Rockefeller University
Digital Commons @ RU

Student Theses and Dissertations

2022

Restriction-Modification and CRISPR-Cas Systems: Cooperation Between Innate and Adaptive Immunity in Prokaryotes

Pascal Maguin

Follow this and additional works at: https://digitalcommons.rockefeller.edu/student_theses_and_dissertations

 Part of the Life Sciences Commons

Restriction-Modification and CRISPR-Cas systems: cooperation between innate and adaptive immunity in prokaryotes

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements

for the degree of Doctor of Philosophy

By

Pascal Maguin

June 2022

© Copyright by Pascal Maguin 2022

RESTRICTION-MODIFICATION AND CRISPR-CAS SYSTEMS: COOPERATION BETWEEN INNATE AND ADAPTIVE IMMUNITY IN PROKARYOTES

Pascal Maguin, Ph.D.
The Rockefeller University 2022

Bacteria have evolved numerous mechanisms to resist the constant assault of viruses (called bacteriophages, or simply phages) that can infect and kill them. Restriction-modification (RM) systems represent one such strategy. Generally, these systems provide defense by coordinating the activities of two distinct enzymes: a restriction endonuclease and a methyltransferase. Both enzymes recognize the same short DNA sequences. The methyltransferase modifies these target sites in the host chromosome, which prevents the restriction endonuclease from cleaving the host's own DNA. In contrast, foreign phage DNA is usually not methylated at these sequences. Consequently, upon injection into the host, the viral DNA is recognized and cleaved by the restriction endonuclease, preventing the progression of the phage's life cycle. Therefore, RM systems are considered a part of the innate immune response because they can provide defense against any phage, including ones that have never been encountered previously, as long as they harbor RM target sites. Clustered regularly interspaced short palindromic repeats (CRISPR) loci and their associated genes (*cas*) form another defense system that destroys foreign DNA. The CRISPR array consists of a series of repetitive DNA sequences separated by unique DNA sequences known as spacers. During phage infection, short DNA fragments are taken from the viral DNA and integrated into the CRISPR locus to form new spacers. These

sequences are then transcribed into CRISPR RNAs (crRNAs). In type II-A CRISPR-Cas systems, the crRNAs guide the Cas9 nuclease to a matching viral DNA target for cleavage. As such, unlike RM systems, CRISPR-Cas systems represent an adaptive immune response because they require an initial exposure to a virus in order to become successfully immunized through the acquisition of new spacer sequences.

CRISPR-Cas and RM are two of the most prevalent types of defense systems found in bacteria and often co-exist together in a single host. Yet, how they may interact with each other in the context of immunity during bacteriophage infection is poorly understood. Here, in my thesis work, I investigate the interplay between RM and type II-A CRISPR-Cas systems. First, I demonstrate that RM systems provide a weak and temporary protection that stimulates CRISPR spacer acquisition, enabling the cells to survive the viral infection. Then, I go on to show that the restriction activity of the RM system is critical for this process and that the rate of spacer acquisition is correlated to the number of RM target sites in the phage genome. To further uncover the mechanistic link between restriction and the acquisition of new spacers, I implement next-generation sequencing to demonstrate that spacers are preferentially extracted at the dsDNA breaks (DSBs) generated by the restriction endonuclease. Additionally, I show that the host DNA repair complex, AddAB, can process these breaks, which further enhances spacer acquisition. Finally, I follow the dynamics between RM and CRISPR-Cas during the chain of events that occur upon viral infection. I demonstrate that although the RM system provides an immediate line of defense due to its ability to recognize a broad range of foreign invaders, it is ultimately overcome by the rapid

emergence of methylated phages, resulting in the death of much of the bacterial population. However, the early RM immune response creates substrates for spacer acquisition by the CRISPR-Cas system in a subset of cells. By using these newly acquired spacers which specify the viral sequences for lethal cleavage by Cas9, these cells can now extinguish the methylated phages, resulting in the survival and regrowth of the population.

Collectively, my thesis reveals the molecular mechanisms connecting RM and CRISPR-Cas systems in providing a synergistic anti-phage defense. Reminiscent of eukaryotic immunity, I demonstrate that RM systems provide an initial, short-lived innate immune response, which stimulates a secondary, more robust adaptive immune response by CRISPR-Cas. This work highlights an example of cooperation between RM and CRISPR-Cas, which are two of the most common bacterial defense systems. However, prokaryotes have been shown to harbor a multitude of other putative anti-phage defense systems, which can often exist together in a single host. I predict that future studies will likely uncover many more fascinating instances of immune interaction among other sets of defense systems

For Mom, Dad, and Peter

ACKNOWLEDGMENTS

First, I would like to express my sincere gratitude to my advisor, Dr. Luciano Marraffini. It was a privilege to be mentored by such a brilliant scientist, and I will be forever grateful for his patience, trust, and commitment to my success. Additionally, I would like to thank Luciano for fostering a positive and inclusive work environment. Joining his laboratory has been one of the best decisions of my scientific career.

Second, I would like to extend my genuine appreciation to the members of my thesis committee, Dr. Charles Rice and Dr. Gabriel Victora, for their insightful comments, helpful discussions, and encouragement throughout my PhD. Also, I would like to thank Dr. Alan Davidson (University of Toronto) for graciously serving as the external examiner for my thesis.

Third, The David Rockefeller graduate program provides a truly wonderful and unique training. Thank you to everyone at the Dean's office that makes this possible. In particular, I would like to thank Dr. Emily Harms, Cristian Rosario, Marta Delgado, Kristen Cullen, and Stephanie Fernandez for their support and guidance.

Fourth, the Marraffini Lab would not be such an amazing place without my incredible labmates. Thank you to every member for helping me in my research and for making the lab such a fun place. Specifically, I am grateful to Dr. Robert Heler, Dr. Jakob Rostol, Dr. Nora Pyenson, Dr. Andrew Varble, Amer Hossain, Dalton Banh, Dr.

Alex Meeske, Dr. Poulami Samai, Dr. Josh Modell, Dr. Philipp Nussenzweig, and Claire Kenney for not only being wonderful co-workers, but also such great friends. I will forever cherish our times together.

Finally, on a personal note, I would like to thank my husband, Peter Cooper, and my family back home in France for their unconditional love and support. Peter's absolute confidence in me has helped me during some of my most challenging times and I could never have achieved this without him.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

bp……base pair

kb……kilobase

BHI…… ……brain heart infusion

Cas……CRISPR-associated

Cascade……CRISPR-associated complex for antiviral defense

CFU……colony forming unit

Cm^R……chloramphenicol-resistant

CRISPR……Clustered Regular Spaced Short Palindromic Repeats

crRNA……CRISPR RNA

Csn……CRISPR-associated endonuclease

pre-crRNA……precursor CRISPR RNA

DNA……deoxyribonucleic acid

DSB……double-stranded DNA break

dsDNA……double-stranded DNA

ssDNA……single-stranded DNA

Erm^R……erythromycin-resistant

IHF……integration host factor

LAS……leader-anchoring sequence

MTase……methyltransferase

MOI……multiplicity of infection

NGS…… next generation sequencing

OD_{600}……optical density at wavelength 600 nanometers

pAgo……prokaryotic Argonaute

PAM……protospacer adjacent motif

pBglII……BglII RM system on the pLZ12 vector

PCR……polymerase chain reaction

pCRISPR……*S. pyogenes* type II-A CRISPR-Cas system on the pC194 vector

PFU……plaque forming units

Phage……bacteriophage

Prespacer……substrate(s) for spacer integration

pSauI……SauI RM system on the pLZ12 vector

REase……restriction endonuclease

RM……Restriction–Modification

RNA……ribonucleic acid

RPM……Reads per million

SR…… single repeat

tracrRNA……trans-activating CRISPR RNA

TRD……target recognition domain

B. globigii……*Bacillus globigii*

E. coli……*Escherichia coli*

S. epidermidis……*Staphylococcus epidermidis*

S. aureus……*Staphylococcus aureus*

CHAPTER 1. INTRODUCTION

1.1 Bacteriophages

Over a century ago, Félix d'Hérelle and Frederick W. Twort independently discovered viruses that can infect bacteria[1,2]. D'Hérelle named these entities "bacteriophages" (or simply "phages," named for their ability "to devour" bacteria)[1,2]. Following their discovery, phages became model organisms during the first half of the 20th century and laid the foundation for many fundamental tenets of molecular biology— from the determination that mutations arise randomly and spontaneously (the Luria-Delbrück experiment in 1943) to the establishment that DNA is the genetic material (the Hershey-Chase experiment in 1952)[3,4]. Studying interactions between phages and their hosts also revolutionized the world of genetic engineering during the second half of the 20th century[5]. Scientists uncovered molecular machinery evolved from these interactions that could be exploited to manipulate DNA, such as the Cre-Lox recombination system[6], restriction enzymes[7], and more recently, the RNA-guided endonuclease Cas9[8].

