Cm^{R}+Erm^{R} CFUs in each culture following treatment with ϕNM1-Erm^{R2}. For comparison, total recipients were quantified as the concentration (CFU / ml) of Cm^{R} CFUs in each untreated culture. Error bars, mean ± s.d. (n = 3, biological replicates). **(C)** Plaquing efficiency of ϕNM1-Erm^{R2} on TB4::ϕNM4 lysogens harboring the different CRISPR-Cas plasmids tested in panel 'B', as indicated by the labels. Dotted line represents the limit of detection under these assay conditions. Error bars, mean ± s.d. (n = 3, technical replicates). **(D)** Colony size comparison of TB4::ϕNM4+ϕNM1-Erm^{R2} double lysogens harboring the different CRISPR-Cas plasmids tested in panel 'B', as indicated by the labels above. Picture is representative of a biological replicate for each strain plated in the presence of chloramphenicol to maintain CRISPR-Cas plasmids. **(E)** Batch competition experiment with unmarked TB4::ϕNM1+ϕNM4 double lysogens harboring the different CRISPR-Cas plasmids tested in panel 'B'. As outlined in the methods, deep sequencing was used to determine targeting spacer frequencies relative to the 35 mm non-targeting spacer within the sample (y-axis), and plotted against the number of transfers (x-axis, one transfer per day). Dashed lines illustrate the change in relative frequency across time points, and are color-coded for their targets according to the schematic in panel 'A'.


A
gp5
gp16
gp43
ϕNM1
ϕNM4
Conserved Target
B
Cm^R
Cm^R+Erm^R
CFU / ml
35 mm
gp5
gp16
gp43
C
Efficiency of plaquing (ϕNM1-Erm^R2)
35 mm
gp5
gp16
gp43
D
35 mm
gp5
gp16
gp43
E
Relative frequency
2 Targets
1 Target (gp5)
1 Target (gp43)
Number of transfers

TB4::ϕNM4+ϕNM1 double lysogens when the same CRISPR-Cas plasmids were introduced via transduction (data not shown). To determine whether these phenotypes reflect discrepancies in relative fitness, I performed pairwise competition assays with marked double lysogens harboring either the *gp5* spacer, the *gp16* spacer, or the *gp43*^{ori} spacer. A control co-culture experiment where differentially marked double lysogens both harbored the same pGG79 parent vector was also performed. Surprisingly, the Tet^{R}-marked double lysogens had an apparent advantage even in control co-cultures (data not shown). Furthermore, although the double lysogens with the *gp16* spacer indeed declined the most rapidly at the outset, all of the fitness costs were offset, and the relative abundance of Tet^{R}-marked lineages would even *increase* by the end of the six-day experiment in cases where sufficient genetic inactivation of the conditionally tolerant CRISPR-Cas plasmid had presumably occurred (data not shown). I speculate that the pE194 plasmid is less stable in this doubly lysogenic background, perhaps due to synergistic or additive effects of harboring both prophages. For example, harboring both prophages might provide additional opportunities for leaky expression of prophage terminase proteins that could license off-target cleavage of plasmids. In this scenario, specific loss of the pE194 plasmid rather than pT181 could be explained by intrinsic terminase sequence preferences (Chen et al., 2015b). Note, however, that neither of these prophages were sufficient to license this effect on their own (**Figures 3-4E & 3-12A**). In order to circumvent this technical constraint, and also to assess the

relative fitness of multiple strains in a less cumbersome manner, I mixed unmarked double lysogens harboring each of the four spacers (35 mm, *gp5*, *gp16*, or *gp43*) into a single batch and competed them in liquid culture for three days. After each day, and also at time 'zero' before the first transfer, spacer abundances were measured by deep sequencing, and the frequency of each targeting spacer was calculated relative to the 35 mm spacer's frequency in the sample. Deep sequencing in a single MiSeq run was performed on small PCR amplicons spanning the spacer of interest, alongside two additional series of three-day batch competitions described later in this chapter. This was enabled by independently barcoding each of the 12 PCR samples, with advice on barcode design provided by Dr. Modell. Spacer quantification was subsequently performed with a sequence-matching Python script that I designed to tally reads containing each of the expected sequences with 100% identity (including their expected barcode sequences). Troubleshooting of the script was carried out with mini files containing a subset of the reads (to reduce processing times). I generated these mini files using another Python script, provided by Phil Nussenzweig and Dr. Varble, which pulls out a user-specified range of reads. During the troubleshooting process, I also generated modified scripts for sampling and tallying aberrant reads (including barcode cross-contaminants that were likely generated during the library preparation steps), which allowed me to confirm that these reads contributed negligibly to my primary tallies. Ultimately, a representative dataset was manually compiled and plotted using

Microsoft Excel. My analysis indicates that the dual-targeting *gp16* spacer declined more rapidly than the spacers targeting only ϕNM1 (**Figure 3-11E**), and its rate of decline was more than double that observed in pairwise competitions with either of the marked TB4::ϕNM1 or TB4::ϕNM4 single lysogens (**Figures 3-4C & 3-12B**). Importantly, this was not observed for double lysogens where only ϕNM1 was targeted with the *gp5* or *gp43* spacers (**Figure 3-11E**), which declined at rates similar to those observed in their

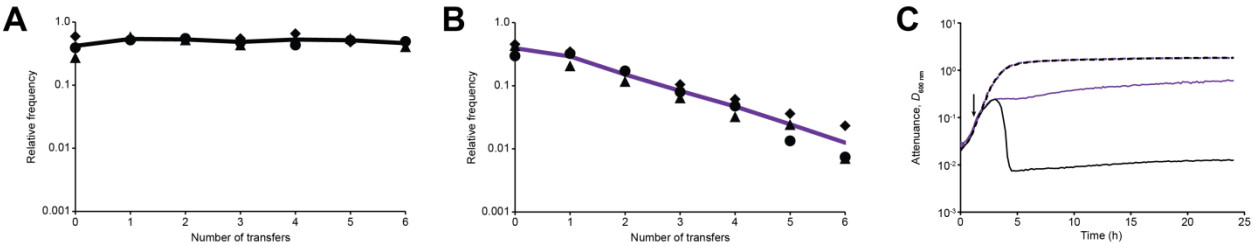


Figure 3-12. Fitness costs associated with conditional tolerance in marked TB4::ϕNM4 single lysogens harboring the *gp16* CRISPR-Cas plasmid.

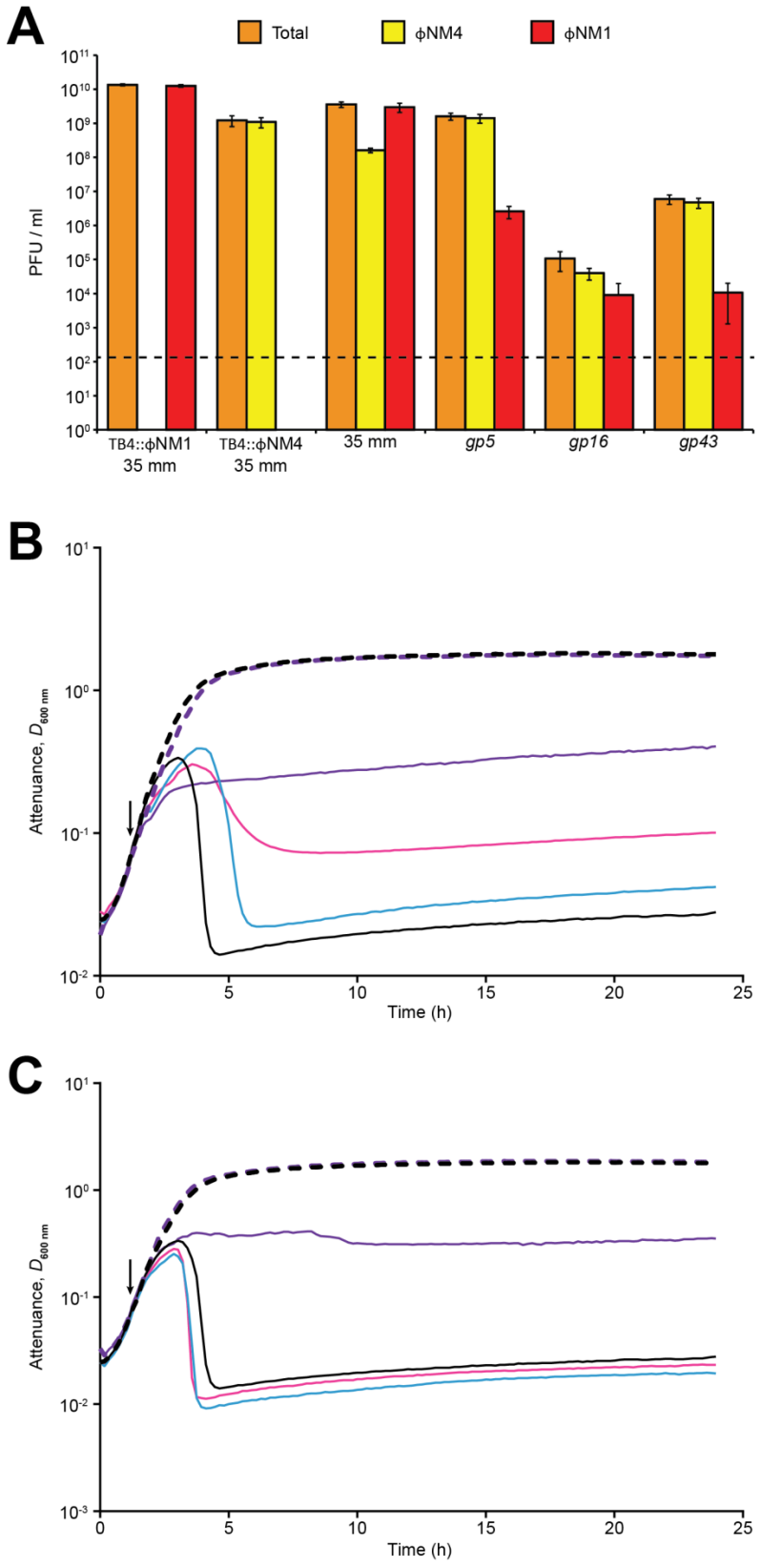
(A) Pairwise competition experiments for measuring the relative fitness of differentially marked TB4::ϕNM4 lysogens each harboring the non-targeting pGG79 plasmid, performed as described in **Figure 3-3**. **(B)** Same as in 'B', except that the perfectly matching *gp16* spacer was tested, and the solid line shares color coding with the ϕNM4 target ORF indicated in **Figure 3-11A**. **(C)** High-resolution growth curves of Tet^R-marked TB4::ϕNM4 single lysogens harboring either the *gp16* or non-targeting pGG79 CRISPR-Cas plasmid. Solid lines represent the average attenuance ($D_{600\,nm}$) measured from three biological replicates with MMC added at the indicated time point (black arrow), and are colored to match the target ORFs highlighted in **Figure 3-11A**, or black for the non-targeting vector. For comparison, average attenuance measured from three biological replicates in the absence of MMC is plotted for two strains (dashed lines).

respective pairwise competitions using marked TB4::ϕNM1 single lysogens (**Figures 3-4B & D and Table 3-1**). Thus, in spite of their capacity for conditional tolerance, type III-A CRISPR-Cas systems can become a greater fitness liability if they target multiple temperate phages. It should be pointed out, however, that these broadly targeting systems are expected to confer superior short-term advantages when guarding against lytic infection by heteroimmune phages.

The results described above confirmed the expectation that conditional tolerance can operate independently on heteroimmune temperate phages during infection. However, the effects of targeting in double lysogens, for example during prophage induction, were not explored. To investigate this, I first measured the number of PFUs released from unmarked doubly lysogenic cultures four hours post-induction with MMC, and plated supernatants on lawns of TB4::ϕNM1, TB4::ϕNM4, and TB4, in order to determine the relative proportion of ϕNM1 and ϕNM4 particles out of the total. For comparison, PFUs were also measured under these assay conditions from unmarked single lysogens of TB4::ϕNM1 and TB4::ϕNM4 harboring the non-targeting (35 mm) CRISPR-Cas plasmid. When double lysogens harboring this non-targeting plasmid are induced, ϕNM1 makes up the majority of the phage particles recovered (**Figure 3-13A**). Such asymmetries have also been observed during induction of *E. coli* double lysogens harboring heteroimmune lambdoid prophages, and are likely the result of intracellular competition (Refardt, 2011). For reasons unclear, ϕNM1 PFUs were specifically depleted

Figure 3-13. Conditionally tolerant type III-A systems can display prophage targeting specificity within double lysogens undergoing induction.

(A) Quantification of PFU concentrations (PFU / ml) in filtered supernatants from subcultures of unmarked TB4::ϕNM1+ϕNM4 double lysogens harboring one of the CRISPR-Cas plasmids tested in **Figure 3-11**, four hours post-treatment with MMC at 2.0 μg/ml. For comparison, single lysogens of TB4::ϕNM1 and TB4::ϕNM4 harboring the 35 mm CRISPR-Cas plasmid were also tested under these assay conditions. All lysates were spotted on TB4 lawns to measure total PFU, as well as TB4::ϕNM1 and TB4::ϕNM4 lawns to measure the fraction of ϕNM4- and ϕNM1-derived PFUs, respectively. Dashed line represents the limit of detection under these assay conditions. Error bars, mean ± s.d. (n = 3, biological replicates). **(B)** High-resolution growth curves of unmarked TB4::ϕNM1+ϕNM4 double lysogens harboring the *gp5*, *gp16*, *gp43*^{ori}, or 35 mm CRISPR-Cas plasmids. Solid lines represent the average attenuance ($D_{600\ nm}$) measured from three biological replicates with MMC added at the indicated time point (black arrow), and are colored to match the target ORFs highlighted in **Figure 3-11A**, or black for the non-targeting vector. For comparison, average attenuance measured from three biological replicates in the absence of MMC is plotted for two strains (dashed lines). **(C)** Same as in 'B', except that the *gp5*(*csm6**), *gp16*(*csm6**), and *gp43*^{ori}(*csm6**) mutant CRISPR-Cas plasmids were tested. For comparison, data from experiments with the 35 mm wild type CRISPR-Cas plasmid are also plotted.


A
Total
ϕNM4
ϕNM1
PFU / ml
TB4::ϕNM1
35 mm
TB4::ϕNM4
35 mm
35 mm
gp5
gp16
gp43
B
Attenuance, $D_{600\ nm}$
Time (h)
C
Attenuance, $D_{600\ nm}$
Time (h)

from the total by approximately one order of magnitude even when both prophages were targeted with the *gp16* spacer. Potentially, ϕNM4 simply fairs better during induction under the *gp16* targeting conditions, but neither prophage was particularly proliferative; PFU measurements were approaching my limit of detection in these experiments. Nevertheless, a much more pronounced effect was observed when only ϕNM1 was targeted with either the *gp5* or *gp43* spacers: ϕNM1 PFUs were specifically depleted from the total by about 3 orders of magnitude in each case. The absolute number of PFUs released during induction was lower with the *gp43* spacer than with the *gp5* spacer. This was initially puzzling, but high-resolution monitoring of these cultures during induction provided some additional insight. Whereas the dual-targeting *gp16* spacer prevented lysis, the *gp5* and *gp43*^{ori} spacers did not (**Figure 3-13B**)—presumably because ϕNM4 is allowed to complete its lytic cycle when it is not targeted. However, lysis profiles were more noticeably perturbed by the *gp43*^{ori} spacer than by the *gp5* spacer. The effect was similarly observed in a triplicate experiment with unmarked double lysogens harboring the *gp43* CRISPR-Cas plasmid (data not shown). This perturbation—suggestive of a simple *delay* in ϕNM4-mediated lysis—might explain why I recovered fewer particles overall with the *gp43* spacer, four hours post-induction. Assessing the number of particles released at a later time point could help to clarify this point. Nevertheless, given that ϕNM1 was specifically depleted to the same extent by the *gp5* spacer without a concomitant strong reduction in total PFUs, the lysis

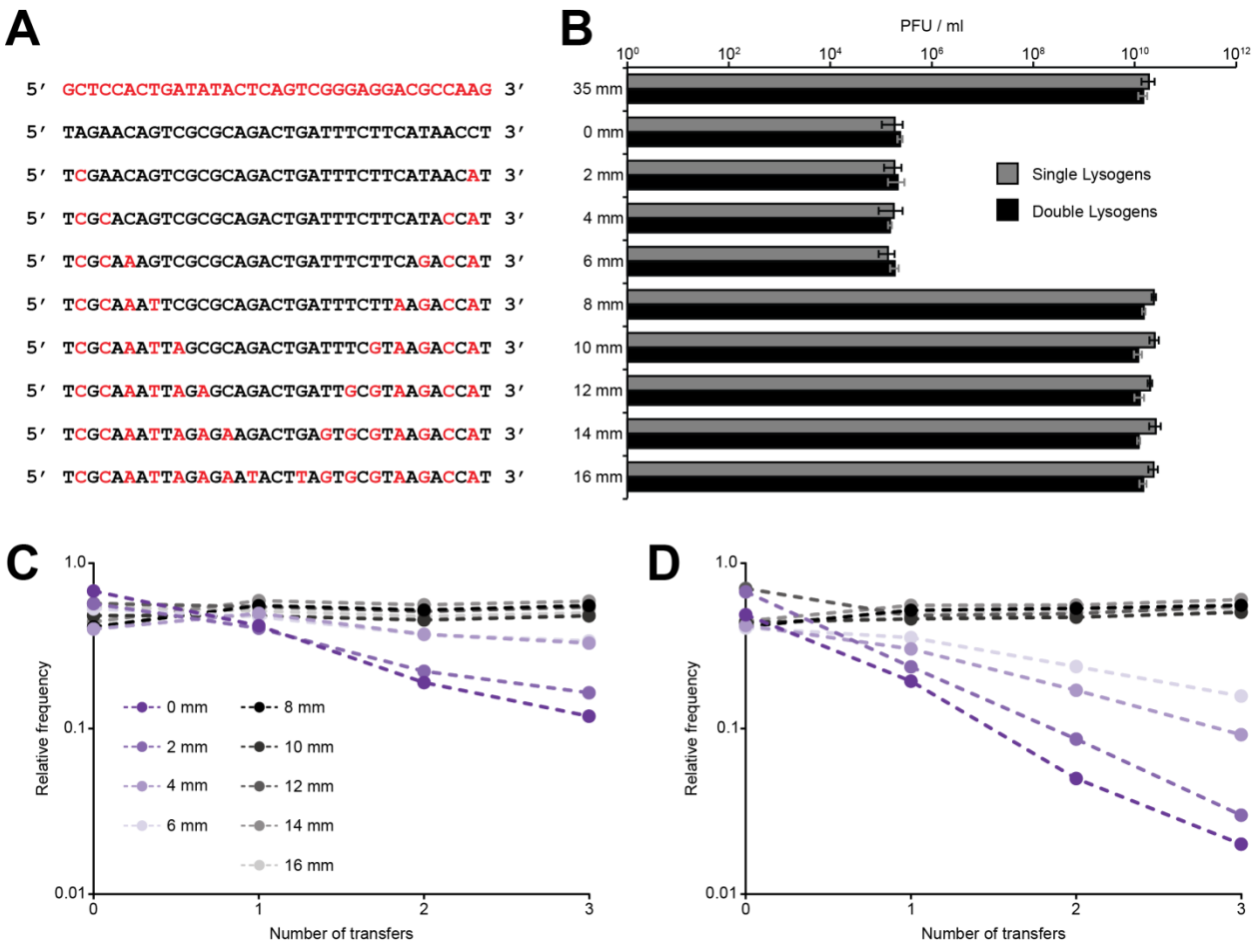
perturbation I observed with *gp43*^{ori} likely reflects a non-specific effect encountered during targeting. Given, also, mounting evidence which suggests that the HEPN RNase domain of type VI effector nucleases—and perhaps other Cas nucleases—can license off-target effects (Jiang et al., 2016; Shmakov et al., 2017), I tested whether an intact HEPN domain active site in Csm6 was required for the lysis perturbations in this assay. Whereas the *gp16* spacer still prevented lysis even in the *csm6*-mutant background (*csm6**), lysis profiles in double lysogens with the *gp5*(*csm6**) and *gp43*^{ori}(*csm6**) CRISPR-Cas plasmids were essentially indistinguishable from double lysogens with the non-targeting plasmid (**Figure 3-13C**). Taken together, these results demonstrate that conditionally tolerant type III-A systems can display prophage targeting specificity within double lysogens. However, the RNase activity of Csm6 can apparently license off-target effects during induction, and potentially perturb lysis by another prophage that is not targeted. Additional insight into the specificity of the type III-A targeting complex during induction could be provided by measuring PFUs derived from *csm6**-mutant cultures. For example, it would be interesting to see if the remaining nuclease activities associated with Cas10 and Csm3 are sufficient to license a specific reduction in ϕNM1-derived PFUs.