Bacteriophages represent the most abundant entities on earth, with about 10^{31} particles estimated to exist on the planet[9]. Even though they are diverse in their morphology and nucleic acid composition, phages are typically composed of a protein-based capsid encapsulating their viral genetic information[10]. Their genomes can be

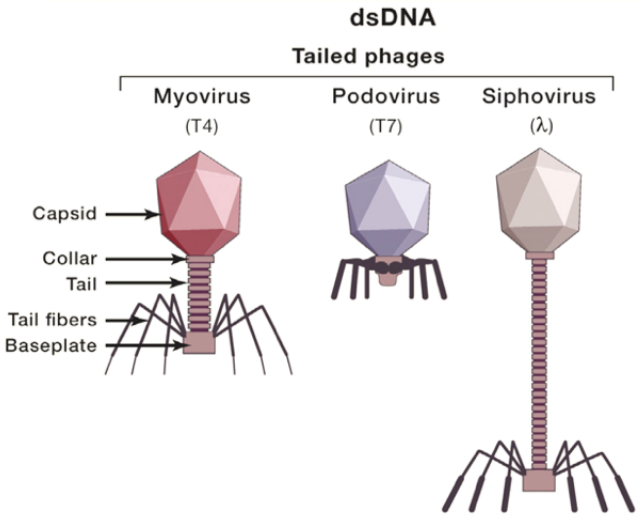


Figure 1.1 Bacteriophages belonging to the *Caudovirales* order.

Representation of the three families of bacteriophage belonging to the *Caudovirales* order. Reproduced with modifications from Ofir and Sorek (2018)[211] with permission from the publisher.

DNA or RNA, and either double- or single-stranded[10]. Thus far, most of the isolated phages belong to the *Siphoviridae* family within the *Caudovirales* order (**Fig. 1.1**)[11]. The classic and well-known *E. coli* phage λ belongs to this family. Viruses in this order have linear dsDNA genomes and a protein-based tube (known as a tail) attached to their capsids. The tail enables attachment to the host surface and delivery of the virus' genetic information inside the host cytosol[10,11]. Once their DNA is injected, strictly lytic phages start replicating their DNA and producing viral proteins, culminating in the lysis of the bacteria and the release of newly assembled viral particles. By contrast, temperate phages, such as phage λ, can enter two different life cycles: lysogenic or lytic[12]. During the lysogenic cycle, temperate phages integrate their genomes in the

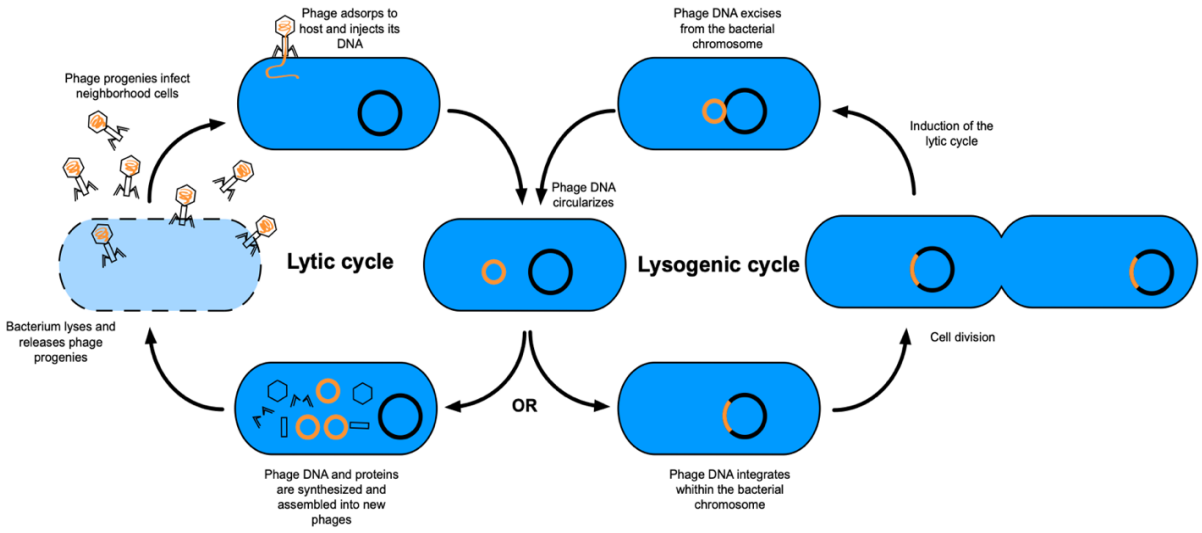


Figure 1.2 Lytic and lysogenic life cycles of bacteriophages.

First, the bacteriophage adsorbs to the bacterium, injects, and circularizes its genome (orange). The bacteriophage can then enter the lytic or lysogenic cycle. During the lytic cycle, the viral genome is replicated and translated. The viral genomes are then packaged into new virions resulting in lysis of the host and release of the newly formed phage progenies. Alternatively, the phage enters the lysogenic cycle and integrates its genome in the bacterial chromosome as a “prophage” (black circle). During cell division, the viral genome is replicated together with the host genome and spreads to the bacterial population. Upon experiencing an environmental stressor, the viral genome excises, and the phage executes the lytic cycle. This causes lysis of the bacterium and the release of new phage particles.

host DNA and are known as prophages. While integrated into the host DNA, the prophage’s lytic genes are repressed, allowing prophages to propagate vertically in the culture through bacterial cell division. Upon environmental stimuli, the lytic cycle is induced, causing viral genome excision from the host chromosome, followed by viral DNA replication and protein production resulting in bacterial lysis and phage progeny release. This lysis-lysogeny decision is governed by a transcriptional repressor, which prevents transcription of the lytic genes necessary for replication and packaging of new viral particles[13]. Proteolytic cleavage of the repressor during stress conditions (such as

UV damage[14]) alleviates this repression, causing the phage to enter the lytic cycle. This results in lysis of the bacterium and the release of newly assembled phage progeny, which can infect neighboring cells (**Fig. 1.2**).

Bacteriophages have been recovered from every biome where bacteria are known to exist[10], from the human body to the open ocean[9,15]. Besides modulating the density and composition of bacterial communities through predation, phages also play important roles in ecology[16,17] and human health[15]. Recently, the presence of specific phages in the gut was shown to be associated with improved executive function and memory in mice and humans[18]. In most environments, phages outnumber bacteria 10:1, but this ratio can be significantly higher in other contexts[19]. As such, there is a constant battle between phages and bacteria, resulting in an evolutionary arms race[20]. One might think that bacteria would ultimately lose this war as phages can release hundreds of new progeny upon infection of a single cell[21]. However, from this host-parasite competition, an abundance of bacterial defense systems evolved to stop an infection at every stage of the phage life cycle[22]. In turn, bacteriophages also have evolved means to circumvent these anti-phage strategies[23]. Consequently, bacteria and phages seem to stably co-exist in a dynamic equilibrium[24]. Restriction-modification (RM) and CRISPR-Cas systems are two of the most well-known and well-studied anti-phage defense systems that this evolutionary arms race has produced. Both systems cleave viral DNA as a strategy to prevent the progression of infection.

1.2 Restriction-modification systems: prokaryotic innate immune systems

In the early 1950s, microbiologists observed and described the phenomenon of host-controlled variation in bacteriophage[25–27]. Luria described this in 1953 as "a restriction of the ability of the phage to grow in some host as a result of one cycle of growth in one type of cell; and a release of this restriction following one cycle of growth in some other" [28]. Ten years later, Werner and Dussoix demonstrated that phage DNA carried an imprint of this host-controlled variation phenomenon described by Luria[29]. This observation eventually led to the discovery of **r**estriction-**m**odification (**RM**) systems, bacterial defense systems capable of recognizing and restricting viral DNA[30]. Following this discovery, scientists quickly realized the practical utility of these systems, particularly the endonuclease activity. Purified restriction enzymes could be used to cut different pieces of genetic material, and these fragments could then be stitched together using a ligase (also purified from phages)[31]. With the development of these simple but powerful tools, molecular cloning was born.

1.2.1 The RM immune response

Present in over 90% of sequenced prokaryotic genomes[32], RM systems are widespread and ubiquitous. Consequently, they are perhaps some of the most characterized bacterial defense systems. Typically, RM immunity is orchestrated by two distinct enzymes with specific activities: a methyltransferase (MTase) and a restriction endonuclease (REase)[30]. Both the MTase and the REase recognize the

same target DNA sequence. These target sites are modified in the host genome by the MTase, and the methylated target DNA is perceived as “self” DNA. In contrast, unmodified target sequences in viral DNA are recognized as “foreign” and consequently cleaved by the REase, preventing the phage from completing its life cycle (**Fig. 1.3**). RM immunity is considered an “innate” immune response because it can immediately provide host protection from any virus as long as the virus harbors unmodified recognition sites.

1.2.2 Type I and II RM systems

RM systems are named after their hosts and are numbered following the chronology of their discovery[33,34]. For example, SauI refers to the first RM system discovered in *S. aureus*. These systems are divided into four types (I, II, III, and IV) according to their cofactor requirements, subunit organization, recognition sites, and their cleavage position within the target DNA[35]. Types I and II are the most abundant, representing about 30% and 43% of RM systems respectively[36]. We will focus on these two types as they are the most well-characterized and relevant to this thesis work.

Type I R-M systems are generally composed of 3 genes: *hsdR, hsdM*, and *hsdS*, coding for a restriction (R), a methyltransferase (M), and a specificity (S) subunit, respectively. They are further subdivided into five families, A-E, based on their amino acid sequences and their gene complementation characteristics[37]. Type I RM can form

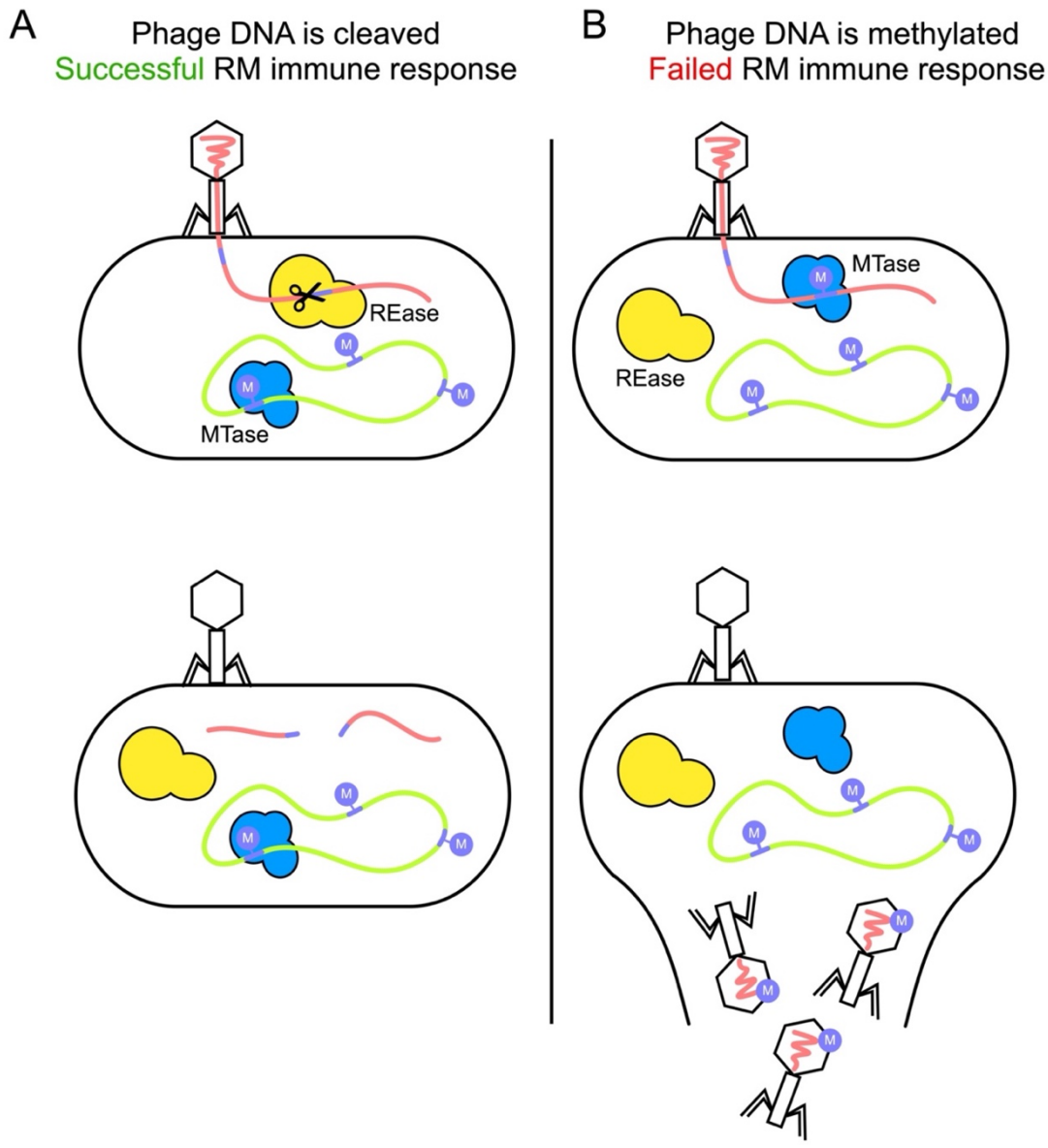


Figure 1.3. The RM immune response and viral escape.

(**A**) Representation of a successful RM immune response. The MTase (blue) methylates the RM recognition sites (purple circles labeled M) in the host chromosome (green) to prevent cleavage by the REase (yellow). Upon injection of a phage's DNA (pink), the REase recognizes the unmethylated RM sites and cleaves the viral genome. (**B**) Representation of a failed RM immune response. At low frequency, the MTase (blue) recognizes unmethylated RM recognition sites (purple) in the incoming viral genome (pink) before cleavage by the REase. The methylated viral genome is no longer recognized by the REase. The phage can complete its life cycle resulting in lysis of the bacterium and release of methylated phage progeny (phages with purple circles labeled M).

two complexes: M_2S_1 and $R_2M_2S_1$[38,39]. The former acts as a MTase; the latter acts as a MTase on hemimethylated sites and an endonuclease on unmethylated sites[40]. The S subunit dictates the specific sequence recognized by the complexes and is composed of two target recognition domains (TRDs) separated by a core domain[41]. Each TRD recognizes a specific half-sequence within the target sequence. As such, type I RM systems recognize asymmetrical bi-partite sequences (e.g., 5'-CCAYNNNNNNTGT-3' for SauI[42]). The core sequence of the S subunit joining the TRDs allows for a random spacer sequence separating the two specific sequences defined by the TRDs (CCAY and TGT in the example above). Type I RM methyltransferases use *S*-adenosylmethionine (SAM) as a methyl group donor to methylate an adenine to N^6-methyladenine (m6A) on both strands of the recognition sequence by flipping the base out of the DNA helix[43].