3.5 Effects of spacer-target mismatches in single and double lysogens

The above results confirmed that type III-A systems which target multiple temperate phages are not a barrier to polylysogenization per se, even if they possess perfectly matching spacers. However, natural examples of prophage targets and type III-A systems with perfectly matching spacers co-existing within staphylococcal isolates have yet to be reported. Having found that the mismatched *gp32* spacer in single lysogens is associated with a lower fitness cost than its perfectly matched *gp32** variant but is still capable of targeting during the lytic cycle (**Figures 3-3 & 3-5**), I wondered if spacer-target mismatches could likewise help to stabilize conditionally tolerant systems in polylysogenic populations. To investigate this, I constructed a series of mismatched spacers based around the *gp16* spacer, which targets an identical sequence found in both ϕNM1 and ϕNM4. It was previously shown that mismatches most readily abrogate type III targeting when they're distributed at both the 5′ and 3′ ends of the spacer (Cao et al., 2016; Manica et al., 2013). Accordingly, I introduced increasing numbers of mutations at both ends of the *gp16* spacer in an effort to sample a range of targeting potentials with a minimal number of mismatches (**Figure 3-14A**). The resulting CRISPR-Cas plasmids were then transduced into unmarked TB4::ϕNM4+ϕNM1 double lysogens, as well as unmarked TB4::ϕNM1 single lysogens for comparison. First, I examined the targeting potential of each spacer in single and double lysogens during MMC-stimulated prophage induction. In both cases, PFU induction was reduced to a similar

Figure 3-14. Effects of spacer-target mismatches in single and double lysogens harboring type III-A CRISPR-Cas systems with dual-targeting potential.

(A) Sequence summary of the *gp16* spacer and its mismatched variants. Black lettering indicates sequences found in the parental *gp16* spacer; red lettering indicates mutated base pairs. A subset of transversion mutations were used to avoid introducing G:U pairing when complementarity between the crRNA and transcribed target RNA is considered. **(B)** Quantification of PFU concentrations (PFU / ml) in filtered supernatants from subcultures of single (TB4::ϕNM1) or double (TB4::ϕNM1+ϕNM4) lysogens harboring the spacers depicted in panel 'A', four hours post-treatment with MMC at 1.0 μg/ml. Error bars, mean ± s.d. (n = 3, biological replicates). **(C)** Batch competition experiment with unmarked TB4::ϕNM1 single lysogens harboring CRISPR-Cas plasmids with the perfectly matched *gp16* spacer or one of several mismatched derivatives shown in panel 'A'. As outlined in the methods, deep sequencing was used to determine targeting spacer frequencies relative to the 35 mm spacer within the sample (y-axis), and plotted against the number of transfers (x-axis, one transfer per day). Dashed lines illustrate the change in relative frequency across time points, and are color-coded according to the labeled inset. **(D)** Same as in panel 'C', except that unmarked TB4::ϕNM1+ϕNM4 double lysogens were used in the experiment.


A
5′ GCTCCACTGATATACTCAGTCGGGAGGACGCCAAG 3′
5′ TAGAACAGTCGCGCAGACTGATTTCTTCATAACCT 3′
5′ TCGAACAGTCGCGCAGACTGATTTCTTCATAACAT 3′
5′ TCGCACAGTCGCGCAGACTGATTTCTTCATACCAT 3′
5′ TCGCAAAGTCGCGCAGACTGATTTCTTCAGACCAT 3′
5′ TCGCAAATTCGCGCAGACTGATTTCTTAAGACCAT 3′
5′ TCGCAAATTAGCGCAGACTGATTTCGTAAGACCAT 3′
5′ TCGCAAATTAGAGCAGACTGATTGCGTAAGACCAT 3′
5′ TCGCAAATTAGAGAAGACTGAGTGCGTAAGACCAT 3′
5′ TCGCAAATTAGAGAATACTTAGTGCGTAAGACCAT 3′
B
PFU / ml
10⁰
10²
10⁴
10⁶
10⁸
10¹⁰
10¹²
35 mm
0 mm
2 mm
4 mm
6 mm
8 mm
10 mm
12 mm
14 mm
16 mm
Single Lysogens
Double Lysogens
C
Relative frequency
1.0
0.1
0.01
0 mm
2 mm
4 mm
6 mm
8 mm
10 mm
12 mm
14 mm
16 mm
0
1
2
3
Number of transfers
D
Relative frequency
1.0
0.1
0.01
0
1
2
3
Number of transfers

degree by spacers with 6 or fewer mismatches, while targeting was undetectable with 8 or more mismatches (**Figure 3-14B**). I next tested each group of single or double lysogens in batch competition experiments lasting three days. Again, spacer abundances were measured by deep sequencing at each time point, and the frequency of each spacer was calculated relative to the 35 mm spacer's frequency in the sample. My analysis indicates that the spacers with 4 and 6 mismatches declined less rapidly than the spacers with 0 or 2 mismatches within their respective populations (**Figures 3-14C & D**). Meanwhile, no decline was discernible for the spacers with 8 or more mismatches in either population, consistent with the complete lack of detectable targeting by these spacers in the prophage induction assay. Furthermore, I observed a similar rate of decline for the perfectly matching *gp16* spacer across experiments with marked and unmarked single lysogens (**Figures 3-4C & 3-14C**), which served to validate my usage of the method in assessing relative fitness. I also found that the costs associated with dual-targeting spacers possessing either 4 or 6 mismatches (**Figure 3-14D**) were comparable to that of the perfectly matched *gp5* spacer targeting only ϕNM1 in double lysogens (**Figure 3-11E**). Thus, mismatches which alleviate fitness costs may help to ensure that spacers targeting multiple prophages are not always more burdensome to their polylysogenic hosts than singly targeting spacers. Moreover, this feature of type III-A systems might improve the long-term stability of spacers that guard against lytic infection by a range of heteroimmune phages with partially

divergent target sequences, given that the 4 and 6 mismatches did not abrogate targeting outright.

3.6 Discussion points and additional proposed experiments

I report here that *S. aureus* lysogens which maintain type III-A CRISPR-Cas systems can incur constitutive fitness costs, despite the genetic stability generally afforded by conditional tolerance. The costs depend on the presence of certain prophage-targeting spacers, and likely result from incomplete prophage tolerance by the system's transcription-dependent targeting machinery. Several lines of evidence support this interpretation.

Previously, transcription-dependent chromosomal targeting by the type III-A system of *S. epidermidis* RP62a was found to cause severe growth defects (Goldberg et al., 2014; Jiang et al., 2016), and similar effects have been observed with other types of CRISPR-Cas systems (Edgar and Qimron, 2010; Jiang et al., 2013b; Vercoe et al., 2013). At least in part, this appears to result from toxicity associated with frequent nicking or cleavage of the host's chromosomal DNA. Minor growth defects, in line with the fitness costs described in this work, might instead be expected if type III-A chromosomal targeting were infrequent; e.g., licensed by transcription in a sufficiently small subpopulation of cells. When lambdoid lysogens are cultured under standard growth conditions, transcription from lytic promoters can occur in a subpopulation of cells

undergoing the lytic cycle via spontaneous prophage induction (Nanda et al., 2015). For SOS-inducible lambdoid phages, the frequency of spontaneous prophage induction in the population depends almost entirely on the host's SOS responsiveness (Fuchs et al., 1999; Little and Michalowski, 2010; Nanda et al., 2015), and this is presumably unaltered by the presence or absence of conditionally tolerant spacers. Given that prophage induction leads to lysis in the absence of targeting spacers, the fitness of conditionally tolerant lysogens would not necessarily be impaired relative to non-targeting lysogens if spontaneous prophage induction events were strictly required for toxicity. Indeed, even when ϕNM1's capacity for detectable spontaneous prophage induction was nearly completely eliminated by mutations in its *cI*-like repressor (**Figures 3-5A & 3-8**), fitness costs were not alleviated (**Figures 3-5F & G**). This suggests that the fitness costs might depend instead on leaky transcription of prophage targets occurring within stably lysogenic cells (**Figure 3-5E**); in other words, targeting occurring in a subpopulation of cells that is not undergoing spontaneous prophage induction. Consistent with a requirement for transcription at prophage target loci, costs associated with the *gp5* spacer were virtually undetectable when the P_{cro} promoter immediately upstream of its target was inactivated (**Figure 3-4G**). Meanwhile, mutation of P_{cro} did not eliminate fitness costs associated with spacers that target other prophage loci (**Figures 3-4H & I**), despite preventing detectable prophage induction (**Figures 3-5A & C**).

The potential for prophage lytic loci to be de-repressed independently of spontaneous prophage induction is not unprecedented for lambdoid temperate phages, and was previously demonstrated using a genetic reporter system (Livny and Friedman, 2004). In this study, furthermore, the authors detected slight de-repression above background even when prophages carried a non-inducible *cI* allele similar to that of my ϕNM1^{ind-} mutant. Conceivably, leaky transcription arising in this manner may license the fitness costs I've detected. This model could also account for the discrepancies I observed with spacers matching to different regions of the ϕNM1 prophage in single lysogens (**Figure 3-4**). For example, different promoters in each region might produce distinct patterns of leaky transcription within stably lysogenic cells, even if upregulation of the promoters is normally coordinated by extrinsic factors—such as the host's SOS response—within cells undergoing prophage induction. Moreover, this model could explain why a combined effect of having both ϕNM1 and ϕNM4 was observed in double lysogens with the dual-targeting *gp16* spacer (**Figure 3-11E**), even though induction of heteroimmune prophages can be coordinated in polylysogens by the SOS response (Refardt, 2011). In this scenario, if instances of leaky transcription in each prophage are poorly correlated within cells that are not undergoing induction, the combined effect of having both targets might be essentially additive at the population level. The cost observed with the *gp16* spacer in double lysogens was indeed comparable to the sum of that observed with single lysogens of

ϕNM1 and ϕNM4 alone (**Figures 3-4C & 3-12B**). Regarding the potential for uncorrelated leaky expression from distinct chromosomal loci in bacteria, there is precedent here as well. It was previously demonstrated that cell-cell variability becomes elevated under transcriptionally repressed conditions, even when expression from two loci with identical promoters and identical repressors is compared (Elowitz et al., 2002). Similar single-cell approaches could be used to clarify whether expression of each prophage is occasionally uncoupled from one another, and from the host's SOS response. The combined effect I observed could also be synergistic, at least in part, particularly if transcription were found to occur simultaneously at both loci in certain cells. Cells which are targeted at both loci within a sufficiently small timeframe might experience more severe toxicity. In this scenario, however, it is assumed that a subset of the target's transcription events do not produce lethal outcomes; otherwise, if a single transcription event at the target site were always sufficient to license lethality, it is difficult to imagine how synergy would arise. Testing the effects of CRISPR arrays with more than one prophage-targeting spacer could perhaps help to clarify whether or not this is the case. I'm proposing, for example, to construct a conditionally tolerant TB4::ϕNM1 single lysogen with a CRISPR array that has two spacers targeting nearby sequences within the *gp16* ORF. If transcriptional surveillance by the type III-A system were saturating with a single spacer, we might find that the presence of the downstream spacer incurs no additional fitness cost. Such a result would imply that all

of the target's transcription events lead to lethal outcomes. Expanded CRISPR arrays could also be used to investigate the effects of targeting at different loci within the *same* prophage. For example, if transcription of the *gp5* and *gp43* loci were poorly coordinated within single cells of TB4::ϕNM1, we might expect a CRISPR array possessing both the *gp5* and *gp43* spacers to license more severe fitness costs (again, perhaps essentially equal the sum of the *gp5*- and *gp43*-associated fitness costs alone).

It is still unclear to me why the fitness cost associated with the *gp16* spacer was reduced in the ϕNM1$^{P_{cro}-}$ background (**Figure 3-4**). One interpretation follows that a weaker downstream promoter which is coordinately transcribed with P_{cro} in wild type ϕNM1 could license these residual fitness costs in the absence of P_{cro} (e.g., if transcription from P_{cro} is effectively 'dominant' to transcription from this downstream promoter, and yet not strictly required for it). Based on my RNA-seq and promoter-prediction analyses (**Figure 3-6 and Table 3-2**), one such downstream promoter might be located within the annotated *gp9* ORF. In this scenario, leaky transcriptional read-through of *gp16* from P_{cro} would explain the higher fitness cost observed in the wild type ϕNM1 background, and potentially also explain why the *gp5* and *gp16* fitness costs were so similar in magnitude. Arguing against this, however, a spacer which targets the intervening *gp8* ORF was not associated with a detectable fitness cost (**Figures 3-4A & 3-9B**). An alternative interpretation follows that leaky transcription from P_{cro} terminates before *gp8* in the lysogenic populations, but gives rise to one or more immediate early

gene products which can partially enhance transcription from a downstream promoter, such as the putative promoter in *gp9*. In this scenario, leaky transcription from P_{cro} and the downstream promoter is not necessarily well-coordinated, and the similarity between the fitness costs observed with the *gp5* and *gp16* spacers in wild type ΦNM1 lysogens is simply fortuitous. Again, testing of a CRISPR-Cas plasmid with multiple ΦNM1-targeting spacers (e.g., targeting both *gp5* and *gp16*), could potentially prove insightful, and help to distinguish between these interpretations.

Additional evidence pointing to transcription-dependent chromosomal targeting as the source of the fitness costs was provided by experiments with mutant type III-A Cas nucleases. Ablation of *csm3*'s RNase activity did not reduce fitness costs (**Figure 3-10D**). Given its role in specific cleavage of target RNAs by type III systems, this was perhaps expected (Tamulaitis et al., 2017); knockdown of prophage lytic transcripts should not be toxic to lysogenic cells per se. Meanwhile, fitness costs appeared to be completely abolished by mutation of *cas10*'s palm polymerase domain (**Figure 3-10F**). In addition, whereas this $cas10^{DD}$ palm domain mutation strongly impaired targeting with the *gp16* spacer in prophage induction assays, ablation of the *csm3* and *csm6* RNase activities did not (**Figures 3-10B & C**). These data were in line with past reports indicating that both transcription-dependent DNA degradation and robust immunity to lytic infections by this system requires an intact palm polymerase domain (Samai et al., 2015), and that its RNase activities are not required to disrupt the lytic cycle if the

phage's target falls within an early-transcribed region, such as *gp16* (Jiang et al., 2016). However, although the *cas10*^{DD} palm domain mutation was not found to prevent target RNA cleavage by the type III-A complex *in vitro* (Samai et al., 2015), the possibility that this mutation could have pleiotropic effects *in vivo* has not been ruled out. Perhaps less expected was the finding that mutation of *csm6*'s HEPN RNase domain almost completely eliminated the costs associated with the *gp16* spacer (**Figures 3-10G & H**). This RNase function is not strictly required to elicit severe growth defects associated with transcription-dependent chromosomal targeting, or targeting of DNA elements in general (Jiang et al., 2016). Hence, the nature of Csm6's contribution to the growth defects described in this work remains unclear. Of potential relevance here, it was recently demonstrated that the HEPN RNase domains of RNA-guided type VI effector nucleases can mediate cleavage of non-target RNAs in *trans*, or 'collateral' RNA cleavage, upon binding of target RNAs (Shmakov et al., 2017). Furthermore, it was shown that type VI systems can cause growth defects in the presence of target RNAs, and this was proposed to result from off-target cleavage of cellular transcripts (Abudayyeh et al., 2016; Smargon et al., 2017). It's possible that I'm observing a similar phenomenon in my experiments (and the results presented in **Figure 3-13** would support this notion), but an analogous mechanism for coupling Csm6's activity to transcription-dependent targeting by the type III-A complex has yet to be established. Current evidence suggests that binding of nascent target transcripts in *cis* is a pre-

requisite for both localizing type III complexes to specific DNA elements and for regulating their DNase activity (Estrella et al., 2016; Goldberg et al., 2014; Kazlauskiene et al., 2016; Samai et al., 2015). Binding of target RNAs, nascent or otherwise, could likewise be a pre-requisite for recruitment or regulation of Csm6's RNase activity, and this recruitment or regulation would presumably involve biophysical (and/or biochemical) interactions between Csm6 and the Cas10-Csm complex during targeting (Jiang et al., 2016). Meanwhile, such interactions—if confirmed—might feed back on *in vivo* DNA targeting by the Cas10-Csm complex in a more direct manner. In other words, the partially Csm6-dependent fitness costs I'm observing might result from enhanced degradation of chromosomal DNA licensed by Csm6-mediated feedback on the complex (at least in instances where nascent target RNAs are bound in *cis*)—rather than Csm6-dependent collateral RNA degradation effects per se. Arguing against this, however, I found almost no reduction in the *gp16* spacer's fitness cost upon mutation of the *cas10* HD domain (**Figure 3-10E**). HD domain mutations are thought to specifically inactivate the DNase activity of Cas10 proteins (Estrella et al., 2016; Kazlauskiene et al., 2016). Testing of a *cas10*^{HD}/*csm6** double mutant would confirm whether the residual fitness costs observed in the *csm6** mutant background are due to the HD domain's DNase activity (and thus presumably the direct result of chromosomal degradation). Note that, in each scenario described above, transcription is assumed to be a pre-requisite for the Csm6- and Cas10-dependent effects, so that the wild type system

would still be expected to tolerate sufficiently repressed prophage targets. However, given the potential for incomplete tolerance of leaky targets, it is tempting to speculate that further layers of regulation could evolve to fine-tune type III-A targeting nuclease activities and provide more complete prophage tolerance while still ensuring robust immunity during lytic infections.

The results presented in this work suggest that, among natural lysogenization-prone lineages of *S. aureus* (Aanensen et al., 2016; Goerke et al., 2009), enrichment for certain temperate-phage-targeting spacers could serve as an ad hoc mechanism of evolutionary fine-tuning for type III-A CRISPR-Cas systems. Different fitness costs were observed for spacers matching different regions of the ϕNM1 prophage (**Figures 3-3, 3-4, & 3-9B**) and fitness costs were potentially more severe in double lysogens if the spacer targeted both prophages (**Figure 3-11E**). Presumably, this latter phenomenon would be particularly relevant in natural environments where encounters with heteroimmune phages are common. Under these conditions, CRISPR-adapted lineages could gain a short-term survival advantage over non-adapted lineages (lysogenic or otherwise) when the population is infected by heteroimmune phages with identical or partially divergent target sequences. However, as my results illustrate, CRISPR-adapted cells that become lysogenized or polylysogenized during these subsequent infections may also incur constitutive fitness costs. In one case, a perfectly matching spacer was found to license curing of its ϕNM1 prophage target (**Figure 3-7**). This effectively

remedied the host's immunopathology in a manner that did not compromise the CRISPR-Cas system itself. While a phenomenon such as this could be detrimental in environments where the host relies on its prophage(s) for beneficial traits, it might enhance the stability of a CRISPR-Cas system in environments where immunity to lytic infection conferred by its spacer(s) is sufficient for survival. In contrast, I found that the other CRISPR-Cas systems with perfectly matching spacers eventually succumbed to genetic inactivation (data not shown). Finally, the prophage-curing phenomenon could also promote the stability of a CRISPR-Cas system by preventing the accumulation of target prophages, which—as I've demonstrated (**Figures 3-11E & 3-14D**)—may further jeopardize host fitness.