The hallmark feature of type I RM systems is that their R subunit endonucleases cleave DNA non-specifically, away from their recognition sequences[44]. The R subunit contains a nuclease domain (PD–(D/E)XK) fused to an SF2 helicase/translocase domain[45]. As such, upon recognizing an unmethylated site, the $R_2M_2S_1$ complex translocates DNA using ATP while remaining bound to the site[46,47]. Stalling of the complex by collision with a physical barrier on the DNA (*e.g.*, another Type I restriction enzyme or another enzyme on the DNA such as a polymerase) triggers restriction. Type I restriction enzymes therefore cleave DNA non-specifically away from their recognition sites but tend to cut preferentially between two recognition sites[44,48]. The type of dsDNA break they generate is not well understood, but one study suggests that

they indiscriminately produce both 5' and 3' single-stranded overhangs of various length[49].

Type II RM systems have a simpler architecture than type I. They typically harbor two genes coding for a MTase and a REase that, unlike type I, act independently from each other. The defining characteristic of type II RM systems is that their REases cleave specifically within their recognition sequences or at a specified distance away from them[35]. They also generate defined DNA breaks with either blunt ends or short single-stranded overhangs[32]. Given their specific cutting tendency, these enzymes are widely applied in molecular cloning. Generally, "classic" type II RM systems (subtype IIP) are composed of a monomeric methyltransferase and a homodimeric endonuclease, both recognizing the same short (4 to 8 bp) palindromic sequence (e.g., 5'-AGATCT-3' for BglII[50]). The MTase modifies a base in the recognition sequence, utilizing SAM to generate N^6-methyladenine (m6A), N^4-methylcytosine (m4A), or C^5-methylcytosine (m5C)[43]. For type II restriction enzymes, each monomer contains an endonuclease motif (generally PD–(D/E)XK but not always)[51,52], and each cleaves one specific DNA strand, together resulting in a dsDNA break[53]. However, type II restriction enzymes that deviate from this classical model have also been reported. For example, the monomeric enzyme BcnI contains one catalytic site and must successively bind and cleave the DNA twice to generate a break[54]. Also, some enzymes must bind to multiple copies of their recognition site to cleave DNA[55,56]. As a consequence, type II RM systems have been subdivided into

11 subtypes based on their recognition sequence (e.g., palindromic or asymmetric) and enzymatic behavior[33].

Ultimately, type I and II RM systems provide phage defense using the same strategy: methylating host DNA and restricting the unmethylated foreign DNA of viruses. The site of DNA cleavage is the major difference between these two types. Type I restriction enzymes translocate and cleave DNA at a random location away from their recognition sites[44]. In contrast, type II endonucleases do not translocate DNA and cut at a specific location[53].

1.2.1 The balancing act of RM systems

RM systems' ability to exploit the modification state of DNA presents two key advantages for successful host defense against phages. First, RM systems can recognize and afford protection against invaders that the host has never encountered before. As such, the RM immune response is considered an example of innate immunity. Second, and crucially, by modifying host DNA, RM systems are able to avoid fatal cleavage of the host genome. However, erroneous methylation of viral DNA can occur, which enables viral escape from restriction[57,58]. Therefore, a tight balance between these systems' methyltransferase and endonuclease activities is needed to enable robust restriction of foreign DNA while also preventing autoimmunity.

Type I RM systems exhibit a phenomenon described as "restriction alleviation" (RA)[59]. This process involves the temporary downregulation of the host type I

restriction activity following DNA damage induced by chemicals[60] or UV radiation[27,59,61]. During DNA repair, unmethylated sites are generated. As such, RA allows time for the cells to repair and methylate those sites[62]. Type I RM systems also exhibit RA when they colonize a new host[63–65]. This buys the methyltransferase time to modify all the targets sites in the new host genome before the restriction enzyme can cleave them. To date, only post-translational regulation has been observed for type I RM systems. In contrast to type I RM, three modes of transcriptional regulation of type II RM systems have been shown[66,67]: antisense RNAs, a dedicated C (controller) transcriptional factor, and methylation of the type II RM promoters by their own MTases. These regulations ensure that type II RM can colonize new hosts but also regulate the level of each enzyme to avoid auto-immunity. While we understand some of their regulation mechanisms, a lot remains unknown about RM regulation.

Despite having multiple layers of regulation, the RM methyltransferase can, at low but appreciable frequencies, localize to invading viral DNA first and modify it before cleavage by the restriction endonuclease[58,68,69]. This is obviously detrimental for that initial host, allowing a lytic bacteriophage to complete its life cycle. Furthermore, this also now poses an uncontrolled threat to the rest of the population, since these methylated phages will avoid recognition by RM, ultimately resulting in lysis of the entire culture. The likelihood of viral escape through this route is inversely correlated to the number of RM recognition sites in the viral genome[70]. Indeed, phages tend to lose RM recognition sites[35,71], a phenomenon known as “restriction site avoidance.” As such, it has been proposed that RM systems afford bacteria a short-lived first-line

defense that may allow them to temporarily colonize new environments containing phage, but not to persist in them given the tendency of viral escape[57]. Perhaps bacteria rely on other phage defense systems, such as CRISPR-Cas systems, which are often found together with RM systems in a single host[36].

1.3 CRISPR-Cas systems: prokaryotic adaptive immune systems

Perhaps one of the most influential findings from the arms race between bacteria and phages is the discovery of bacterial "adaptive" immune systems termed CRISPR-Cas. In the late 80s and early 90s, microbiologists observed in prokaryotic genomes unusual DNA regions composed of short semi-palindromic sequences repeating multiple times and separated by variable sequences[72,73]. These genomic loci were named **CRISPR** (**c**lustered of **r**egularly **i**nterspaced **s**hort **p**alindromic **r**epeats) and were subsequently found to be frequently associated with signature *cas* genes[74]. In the early 2000s, the variable sequences (termed "spacers") were bioinformatically mapped to mobile genetic elements, such as phages[75]. This soon led to the hypothesis[75,76], and subsequent experimental demonstrations[77], that CRISPR-Cas systems provide defense against foreign nucleic acids including plasmids and bacteriophages[8,77–81].

Just as with RM systems 40 years prior, the RNA-guided endonucleases from these systems did not go unnoticed by scientists interested in repurposing these enzymes into tools for biotechnology applications. In 2012, it was demonstrated that

Cas9 (the signature Cas enzyme from Type II-A CRISPR-Cas systems), paired together with a short RNA guide, could cleave DNA in vitro[8,81]. Shortly after, Cas9 was used to cut DNA and mediate genome editing in human cells[82,83]. The adoption of CRISPR-Cas9 by the scientific community was instantaneous because the method is programmable and far more efficient to previous gene-editing tools, while remaining relatively simple and inexpensive. Today, the technique is used in a wide range of cell types and organisms in laboratories to characterize and study specific genes. CRISPR technologies also hold tremendous promise in clinical settings, particularly for diseases like cancer or sickle cell anemia, where a patient's cells can be edited *ex vivo* and transplanted back into the body to functionally cure the condition. Although a long-term goal with many far-reaching implications and ethical considerations, treating human genetic disorders with CRISPR technology is likely on the horizon.

1.3.1 The CRISPR immune response

CRISPR-Cas systems are present in about 40% of bacteria and in 85% of archaea[84]. They are composed of a CRISPR locus together with CRISPR-associated (*cas*) genes (**Fig. 1.4**). CRISPR loci are composed of an array of short partially palindromic repeat sequences (about 30 to 40 bp) separated by short unique DNA sequences matching foreign DNA, called spacers[73–75,85,86]. Additionally, a DNA sequence, known as the leader, is found upstream of the CRISPR array[74]. The spacer sequences represent an immunological record of past invaders. CRISPR-Cas systems utilize these spacer

sequences together with the Cas proteins to recognize and provide defense against invaders in future infections[77].

Despite structural and mechanistic differences between CRISPR-Cas systems, the CRISPR immune response can be generally divided into two major stages: spacer acquisition (sometimes referred to as “adaptation”) and interference (**Fig. 1.5**)[87]. During spacer acquisition, foreign DNA (e.g., phage or plasmid DNA) is sampled, and a short piece of DNA (known as a “prespacer”) is selected, processed, and integrated into the CRISPR array between two repeats, forming a new spacer[77,88]. Regardless of the CRISPR type, the Cas1-Cas2 integrase complex is always involved in this step[84,89–93]. Consequently, these two proteins are conserved in almost all CRISPR-Cas systems[84]. During the second phase of CRISPR interference, the CRISPR array is transcribed into a long precursor CRISPR RNA (a pre-crRNA) and processed into smaller mature crRNAs (CRISPR RNAs)[79,94,95]. One or multiple Cas proteins assemble with a crRNA to form an effector complex[84,96]. Upon subsequent infection with the same invader, the

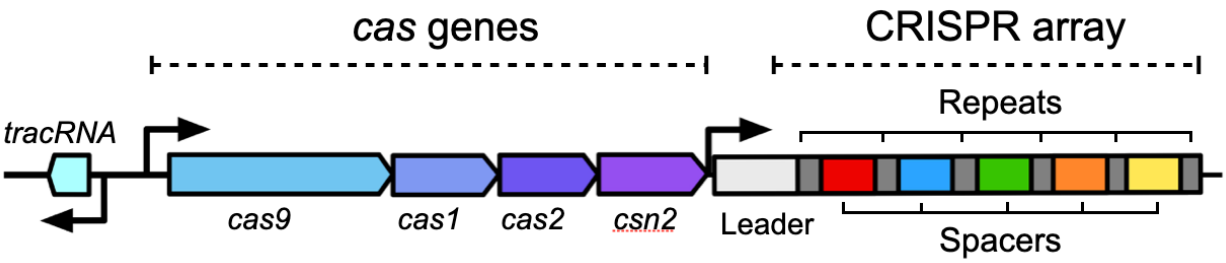


Figure 1.4 Schematic representation of the *Streptococcus pyogenes* type II-A CRISPR-Cas locus.

The CRISPR array is comprised of five spacers (in red, blue, green, orange, and yellow) separated by repeats (dark grey). Upstream of the CRISPR array is the leader sequence which is important for the integration of the new spacers. This system also harbors four *cas* genes (*cas9*, *cas1*, *cas2*, and *csn2*) and a small non-coding RNA, tracRNA.

effector complex is guided by its crRNA to the matching DNA in the invader's genome, also known as the "protospacer". Base pairing of the crRNA to the protospacer sequence triggers DNA cleavage by the effector complex, preventing completion of the infection[8,78–81]. Therefore, CRISPR-Cas systems require an initial exposure to a virus to first become "immunized" (through the acquisition of new spacers), to mediate defense during a later exposure (**Fig. 1.5**). As such, unlike RM systems which are considered innate defenses, CRISPR-Cas systems are described as adaptive immune systems.

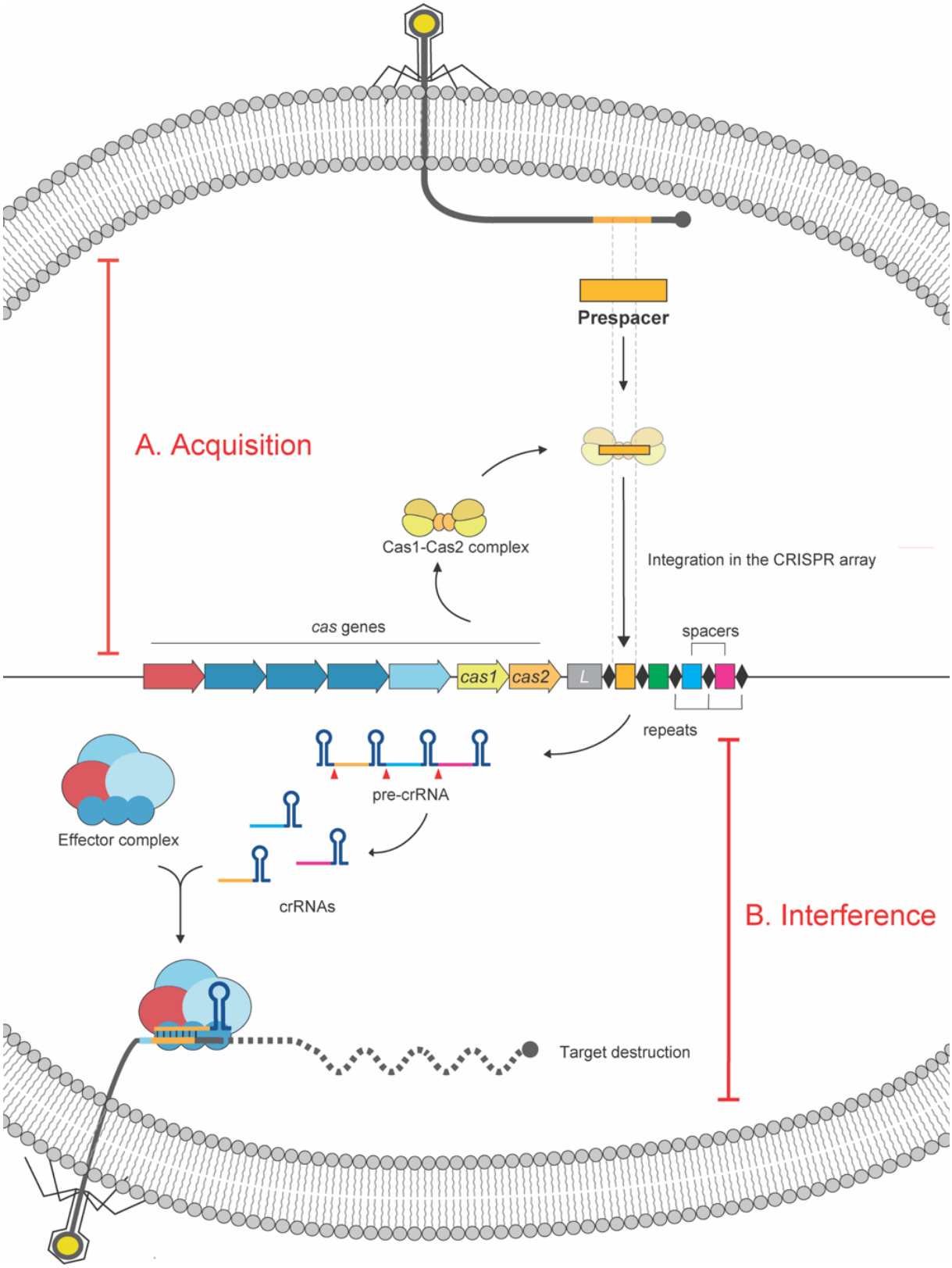

Prespacer
A. Acquisition
Cas1-Cas2 complex
Integration in the CRISPR array
cas genes
spacers
cas1
cas2
L
repeats
pre-crRNA
Effector complex
crRNAs
B. Interference
Target destruction

Figure 1.5 Representation of a canonical CRISPR immune response against bacteriophage.

The CRISPR immune response is generally broken down into two stages: spacer acquisition and interference. (**A**) During acquisition, a prespacer (yellow rectangle) is selected from the viral DNA. The prespacer is integrated into the CRISPR array by the Cas1-Cas2 integrase forming a new spacer. (**B**) During interference, the CRISPR array is transcribed into long precursor CRISPR RNAs (pre-crRNAs). which are further processed into single CRISPR RNAs (crRNAs). The mature crRNAs associated with one or multiple Cas proteins to form effector complexes. Guided by their crRNAs, the effector complexes recognize and destroy the viral DNA. Reproduced with modification from Nussenzweig and Marraffini (2020)[96] with permission from the publisher.

1.3.2 A brief overview of CRISPR-Cas types and interference mechanisms

CRISPR-cas systems are divided into two classes (I and II), six types (I-VI), and multiple subtypes based on their *cas* gene content[84]. Systems belonging to class I mediate interference using multiple Cas proteins in a complex[84]. In contrast, class II systems provide defense using a single Cas protein[84]. Additionally, some systems target DNA (type I, II, and V), RNA (type VI), or both (type III) (**Fig. 1.6**).