Importantly, costs were found to be reduced in the presence of sufficient spacer-target mismatching, and in certain cases even when the mismatching did not abrogate immunity during the lytic cycle (**Figures 3-3, 3-5A, 3-5B & 3-14**). For these reasons, I speculate that conditionally tolerant lysogens harboring mismatched spacers could be generally more stable in natural populations than those with perfectly matched spacers. In line with this possibility, a putative conditionally tolerant isolate of *S. aureus* was recently identified, and it contains mismatches in each prophage target relative to a spacer in its type III-A CRISPR-Cas system (Cao et al., 2016). Type III-A systems have also been found to display signatures of horizontal acquisition within *S. aureus* genomes (Cao et al., 2016; Golding et al., 2010; Kinnevey et al., 2013). Hence, I further speculate

that the genetic stability afforded by conditional tolerance, and by extension spacer-target mismatches, could facilitate horizontal dissemination of prophage-targeting type III-A systems throughout natural lysogenic populations. Currently, it is unclear to me why the type III-A CRISPR-Cas plasmids employed in this work do not undergo detectable CRISPR adaptation, despite various conditions tested by me and other members of the lab (data not shown). Once rectified, investigation of type III-A CRISPR adaptation could be useful in clarifying the extent to which spacer diversity and incomplete prophage tolerance influence the stability of these systems in temperate-phage-infected populations at large.

Lastly, it should be emphasized that additional phenotypic consequences of conditional tolerance, or of incomplete tolerance, have not been ruled out. For example, it was previously found that low-grade chromosomal targeting by a type I CRISPR-Cas system, made possible by certain spacer-target mismatches, is sufficient to trigger a host-controlled response in *Pseudomonas aeruginosa* PA14 which prevents biofilm formation without an overt growth defect (Heussler et al., 2015). This phenotype was originally identified using a lysogenic derivative where the target was chromosomally integrated via a prophage, although these mismatches also abolished immunity during lytic infections (Cady et al., 2012; Cady et al., 2011). It remains to be seen whether immunopathological 'costs' associated with type III-A systems in lysogenic hosts could

be similarly exploited to provide a phenotype in certain environments, without necessarily compromising resistance to lytic infections.

3.7 Tables associated with Chapter 3

Table 3-1. Selection coefficients determined in fitness assays.

Prophage(s)	Spacer[(a)]	Figure	Mean *s*[(b)]	s.d.[(c)]	Median *s*
ϕNM1	*gp32**	3-3B	0.1088	0.0561 (*n* = 3)	0.1162
ϕNM1	*gp32* [5 mm]	3-3C	0.0522	0.0476 (*n* = 3)	0.0530
ϕNM1	*gp5*	3-4B	0.1036	0.0394 (*n* = 6)	0.1062
ϕNM1	*gp16*	3-4C	0.0822	0.0663 (*n* = 6)	0.0680
ϕNM1	*gp43*	3-4D	0.0327	0.0283 (*n* = 6)	0.0192
ϕNM1	Parent vector	3-4E	0.0077	0.0344 (*n* = 6)	0.0005
ϕNM1^{Pcro-}	*gp5*	3-4G	0.0137	0.0256 (*n* = 6)	0.0168
ϕNM1^{Pcro-}	*gp16*	3-4H	0.0531	0.0098 (*n* = 6)	0.0557
ϕNM1^{Pcro-}	*gp43*	3-4I	0.0334	0.0262 (*n* = 6)	0.0373
ϕNM1^{ind-}	*gp16*	3-5F	0.1029	0.0498 (*n* = 6)	0.0981
ϕNM1^{ind-}	*gp43*	3-5G	0.0356	0.0118 (*n* = 6)	0.0335
ϕNM1^{Pcro-}	*gp32**	3-7B	0.1449	0.0644 (*n* = 3)	0.1117
ϕNM1^{Pcro-}	*gp32**	3-7B	0.1051	0.0682 (*n* = 6)	0.1078
ϕNM1	*gp8*	3-9B	0.0045	0.0074 (*n* = 6)	0.0047
ϕNM1	*gp16* (*csm3*)	3-10D	0.0745	0.0504 (*n* = 6)	0.0844
ϕNM1	*gp16* (*cas10*^{HD})	3-10E	0.0667	0.0166 (*n* = 6)	0.0681
ϕNM1	*gp16* (*cas10*^{DD})	3-10F	0.0055	0.0170 (*n* = 6)	0.0083
ϕNM1	*gp16* (*csm6*)	3-10G	0.0113	0.0177 (*n* = 6)	0.0088
ϕNM1	*gp16* (*csm3*/*csm6*)	3-10H	0.0258	0.0189 (*n* = 6)	0.0233
ϕNM4+ϕNM1	*gp5*	3-11E	0.1059	0.0441 (*n* = 3)	0.1288
ϕNM4+ϕNM1	*gp16*	3-11E	0.3341	0.0775 (*n* = 3)	0.3396
ϕNM4+ϕNM1	*gp43*	3-11E	0.0581	0.0528 (*n* = 3)	0.0300
ϕNM4	*gp16*	3-12A	-0.0030	0.0262 (*n* = 6)	0.0045
ϕNM4	*gp16*	3-12B	0.0868	0.0209 (*n* = 6)	0.0919
ϕNM1	*gp16*	3-14C	0.1080	0.0404 (*n* = 3)	0.1076
ϕNM1	*gp16* [2 mm]	3-14C	0.0722	0.0320 (*n* = 3)	0.0689
ϕNM1	*gp16* [4 mm]	3-14C	0.0115	0.0474 (*n* = 3)	0.0198
ϕNM1	*gp16* [6 mm]	3-14C	0.0268	0.0163 (*n* = 3)	0.0210
ϕNM1	*gp16* [8 mm]	3-14C	-0.0191	0.0355 (*n* = 3)	-0.0133
ϕNM1	*gp16* [10 mm]	3-14C	0.0002	0.0111 (*n* = 3)	0.0003
ϕNM1	*gp16* [12 mm]	3-14C	0.0038	0.0147 (*n* = 3)	0.0110
ϕNM1	*gp16* [14 mm]	3-14C	-0.0206	0.0393 (*n* = 3)	-0.0122
ϕNM1	*gp16* [16 mm]	3-14C	0.0056	0.0139 (*n* = 3)	0.0115
ϕNM4+ϕNM1	*gp16*	3-14D	0.2155	0.0780 (*n* = 3)	0.1902
ϕNM4+ϕNM1	*gp16* [2 mm]	3-14D	0.2090	0.0289 (*n* = 3)	0.1949
ϕNM4+ϕNM1	*gp16* [4 mm]	3-14D	0.0812	0.0224 (*n* = 3)	0.0939
ϕNM4+ϕNM1	*gp16* [6 mm]	3-14D	0.0494	0.0238 (*n* = 3)	0.0604
ϕNM4+ϕNM1	*gp16* [8 mm]	3-14D	-0.0192	0.0201 (*n* = 3)	-0.0098
ϕNM4+ϕNM1	*gp16* [10 mm]	3-14D	-0.0089	0.0043 (*n* = 3)	-0.0090
ϕNM4+ϕNM1	*gp16* [12 mm]	3-14D	0.0201	0.0604 (*n* = 3)	-0.0051
ϕNM4+ϕNM1	*gp16* [14 mm]	3-14D	-0.0212	0.0216 (*n* = 3)	-0.0181
ϕNM4+ϕNM1	*gp16* [16 mm]	3-14D	-0.0145	0.0139 (*n* = 3)	-0.0117

Table 3-1. Selection coefficients determined in fitness assays (continued).

(a) For spacers possessing mismatches relative to their prophage target, the number of mismatches (mm) is written in brackets. For spacers present in CRISPR-Cas plasmids with one or more mutant *cas* genes, the mutated gene is written in parentheses.

(b) Selection coefficients, '*s*', determined from the change or average change (where available) in relative frequency after each transfer interval in fitness assays. Values were extracted from the equation, $dq/dt = -q(1 - q)s$, where q is the relative frequency of the spacer being assayed, and $t = 9.97$ generations is assumed for 1:1000 dilutions at each transfer interval. Positive selection coefficients ($s > 0$) suggest that the spacer is associated with a fitness cost.

(c) Standard deviation.

Table 3-2. Sequences of putative ϕNM1 promoter elements depicted in **Figure 3-6**.

Promoter[(a)]	Sequence (5'-3')[(b)]
P_{cro}	AAAATAATTACGAAAAATACTTGCA*ATCGTATTCTAATTACGAT*ATACTTTGATCAGAACTTAA
$P_{ssb(1)}$	AGTCACGATCTAAAAGATATTTAACTGTATTCAAAATTTTCATATCTTGTTGAGCTTTTAAGC
$P_{ssb(2)}$	TTTCCAAGAAAAACAATTGATTGAATTAGTTACTCGATTAGGTATTAAGTTAAATCTTCCTAG
P_{rep}	TTAAATTGTTCTTTGAATTTTTCAAATTCTACTTCTCTTTGATAAATAACTTTATCCACATAA
P_{dut}	ACCAACAAGCAATTGTTATAGTGATAGACATAGCTTACAACTTATATTCTATCGACCAACTCA
P_{terS}	**TGACAGT**AAAA**TGACAGT**TTTTGACACCTATAACGAGGTATTATGATAGCGTAAGATATTGA
P_{portal}	GAAAAAGAAGTTATACATCATTGAAGAGTATGTTAAACAAGGTATGCTGAATGATGAAATAGC
P_{capsid}	ACAGCGTAAGGTCGGCAGTATTAGCCGATAAAGAAAAATCGAAATATAATGAACCTCTCTTTTAA

(a) Promoters are named after the nearest encompassing or downstream ORF with a known or predicted function, according to the ϕNM1 annotation (Accession: NC_008583.1).

(b) Candidates were selected from among prokaryotic promoters predicted using a neural network promoter prediction (Reese, 2001) tool available online at: http://www.fruitfly.org/seq_tools/promoter.html. Published *E. coli* promoter consensuses (Shimada et al., 2014; Shultzaberger et al., 2007) were consulted for additional reference to manually curate the initial hits. Putative -35 and -10 elements are underlined. A putative operator binding site palindrome is italicized in the P_{cro} promoter. Putative RinA-interacting direct repeat elements (Ferrer et al., 2011; Quiles-Puchalt et al., 2013) adjacent to the P_{terS} promoter are bold-faced.

Table 3-3. Spacers tested in Chapter 3.

Spacer[(a)]	Sequence (5'-3')[(b)]	Plasmid Name
Parent vector	TGAGACCAGTCTCGGAAGCTCAAAGGTCTC	pGG79
gp5	ATGTCTTAGCAATTCTAAAAGCATCTCTAGGTTTA	pGG126
gp8	TTGCCAGGTTCACTACGTTGAATCATTGCAATCTC	pGG124
gp16	TAGAACAGTCGCGCAGACTGATTTCTTCATAACCT	pGG102
gp32 [5 mm]	TTAAATCTTTGATTGCTCTTAGCTCTAGTTATGTAT	pGG122
*gp32**	GTAAACCTTTGATTGCTCTTAGCTCGAGTTATGTGC	pGG100
gp43^{ori}	ATTCGTCATCTTCAAGTAATGCCTCTAAATCAATAA	pGG104
gp43	ATTCGTCATCTTCAAGTAATGCCTCTAAATCAATA	pGG165
gp16 [35 mm]	GCTCCACTGATATACTCAGTCGGGAGGACGCCAAG	pGG153
gp16 [2 mm]	TCGAACAGTCGCGCAGACTGATTTCTTCATAACAT	pGG154
gp16 [4 mm]	TCGCACAGTCGCGCAGACTGATTTCTTCATACCAT	pGG155
gp16 [6 mm]	TCGCAAAGTCGCGCAGACTGATTTCTTCAGACCAT	pGG156
gp16 [8 mm]	TCGCAAATTCGCGCAGACTGATTTCTTAAGACCAT	pGG157
gp16 [10 mm]	TCGCAAATTAGCGCAGACTGATTTCGTAAGACCAT	pGG158
gp16 [12 mm]	TCGCAAATTAGAGCAGACTGATTGCGTAAGACCAT	pGG160
gp16 [14 mm]	TCGCAAATTAGAGAAGACTGAGTGCGTAAGACCAT	pGG161
gp16 [16 mm]	TCGCAAATTAGAGAATACTTAGTGCGTAAGACCAT	pGG162

(a) '*gp*' numbers correspond to open reading frames with a matching target sequence in ϕNM1 (Accession: NC_008583.1). Mismatched variants are also listed, with the number of mismatches (mm) denoted within brackets. Each spacer is located between repeats 2 and 3 of the CRISPR array. A placeholder sequence with two opposing BsaI restriction sites is present at this position in the parent vector.

(b) In order to avoid excessive matching between the 5' crRNA tag and the target's corresponding flanking sequence, which has been shown to abrogate DNA targeting by the type III-A system (Marraffini and Sontheimer, 2010b; Samai et al., 2015), spacer-target combinations were selected such that no more than 1 bp of complementarity was present at flanking sequences in each case.

Table 3-4. Oligonucleotides used in Chapter 3.

Name	Sequence	Primary Purpose
oGG281	GTATCGATCGAGACCTTTGAGCTTCCGAGAC	Construction of pGG79 control/parent vector
oGG282	CCACCCCGAAGAAAAGGGGACGAGAACTAAATCTAACAACACTCTAAAAAATTG	Construction of pGG79 control/parent vector
oGG330	GAACGTAAACCTTTGATTGCTCTTAGCTCGAGTTATGTGCG	Construction of pGG100
oGG331	CGATCGCACATAACTCGAGCTAAGAGCAATCAAAGGTTTAC	Construction of pGG100
oGG332	GAACTAGAACAGTCGCGCAGACTGATTTCTTCATAACCTG	Construction of pGG102
oGG333	CGATCAGGTTATGAAGAAATCAGTCTGCGCGACTGTTCTA	Construction of pGG102
oGG334	GAACATTCGTCATCTTCAAGTAATGCCTCTAAATCAATAAG	Construction of pGG104
oGG335	CGATCTTATTGATTTAGAGGCATTACTTGAAGATGACGAAT	Construction of pGG104
oGG351	GAACCAGGTGCAGTTACACCAATCATCAAACCTTTACCAAG	Construction of pGG115
oGG352	CGATCTTGGTAAAGGTTTGATGATTGGTGTAACTGCACCTG	Construction of pGG115
oGG385	GAACTTAAATCTTTGATTGCTCTTAGCTCTAGTTATGTATG	Construction of pGG122
oGG386	CGATCATACATAACTAGAGCTAAGAGCAATCAAAGATTTAA	Construction of pGG122
oGG393	GAACTTGCCAGGTTCACTACGTTGAATCATTGCAATCTCG	Construction of pGG124
oGG394	CGATCGAGATTGCAATGATTCAACGTAGTGAACCTGGCAA	Construction of pGG124
oGG397	GAACATGTCTTAGCAATTCTAAAAGCATCTCTAGGTTTAG	Construction of pGG126
oGG398	CGATCTAAACCTAGAGATGCTTTTAGAATTGCTAAGACAT	Construction of pGG126
oGG435	GAACGCTCCACTGATATACTCAGTCGGGAGGACGCCAAGG	Construction of pGG153
oGG436	CGATCCTTGGCGTCCTCCCGACTGAGTATATCAGTGGAGC	Construction of pGG153
oGG437	GAACTCGAACAGTCGCGCAGACTGATTTCTTCATAACATG	Construction of pGG154
oGG438	CGATCATGTTATGAAGAAATCAGTCTGCGCGACTGTTCGA	Construction of pGG154
oGG439	GAACTCGCACAGTCGCGCAGACTGATTTCTTCATACCATG	Construction of pGG155
oGG440	CGATCATGGTATGAAGAAATCAGTCTGCGCGACTGTGCGA	Construction of pGG155
oGG441	GAACTCGCAAAGTCGCGCAGACTGATTTCTTCAGACCATG	Construction of pGG156
oGG442	CGATCATGGTCTGAAGAAATCAGTCTGCGCGACTTTGCGA	Construction of pGG156
oGG443	GAACTCGCAAATTCGCGCAGACTGATTTCTTAAGACCATG	Construction of pGG157
oGG444	CGATCATGGTCTTAAGAAATCAGTCTGCGCGAATTTGCGA	Construction of pGG157
oGG445	GAACTCGCAAATTAGCGCAGACTGATTTCGTAAGACCATG	Construction of pGG158
oGG446	CGATCATGGTCTTACGAAATCAGTCTGCGCTAATTTGCGA	Construction of pGG158
oGG447	GAACTCGCAAATTAGAGCAGACTGATTGCGTAAGACCATG	Construction of pGG160
oGG448	CGATCATGGTCTTACGCAATCAGTCTGCTCTAATTTGCGA	Construction of pGG160
oGG449	GAACTCGCAAATTAGAGAAGACTGAGTGCGTAAGACCATG	Construction of pGG161
oGG450	CGATCATGGTCTTACGCACTCAGTCTTCTCTAATTTGCGA	Construction of pGG161
oGG451	GAACTCGCAAATTAGAGAATACTTAGTGCGTAAGACCATG	Construction of pGG162
oGG452	CGATCATGGTCTTACGCACTAAGTATTCTCTAATTTGCGA	Construction of pGG162
oGG461	GAACATTCGTCATCTTCAAGTAATGCCTCTAAATCAATAG	Construction of pGG165
oGG462	CGATCTATTGATTTAGAGGCATTACTTGAAGATGACGAAT	Construction of pGG165
oGG320	GAACGTTTTGGTCGTAGAGCACACGGTTTAACGACTTAAG	Construction of pGG91
oGG321	CGATCTTAAGTCGTTAAACCGTGTGCTCTACGACCAAAAC	Construction of pGG91
oGG432	GAACCCAGTTGCACCACATGCAATATACGATACTAGTTTG	Construction of pGG152
oGG433	CGATCAAACTAGTATCGTATATTGCATGTGGTGCAACTGG	Construction of pGG152
L6	AAAGGTACCAAATTTAATGCTATTTTCCTTCGC	Type III CRISPR array sequence verification
L50	AAAAGATCTAATAATGTATTTACGCTGGGGC	Type III CRISPR array sequence verification
PS153	GGTAAATCAAAACTAACTAACAAATACATTAGTTTCCCACCTCTATCATC	Construction of pPS95
PS154	GATGATAGAGGTGGGAAACTAATGTATTTGTTAGTTAGTTTTGATTTACC	Construction of pPS95
PS465	GAATCTAGTATGATTGGAGCAATTGCTTCTCCTGTAGTTAGAGATTTGCAAACC	Construction of pPS95
PS466	GGTTTGCAAATCTCTAACTACAGGAGAAGCAATTGCTCCAATCATACTAGATTC	Construction of pPS95
W852	CCAACAAACGACTTTTAGTATAACC	Construction of other mutant type III plasmids
W614	GGTTATACTAAAAGTCGTTTGTTGG	Construction of other mutant type III plasmids
PS565	GTTAGAGTATTATATTTATCAAATAAGGGAG	Construction of pGG99 and pGG89
PS566	CTCCCTTATTTGATAAATATAATACTCTAAC	Construction of pGG99 and pGG89
oGG457	GAGCACAAATTAAGTATTAAAGAAG	Construction of pGG167
oGG458	CTTCTTTAATACTTAATTTGTGCTC	Construction of pGG167
oGG424	CATATTGCCTGATGAAGTGAATAG	Construction of pGG139
oGG425	CTATTCACTTCATCAGGCAATATG	Construction of pGG139
oGG270	TAAAATGTGGTTTGACAAACGAAAATTGGATAAAGTGGG	Additional type III plasmid sequencing
L36	AAAGGTACCTTATTACTACCTAAGATGATAGAGG	Additional type III plasmid sequencing
W14	ATCAATTTTTGTCCCAATTTTCAG	Additional type III plasmid sequencing
W15	CAAATTACTGCTATATATTCAGGC	Additional type III plasmid sequencing
W16	TTAAATTTTATTATGAAGCAGGACG	Additional type III plasmid sequencing
W19	CTACTTTAATAATTGAAAAAGATGG	Additional type III plasmid sequencing
PS566	CTCCCTTATTTGATAAATATAATACTCTAAC	Additional type III plasmid sequencing
W125	ATTTAGCGTCTATAGTAAAGGTG	Construction of pAV71
PS557	CCATGCACCGATTAAAAATAAAGCGGCACCGCCTGAATATATAGCAG	Construction of pAV71
PS556	CTGCTATATATTCAGGCGGTGCCGCTTTATTTTTAATCGGTGCATGG	Construction of pAV71
W762	CAAATCACCTTTACTATAGACGC	Construction of pAV71
W1169	CCGATTAAAAATAAAGCTGCACCGCCTGAATATATAGCAGTAATTTG	Construction of pWJ291
W1170	TATTCAGGCGGTGCAGCTTTATTTTTAATCGGTGCATGGGATG	Construction of pWJ291
JW398	GAACGTCTCATAATTTGAAATAGTAGCTTTTTTAGTGTT	Construction of JW233's type III plasmid
JW399	TTTAAACACTAAAAAAGCTACTATTTCAAATTATGAGAC	Construction of JW233's type III plasmid
JW443	CTTTAGGTAACTCATAAGTGAATGGTTG	Sequencing of the ϕ11γ2 deletion
JW444	GCCACTCTGTTAAATCAGTAACTTTG	Sequencing of the ϕ11γ2 deletion
oGG387	GTGTTGAACTTGAACAAAGTCAC	Sequencing of the ϕ11γ2β1 deletion
oGG388	TCATCAACTTTTATCCACGAGTC	Sequencing of the ϕ11γ2β1 deletion
AV205	ATACAAAACCACATACCTATCAACGTGATGAGCTTATTGGGG	Construction of pAV43

Table 3-4. Oligonucleotides used in Chapter 3 (continued).

AV202	GAAGTATATAAATCATCAGTACAAAGGTTTACGTCCTGTTGAATCTTTG	Construction of pAV43
AV203	GATGATTTATATACTTCGGCATACGTAGTCGTTGAGGCAGAGAGTA	Construction of pAV43
AV185	CTGTTATGTGGTTATCGATTCCTAGCTTGTATGTCTGCGC	Construction of pAV43
AV186	AATCGATAACCACATAACAGTCATAAAAC	Construction of pAV43
AV204	ATAGGTATGTGGTTTTGTATTGGAAT	Construction of pAV43
oGG38	AAGATAAAGAATTTGCTCAAGACG	ϕNM4γ4α2 sequence verification
oGG40	ACCATTAAAACTCGTCATTCTTTC	ϕNM4γ4α2 sequence verification
oGG297	GCAAAATTCATATAAACAAATTAAAGAGGGTTATAATGAACG	Construction of pE194-*tyc1*
oGG298	CCCTCTTTAATTTGTTTATATGAATTTTGCTTATTAACGATTC	Construction of pE194-*tyc1*
oGG315	AGTGATCGTTAAATTTTAATCAAATATGTTCATAGCTTGATG	Construction of pGG90
oGG316	CGCATCCGATTGCAGCGATGCAGGTCTTTCTGTTGGC	Construction of pGG90
oGG314	AACATATTTGATTAAAATTTAACGATCACTCATCATGTTC	Construction of pGG90
oGG317	GAAAGACCTGCATCGCTGCAATCGGATGCGATTATTG	Construction of pGG90
oGG192	TCTACTTAATCTGATAAGTGAGC	ϕNM1-ErmR2 sequence verification
oGG191	GAAGCTTTAGCTTTGCAGTGG	ϕNM1-ErmR2 sequence verification
W277	CTGTAATAGACATCGTTCGCAG	TB4::ϕ11 *attL* junction verification
oGG342	CATCTTAAAGGAGACATAACAAATG	ϕ11-specific amplification
oGG343	CCAGCTCTCGCTCACATCTG	ϕ11-specific amplification
oGG338	CTTGTTCAAATTACTTTCGTCTG	Multiplex PCR for SaPI1*tst*::*tetM*
oGG339	CTTAAAATTATTCGTTGATGCAGG	Multiplex PCR for SaPI1*tst*::*tetM*
oGG340	CCTTGAACTTCCTTGATAATCTG	Multiplex PCR for SaPI1*tst*::*tetM*
oGG341	ATTCTCACGTTACTGAACAGATG	Multiplex PCR for SaPI1*tst*::*tetM*
oGG6	TACCCTAGTTAACGTCTCTTG	ϕNM1 prophage genotyping
oGG7	GATATCAACTTGTAGTGCATCG	ϕNM1 prophage genotyping
oGG10	GCTGACTTACAAGAAGGTGGAC	ϕNM4 prophage genotyping
oGG11	GTTGTAATTGGATTAAATTCAGTC	ϕNM4 prophage genotyping
JW809	GTTGTACAAGGGTTACAATTCTTAATGC	Construction of pGG170 and pGG172
JW810	CCATCTTCATCAGCTGACATTAGAG	Construction of pGG170 and pGG172
oGG467	AATAACTCTAATGTCAGCTGATGAAGATGGGTTTACGACCGTTAAAAGCAAATG	Construction of pGG170
oGG468	AGTTCTGATCAACGTACATCGTAATTAGAATACGATTGCAAG	Construction of pGG170
oGG469	TTCTAATTACGATGTACGTTGATCAGAACTTAACAAGGAGG	Construction of pGG170
oGG470	TTGCATTAAGAATTGTAACCCTTGTACAACGAAAACCCTTTAACTTTTGATGTTG	Construction of pGG170
oGG471	AATAACTCTAATGTCAGCTGATGAAGATGGTTCAACTACATATTTGTAGCCTTC	Construction of pGG172
oGG472	AATAAACGGAGATGCCATGAATAAAATACTCGCTAACGGTTC	Construction of pGG172
oGG473	GTATTTTATTCATGGCATCTCCGTTTATTTTTAGTGCTATAC	Construction of pGG172
oGG474	TTGCATTAAGAATTGTAACCCTTGTACAACGCTTTCAAAAATTCCATAAAAGAAATC	Construction of pGG172
W1250	TTCAAAGAGTTGGTAGCTCAGAG	pGG170/pGG172 homology arm sequencing
oGG455	CATACCCATTCTAGGTTCAGTTC	pGG170 homology arm sequencing
W1245	GACAAAAATCACCTTGCGCTAATGCTCTGTTACAGCTGTTAGATTATGAAAGCCGATG	Sequencing of the *repA*(ts) region
W354	GTTTTTCAAAATCTGCGGTTGCG	Sequencing of the *repA*(ts) region
B266	TGGAATAATAGTAATATTATACAAAATGGA	Sequencing of the *repA*(ts) region
W653	ATTTACCGCTATCTTTACAGGTAC	Sequencing of the *repA*(ts) region
oGG509	TGGAAAGAAAAATTAGAGAGTTTGGGCGTATCTATGGC	Construction of pGG170ts and pGG172ts
oGG510	TACGCCCAAACTCTCTAATTTTTCTTTCCAATCATTAGGAATTG	Construction of pGG170ts and pGG172ts
oGG32	GTATTTTAAGGGTTGCAATTACG	Verification of pGG172ts integrants
oGG33	GTATCTGTTTTCAACTACATATTTG	Sequencing of GWG7
oGG475	AATAACTCTAATGTCAGCTGATGAAGATGGCGTAGTAAATAGATGTGTAATAAATG	Sequencing of GWG9
oGG31	TACCATCAACAGTTAAAGACAATG	Verification of ϕNM1*ind*- escaper plaques

3.8 Materials and methods used in Chapter 3

Bacterial strains and growth conditions.

Unless otherwise noted, cultivation of RN4220 (Nair et al., 2011) or TB4 (Bae et al., 2006) and their derivatives was carried out in sterile 15 or 50 ml conical-bottom Falcon tubes (Corning) containing Difco™ TSB liquid media from BD (volume ≤ 14% tube capacity), or on sterile petri dish plates containing Difco™ TSA solid media from BD. Media was supplemented with erythromycin (Erm) at 10 μg/ml or tetracycline (Tet) at 5 μg/ml only when streaking out single colonies of marked TB4 lysogens containing pE194 (Horinouchi and Weisblum, 1982b) or pT181 (Khan and Novick, 1983) plasmids, respectively, and during the construction of these strains as needed. When cultivating Tet^R strains that harbor the SaPI1*tst*::*tetM* element in their chromosome, media was supplemented with tetracycline at 2.5 μg/ml instead of 5 μg/ml. All media was supplemented with chloramphenicol (Cm) at 10 μg/ml when cultivating strains with CRISPR-Cas plasmids. Unless otherwise noted, plates were incubated at 37 °C for 12-18 hours and then stored at 4 °C for up to a week. Overnight liquid cultures were inoculated from single colonies (biological replicates), grown for 12-16 hours at 37 °C with shaking, and then stored at 4 °C for 3 hours or less before briefly vortexing and subculturing at a 1:100 dilution as needed. Where applicable, subcultures were treated with Mitomycin C (AG Scientific) to a final concentration of either 0.5, 1.0, or 2.0 μg/ml, as indicated in the figure legends. When the overnight cultures were to be used in

competition experiments, 24 hours of growth was instead allowed, followed by immediate vortexing, mixing of aliquots, and passaging at a 1:1000 dilution. When streaking out lysogens with targeting spacers, aberrantly large colonies were presumed to have lost conditional tolerance either via genetic inactivation of CRISPR-Cas plasmids or prophage deletion, and not studied further in this work.

DNA preparation and cloning.

Plasmid DNA of *E. coli* DH5*α* was purified from 4-6 ml overnight cultures using plasmid miniprep reagents from Qiagen, according to the manufacturer's protocol. DNA from *S. aureus* RN4220 or TB4 was purified similarly except that 2 ml overnight cultures were used, and cells were treated with 10-15 μl lysostaphin (1 mg ml^{-1}) at 37 °C for 1.5 h immediately after resuspension in P1 buffer. Minipreps were carried out with either Qiagen or EconoSpin columns. For PCR-based cloning procedures, DNA was amplified with Phusion polymerase (Thermo) and purified using Qiagen reagents and EconoSpin columns. When generating amplicons for Sanger sequencing, TopTaq polymerase (Qiagen) was often used for amplicons smaller than 750 bp, and DNA was not necessarily purified after the PCRs.

Electrocompetent RN4220 cells (described previously, (Goldberg et al., 2014)) were used for cloning in *S. aureus*. The pGG79 parent vector is a derivative of pGG3-BsaI, (Goldberg et al., 2014) with the BsaI placeholder sequence at position two trimmed

at its 3' end by one base pair, and a CRISPR repeat inserted immediately downstream. This was accomplished by 'round-the-horn PCR (Moore and Prevelige, 2002) using primers oGG281/oGG282, followed by blunt ligation. The pGG78 plasmid was constructed similarly, except that the pGG-BsaI plasmid was used in place of pGG3-BsaI to template the PCR. Their CRISPR arrays were sequenced by Sanger as described previously (Goldberg et al., 2014), and expected plasmid sizes were verified via analytical digestion during subsequent manipulations. The immune functionality of pGG79 in *S. aureus* was later also confirmed by resistance to infection with ϕNM4γ4α2, a derivative of ϕNM4γ4 (Heler et al., 2015) possessing a target for the *nes* spacer in position 1 of its CRISPR array. Except where noted, all other CRISPR-Cas plasmids with wild type *cas* genes were constructed via scarless insertion of spacers between CRISPR repeats 2 and 3 of the parent vector, pGG79. This was accomplished by restriction digesting the two oppositely oriented BsaI sites and then ligating them with an annealed oligonucleotide pair possessing compatible overhangs and the desired spacer sequence (see Supplementary Table 3). CRISPR array modifications were then verified again as described previously (Goldberg et al., 2014). CRISPR-Cas plasmids with the *gp16* spacer and one or more *cas* gene mutations were also constructed by the annealed oligo cloning method in most cases, and checked similarly, except that mutant BsaI parent vectors were used as backbones for insertion of the spacer. The *csm3** (D32A) mutant vector, pPS95, was constructed via 2-piece Gibson assembly of PCR products

(Gibson et al., 2009) amplified from pGG79 using primers PS153/PS465 and PS154/PS466. The *csm6** (R364A,H369A) mutant vector, pGG99, as well as the *csm3**/*csm6** double mutant vector, pGG89, were also constructed via 2-piece Gibson. For pGG99, the fragment containing the CRISPR array was amplified from pGG79 using primers W852/PS566, while the rest of the backbone including the *csm6** mutations was amplified from pWJ241 (Jiang et al., 2016) using primers PS565/W614. The same was done for pGG89, except that pWJ242 (Jiang et al., 2016) was used for templating the backbone with both *csm3** and *csm6** mutations. To construct the pGG164 *cas10*^{HD} (H14A,D15A) HD domain mutant plasmid harboring the *gp16* spacer, 2-piece Gibson was performed with a CRISPR-containing fragment amplified from pGG141 using primers W852/oGG458, and a backbone fragment amplified from pGG79 using primers oGG457/W614. The pGG167 *cas10*^{DD} (D586A,D587A) palm domain mutant plasmid harboring the *gp16* spacer was constructed similarly, except that the CRISPR-containing fragment was amplified from pGG102, and the backbone fragment was amplified from pGG139. The pGG139 BsaI parent vector contains *cas10* mutations in both the HD (H14A,D15A) and palm polymerase (D586A,D587A) domains, and was itself constructed via 2-piece Gibson of a fragment amplified from pGG79 with W852/oGG425 and a fragment amplified from pAV71 using primers oGG424/W614. Following Gibson assembly of these vectors, the expected plasmid size was confirmed by restriction digest, and regions of interest were sequenced by Sanger using primers

listed in Supplementary Table 3 (Gibson assembly junctions, CRISPR arrays, and in select cases the relevant *cas* nuclease active sites). The pAV71 vector was constructed via 2-piece Gibson of a fragment containing the HD domain mutation amplified from pLM546 (Hatoum-Aslan et al., 2014) with W125/PS557, and a fragment containing the palm domain mutation amplified from pWJ291 with PS556/W762. The pWJ291 vector was constructed via 2-piece Gibson of fragments amplified from pWJ30β (Goldberg et al., 2014) using primer pairs W852/W1169 and W1170/W614. In turn, the pGG141 plasmid was constructed by inserting the *gp16* spacer into the pGG139 vector via the oligo cloning method described above. JW233 is a derivative of RN4220 with a type III CRISPR-Cas plasmid that targets ϕ11's *cI*-like repressor gene. This plasmid was constructed from pGG3-BsaI according to the annealed oligo cloning procedure described previously (Goldberg et al., 2014), using oligos JW398 and JW399 for insertion of the spacer. The pAV43 plasmid was constructed by 3-piece Gibson assembly of two homology fragments amplified from ϕNM4γ4 (AV205/AV202 and AV203/AV185), along with a backbone fragment amplified from pC194 (AV186/AV204). The pE194-*tyc1* plasmid was constructed by 1-piece Gibson assembly of PCR products generated with primers oGG297/oGG298. The presence of an intact assembly junction, including the 5' UTR SNP in *ermC*, was subsequently confirmed by Sanger sequencing with primers oGG191/W235. The pGG90 plasmid was constructed by 2-piece Gibson assembly of PCR products; its integrase homology fragment was amplified from ϕNM1 with

primers oGG315/oGG316, while its backbone fragment including *ermC* homology was amplified from pE194-*tyc1* using primers oGG314/oGG317. Intact assembly junctions were verified by PCR and Sanger sequencing with primers W234/W235.

Chemically competent DH5α cells were used for cloning pWJ327-derived vectors in *E. coli*, each constructed via 3-piece Gibson assembly of PCR products. The pGG170 plasmid contains homology to ϕNM1 with mutations in the P_{cro} promoter, while the pGG172 plasmid contains homology to ϕNM1 with mutations in its *cI*-like repressor. A backbone fragment was amplified from pWJ327 using primers JW809/JW810, while the homology arms were amplified from ϕNM1 using primer pairs oGG467/oGG468 and oGG469/oGG470 for pGG170, or oGG471-oGG472 and oGG473-oGG474 for pGG172. The expected size of the resulting plasmids was confirmed by restriction digest, and homology arm sequences were confirmed by Sanger using primers listed in Supplementary Table 3 (W1250, oGG455, and oGG469 or W1250, oGG471, and oGG473). Unexpectedly, the plasmids failed to transform RN4220. PCR and Sanger sequencing of the region encoding elements required for replication in *S. aureus* at 28 °C revealed a SNP within the temperature-sensitive *repA* gene, resulting in a premature stop codon. The SNP was corrected in each case to create "pGG170ts" or "pGG172ts" via 1-piece Gibson of PCR products amplified from pGG170 or pGG172 using primers oGG509/oGG510, followed by electroporation directly into RN4220 and plating at 28 °C.

Estimation of phage lysate titers.

As needed, phage lysate titers were estimated according to the procedure described previously (Goldberg et al., 2014).

Preparation of transducing lysates.

Overnight culture aliquots (100 ul) of RN4220 or TB4 derivatives harboring a plasmid of interest were infected with 'clear' mutant (lytic) derivatives of either ϕNM1, ϕNM2, ϕNM4, or ϕ11 (MOI ~0.1-10) in HIA soft agar supplemented with $CaCl_2$ at 5mM, and incubated overnight at 37 °C to produce a lysed lawn. Within 48 hours of storage at 4 °C, soft agar was hydrated with 600 µl fresh HIB, scraped and decanted into a 50 ml conical-bottom Falcon tube, supplemented with an additional 600 µl fresh HIB, and then centrifuged at 4303*g* for 8-10 min. Filtered supernatants containing the phage and transducing particles were stored in autoclaved 1.5 ml tubes (Eppendorf) at 4 °C. When the plasmid of interest carried a CRISPR-Cas system, care was taken to lyse its host with a suitable non-targeted phage, among the following: ϕNM1γ6 (Goldberg et al., 2014), ϕNM2γ1, ϕNM1γ6α1, ϕNM4γ4 (Heler et al., 2015), or ϕ11γ2β1. Phages not described previously were procured via CRISPR-assisted editing as outlined below.

CRISPR-assisted editing of phages.

Recombinant phages were isolated by using type III CRISPR-Cas plasmids to counter-select against different target sequences, essentially as described for the isolation of CRISPR-escape mutant phages in Chapter 2. The ϕ11γ2β1 phage is a mutant of ϕ11γ2 isolated during infection of lawns harboring the pGG102 plasmid (*gp16*). After an additional round of counter-selection with pGG102, the presence of a 357 bp deletion encompassing the target site was confirmed by PCR and Sanger sequencing with primers oGG387/oGG388. ϕ11γ2 is a clear mutant of ϕ11 that was isolated during infection of JW233 lawns harboring a type III plasmid targeting its *cI*-like repressor gene. Following plaque re-isolations, a 791 bp deletion was confirmed by PCR and Sanger sequencing with primers JW443/JW444. The ϕNM1γ6α1 phage is a mutant of ϕNM1γ6 that was similarly isolated while infecting lawns harboring the pGG100 plasmid (*gp32**), although the presence of a target site deletion was not confirmed after plaque re-isolations. ϕNM2γ1 is a clear mutant of ϕNM2 that was isolated during infection of lawns harboring the pGG152 type III plasmid targeting its *cI*-like repressor gene, although the presence of a target site deletion was not confirmed after plaque re-isolations. The original two ϕNM2 γ-series escapers of spacer 2B described in Chapter 2 were never analyzed for plaque phenotypes, and were ultimately discarded. For homology-directed editing of phages, we used a variation of phage editing methods described previously (Kiro et al., 2014; Martel and Moineau, 2014). Briefly, phages were

propagated on hosts containing homology plasmids for recombination with phage genomes, and the resultant lysates were used to infect soft agar lawns containing a suitable CRISPR-Cas plasmid that can counter-select against parental genotypes. Homologies were designed such that recombinant phages would lack a target sequence for the type III CRISPR-Cas plasmid used in counter-selection. The ϕNM4γ4α2 phage was created by replacing a 40 bp sequence containing the *gp32* spacer's partially matched target in ϕNM4γ4 with a perfectly matched target sequence for the *nes* spacer (Marraffini and Sontheimer, 2008). ϕNM4γ4 was first propagated on lawns of RN4220 containing the homology plasmid, pAV43, and recombinants were subsequently selected during infection of lawns containing the pGG12 (Goldberg et al., 2014) CRISPR-Cas plasmid. After an additional round of counter-selection with pGG12, a single ϕNM4γ4α2 plaque was isolated, and the presence of its recombinant target was confirmed by Sanger sequencing of PCR products generated with primers oGG38/oGG40. The ϕNM1-Erm^{R2} phage was constructed similarly by first propagating the ϕNM1-Erm^{R} phage (Goldberg et al., 2014) on lawns containing the pGG90 homology plasmid, which includes the *tyc-1* allele SNP in the 5′ UTR of *ermC*, and subsequently selecting for recombinants lacking a nearby 182 bp fragment that encompasses the target for the pGG91 CRISPR-Cas plasmid. The presence of the recombinant sequences in ϕNM1-Erm^{R2} was confirmed by PCR and Sanger sequencing with primers oGG192/oGG191.

Transduction.