Class I CRISPR-Cas are grouped into three types: I, III, and IV[84]. Type I systems target viral DNA using a multi-subunit crRNA Cas complex called the Cascade complex[79]. Guided by the crRNA, the Cascade complex recognizes the protospacer in the viral DNA[97]. This recruits the helicase/nuclease Cas3, which cleaves the target DNA[98]. Why doesn't Cascade bind to the CRISPR locus, which contains matching spacer sequences? In addition to the requirement for sequence complementarity between the protospacer and the crRNA, a conserved protospacer-adjacent motif

(PAM) on one side of the target is required for DNA cleavage[97]. Because the CRISPR repeats lack the PAM, the requirement for a PAM for targeting prevents autoimmunity. In contrast to type I, the type III systems cleave both DNA and RNA in a PAM independent manner and require transcription across the target for immunity[99–102]. One unique feature of type III systems involves an oligoadenylate second messenger signaling system that activates various ancillary effectors with diverse functions that further contribute to immunity[103–107]. The last Class I system is type IV[84]. Little is known about the biology of this CRISPR type. These systems are predicted to form an effector complex comprised of multiple Cas proteins and may target DNA[84]. Additionally, a bioinformatic search of the spacer sequences they contain suggests that they may preferentially target plasmids[108]. Indeed, one recent study showed that a type IV

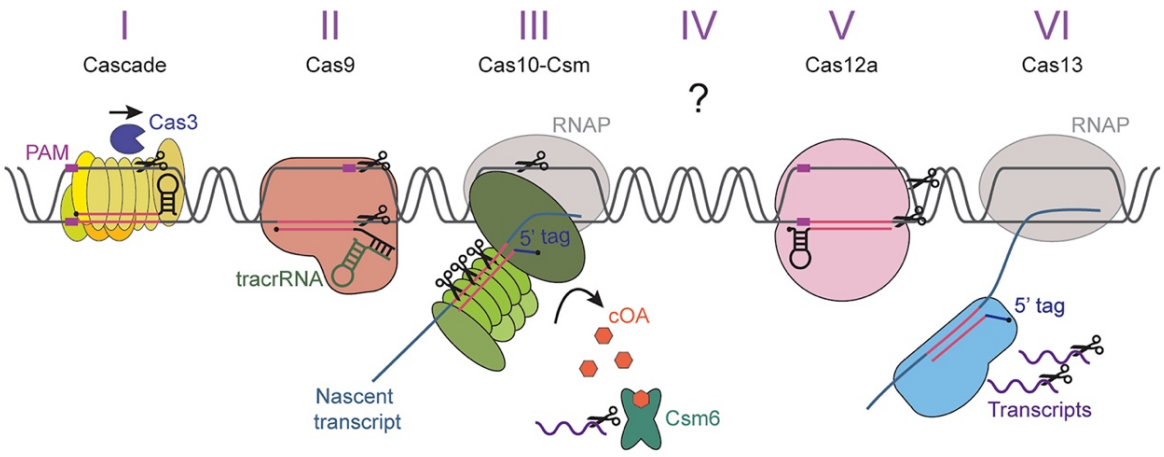


Figure 1.6 Different interference mechanisms by each CRISPR type.

On top are each CRISPR type and the name of their main effector protein/complex. Type I, II, and V systems recognize and cleave foreign DNA in a PAM (purple boxes) dependent manner. Type III systems recognize foreign RNA and cleaves both DNA and RNA without a PAM requirement. Type VI systems recognize viral RNA and cleave both viral RNA transcripts and host transcripts upon activation. As with type III systems, the RNA cleavage activity of type VI systems is PAM independent. The mechanism of DNA interference by type IV systems is unknown. Reproduced with modification from Rostøl and Marraffini (2019)[22] with permission from the publisher.

CRISPR-Cas system from *Pseudomonas aeruginosa* could prevent plasmid uptake when harboring a spacer matching the plasmid[109].

Class II CRISPR-cas systems are divided into three types: II, V, and VI[84]. These types use a crRNA complexed with a single effector protein to mediate immunity[84]. The effector proteins for type II, V, and VI are Cas9, Cas12, and Cas13, respectively[110–112]. An additional trans-activating small RNA (tracrRNA) annealed to the crRNA is required for Cas9 and some Cas12 nucleases to mediate defense[84,113,114]. Both Cas9 and Cas12 mediate the cleavage of DNA matching their associated crRNAs[8,81,111]. As with type I systems, recognition of the protospacer by Cas9 and Cas12 require a PAM sequence on one side of the target[111,115,116]. Finally, whereas Cas9 and Cas12 cleave DNA, Cas13 recognizes and targets RNA in a PAM independent manner[112,117]. As such, Cas13 requires transcription of the foreign DNA for base pairing of the crRNA to the target transcript, which triggers cleavage of the bound RNA as well as of non-target transcripts[112,117].

The remaining sections of this chapter will now focus on detailed aspects of type II-A CRISPR-Cas systems because they are the focus of this thesis work and a primary interest of the laboratory.

1.3.3 Type II-A CRISPR-Cas systems

Type II systems represent about 13% of all CRISPR-Cas systems in bacteria and are absent in archaea[118]. Despite being one of the least represented CRISPR types, type II systems are one of the most studied since their signature effector protein, Cas9, has been widely repurposed for genome editing[31,82,83,119]. All type II systems contain *cas9*, *cas1,* and *cas2 genes* and a trans-activating CRISPR RNA (tracrRNA)[84,113,114]. Depending on their subtype, they can carry additional genes, such as *csn2* or *cas4*[84]. Type II-A harbors *csn2*, whose product is essential for the spacer acquisition stage of the CRISPR immune response, although the exact role and mechanism remain to be fully elucidated (**Fig. 1.4**)[91,92].

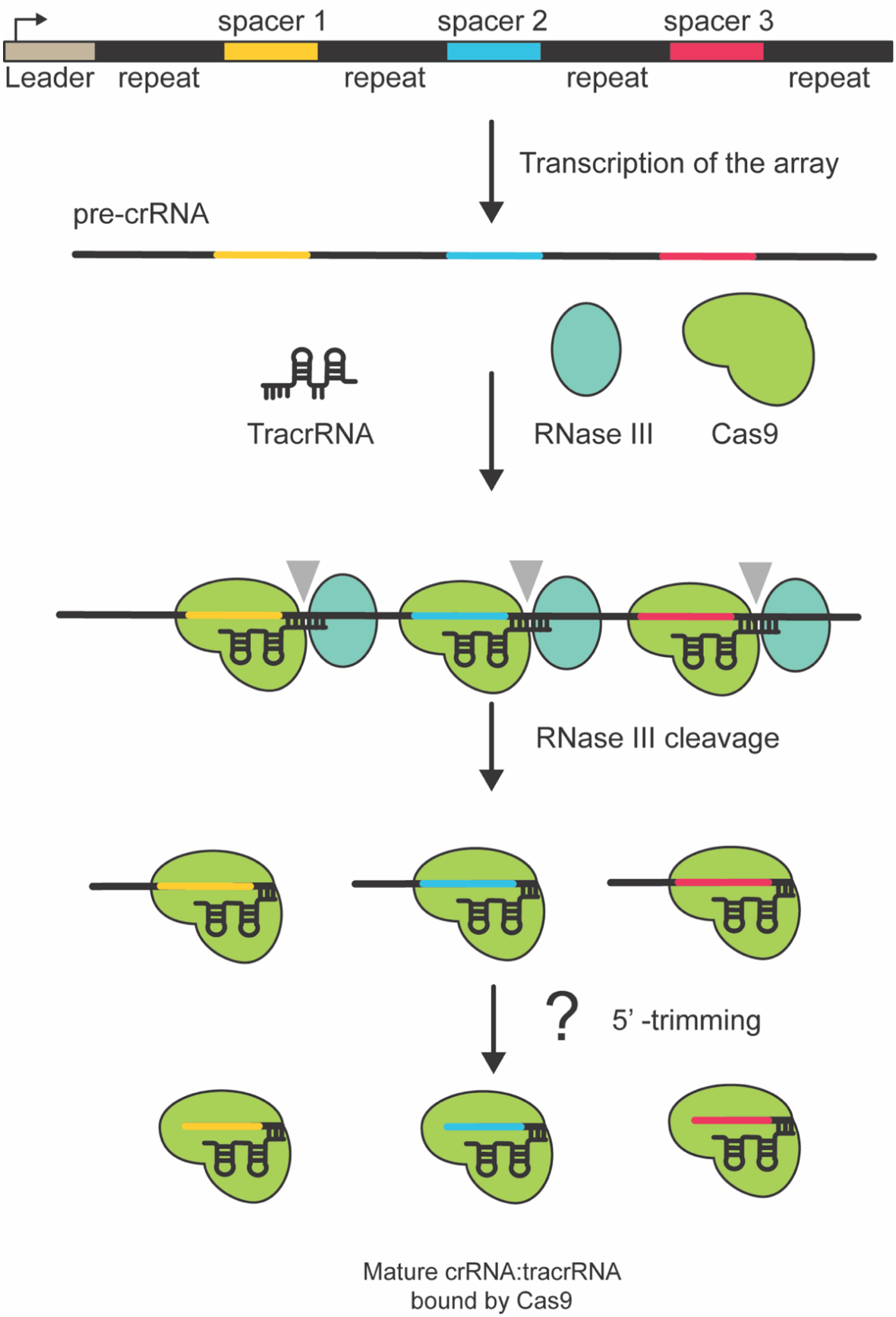
1.3.4 crRNA maturation in type II-A CRISPR-Cas systems