Subcultures were grown at 37 °C in TSB supplemented with $CaCl_2$ at 5mM for 1 hour 25 minutes (~0.5-1.0 attenuance, $D_{600\ nm}$), and 990 μl aliquots were subsequently mixed with 10 μl of transducing lysate containing the plasmid or SaPI of interest in autoclaved 1.5 ml tubes (Eppendorf). After 15 minutes of growth at 37 °C, infected cultures were treated with filter-sterilized sodium citrate to a final concentration of 40 mM, and then pelleted by centrifugation at 16100*g* for 2 min with refrigeration (4 °C). When transducing pE194, pT181, or SaPI1*tst*::*tetM*, an additional 1 hour 45 minutes of growth was allowed immediately after treatment with citrate and prior to centrifugation. Following centrifugation, fresh TSB supplemented with sodium citrate (40 mM) was used to resuspend and dilute pellets for plating on solid media supplemented with both sodium citrate at 20 mM and antibiotics to select for the transductants. To avoid residual free phage, an additional re-streak was performed on solid media in the presence of citrate (20 mM) and antibiotics to select for the strain's plasmid(s).

Streak-test (phage-sensitivity) assays.

Throughout cloning and strain construction procedures, or as otherwise needed, clones were readily checked for sensitivity to a phage of interest according to the streak method previously performed with ϕNM2 (Goldberg et al., 2014), except that clear mutants described above were used in place of their parental temperate phages to

facilitate scoring, and chloramphenicol-supplemented TSA without added $CaCl_2$ was usually used in place of HIA.

Construction of lysogens and derivative strains.

TB4::ϕNM1-Erm^{R2} and TB4::ϕ11 single lysogens were obtained by mixing 100 μl of a TB4 overnight culture with either ϕNM1-Erm^{R2} or ϕ11 (MOI ~0.1-10) in HIA soft agar supplemented with $CaCl_2$ at 5mM, incubating overnight at 37 °C, and re-streaking from the resulting turbid lysate lawns. After an additional re-streak from single colonies (plates were supplemented with erythromycin in the case of TB4::ϕNM1-Erm^{R2}), a clone was selected for each and saved. The TB4::ϕ11 single lysogen was verified further via PCR with primers oGG342 and oGG343 to amplify a specific internal sequence not present in my other phages, and also via PCR with primers oGG191 and W277 to amplify across its integrated *attL* junction. Other TB4::ϕNM1-Erm^{R2} single lysogens were obtained by lysogenizing TB4 derivatives harboring CRISPR-Cas plasmids via the erythromycin selection protocol for lysogenization described in another section below. TB4::ϕNM1 or TB4::ϕNM4 single lysogens were likewise obtained by mixing 100 μl of a TB4 overnight culture with either ϕNM1 or ϕNM4 (MOI ~0.1-10) in HIA soft agar supplemented with $CaCl_2$ at 5mM, incubating overnight at 37 °C, and re-streaking from the resulting turbid lysate lawns. After an additional re-streak from single colonies, a clone was selected for each, and immunity groups were confirmed by assaying for

sensitivity to clear mutants (ϕNM1γ6 (Goldberg et al., 2014) or ϕNM4γ4 (Heler et al., 2015)). Both clones were also typed via PCR using ϕNM1- or ϕNM4-specific primer pairs, oGG6/oGG7 or oGG10/oGG11, respectively. The TB4::ϕNM1+ϕNM4 double lysogen was constructed and typed similarly, except that an overnight culture of TB4::ϕNM4 was mixed with ϕNM1 in soft agar to produce the turbid lysate lawn. ϕNM1 prophage mutants were constructed via CRISPR-assisted genome editing as outlined below. All plasmid-containing derivatives of TB4::ϕNM1, TB4::ϕNM4, or TB4::ϕNM1+ϕNM4 were subsequently generated via transduction.

CRISPR-assisted genome editing of *S. aureus*.

Homology-directed allelic exchange coupled with CRISPR-Cas counter selection was performed essentially as described previously for the pWJ327 allelic replacement system (Modell et al., 2017), with modifications. pWJ327-derived vectors containing homology to ϕNM1 and the prophage mutation of interest were used to transform electrocompetent RN4220 non-lysogens instead of RN4220::ϕ12, and a subsequent transduction step was used to transfer the plasmids from RN4220 to TB4::ϕNM1. Transduction was performed with ϕNM4γ4 as described above, except that transducing lysates were raised and injected at 28 °C, with 2 hours 30 minutes of growth allowed for the recipient subculture. After isolating putative co-integrants via two consecutive re-streaks at 37 °C in the presence of chloramphenicol and verifying them with PCR using

primers W1250/oGG7 (pGG170ts) or W1250/oGG32 (pGG172ts), clones were inoculated overnight in plain media at 37 °C instead of 28 °C. Subcultures were grown at 28 °C in plain media supplemented with $CaCl_2$ at 5mM to obtain logarithmic-phase cultures for treatment with the pWJ326 phagemid, but treated cultures were plated at 30 °C on media containing erythromycin after 3 hours of growth at 30 °C rather than 1 hour. An additional re-streak at 30 °C was performed in the presence of erythromycin before clones were checked by PCR and Sanger sequencing with primers oGG455/oGG470 (pGG170ts) or oGG471/oGG474 (pGG172ts) for the presence of the desired mutation (ϕNM1^{Pcro} or ϕNM1^{ind-}, respectively). Clones which had also lost the integrated pWJ327 amplicon (again determined by PCR with primers W1250/oGG7 or W1250/oGG32) were inoculated overnight in plain media but grown at 42 °C instead of 37 °C in order to cure the strain of pWJ326. Dilutions were plated on plain media at 37 °C, and single colonies were replica-plated at 37 °C on both plain and erythromycin-supplemented media to confirm plasmid loss. After re-streaking an erythromycin-sensitive clone on plain media, a PCR amplicon spanning the entire homology region was procured with primers oGG33/oGG7 (pGG170ts) or oGG475/oGG32 (pGG172ts), and sequenced by Sanger using additional primers listed in Supplementary Table 3. The clone was also re-streaked on media supplemented with chloramphenicol as a final check for sensitivity to the antibiotic; and, checked for sensitivity to ϕNM1γ6 to ensure that lysogenic immunity had not been lost. The pGG170ts- and pGG172ts-derived strains were

renamed GWG7 and GWG9, respectively. TSB and TSA media were used in place of BHI throughout. The pWJ326 phagemid lysate used in this work was procured using the method described previously (Modell et al., 2017).

Enumeration of plaque-forming units liberated from lysogenic cultures.

'Spontaneous' particle release from single lysogens was measured essentially as described previously (Goldberg et al., 2014), except that filtered supernatants were collected immediately after 12 hours of growth overnight (or as otherwise noted in the figure legend). For quantification of particles released from MMC-induced cultures, overnight cultures were grown for 15 hours and subcultures were grown for 1 hour 15 minutes (~0.3-0.6 attenuance, $D_{600\ nm}$) before treatment with MMC (or as otherwise noted in the figure legend). Following 4 hours of growth post-treatment, subcultures were pelleted by centrifugation at 4303*g* for 6 min, and filtered supernatants were collected for serial dilution and spotting on lawns of TB4 harboring the CRISPR-Cas parent vector. Where applicable, filtered supernatants derived from double lysogens were spotted in parallel on TB4::ϕNM4 and TB4::ϕNM1 lawns harboring the CRISPR-Cas parent vector in order to quantify the fraction of PFU derived from ϕNM1 or ϕNM4, respectively.

Plate reader high-resolution growth curves.

Biological replicates were subcultured as described above, except that 200ul culture volumes were used in 96-well microplates, and incubation at 37 °C with shaking was performed with an Infinite M200 PRO plate reader (TECAN) measuring attenuance ($D_{600\ nm}$) every 10 minutes. Following 70 minutes of growth, subcultures were treated with MMC to a final concentration of 2.0 μg/ml. For each biological replicate, an additional subculture was performed in parallel to monitor growth in the absence of MMC treatment.

Pairwise competition assays with marked lysogens.

Aliquots of overnight cultures were mixed 1:1 by volume and then diluted for passaging and selective plating on tetracycline (5 μg/ml) and chloramphenicol or erythromycin (5 μg/ml) and chloramphenicol. Passaging and plating was repeated in this manner every 24 hours for 6 days, and relative frequencies were calculated at each interval as the number of Tet^R+Cm^R colony forming units (CFU) divided by the sum of Erm^R+Cm^R CFU plus Tet^R+Cm^R CFU.

Batch competition of unmarked lysogens for deep sequencing.

Aliquots of overnight cultures were mixed in equal proportions by volume and then diluted for passaging as described above. The remaining culture (~2 ml) was pelleted by

centrifugation at 4303*g* for 6 min. Supernatants were discarded, and pellets were stored at -20 °C after drying for 10 min at benchtop. Passaging and storage of pellets was repeated in this manner every 24 hours for 3 days.

Preparation of DNA for deep sequencing.

Frozen pellets were thawed on ice for 20 minutes and then miniprepped as described above. 100 ng of miniprepped DNA from each sample was used as the template for barcoded PCRs to amplify a region of the CRISPR array that fully spans spacer position 2, using forward and reverse primers with identical barcode pairs. Barcoded PCR products were gel purified and then pooled at roughly equal proportions by normalizing to the least-concentrated NanoDrop Spectrophotometer (Thermo) reading. Library preparation was subsequently carried out on the pooled sample using a TruSeq Nano DNA LT kit (Illumina).

Deep sequencing and analysis.

Single-read deep sequencing (321 cycles) was performed on an Illumina MiSeq essentially as described in the manufacturer's protocol for low-complexity amplicon sequencing with the v3 reagent kit (~25% PhiX spike-in). Using a custom Python script, output files (.fastq) were parsed using the Biopython package for Python, and reads containing the expected sequences with 100% identity (either forward or reverse) were

tallied for each barcode set. Relative frequency datapoints are plotted as the number of reads containing a particular barcode and the perfect or partially matching spacer of interest divided by the sum of that count plus reads containing the fully mismatched spacer with the same barcode.

Efficiency of plaquing.

Assays were performed as described previously (Goldberg et al., 2014), except that the sensitive control lawns were prepared with a TB4 derivative harboring the pGG79 (non-targeting) parent vector.

Lysogenization with *ermC(tyc-1)*-marked ϕNM1.

Lysogenization was performed essentially as described previously (Goldberg et al., 2014), except that overnight cultures were always grown for 14 hours, TSB broth was used in place of HIB, multiplicity of infection was ~1, and incubation times were reduced to 10 minutes on ice and 20 minutes at 37 °C. Following the 20 minute incubation, cultures were supplemented and plated with citrate as described for the transduction protocol above, except that chloramphenicol was maintained in all media and plates were also supplemented with erythromycin (5 μg/ml) to quantify acquisition of the marked prophage. The ϕNM1-Erm^{R2} phage used for infections is a derivative of ϕNM1-Erm^{R} (ref. (Goldberg et al., 2014)) with a modified *ermC* 5′ UTR that includes the

SNP found in the *tyc-1* allele previously reported (Gryczan et al., 1980) to improve constitutive expression of the cassette's gene product in *B. subtilis*. To estimate the concentration of total recipients harboring CRISPR-Cas systems, serial dilutions of the untreated overnight cultures were plated in the presence of chloramphenicol alone.

Mitomycin C survival assays.

Overnight cultures were grown for 14 hours and then subcultured for 1 hour 15 minutes (~0.3-0.6 attenuance, $D_{600\ nm}$). Subcultures were split into an untreated or treated tube where MMC was added to a final concentration of 0.5 μg/ml. After an additional 45 minutes of growth, treated and untreated cultures were vortexed, diluted, and plated to quantify CFUs.

CHAPTER 4: PERSPECTIVES ON TOLERANCE TO FOREIGN ELEMENTS THROUGH THE LENS OF PROKARYOTIC IMMUNE SYSTEMS

Although an organism may be defined by its genomic content in a strict genetic sense, the classical evolutionary definition is concerned most with its phenotype (Doolittle and Sapienza, 1980). The descriptive power of a genome sequence is therefore limited, in part, to the extent that phenotypes can vary independently of a particular genotype. Presumably, phenotypic variability can more predictably be accounted for, given a better understanding of how multicellular organisms are shaped by environmental factors, including the microbiota that colonizes them (Amaral et al., 2008; Backhed et al., 2004; Li et al., 2008; Rakoff-Nahoum et al., 2004; Stappenbeck et al., 2002). Collectively, the genetic repertoire of a multicellular organism, along with the microbiomes of its resident microbiota, can be conceptualized as a metagenome. Hence, there is increasing effort to define metagenomic information, such as microbiomes, that could be correlated with organismal phenotypes (Bruls and Weissenbach, 2011).

Similar in a sense to how microbiome analyses have uncovered substantial metagenomic variation among multicellular organisms within a species, comparative genome analyses of prokaryotes have revealed a surprising degree of genomic variability among related strains that were traditionally classified as members of the same species (Murray et al., 2001; Tettelin et al., 2005). Accordingly, a modern view of

prokaryotic genomes has emerged that distinguishes 'core' genome sequences, which are common to most or all strains of a particular group, from the 'accessory' genome sequences, which are not universally sampled (Lan and Reeves, 2000; Mira et al., 2010). Accessory sequences typically comprise no more than 10–20% of a given genome (Makarova et al., 2014), but their gene diversity across strains can be substantially greater than that observed for core sequences (Koonin and Wolf, 2008). Furthermore, it has been found that prophages, plasmids and various predicted MGEs are usually associated with accessory rather than core sequences of the genome (Cortez et al., 2009; den Bakker et al., 2013; Di Nocera et al., 2011; Lindsay and Holden, 2006; Ozer et al., 2014). Thus, by analogy to mammal-associated microbiomes, the accessory genomes of prokaryotes may represent a transient repository for horizontally derived foreign genetic information that can contribute to adaptability. However, as both microbiomes and prokaryotic accessory genomes are also liable to harbor parasitic elements, the host organism may employ its selective defenses to keep these elements in check while participating in symbiotic interactions or HGT. In turn, the flux of genetic information in prokaryotic genomes can be selectively moderated by both CRISPR–Cas and R–M systems through their influence on MGEs and accessory genomic content encountered via HGT. Although the selectivity of these systems is tied to their resistance mechanisms, it can be reinforced by tolerance in certain contexts, as I explain below.

4.1 CRISPR-Cas

As outlined in **Chapter 1**, a CRISPR–Cas system resists diverse MGEs according to its spacer content, and this can have evolutionary consequences for lineages that acquire different spacers. For example, cells that acquire immunity to parasitic elements can gain a selective advantage (Barrangou et al., 2007), whereas those that target favorable or 'self' elements can be put at a disadvantage and potentially be lost from the population (Bikard et al., 2014; Bikard et al., 2012; Jiang et al., 2013b; Marraffini and Sontheimer, 2008). However, just as the risk imposed by particular microorganisms can be niche- or context-dependent within a mammalian host (Hube, 2004; Stecher et al., 2007; von Eiff et al., 2001), the fitness contributions associated with a particular MGE are not always clear-cut in the prokaryotic domain. Temperate phages are a prime example (**Figure 1-5**). Although toxic during lytic infections, they can be maintained as prophages in an alternative, lysogenic state that does not necessarily reduce their host's fitness and can even be advantageous (Brüssow et al., 2004; Dykhuizen et al., 1978; Edlin et al., 1977). Notwithstanding, indiscriminate CRISPR–Cas targeting of temperate phage DNA compromises the stability of the lysogenic state in addition to preventing lytic infection (Edgar and Qimron, 2010; Goldberg et al., 2014). This is because the transition between the two states does not involve genetic alteration of the phage DNA sequence (Lwoff, 1953) but instead results from changes in its transcriptional activity

within the host; in fact, most of a lambdoid prophage's genes are typically repressed during lysogeny (Ptashne, 2011).

These properties of lambdoid temperate phages, it would appear, allow a staphylococcal type III CRISPR–Cas system to distinguish between each of their infection states (that is, lytic versus lysogenic) via transcription-dependent targeting (Goldberg et al., 2014). Type III CRISPR–Cas systems only initiate an immune response

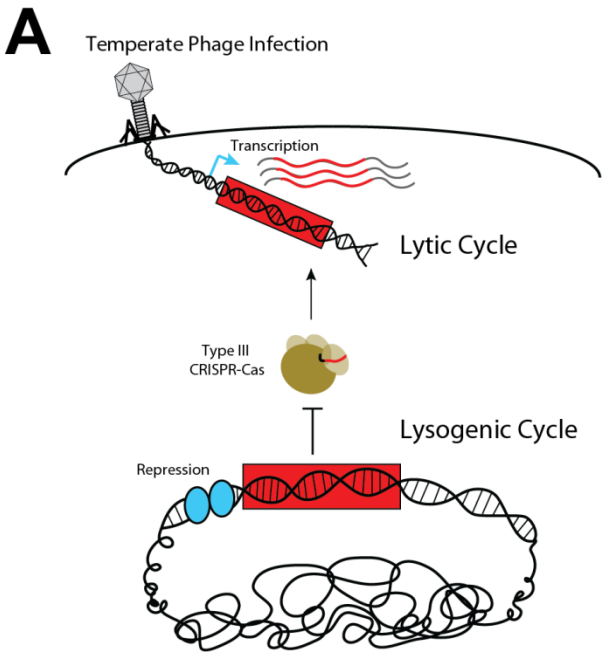


Figure 4-1. Schematic diagram summarizing selectivity associated with conditional tolerance by type III CRISPR-Cas systems.

(A) Transcription-dependent targeting by type III CRISPR-Cas systems is only directed at DNA elements with actively transcribed target sequences. In this example, type III targeting licenses conditional tolerance of a lambdoid temperate phage. During the phage's lytic cycle, when most of its genome is transcribed, type III targeting occurs readily. During its lysogenic cycle, most prophage sequences are repressed in the host chromosome, and the prophage DNA may be spared from type III targeting.

when their target sequences are transcribed (**Figure 4-1**). As such, they can tolerate non-transcribed prophage targets during lysogeny without withdrawing resistance to lytic infection by the same phage. Tolerance, at least for temperate phages, may thus be achieved in a context-dependent, conditional manner, as it is for microbiota in the gut of a healthy mammal (Belkaid and Hand, 2014). This prokaryotic phenomenon could be further viewed as a 'disease tolerance' paradigm (Gozzelino et al., 2012; Medzhitov et al., 2012; Soares et al., 2014), insofar as it averts the 'immunopathology' associated with targeting of prophages in the host chromosome without elimination of the target element. Alternative strategies for disease tolerance, aimed at neutralizing parasite-derived toxins (Playfair et al., 1990; Schofield et al., 2002) rather than the parasites themselves, might also exist in the prokaryotic domain. Past and recent work have demonstrated that type III CRISPR–Cas systems can cleave RNA targets both *in vitro* (Hale et al., 2009; Jiang et al., 2016; Samai et al., 2015; Staals et al., 2014; Tamulaitis et al., 2014; Zhang et al., 2012) and *in vivo* (Hale et al., 2012; Jiang et al., 2016; Samai et al., 2015; Tamulaitis et al., 2014; Zebec et al., 2014) and thus offer the potential to reduce the toxicity associated with certain viral transcripts. It is currently unclear whether DNA degradation can be uncoupled from RNA cleavage during canonical type III targeting circumstances. If this were found to occur, however, cleavage of RNA in the absence of DNA degradation could potentially represent another avenue for tolerating certain types of MGEs. To this end, the potential of the recently discovered type VI systems

might also be explored, since they do not possess any known DNase activities (Shmakov et al., 2017).

4.2 Restriction-Modification

In certain contexts, R–M systems can also tolerate the insertion of foreign DNA into the chromosome. Owing to their intracellular, double-strand cleavage mechanism, REases do not eliminate the DNA of unmodified target elements per se. In fact, it has been shown that fragments of restriction-sensitive DNA encountered through transduction or conjugation can be rescued by recombination with the chromosome (McKane and Milkman, 1995; Milkman et al., 1999; Wood, 1966). This was proposed to contribute to the genomic mosaicism observed among natural isolates of *Escherichia coli* (Milkman and Bridges, 1993). Furthermore, evidence indicates that R–M systems are ineffective at blocking natural transformation with otherwise restriction-sensitive DNA, at least when homologous sequences are introduced (Bron et al., 1980; Cohan et al., 1991; Harris-Warrick and Lederberg, 1978; Lacks and Springhorn, 1984; Trautner et al., 1974). These results have been explained in light of the findings that DNA enters the cell through a single-stranded intermediate during natural transformation (Lacks, 1962; Piechowska and Fox, 1971), which remains stable (Eisenstadt et al., 1975; Morrison and Mannarelli, 1979; Mortier-Barriere et al., 2007) before recombination with a methylated complementary strand (**Figure 4-2**). Interestingly, an R–M system in *Streptococcus*

pneumoniae was found to encode an auxiliary MTase that is upregulated during competence and allows even non-homologous sequences to be protected from restriction during natural transformation (Cerritelli et al., 1989; Johnston et al., 2013a). Known as modification methylase DpnIIB (encoded by *dpnA*), this MTase

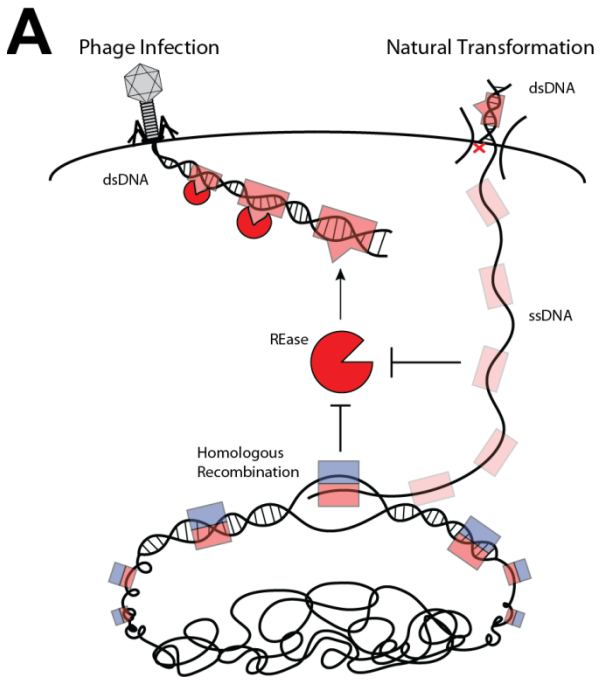


Figure 4-2. Schematic diagram summarizing the selective resistance and tolerance activities associated with Restriction-Modification.

(A) Intracellular restriction endonuclease enzymes (REases) can recognize and inactivate double-stranded DNA (dsDNA) of unmodified phages and other mobile genetic elements but do not recognize single-stranded DNA (ssDNA) intermediates of natural transformation. Successful homologous recombination may result from pairing of unmethylated (red) ssDNA with a methylated (red & blue) homologous sequence in the host chromosome, which promotes rapid modification and tolerance of the newly incorporated strand. The red 'x' represents arbitrary degradation of a donor DNA strand upon uptake of dsDNA during natural transformation.

preferentially methylates single-stranded DNA and may thus promote natural transformation without compromising the ability of the system to resist phages that enter double stranded. In its absence, replication of non-homologous, unmethylated single-stranded DNA that has integrated into the chromosome can give rise to unmethylated double-stranded DNA that is susceptible to the REase (Johnston et al., 2013b). Hence, I'm suggesting that the specialized action of DpnIIB exemplifies another disease tolerance strategy in that it is only required to protect the chromosome from immunopathological damage when foreign, non-homologous sequences are introduced. Non-homologous fragments of DNA that are introduced through phage transduction might also be processed to single strands before recombination. The effect of DpnIIB on transduction efficiency from an unmodified donor should therefore be examined in future work. Finally, it should be emphasized that the tolerance scenario observed with DpnIIB closely mirrors that observed with the type III CRISPR–Cas system, in which a substantial stretch of non-homologous DNA—the prophage—is allowed to integrate into the chromosome.

4.3 General implications and concluding remarks

Surveillance by both R–M and CRISPR–Cas systems can be optimized for selective incorporation of foreign genetic information in a manner that reduces the risks of parasitism and immunopathology. This is similar to the pattern observed for

multicellular organisms, in which an optimal balance of resistance and tolerance must be struck to accommodate commensal microbiota without succumbing to infection or compromising health (Ayres and Schneider, 2012). Mammalian strategies for pathogen resistance have been studied extensively in the field of immunology. Meanwhile, host strategies for tolerance of non-self elements, especially as an alternative or, at least, auxiliary immune function with respect to resistance, have been far less explored in animal models (Ayres and Schneider, 2012; Howick and Lazzaro, 2014; Raberg et al., 2009). Efforts to understand tolerance mechanisms hold the promise of revitalizing our grasp on clinical problems, in which vaccination, antibiotics and other therapeutic interventions aimed at bolstering resistance have fallen short. Alternative therapeutic approaches could be aimed at exploiting or reinforcing tolerance (Soares et al., 2014). Of particular relevance to this discussion, it has been postulated that tolerance strategies can lead to more stable (or at least more homogeneous) evolutionary outcomes(Miller et al., 2006; Roy and Kirchner, 2000; Schafer, 1971) that are distinct from the arms race dynamics that typically result from resistance to pathogens (for further discussion of the latter, see (Brockhurst and Koskella, 2013)). The outcomes observed for CRISPR–Cas targeting of temperate phages in prokaryotes are ostensibly consistent with this notion, as tolerance during lysogeny does not lead to genetic alteration of either the host or its target prophage. In the absence of tolerance, lysogenization was less frequent and only occurred in genetic mutants that had either lost their prophage target sequence or

dismantled their CRISPR–Cas immunity (Goldberg et al., 2014). This latter propensity to abandon CRISPR–Cas immunity (Bikard et al., 2012; Jiang et al., 2013b; Vercoe et al., 2013; Yosef et al., 2011) might in part explain the absence of these systems in about half of sequenced bacteria, especially considering that mechanisms for non-self tolerance have yet to be identified for most of the CRISPR–Cas types. In these cases, other mechanisms that allow for phage defense and resistance to MGEs could be helping these bacteria to survive without CRISPR-Cas systems (Labrie et al., 2010; Ram et al., 2014). Another point to consider is that the abundance of CRISPR-Cas systems within a population could also be diminished by fitness costs associated with CRISPR-Cas immunopathology, even if the systems are intrinsically capable of tolerance. This point was expressly demonstrated in **Chapter 3** for type III-A CRISPR-Cas systems in lysogenic populations of *S. aureus*. **Chapter 3** also established that type III-A systems which provide broad-spectrum immunity to diverse temperate phages can put their host at a greater risk of immunopathology in the event of (poly)lysogenization. At least in part, however, this increased risk was found to be alleviated by the presence of spacer-target mismatches that are expected to occur in natural populations. Analogous immunopathological complications faced by mammalian lineages are dealt with through well-conserved central and peripheral tolerance mechanisms. The latter tolerance mechanisms appear to be particularly crucial during encounters with innocuous non-self elements, including food-derived products and commensal

microorganisms that are highly abundant in the gut (Pabst and Mowat, 2012). Insofar as tolerance strategies can work to curb — or even replace — resistance strategies that would otherwise reduce a particular microbial burden, tolerance may, in turn, influence microbiome compositions and stability. Therefore, in addition to leading us towards novel therapeutics for dealing with the complications of infection and immunopathology (Soares et al., 2014), knowledge of tolerance mechanisms that fulfil these criteria could prove insightful when attempting to develop clinical interventions aimed at manipulating specific members of the commensal microbiota.

REFERENCES

Aanensen, D.M., Feil, E.J., Holden, M.T., Dordel, J., Yeats, C.A., Fedosejev, A., Goater, R., Castillo-Ramirez, S., Corander, J., Colijn, C., *et al.* (2016). Whole-Genome Sequencing for Routine Pathogen Surveillance in Public Health: a Population Snapshot of Invasive Staphylococcus aureus in Europe. MBio *7*.

Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B., Shmakov, S., Makarova, K.S., Semenova, E., Minakhin, L., *et al.* (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science *353*, aaf5573.

Amaral, F.A., Sachs, D., Costa, V.V., Fagundes, C.T., Cisalpino, D., Cunha, T.M., Ferreira, S.H., Cunha, F.Q., Silva, T.A., Nicoli, J.R., *et al.* (2008). Commensal microbiota is fundamental for the development of inflammatory pain. Proc Natl Acad Sci U S A *105*, 2193-2197.

Amitai, G., and Sorek, R. (2016). CRISPR-Cas adaptation: insights into the mechanism of action. Nat Rev Microbiol *14*, 67-76.

Anantharaman, V., Makarova, K.S., Burroughs, A.M., Koonin, E.V., and Aravind, L. (2013). Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra genomic conflicts, defense, pathogenesis and RNA processing. Biol Direct *8*, 15.

Anderson, K.L., Roberts, C., Disz, T., Vonstein, V., Hwang, K., Overbeek, R., Olson, P.D., Projan, S.J., and Dunman, P.M. (2006). Characterization of the Staphylococcus aureus heat shock, cold shock, stringent, and SOS responses and their effects on log-phase mRNA turnover. J Bacteriol *188*, 6739-6756.

Andersson, J.O., Doolittle, W.F., and Nesbo, C.L. (2001). Genomics. Are there bugs in our genome? Science *292*, 1848-1850.

Arber, W., and Linn, S. (1969). DNA modification and restriction. Annu Rev Biochem *38*, 467-500.

Ayres, J.S., and Schneider, D.S. (2012). Tolerance of infections. Annu Rev Immunol *30*, 271-294.

Baba, T., Bae, T., Schneewind, O., Takeuchi, F., and Hiramatsu, K. (2008). Genome sequence of *Staphylococcus aureus* strain Newman and comparative analysis of staphylococcal

genomes: polymorphism and evolution of two major pathogenicity islands. J Bacteriol *190*, 300-310.

Backhed, F., Ding, H., Wang, T., Hooper, L.V., Koh, G.Y., Nagy, A., Semenkovich, C.F., and Gordon, J.I. (2004). The gut microbiota as an environmental factor that regulates fat storage. Proc Natl Acad Sci U S A *101*, 15718-15723.

Bae, T., Baba, T., Hiramatsu, K., and Schneewind, O. (2006). Prophages of *Staphylococcus aureus* Newman and their contribution to virulence. Mol Microbiol *62*, 1035-1047.

Bae, T., Banger, A.K., Wallace, A., Glass, E.M., Aslund, F., Schneewind, O., and Missiakas, D.M. (2004). Staphylococcus aureus virulence genes identified by bursa aurealis mutagenesis and nematode killing. Proc Natl Acad Sci U S A *101*, 12312-12317.

Bae, T., and Schneewind, O. (2006). Allelic replacement in Staphylococcus aureus with inducible counter-selection. Plasmid *55*, 58-63.

Bailone, A., Levine, A., and Devoret, R. (1979). Inactivation of prophage lambda repressor in vivo. J Mol Biol *131*, 553-572.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709-1712.

Barrangou, R., and Marraffini, L.A. (2014). CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. Mol Cell *54*, 234-244.

Beguin, P., Charpin, N., Koonin, E.V., Forterre, P., and Krupovic, M. (2016). Casposon integration shows strong target site preference and recapitulates protospacer integration by CRISPR-Cas systems. Nucleic Acids Res *44*, 10367-10376.

Belkaid, Y., and Hand, T.W. (2014). Role of the microbiota in immunity and inflammation. Cell *157*, 121-141.

Bertani, G., and Weigle, J.J. (1953). Host controlled variation in bacterial viruses. J Bacteriol *65*, 113-121.

Bikard, D., Euler, C.W., Jiang, W., Nussenzweig, P.M., Goldberg, G.W., Duportet, X., Fischetti, V.A., and Marraffini, L.A. (2014). Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. Nat Biotechnol *32*, 1146-1150.

Bikard, D., Hatoum-Aslan, A., Mucida, D., and Marraffini, L.A. (2012). CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. Cell Host Microbe *12*, 177-186.

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L.A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res *41*, 7429-7437.

Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology *151*, 2551-2561.

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

Bondy-Denomy, J., Pawluk, A., Maxwell, K.L., and Davidson, A.R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature *493*, 429-432.

Brockhurst, M.A., and Koskella, B. (2013). Experimental coevolution of species interactions. Trends Ecol Evol *28*, 367-375.

Brodt, A., Lurie-Weinberger, M.N., and Gophna, U. (2011). CRISPR loci reveal networks of gene exchange in archaea. Biology direct *6*, 65.

Bron, S., Luxen, E., and Trautner, T.A. (1980). Restriction and modification in B. subtilis: the role of homology between donor and recipient DNA in transformation and transfection. Mol Gen Genet *179*, 111-117.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960-964.

Bruls, T., and Weissenbach, J. (2011). The human metagenome: our other genome? Hum Mol Genet *20*, R142-148.

Brüssow, H., Canchaya, C., and Hardt, W.D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol Mol Biol Rev *68*, 560-602.

Bzymek, M., and Lovett, S.T. (2001). Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. Proc Natl Acad Sci U S A *98*, 8319-8325.

Cady, K.C., Bondy-Denomy, J., Heussler, G.E., Davidson, A.R., and O'Toole, G.A. (2012). The CRISPR/Cas Adaptive Immune System of *Pseudomonas aeruginosa* Mediates Resistance to Naturally Occurring and Engineered Phages. J Bacteriol *194*, 5728-5738.

Cady, K.C., White, A.S., Hammond, J.H., Abendroth, M.D., Karthikeyan, R.S., Lalitha, P., Zegans, M.E., and O'Toole, G.A. (2011). Prevalence, conservation and functional analysis of Yersinia and Escherichia CRISPR regions in clinical Pseudomonas aeruginosa isolates. Microbiology *157*, 430-437.

Canchaya, C., Proux, C., Fournous, G., Bruttin, A., and Brussow, H. (2003). Prophage genomics. Microbiol Mol Biol Rev *67*, 238-276, table of contents.

Cao, L., Gao, C.H., Zhu, J., Zhao, L., Wu, Q., Li, M., and Sun, B. (2016). Identification and functional study of type III-A CRISPR-Cas systems in clinical isolates of Staphylococcus aureus. Int J Med Microbiol *306*, 686-696.

Cao, R., Zeaki, N., Wallin-Carlquist, N., Skandamis, P.N., Schelin, J., and Radstrom, P. (2012). Elevated enterotoxin A expression and formation in Staphylococcus aureus and its association with prophage induction. Appl Environ Microbiol *78*, 4942-4948.

Carte, J., Wang, R., Li, H., Terns, R.M., and Terns, M.P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. Genes Dev *22*, 3489-3496.

Casjens, S. (2003). Prophages and bacterial genomics: what have we learned so far? Mol Microbiol *49*, 277-300.

Casjens, S.R., and Hendrix, R.W. (2015). Bacteriophage lambda: Early pioneer and still relevant. Virology *479-480*, 310-330.

Cerritelli, S., Springhorn, S.S., and Lacks, S.A. (1989). DpnA, a methylase for single-strand DNA in the Dpn II restriction system, and its biological function. Proc Natl Acad Sci U S A *86*, 9223-9227.

Chakraborty, S., Snijders, A.P., Chakravorty, R., Ahmed, M., Tarek, A.M., and Hossain, M.A. (2010). Comparative network clustering of direct repeats (DRs) and cas genes confirms

the possibility of the horizontal transfer of CRISPR locus among bacteria. Mol Phylogenet Evol *56*, 878-887.

Charpentier, E., Richter, H., van der Oost, J., and White, M.F. (2015). Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. FEMS Microbiol Rev *39*, 428-441.

Chayot, R., Montagne, B., Mazel, D., and Ricchetti, M. (2010). An end-joining repair mechanism in Escherichia coli. Proc Natl Acad Sci U S A *107*, 2141-2146.

Chen, J., Carpena, N., Quiles-Puchalt, N., Ram, G., Novick, R.P., and Penades, J.R. (2015a). Intra- and inter-generic transfer of pathogenicity island-encoded virulence genes by cos phages. ISME J *9*, 1260-1263.

Chen, J., and Novick, R.P. (2009). Phage-mediated intergeneric transfer of toxin genes. Science *323*, 139-1341.

Chen, J., Ram, G., Penades, J.R., Brown, S., and Novick, R.P. (2015b). Pathogenicity island-directed transfer of unlinked chromosomal virulence genes. Mol Cell *57*, 138-149.

Chopin, M.C., Chopin, A., and Bidnenko, E. (2005). Phage abortive infection in lactococci: variations on a theme. Curr Opin Microbiol *8*, 473-479.

Christie, G.E., and Dokland, T. (2012). Pirates of the Caudovirales. Virology *434*, 210-221.

Christie, G.E., Matthews, A.M., King, D.G., Lane, K.D., Olivarez, N.P., Tallent, S.M., Gill, S.R., and Novick, R.P. (2010). The complete genomes of Staphylococcus aureus bacteriophages 80 and 80alpha--implications for the specificity of SaPI mobilization. Virology *407*, 381-390.

Cohan, F.M., Roberts, M.S., and King, E.C. (1991). The Potential for Genetic Exchange by Transformation within a Natural Population of Bacillus subtilis. Evolution *45*, 1393-1421.

Coleman, D.C., Sullivan, D.J., Russell, R.J., Arbuthnott, J.P., Carey, B.F., and Pomeroy, H.M. (1989). Staphylococcus aureus bacteriophages mediating the simultaneous lysogenic conversion of beta-lysin, staphylokinase and enterotoxin A: molecular mechanism of triple conversion. J Gen Microbiol *135*, 1679-1697.

Corrigan, R.M., and Foster, T.J. (2009). An improved tetracycline-inducible expression vector for Staphylococcus aureus. Plasmid *61*, 126-129.

Cortez, D., Forterre, P., and Gribaldo, S. (2009). A hidden reservoir of integrative elements is the major source of recently acquired foreign genes and ORFans in archaeal and bacterial genomes. Genome Biol *10*, R65.

Croucher, N.J., Harris, S.R., Fraser, C., Quail, M.A., Burton, J., van der Linden, M., McGee, L., von Gottberg, A., Song, J.H., Ko, K.S., *et al.* (2011). Rapid pneumococcal evolution in response to clinical interventions. Science *331*, 430-434.

Croucher, N.J., Mostowy, R., Wymant, C., Turner, P., Bentley, S.D., and Fraser, C. (2016). Horizontal DNA Transfer Mechanisms of Bacteria as Weapons of Intragenomic Conflict. PLoS Biol *14*, e1002394.

Cumby, N., Davidson, A.R., and Maxwell, K.L. (2012). The moron comes of age. Bacteriophage *2*, 225-228.

Damian, R.T. (1964). Molecular Mimicry: Antigen Sharing by Parasite and Host and Its Consequences. The American Naturalist *98*, 129-149.

Daniel, A., Bonnen, P.E., and Fischetti, V.A. (2007). First complete genome sequence of two *Staphylococcus epidermidis* bacteriophages. J Bacteriol *189*, 2086-2100.

Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K., and Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat Commun *3*, 945.

Datta, S., Costantino, N., Zhou, X., and Court, D.L. (2008). Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. Proc Natl Acad Sci U S A *105*, 1626-1631.

de Haas, C.J., Veldkamp, K.E., Peschel, A., Weerkamp, F., Van Wamel, W.J., Heezius, E.C., Poppelier, M.J., Van Kessel, K.P., and van Strijp, J.A. (2004). Chemotaxis inhibitory protein of Staphylococcus aureus, a bacterial antiinflammatory agent. J Exp Med *199*, 687-695.

Dearborn, A.D., and Dokland, T. (2012). Mobilization of pathogenicity islands by Staphylococcus aureus strain Newman bacteriophages. Bacteriophage *2*, 70-78.

den Bakker, H.C., Desjardins, C.A., Griggs, A.D., Peters, J.E., Zeng, Q., Young, S.K., Kodira, C.D., Yandava, C., Hepburn, T.A., Haas, B.J., *et al.* (2013). Evolutionary dynamics of the accessory genome of Listeria monocytogenes. PLoS One *8*, e67511.

Deng, L., Garrett, R.A., Shah, S.A., Peng, X., and She, Q. (2013). A novel interference mechanism by a type IIIB CRISPR-Cmr module in *Sulfolobus*. Mol Microbiol *87*, 1088-1099.

Depardieu, F., Didier, J.P., Bernheim, A., Sherlock, A., Molina, H., Duclos, B., and Bikard, D. (2016). A Eukaryotic-like Serine/Threonine Kinase Protects Staphylococci against Phages. Cell Host Microbe *20*, 471-481.

Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J Bacteriol *190*, 1390-1400.