To mediate interference, the spacer sequences need to be transcribed and processed into mature crRNAs to guide Cas9 to the matching protospacers in the foreign DNA. First, the CRISPR array is transcribed (usually from an upstream promoter in the leader region) into one long precursor RNA called a pre-crRNA[113,114]. Then, the pre-crRNA is further processed into individual crRNAs representing each spacer sequence in the array (**Fig. 1.7**)[113]. Unlike Class I CRISPR-Cas systems, Class II systems do not have a Cas protein dedicated to this process[94]. Instead, they rely on their effector protein (Cas9 for type II-A systems) to mediate crRNA processing. Additionally, in order to recognize and process the pre-crRNA among all the other

RNAs in the cell, type II-A systems utilize a trans-activating CRISPR RNA (tracrRNA)[113,114,120]. TracrRNA is a small non-coding RNA forming stem-loop structures and containing a short region complementary to the repeat sequence of the CRISPR array[113,114]. Upon transcription of the array into the long pre-crRNA, tracrRNAs anneal to each repeat unit in the pre-crRNA, forming RNA duplexes. Following the formation of tracrRNA:pre-crRNA repeat duplexes, the host RNase III, in the presence of Cas9, cleaves each duplex (**Fig. 1.7**)[113]. This releases individual immature crRNAs containing the full-length spacer flanked on either side by half of a repeat. The released immature crRNAs are then further trimmed on the 5' side, resulting in a mature crRNA with spacer-derived 5' sequence and a repeat-derived 3' sequence.[113]. It is hypothesized that host ribonucleases perform this last trimming step but the actors governing this step are still unknown[113]. Following crRNA maturation, Cas9 bound to the newly formed crRNA:tracrRNA duplex can mediate target recognition and cleavage.

1.3.5 Cas9 DNA cleavage

To recognize and cleave DNA, Cas9 must be bound to a crRNA:tracrRNA duplex and the target DNA must harbor a complementary sequence to the crRNA along with a protospacer-adjacent motif (PAM) [8,81,113]. The PAM requirement allows Cas9 to



Mature crRNA:tracrRNA
bound by Cas9

Figure 1.7. crRNA biogenesis in type II-A CRISPR-Cas systems.

The CRISPR array is transcribed into a precursor CRISPR RNA (pre-crRNA). The tracrRNAs bind to each repeat sequence (black lines) in the pre-crRNA. In the presence of Cas9 (green shape), the tracRNA:pre-crRNA duplexes are cleaved by the host RNase III (turquoise shape). The 5' side of each immature crRNA is further trimmed by an unknown enzyme to form a mature crRNA. Cas9 bound to a tracrRNA:crRNA can now mediate DNA recognition and cleavage.

differentiate a protospacer in the invading DNA from the spacers in the CRISPR array (the repeats flanking each spacer in the locus do not have a PAM). This prevents Cas9 from binding and cleaving the CRISPR array. For *S. pyogenes* Cas9, the PAM sequence is 5'-NGG-3' located directly on the 3' side of the protospacer[8,115]. Cas9 is a multidomain enzyme adopting a bilobed architecture with a nuclease (NUC) lobe and a recognition (REC) lobe[121,122]. The NUC lobe consists of 3 domains: two nuclease domains (RuvC and HNH) and a PAM-interacting domain[121,122]. The HNH and RuvC nuclease domain cleave the target and non-target strands, respectively, resulting in a dsDNA break[8,81]. Cas9 recognition of an invader progresses through two steps[116]. First, it searches for PAM sequences. Second, upon locating a PAM sequence, Cas9 checks for complementarity between the crRNA and the DNA sequence just upstream of the PAM.

ApoCas9 is inactive but upon forming a complex with a crRNA:tracrRNA duplex, Cas9 goes through a large conformational change enabling PAM search and target recognition(**Fig. 1.8**)[122,123]. Through 3-dimensional collision, Cas9 binds DNA and searches for PAMs[116]. Recognition of a PAM sequence on the non-target strand results in a local DNA melting directly next to the PAM, where Cas9 starts interrogating

the DNA using the crRNA (**Fig. 1.8**)[116,124]. Formation of an R-loop by base pairing of the crRNA to the target DNA while displacing the non-target strand proceeds stepwise from the PAM in the 3' to 5' direction through the target sequence[116]. If mismatches are present in the first 8 to 12 PAM-proximal nucleotides (a region called the seed), Cas9 does not cleave[8,115,116]. Otherwise, RNA strand invasion proceeds stepwise until the end of the crRNA. This progression causes structural changes in the HNH domain[125] resulting in simultaneous DNA cleavage by the RuvC and HHN domain[125].

The initial recognition of a short PAM sequence followed by recognition of a seed region in the target sequence allows Cas9 to reduce the complexity of the search and to scan foreign DNA much more efficiently. However, this efficient search strategy comes at a price. Bacteriophage containing single-nucleotide mutations in PAM or seed sequences can escape Cas9 targeting[77,88]. As such, the type II CRISPR-Cas systems require the frequent acquisition of new spacers to diversify the spacer pool in the culture and to prevent mutant bacteriophage from completely escaping Cas9 targeting[126,127]