Di Nocera, P.P., Rocco, F., Giannouli, M., Triassi, M., and Zarrilli, R. (2011). Genome organization of epidemic Acinetobacter baumannii strains. BMC Microbiol *11*, 224.

Dobrindt, U., Hochhut, B., Hentschel, U., and Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. Nat Rev Microbiol *2*, 414-424.

Dodd, I.B., Shearwin, K.E., and Egan, J.B. (2005). Revisited gene regulation in bacteriophage lambda. Curr Opin Genet Dev *15*, 145-152.

Dohlsten, M., Bjorklund, M., Sundstedt, A., Hedlund, G., Samson, D., and Kalland, T. (1993). Immunopharmacology of the superantigen staphylococcal enterotoxin A in T-cell receptor V beta 3 transgenic mice. Immunology *79*, 520-527.

Doolittle, W.F., and Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. Nature *284*, 601-603.

Dunning Hotopp, J.C. (2011). Horizontal gene transfer between bacteria and animals. Trends Genet *27*, 157-163.

Dupuis, M.E., Villion, M., Magadan, A.H., and Moineau, S. (2013). CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance. Nat Commun *4*, 2087.

Dykhuizen, D., Campbell, J.H., and Rolfe, B.G. (1978). The influences of a lambda prophage on the growth rate of Escherichia coli. Microbios *23*, 99-113.

Edgar, R., and Qimron, U. (2010). The *Escherichia coli* CRISPR system protects from lambda lysogenization, lysogens, and prophage induction. J Bacteriol *192*, 6291-6294.

Edlin, G., Lin, L., and Bitner, R. (1977). Reproductive fitness of P1, P2, and Mu lysogens of Escherichia coli. J Virol *21*, 560-564.

Eisenstadt, E., Lange, R., and Willecke, K. (1975). Competent Bacillus subtilis cultures synthesize a denatured DNA binding activity. Proc Natl Acad Sci U S A *72*, 323-327.

Elmore, J.R., Sheppard, N.F., Ramia, N., Deighan, T., Li, H., Terns, R.M., and Terns, M.P. (2016). Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system. Genes Dev *30*, 447-459.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. Science *297*, 1183-1186.

Erill, I., Campoy, S., and Barbe, J. (2007). Aeons of distress: an evolutionary perspective on the bacterial SOS response. FEMS Microbiol Rev *31*, 637-656.

Estrella, M.A., Kuo, F.T., and Bailey, S. (2016). RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex. Genes Dev *30*, 460-470.

Felsenstein, J. (1974). The evolutionary advantage of recombination. Genetics *78*, 737-756.

Ferrer, M.D., Quiles-Puchalt, N., Harwich, M.D., Tormo-Mas, M.A., Campoy, S., Barbe, J., Lasa, I., Novick, R.P., Christie, G.E., and Penades, J.R. (2011). RinA controls phage-mediated packaging and transfer of virulence genes in Gram-positive bacteria. Nucleic Acids Res *39*, 5866-5878.

Fineran, P.C., Gerritzen, M.J., Suarez-Diez, M., Kunne, T., Boekhorst, J., van Hijum, S.A., Staals, R.H., and Brouns, S.J. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. Proc Natl Acad Sci USA *111*, E1629-1638.

Fischer, S., Maier, L.K., Stoll, B., Brendel, J., Fischer, E., Pfeiffer, F., Dyall-Smith, M., and Marchfelder, A. (2012). An archaeal immune system can detect multiple Protospacer Adjacent Motifs (PAMs) to target invader DNA. J Biol Chem *287*, 33351-33363.

Fuchs, S., Muhldorfer, I., Donohue-Rolfe, A., Kerenyi, M., Emody, L., Alexiev, R., Nenkov, P., and Hacker, J. (1999). Influence of RecA on in vivo virulence and Shiga toxin 2 production in Escherichia coli pathogens. Microb Pathog *27*, 13-23.

Ganguly, T., Das, M., Bandhu, A., Chanda, P.K., Jana, B., Mondal, R., and Sau, S. (2009). Physicochemical properties and distinct DNA binding capacity of the repressor of temperate Staphylococcus aureus phage phi11. FEBS J *276*, 1975-1985.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67-71.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci USA *109*, E2579-2586.

Geissendorfer, M., and Hillen, W. (1990). Regulated expression of heterologous genes in Bacillus subtilis using the Tn10 encoded tet regulatory elements. Appl Microbiol Biotechnol *33*, 657-663.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods *6*, 343-345.

Gill, S.R., Fouts, D.E., Archer, G.L., Mongodin, E.F., Deboy, R.T., Ravel, J., Paulsen, I.T., Kolonay, J.F., Brinkac, L., Beanan, M., *et al.* (2005). Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain. J Bacteriol *187*, 2426-2438.

Godde, J.S., and Bickerton, A. (2006). The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. J Mol Evol *62*, 718 729.

Goerke, C., Koller, J., and Wolz, C. (2006). Ciprofloxacin and trimethoprim cause phage induction and virulence modulation in Staphylococcus aureus. Antimicrob Agents Chemother *50*, 171-177.

Goerke, C., Pantucek, R., Holtfreter, S., Schulte, B., Zink, M., Grumann, D., Broker, B.M., Doskar, J., and Wolz, C. (2009). Diversity of prophages in dominant Staphylococcus aureus clonal lineages. J Bacteriol *191*, 3462-3468.

Goldberg, G.W., Jiang, W., Bikard, D., and Marraffini, L.A. (2014). Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. Nature *514*, 633-637.

Goldberg, G.W., and Marraffini, L.A. (2015). Resistance and tolerance to foreign elements by prokaryotic immune systems - curating the genome. Nat Rev Immunol *15*, 717-724.

Golding, G.R., Bryden, L., Levett, P.N., McDonald, R.R., Wong, A., Wylie, J., Graham, M.R., Tyler, S., Van Domselaar, G., Simor, A.E., *et al.* (2010). Livestock-associated methicillin-resistant *Staphylococcus aureus* sequence type 398 in humans, Canada. Emerg Infect Dis *16*, 587-594.

Gophna, U., Kristensen, D.M., Wolf, Y.I., Popa, O., Drevet, C., and Koonin, E.V. (2015). No evidence of inhibition of horizontal gene transfer by CRISPR-Cas on evolutionary timescales. ISME J *9*, 2021-2027.

Goren, M., Yosef, I., Edgar, R., and Qimron, U. (2012). The bacterial CRISPR/Cas system as analog of the mammalian adaptive immune system. RNA Biol *9*, 549-554.

Gottesman, M.E., and Weisberg, R.A. (2004). Little lambda, who made thee? Microbiol Mol Biol Rev *68*, 796-813.

Gozzelino, R., Andrade, B.B., Larsen, R., Luz, N.F., Vanoaica, L., Seixas, E., Coutinho, A., Cardoso, S., Rebelo, S., Poli, M., *et al.* (2012). Metabolic adaptation to tissue iron overload confers tolerance to malaria. Cell Host Microbe *12*, 693-704.

Griffith, F. (1928). The Significance of Pneumococcal Types. J Hyg (Lond) *27*, 113-159.

Grindstaff, J.L., Brodie, E.D., 3rd, and Ketterson, E.D. (2003). Immune function across generations: integrating mechanism and evolutionary process in maternal antibody transmission. Proc Biol Sci *270*, 2309-2319.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007a). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics *8*, 172.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007b). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res *35*, W52-57.

Grohmann, E., Muth, G., and Espinosa, M. (2003). Conjugative plasmid transfer in gram-positive bacteria. Microbiol Mol Biol Rev *67*, 277-301, table of contents.

Groman, N.B. (1955). Evidence for the active role of bacteriophage in the conversion of nontoxigenic Corynebacterium diphtheriae to toxin production. J Bacteriol *69*, 9-15.

Grundling, A., and Schneewind, O. (2007). Genes required for glycolipid synthesis and lipoteichoic acid anchoring in Staphylococcus aureus. J Bacteriol *189*, 2521-2530.

Gryczan, T.J., Grandi, G., Hahn, J., Grandi, R., and Dubnau, D. (1980). Conformational alteration of mRNA structure and the posttranscriptional regulation of erythromycin-induced drug resistance. Nucleic Acids Res *8*, 6081-6097.

Gudbergsdottir, S., Deng, L., Chen, Z., Jensen, J.V., Jensen, L.R., She, Q., and Garrett, R.A. (2011). Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. Mol Microbiol *79*, 35-49.

Haerter, J.O., and Sneppen, K. (2012). Spatial structure and Lamarckian adaptation explain extreme genetic diversity at CRISPR locus. mBio *3*, e00126-00112.

Hale, C.R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A.M., Glover, C.V., 3rd, Graveley, B.R., Terns, R.M., *et al.* (2012). Essential Features and Rational Design of CRISPR RNAs that Function with the Cas RAMP Module Complex to Cleave RNAs. Mol Cell *45*, 292-302.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell *139*, 945-956.

Han, W., Li, Y., Deng, L., Feng, M., Peng, W., Hallstrom, S., Zhang, J., Peng, N., Liang, Y.X., White, M.F., *et al.* (2017). A type III-B CRISPR-Cas effector complex mediating massive target DNA destruction. Nucleic Acids Res *45*, 1983-1993.

Harris-Warrick, R.M., and Lederberg, J. (1978). Interspecies transformation in Bacillus: sequence heterology as the major barrier. J Bacteriol *133*, 1237-1245.

Hatoum-Aslan, A., Maniv, I., and Marraffini, L.A. (2011). Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. Proc Natl Acad Sci USA *108*, 21218-21222.

Hatoum-Aslan, A., Maniv, I., Samai, P., and Marraffini, L.A. (2014). Genetic Characterization of Antiplasmid Immunity through a Type III-A CRISPR-Cas System. J Bacteriol *196*, 310-317.

Hatoum-Aslan, A., Samai, P., Maniv, I., Jiang, W., and Marraffini, L.A. (2013). A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. J Biol Chem *288*, 27888-27897.

Heler, R., Samai, P., Modell, J.W., Weiner, C., Goldberg, G.W., Bikard, D., and Marraffini, L.A. (2015). Cas9 specifies functional viral targets during CRISPR-Cas adaptation. Nature *519*, 199-202.

Heler, R., Wright, A.V., Vucelja, M., Bikard, D., Doudna, J.A., and Marraffini, L.A. (2017). Mutations in Cas9 Enhance the Rate of Acquisition of Viral Spacer Sequences during the CRISPR-Cas Immune Response. Mol Cell *65*, 168-175.

Helle, L., Kull, M., Mayer, S., Marincola, G., Zelder, M.E., Goerke, C., Wolz, C., and Bertram, R. (2011). Vectors for improved Tet repressor-dependent gradual gene induction or silencing in *Staphylococcus aureus*. Microbiology *157*, 3314-3323.

Heussler, G.E., Cady, K.C., Koeppen, K., Bhuju, S., Stanton, B.A., and O'Toole, G.A. (2015). Clustered Regularly Interspaced Short Palindromic Repeat-Dependent, Biofilm-Specific Death of Pseudomonas aeruginosa Mediated by Increased Expression of Phage-Related Genes. MBio *6*, e00129-00115.

Holt, D.C., Holden, M.T., Tong, S.Y., Castillo-Ramirez, S., Clarke, L., Quail, M.A., Currie, B.J., Parkhill, J., Bentley, S.D., Feil, E.J., *et al.* (2011). A very early-branching *Staphylococcus aureus* lineage lacking the carotenoid pigment staphyloxanthin. Genome Biol Evol *3*, 881-895.

Horinouchi, S., and Weisblum, B. (1982a). Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. J Bacteriol *150*, 815-825.

Horinouchi, S., and Weisblum, B. (1982b). Nucleotide sequence and functional map of pE194, a plasmid that specifies inducible resistance to macrolide, lincosamide, and streptogramin type B antibodies. J Bacteriol *150*, 804-814.

Horton, R.M. (1993). In vitro recombination and mutagenesis of DNA: SOEing together tailor-made genes. Methods Mol Biol *15*, 251-261.

Horvath, P., Coute-Monvoisin, A.C., Romero, D.A., Boyaval, P., Fremaux, C., and Barrangou, R. (2009). Comparative analysis of CRISPR loci in lactic acid bacteria genomes. Int J Food Microbiol *131*, 62-70.

Howick, V.M., and Lazzaro, B.P. (2014). Genotype and diet shape resistance and tolerance across distinct phases of bacterial infection. BMC Evol Biol *14*, 56.

Hube, B. (2004). From commensal to pathogen: stage- and tissue-specific gene expression of Candida albicans. Curr Opin Microbiol *7*, 336-341.

Hynes, A.P., Villion, M., and Moineau, S. (2014). Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages. Nat Commun *5*, 4399.

Iandolo, J.J., Worrell, V., Groicher, K.H., Qian, Y., Tian, R., Kenton, S., Dorman, A., Ji, H., Lin, S., Loh, P., *et al.* (2002). Comparative analysis of the genomes of the temperate bacteriophages phi 11, phi 12 and phi 13 of Staphylococcus aureus 8325. Gene *289*, 109-118.

Jackson, S.A., McKenzie, R.E., Fagerlund, R.D., Kieper, S.N., Fineran, P.C., and Brouns, S.J. (2017). CRISPR-Cas: Adapting to change. Science *356*.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013a). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol *31*, 233-239.

Jiang, W., Maniv, I., Arain, F., Wang, Y., Levin, B.R., and Marraffini, L.A. (2013b). Dealing with the evolutionary downside of CRISPR immunity: bacteria and beneficial plasmids. PLoS Genet *9*, e1003844.

Jiang, W., Samai, P., and Marraffini, L.A. (2016). Degradation of Phage Transcripts by CRISPR-Associated RNases Enables Type III CRISPR-Cas Immunity. Cell *164*, 710-721.

Jin, T., Bokarewa, M., Foster, T., Mitchell, J., Higgins, J., and Tarkowski, A. (2004). *Staphylococcus aureus* resists human defensins by production of staphylokinase, a novel bacterial evasion mechanism. J Immunol *172*, 1169-1176.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Johnson, A.D., Poteete, A.R., Lauer, G., Sauer, R.T., Ackers, G.K., and Ptashne, M. (1981). lambda Repressor and cro--components of an efficient molecular switch. Nature *294*, 217-223.

Johnston, C., Martin, B., Fichant, G., Polard, P., and Claverys, J.P. (2014). Bacterial transformation: distribution, shared mechanisms and divergent control. Nat Rev Microbiol *12*, 181-196.

Johnston, C., Martin, B., Granadel, C., Polard, P., and Claverys, J.P. (2013a). Programmed protection of foreign DNA from restriction allows pathogenicity island exchange during pneumococcal transformation. PLoS Pathog *9*, e1003178.

Johnston, C., Martin, B., Polard, P., and Claverys, J.P. (2013b). Postreplication targeting of transformants by bacterial immune systems? Trends Microbiol *21*, 516-521.

Juhala, R.J., Ford, M.E., Duda, R.L., Youlton, A., Hatfull, G.F., and Hendrix, R.W. (2000). Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J Mol Biol *299*, 27-51.

Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, C., and Siksnys, V. (2016). Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. Mol Cell *62*, 295-306.

Keeling, P.J., and Palmer, J.D. (2008). Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet *9*, 605-618.

Khan, S.A., and Novick, R.P. (1983). Complete nucleotide sequence of pT181, a tetracycline-resistance plasmid from Staphylococcus aureus. Plasmid *10*, 251-259.

Kinnevey, P.M., Shore, A.C., Brennan, G.I., Sullivan, D.J., Ehricht, R., Monecke, S., Slickers, P., and Coleman, D.C. (2013). Emergence of sequence type 779 methicillin-resistant Staphylococcus aureus harboring a novel pseudo staphylococcal cassette chromosome mec (SCCmec)-SCC-SCCCRISPR composite element in Irish hospitals. Antimicrob Agents Chemother *57*, 524-531.

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

Koonin, E.V., and Wolf, Y.I. (2008). Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res *36*, 6688-6719.

Korona, R., and Levin, B.R. (1993). Phage-Mediated Selection and the Evolution and Maintenance of Restriction-Modification. Evolution *47*, 556-575.

Kotewicz, M., Chung, S., Takeda, Y., and Echols, H. (1977). Characterization of the integration protein of bacteriophage lambda as a site-specific DNA-binding protein. Proc Natl Acad Sci U S A *74*, 1511-1515.

Krupovic, M., Makarova, K.S., Forterre, P., Prangishvili, D., and Koonin, E.V. (2014). Casposons: a new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-Cas immunity. BMC Biol *12*, 36.

Kyewski, B., and Klein, L. (2006). A central role for central tolerance. Annu Rev Immunol *24*, 571-606.

Labrie, S.J., Samson, J.E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. Nat Rev Microbiol *8*, 317-327.

Lacks, S. (1962). Molecular fate of DNA in genetic transformation of Pneumococcus. J Mol Biol *5*, 119-131.

Lacks, S.A., and Springhorn, S.S. (1984). Transfer of recombinant plasmids containing the gene for DpnII DNA methylase into strains of Streptococcus pneumoniae that produce DpnI or DpnII restriction endonucleases. J Bacteriol *158*, 905-909.

Lan, R., and Reeves, P.R. (2000). Intraspecies variation in bacterial genomes: the need for a species genome concept. Trends Microbiol *8*, 396-401.

Lee, C.Y., and Buranen, S.L. (1989). Extent of the DNA sequence required in integration of staphylococcal bacteriophage L54a. J Bacteriol *171*, 1652-1657.

Lee, C.Y., Buranen, S.L., and Ye, Z.H. (1991). Construction of single-copy integration vectors for *Staphylococcus aureus*. Gene *103*, 101-105.

Levy, A., Goren, M.G., Yosef, I., Auster, O., Manor, M., Amitai, G., Edgar, R., Qimron, U., and Sorek, R. (2015). CRISPR adaptation biases explain preference for acquisition of foreign DNA. Nature *520*, 505-510.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Li, M., Wang, B., Zhang, M., Rantalainen, M., Wang, S., Zhou, H., Zhang, Y., Shen, J., Pang, X., Zhang, M., *et al.* (2008). Symbiotic gut microbes modulate human metabolic phenotypes. Proc Natl Acad Sci U S A *105*, 2117-2122.

Lindsay, J.A., and Holden, M.T. (2004). Staphylococcus aureus: superbug, super genome? Trends Microbiol *12*, 378-385.

Lindsay, J.A., and Holden, M.T. (2006). Understanding the rise of the superbug: investigation of the evolution and genomic variation of Staphylococcus aureus. Funct Integr Genomics *6*, 186-201.

Lindsay, J.A., Ruzin, A., Ross, H.F., Kurepina, N., and Novick, R.P. (1998). The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in Staphylococcus aureus. Mol Microbiol *29*, 527-543.

Little, J.W., and Michalowski, C.B. (2010). Stability and instability in the lysogenic state of phage lambda. J Bacteriol *192*, 6064-6076.

Livny, J., and Friedman, D.I. (2004). Characterizing spontaneous induction of Stx encoding phages using a selectable reporter system. Mol Microbiol *51*, 1691-1704.

Llosa, M., Gomis-Ruth, F.X., Coll, M., and de la Cruz Fd, F. (2002). Bacterial conjugation: a two-step mechanism for DNA transport. Mol Microbiol *45*, 1-8.

Lwoff, A. (1953). Lysogeny. Bacteriol Rev *17*, 269-337.

Makarova, K.S., Anantharaman, V., Aravind, L., and Koonin, E.V. (2012). Live virus-free or die: coupling of antivirus immunity and programmed suicide or dormancy in prokaryotes. Biol Direct *7*, 40.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H., *et al.* (2015). An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol *13*, 722-736.

Makarova, K.S., Wolf, Y.I., Forterre, P., Prangishvili, D., Krupovic, M., and Koonin, E.V. (2014). Dark matter in archaeal genomes: a rich source of novel mobile elements, defense systems and secretory complexes. Extremophiles *18*, 877-893.

Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2013). Comparative genomics of defense systems in archaea and bacteria. Nucleic Acids Res *41*, 4360-4377.

Manica, A., Zebec, Z., Steinkellner, J., and Schleper, C. (2013). Unexpectedly broad target recognition of the CRISPR-mediated virus defence system in the archaeon *Sulfolobus solfataricus*. Nucleic Acids Res *41*, 10509-10517.

Maniv, I., Jiang, W., Bikard, D., and Marraffini, L.A. (2016). Impact of Different Target Sequences on Type III CRISPR-Cas Immunity. J Bacteriol *198*, 941-950.

Marraffini, L.A. (2013). CRISPR-Cas Immunity against Phages: Its Effects on the Evolution and Survival of Bacterial Pathogens. PLoS Pathog *9*, e1003765.

Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science *322*, 1843-1845.

Marraffini, L.A., and Sontheimer, E.J. (2010a). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet *11*, 181-190.

Marraffini, L.A., and Sontheimer, E.J. (2010b). Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature *463*, 568-571.

Martel, B., and Moineau, S. (2014). CRISPR-Cas: an efficient tool for genome engineering of virulent bacteriophages. Nucleic Acids Res *42*, 9504-9513.

Martinsohn, J.T., Radman, M., and Petit, M.A. (2008). The lambda red proteins promote efficient recombination between diverged sequences: implications for bacteriophage genome mosaicism. PLoS Genet *4*, e1000065.

McGinn, J., and Marraffini, L.A. (2016). CRISPR-Cas Systems Optimize Their Immune Response by Specifying the Site of Spacer Integration. Mol Cell *64*, 616-623.

McGuckin, M.A., Linden, S.K., Sutton, P., and Florin, T.H. (2011). Mucin dynamics and enteric pathogens. Nat Rev Microbiol *9*, 265-278.

McKane, M., and Milkman, R. (1995). Transduction, restriction and recombination patterns in Escherichia coli. Genetics *139*, 35-43.

Medzhitov, R. (2007). Recognition of microorganisms and activation of the immune response. Nature *449*, 819-826.

Medzhitov, R., Schneider, D.S., and Soares, M.P. (2012). Disease tolerance as a defense strategy. Science *335*, 936-941.

Milkman, R., and Bridges, M.M. (1993). Molecular evolution of the Escherichia coli chromosome. IV. Sequence comparisons. Genetics *133*, 455-468.

Milkman, R., Raleigh, E.A., McKane, M., Cryderman, D., Bilodeau, P., and McWeeny, K. (1999). Molecular evolution of the Escherichia coli chromosome. V. Recombination patterns among strains of diverse origin. Genetics *153*, 539-554.

Miller, M.R., White, A., and Boots, M. (2006). The evolution of parasites in response to tolerance in their hosts: the good, the bad, and apparent commensalism. Evolution *60*, 945-956.

Mira, A., Martin-Cuadrado, A.B., D'Auria, G., and Rodriguez-Valera, F. (2010). The bacterial pan-genome:a new paradigm in microbiology. Int Microbiol *13*, 45-57.

Mo, C.Y., Birdwell, L.D., and Kohli, R.M. (2014). Specificity determinants for autoproteolysis of LexA, a key regulator of bacterial SOS mutagenesis. Biochemistry *53*, 3158-3168.

Modell, J.W., Jiang, W., and Marraffini, L.A. (2017). CRISPR-Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. Nature.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol *60*, 174-182.

Moore, S.D., and Prevelige, P.E., Jr. (2002). A P22 scaffold protein mutation increases the robustness of head assembly in the presence of excess portal protein. J Virol *76*, 10245-10255.

Morrison, D.A., and Mannarelli, B. (1979). Transformation in pneumococcus: nuclease resistance of deoxyribonucleic acid in the eclipse complex. J Bacteriol *140*, 655-665.

Mortier-Barriere, I., Velten, M., Dupaigne, P., Mirouze, N., Pietrement, O., McGovern, S., Fichant, G., Martin, B., Noirot, P., Le Cam, E., *et al.* (2007). A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA. Cell *130*, 824-836.

Muller, H.J. (1932). Some Genetic Aspects of Sex. The American Naturalist *66*, 118-138.

Muller, H.J. (1964). The Relation of Recombination to Mutational Advance. Mutat Res *106*, 2-9.

Murray, A.E., Lies, D., Li, G., Nealson, K., Zhou, J., and Tiedje, J.M. (2001). DNA/DNA hybridization to microarrays reveals gene-specific differences between closely related microbial genomes. Proc Natl Acad Sci U S A *98*, 9853-9858.

Musharova, O., Klimuk, E., Datsenko, K.A., Metlitskaya, A., Logacheva, M., Semenova, E., Severinov, K., and Savitskaya, E. (2017). Spacer-length DNA intermediates are associated with Cas1 in cells undergoing primed CRISPR adaptation. Nucleic Acids Res *45*, 3297-3307.

Nair, D., Memmi, G., Hernandez, D., Bard, J., Beaume, M., Gill, S., Francois, P., and Cheung, A.L. (2011). Whole-genome sequencing of Staphylococcus aureus strain RN4220, a key laboratory strain used in virulence research, identifies mutations that affect not only virulence factors but also the fitness of the strain. J Bacteriol *193*, 2332-2335.

Nanda, A.M., Thormann, K., and Frunzke, J. (2015). Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions. J Bacteriol *197*, 410-419.

Novick, R. (1967). Properties of a cryptic high-frequency transducing phage in Staphylococcus aureus. Virology *33*, 155-166.

Novick, R.P., Christie, G.E., and Penades, J.R. (2010). The phage-related chromosomal islands of Gram-positive bacteria. Nat Rev Microbiol *8*, 541-551.

Ochman, H., Lawrence, J.G., and Groisman, E.A. (2000). Lateral gene transfer and the nature of bacterial innovation. Nature *405*, 299-304.

Oliveira, P.H., Touchon, M., and Rocha, E.P. (2014). The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. Nucleic Acids Res *42*, 10618-10631.

Orgel, L.E., and Crick, F.H. (1980). Selfish DNA: the ultimate parasite. Nature *284*, 604-607.

Ozer, E.A., Allen, J.P., and Hauser, A.R. (2014). Characterization of the core and accessory genomes of Pseudomonas aeruginosa using bioinformatic tools Spine and AGEnt. BMC Genomics *15*, 737.

Pabst, O., and Mowat, A.M. (2012). Oral tolerance to food protein. Mucosal Immunol *5*, 232-239.

Palmer, K.L., and Gilmore, M.S. (2010). Multidrug-resistant enterococci lack CRISPR-cas. mBio *1*, e00227-00210.

Park, K.H., An, Y., Jung, T.Y., Baek, I.Y., Noh, H., Ahn, W.C., Hebert, H., Song, J.J., Kim, J.H., Oh, B.H., *et al.* (2017). RNA activation-independent DNA targeting of the Type III CRISPR-Cas system by a Csm complex. EMBO Rep *18*, 826-840.

Pawluk, A., Bondy-Denomy, J., Cheung, V.H., Maxwell, K.L., and Davidson, A.R. (2014). A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of Pseudomonas aeruginosa. MBio *5*, e00896.

Penades, J.R., Chen, J., Quiles-Puchalt, N., Carpena, N., and Novick, R.P. (2015). Bacteriophage-mediated spread of bacterial virulence genes. Curr Opin Microbiol *23*, 171-178.

Piechowska, M., and Fox, M.S. (1971). Fate of transforming deoxyribonucleate in Bacillus subtilis. J Bacteriol *108*, 680-689.

Playfair, J.H., Taverne, J., Bate, C.A., and de Souza, J.B. (1990). The malaria vaccine: anti-parasite or anti-disease? Immunol Today *11*, 25-27.

Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology *151*, 653-663.

Proksch, E., Brandner, J.M., and Jensen, J.M. (2008). The skin: an indispensable barrier. Exp Dermatol *17*, 1063-1072.

Ptashne, M. (2011). Principles of a switch. Nat Chem Biol *7*, 484-487.

Quiles-Puchalt, N., Tormo-Mas, M.A., Campoy, S., Toledo-Arana, A., Monedero, V., Lasa, I., Novick, R.P., Christie, G.E., and Penades, J.R. (2013). A super-family of transcriptional activators regulates bacteriophage packaging and lysis in Gram-positive bacteria. Nucleic Acids Res *41*, 7260-7275.

Raberg, L., Graham, A.L., and Read, A.F. (2009). Decomposing health: tolerance and resistance to parasites in animals. Philos Trans R Soc Lond B Biol Sci *364*, 37-49.

Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S., and Medzhitov, R. (2004). Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. Cell *118*, 229-241.

Ram, G., Chen, J., Ross, H.F., and Novick, R.P. (2014). Precisely modulated pathogenicity island interference with late phage gene transcription. Proc Natl Acad Sci U S A *111*, 14536-14541.

Reese, M.G. (2001). Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. Comput Chem *26*, 51-56.

Refardt, D. (2011). Within-host competition determines reproductive success of temperate bacteriophages. ISME J *5*, 1451-1460.

Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J., Blumenthal, R.M., Degtyarev, S., Dryden, D.T., Dybvig, K., *et al.* (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res *31*, 1805-1812.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol *29*, 24-26.

Rocha, E.P., Danchin, A., and Viari, A. (2001). Evolutionary role of restriction/modification systems as revealed by comparative genome analysis. Genome Res *11*, 946-958.

Rooijakkers, S.H., Ruyken, M., Roos, A., Daha, M.R., Presanis, J.S., Sim, R.B., van Wamel, W.J., van Kessel, K.P., and van Strijp, J.A. (2005). Immune evasion by a staphylococcal complement inhibitor that acts on C3 convertases. Nat Immunol *6*, 920-927.

Roy, B.A., and Kirchner, J.W. (2000). Evolutionary dynamics of pathogen resistance and tolerance. Evolution *54*, 51-63.

Samai, P., Pyenson, N., Jiang, W., Goldberg, G.W., Hatoum-Aslan, A., and Marraffini, L.A. (2015). Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. Cell *161*, 1164-1174.

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. Nucleic Acids Res *39*, 9275-9282.

Schafer, J.F. (1971). Tolerance to Plant Disease. Annu Rev Phytopathol *9*, 235-252.

Schneewind, O., and Missiakas, D.M. (2012). Protein secretion and surface display in Gram-positive bacteria. Philos Trans R Soc Lond B Biol Sci *367*, 1123-1139.

Schofield, L., Hewitt, M.C., Evans, K., Siomos, M.A., and Seeberger, P.H. (2002). Synthetic GPI as a candidate anti-toxic vaccine in a model of malaria. Nature *418*, 785-789.

Schubert, R.A., Dodd, I.B., Egan, J.B., and Shearwin, K.E. (2007). Cro's role in the CI Cro bistable switch is critical for {lambda}'s transition from lysogeny to lytic development. Genes Dev *21*, 2461-2472.

Seed, K.D., Lazinski, D.W., Calderwood, S.B., and Camilli, A. (2013). A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. Nature *494*, 489-491.

Seehausen, O., Butlin, R.K., Keller, I., Wagner, C.E., Boughman, J.W., Hohenlohe, P.A., Peichel, C.L., Saetre, G.P., Bank, C., Brannstrom, A., *et al.* (2014). Genomics and the origin of species. Nat Rev Genet *15*, 176-192.

Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J., and Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc Natl Acad Sci USA *108*, 10098-10103.

Shen, P., and Huang, H.V. (1986). Homologous recombination in Escherichia coli: dependence on substrate length and homology. Genetics *112*, 441-457.

Shimada, T., Yamazaki, Y., Tanaka, K., and Ishihama, A. (2014). The whole set of constitutive promoters recognized by RNA polymerase RpoD holoenzyme of Escherichia coli. PLoS One *9*, e90447.

Shimizu, T., Ohta, Y., and Noda, M. (2009). Shiga toxin 2 is specifically released from bacterial cells by two different mechanisms. Infect Immun *77*, 2813-2823.

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I., *et al.* (2017). Diversity and evolution of class 2 CRISPR-Cas systems. Nat Rev Microbiol *15*, 169-182.

Shultzaberger, R.K., Chen, Z., Lewis, K.A., and Schneider, T.D. (2007). Anatomy of Escherichia coli sigma70 promoters. Nucleic Acids Res *35*, 771-788.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J., and Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. Genome Res *19*, 1117-1123.

Sloane, R., de Azavedo, J.C., Arbuthnott, J.P., Hartigan, P.J., Kreiswirth, B., Novick, R., and Foster, T.J. (1991). A toxic shock syndrome toxin mutant of Staphylococcus aureus isolated by allelic replacement lacks virulence in a rabbit uterine model. FEMS Microbiol Lett *62*, 239-244.

Smargon, A.A., Cox, D.B., Pyzocha, N.K., Zheng, K., Slaymaker, I.M., Gootenberg, J.S., Abudayyeh, O.A., Essletzbichler, P., Shmakov, S., Makarova, K.S., *et al.* (2017). Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. Mol Cell *65*, 618-630 e617.

Soares, M.P., Gozzelino, R., and Weis, S. (2014). Tissue damage control in disease tolerance. Trends Immunol *35*, 483-494.

Staals, R.H., Zhu, Y., Taylor, D.W., Kornfeld, J.E., Sharma, K., Barendregt, A., Koehorst, J.J., Vlot, M., Neupane, N., Varossieau, K., *et al.* (2014). RNA Targeting by the Type III-A CRISPR-Cas Csm Complex of *Thermus thermophilus*. Mol Cell *56*, 518-530.

Stappenbeck, T.S., Hooper, L.V., and Gordon, J.I. (2002). Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. Proc Natl Acad Sci U S A *99*, 15451-15455.

Stary, E., Gaupp, R., Lechner, S., Leibig, M., Tichy, E., Kolb, M., and Bertram, R. (2010). New architectures for Tet-on and Tet-off regulation in Staphylococcus aureus. Appl Environ Microbiol *76*, 680-687.

Stecher, B., Robbiani, R., Walker, A.W., Westendorf, A.M., Barthel, M., Kremer, M., Chaffron, S., Macpherson, A.J., Buer, J., Parkhill, J., *et al.* (2007). Salmonella enterica serovar

typhimurium exploits inflammation to compete with the intestinal microbiota. PLoS Biol *5*, 2177-2189.

Sumby, P., and Waldor, M.K. (2003). Transcription of the toxin genes present within the Staphylococcal phage phiSa3ms is intimately linked with the phage's life cycle. J Bacteriol *185*, 6841-6851.

Takeuchi, N., Kaneko, K., and Koonin, E.V. (2014). Horizontal gene transfer can rescue prokaryotes from Muller's ratchet: benefit of DNA from dead cells and population subdivision. G3 (Bethesda) *4*, 325-339.

Tamulaitis, G., Kazlauskiene, M., Manakova, E., Venclovas, C., Nwokeoji, A.O., Dickman, M.J., Horvath, P., and Siksnys, V. (2014). Programmable RNA Shredding by the Type III-A CRISPR-Cas System of *Streptococcus thermophilus*. Mol Cell *56*, 506-517.

Tamulaitis, G., Venclovas, C., and Siksnys, V. (2017). Type III CRISPR-Cas Immunity: Major Differences Brushed Aside. Trends Microbiol *25*, 49-61.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., *et al.* (2005). Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan genome". Proc Natl Acad Sci U S A *102*, 13950-13955.

Thomas, C.M., and Nielsen, K.M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat Rev Microbiol *3*, 711-721.

Tock, M.R., and Dryden, D.T. (2005). The biology of restriction and anti-restriction. Curr Opin Microbiol *8*, 466-472.

Tormo-Mas, M.A., Mir, I., Shrestha, A., Tallent, S.M., Campoy, S., Lasa, I., Barbe, J., Novick, R.P., Christie, G.E., and Penades, J.R. (2010). Moonlighting bacteriophage proteins derepress staphylococcal pathogenicity islands. Nature *465*, 779-782.

Touchon, M., Bernheim, A., and Rocha, E.P. (2016). Genetic and life-history traits associated with the distribution of prophages in bacteria. ISME J *10*, 2744-2754.

Trautner, T.A., Pawlek, B., Bron, S., and Anagnostopoulos, C. (1974). Restriction and modification in B. subtilis. Biological aspects. Mol Gen Genet *131*, 181-191.

Tyler, J.S., Beeri, K., Reynolds, J.L., Alteri, C.J., Skinner, K.G., Friedman, J.H., Eaton, K.A., and Friedman, D.I. (2013). Prophage induction is enhanced and required for renal disease and lethality in an EHEC mouse model. PLoS Pathog *9*, e1003236.

van der Oost, J., Westra, E.R., Jackson, R.N., and Wiedenheft, B. (2014). Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nat Rev Microbiol *12*, 479-492.

van Houte, S., Buckling, A., and Westra, E.R. (2016). Evolutionary Ecology of Prokaryotic Immune Mechanisms. Microbiol Mol Biol Rev *80*, 745-763.

van Wamel, W.J., Rooijakkers, S.H., Ruyken, M., van Kessel, K.P., and van Strijp, J.A. (2006). The innate immune modulators staphylococcal complement inhibitor and chemotaxis inhibitory protein of Staphylococcus aureus are located on beta-hemolysin-converting bacteriophages. J Bacteriol *188*, 1310-1315.

Vasu, K., and Nagaraja, V. (2013). Diverse functions of restriction-modification systems in addition to cellular defense. Microbiol Mol Biol Rev *77*, 53-72.

Vercoe, R.B., Chang, J.T., Dy, R.L., Taylor, C., Gristwood, T., Clulow, J.S., Richter, C., Przybilski, R., Pitman, A.R., and Fineran, P.C. (2013). Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. PLoS Genet *9*, e1003454.

Verkaik, N.J., Benard, M., Boelens, H.A., de Vogel, C.P., Nouwen, J.L., Verbrugh, H.A., Melles, D.C., van Belkum, A., and van Wamel, W.J. (2011). Immune evasion cluster-positive bacteriophages are highly prevalent among human Staphylococcus aureus strains, but they are not essential in the first stages of nasal colonization. Clin Microbiol Infect *17*, 343-348.

von Eiff, C., Becker, K., Machka, K., Stammer, H., and Peters, G. (2001). Nasal Carriage as a Source of Staphylococcus aureus Bacteremia. New Engl J Med *344*, 11-16.

Vos, M. (2009). Why do bacteria engage in homologous recombination? Trends Microbiol *17*, 226-232.

Wabl, M., and Steinberg, C. (1996). Affinity maturation and class switching. Curr Opin Immunol *8*, 89-92.

Wagner, P.L., Livny, J., Neely, M.N., Acheson, D.W., Friedman, D.I., and Waldor, M.K. (2002). Bacteriophage control of Shiga toxin 1 production and release by Escherichia coli. Mol Microbiol *44*, 957-970.

Weinstock, G.M. (2002). Transduction in Gram-Negative Bacteria. In Modern Microbial Genetics (John Wiley & Sons, Inc.), pp. 561-579.

Westra, E.R., van Erp, P.B., Kunne, T., Wong, S.P., Staals, R.H., Seegers, C.L., Bollen, S., Jore, M.M., Semenova, E., Severinov, K., *et al.* (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Mol Cell *46*, 595-605.

Wiedenheft, B., Sternberg, S.H., and Doudna, J.A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. Nature *482*, 331-338.

Wood, W.B. (1966). Host specificity of DNA produced by Escherichia coli: bacterial mutations affecting the restriction and modification of DNA. J Mol Biol *16*, 118-133.

Xia, G., and Wolz, C. (2014). Phages of Staphylococcus aureus and their impact on host evolution. Infect Genet Evol *21*, 593-601.

Yosef, I., Goren, M.G., Kiro, R., Edgar, R., and Qimron, U. (2011). High-temperature protein G is essential for activity of the *Escherichia coli* clustered regularly interspaced short palindromic repeats (CRISPR)/Cas system. Proc Natl Acad Sci USA *108*, 20136-20141.

Yosef, I., Goren, M.G., and Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. Nucleic Acids Res *40*, 5569-5576.

Yu, D., Ellis, H.M., Lee, E.C., Jenkins, N.A., Copeland, N.G., and Court, D.L. (2000). An efficient recombination system for chromosome engineering in Escherichia coli. Proc Natl Acad Sci U S A *97*, 5978-5983.

Zebec, Z., Manica, A., Zhang, J., White, M.F., and Schleper, C. (2014). CRISPR-mediated targeted mRNA degradation in the archaeon Sulfolobus solfataricus. Nucleic Acids Res *42*, 5280-5288.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., *et al.* (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell *163*, 759-771.

Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S.V., *et al.* (2012). Structure and Mechanism of the CMR Complex for CRISPR-Mediated Antiviral Immunity. Mol Cell *45*, 303-313.

Zinder, N.D., and Lederberg, J. (1952). Genetic exchange in Salmonella. J Bacteriol *64*, 679-699.