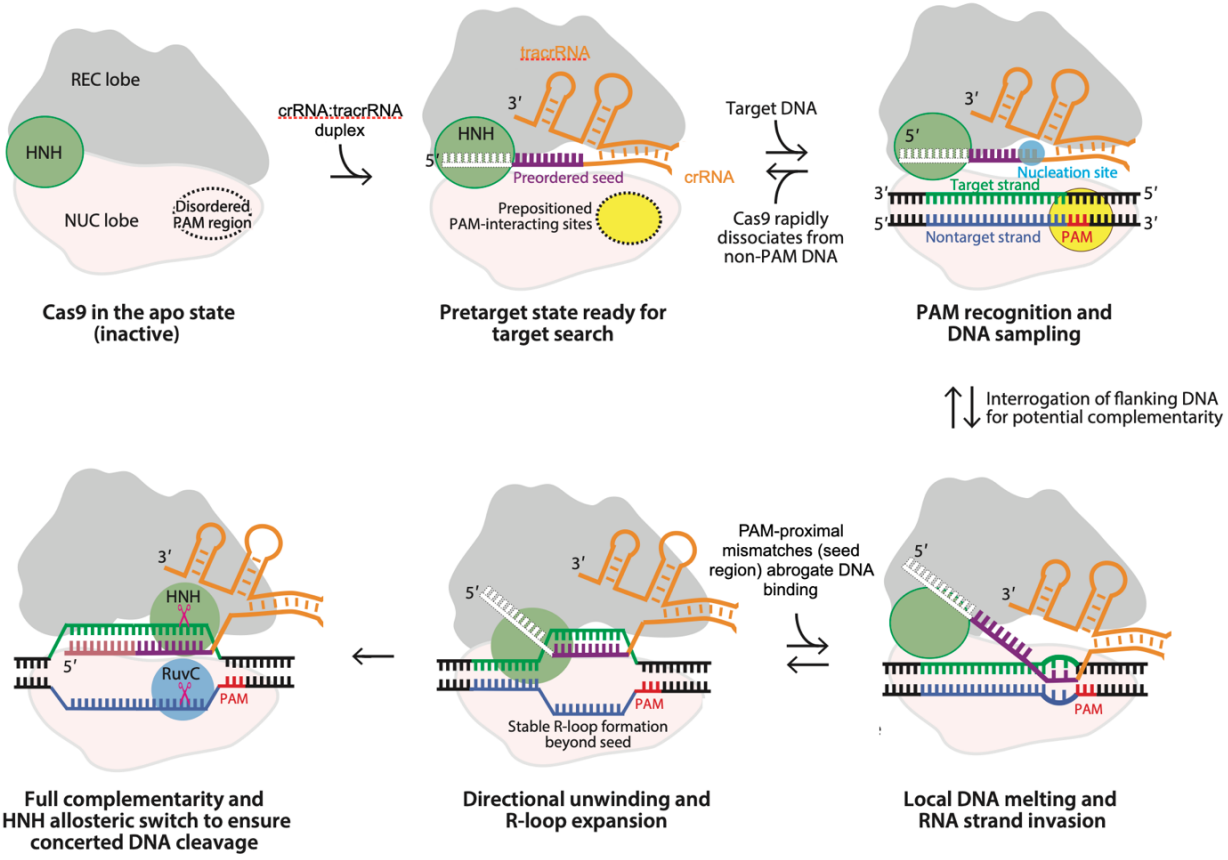


Figure 1.8 Cas9 DNA cleavage mechanism.

ApoCas9 is inactive, and its PAM-interacting domain is disordered. Once bound to a tracrRNA:crRNA duplex, Cas9 goes through a conformational change, which allows for target search. Cas9 binding to a PAM sequence (in red) in the target DNA incudes local melting of the DNA just upstream of the PAM. This allows for base pairing of the seed region of the crRNA (purple) with the DNA target strand (in blue), which displace the non-target strand (green) forming an R-loop. Mismatches in the seed region abrogate Cas9 binding and cleavage. Annealing of the whole crRNA triggers DNA cleavage by the RuvC and HNH domain. Reproduced from Fuguo and Doudna (2017)[212] with permission from the publisher.

1.3.6 Spacer acquisition in type II-A CRISPR-Cas systems

Spacer acquisition is the hallmark of the CRISPR-Cas adaptive immune response. The ability of CRISPR-Cas systems to keep a record of past infection and to continuously update it through the acquisition of new spacers is truly unique among all known prokaryotic defense systems. While the molecular events orchestrating interference by type II-A systems are now well defined, some aspects of spacer acquisition remains poorly understood. In contrast to the interference stage, which requires only Cas9 and a tracrRNA:crRNA duplex, all the components of these systems are needed for acquisition (*cas9*, *cas1*, *cas2*, *csn2*, and tracrRNA)[91,92]. Spacer acquisition can be separated into two broad steps[128]. The first is the selection and capture of protospacer sequences from foreign DNA, such as bacteriophages or plasmids. The second is the integration of "prespacers" into the CRISPR array to form new spacers.

1.3.7 Prespacer integration

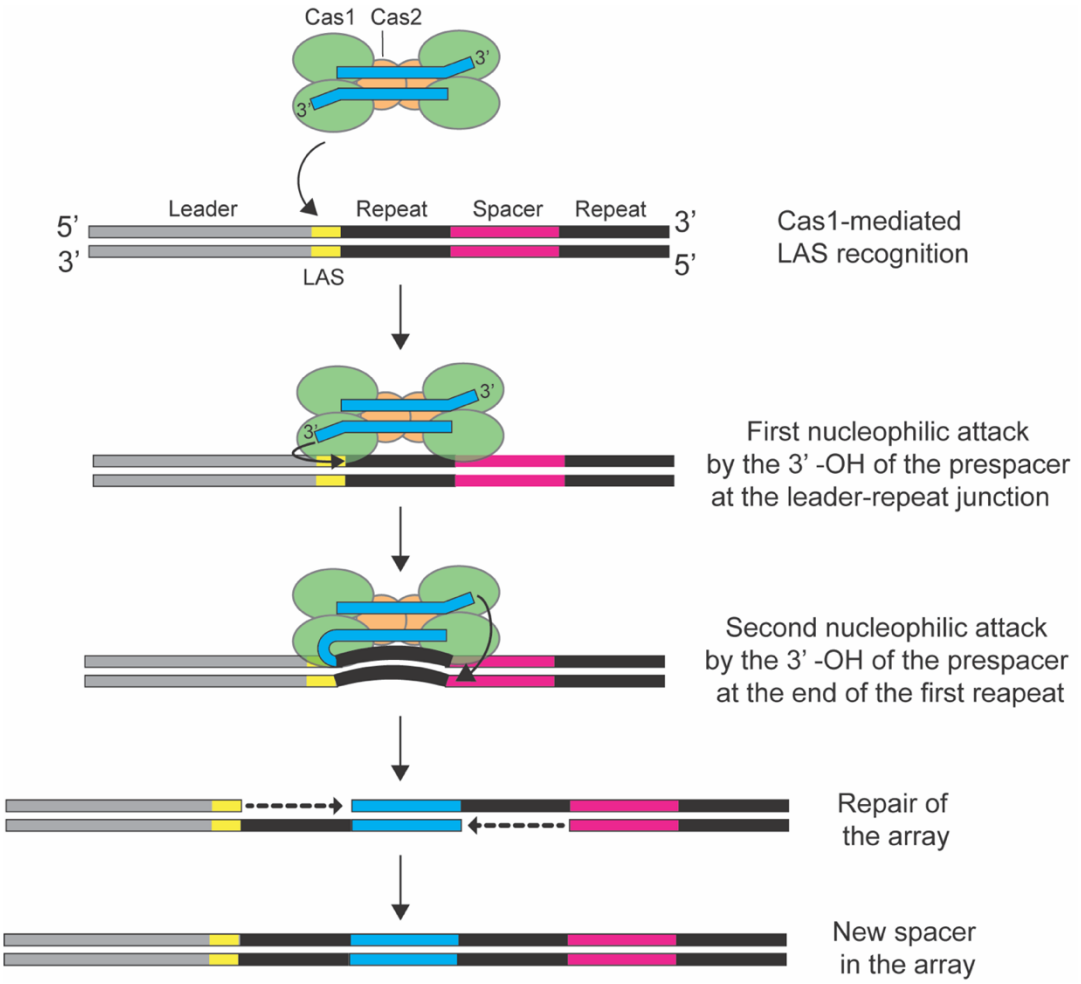
While *in vivo* all the *cas* genes are required to orchestrate full spacer acquisition (from prespacer selection to integration) in type II-A CRISPR-Cas systems[91,92], *in vitro* *cas1* and *cas2* are sufficient for spacer integration[129–131]. Moreover, the addition of Csn2 and Cas9 inhibited prespacer integration *in vitro*[130]. This suggests Cas1 and Cas2 are responsible for integration and that Csn2 and Cas9 are involved in steps prior to integration, during prespacer selection from foreign DNA.

Cas1 and Cas2 are conserved across all the CRISPR types[84]. Together they form a heterohexameric integrase complex comprised of two Cas1 dimers on each side of one Cas2 dimer[93,131] (**Fig. 1.9**). *In vitro*, the Cas1-Cas2 complex preferentially binds and integrates prespacers composed of a short DNA duplex with small 3' overhangs (4 to 5 nt)[129–133]. Spacer integration is a two-step integration process mediated by two separate nucleophilic attacks by Cas1 using the 3'-OH on each strand of the prespacer (**Fig. 1.9**)[130]. The first attack is at the leader-repeat junction and the second at the end of the first repeat. As such, spacer acquisition is polarized with a prespacer always being integrated as the first spacer. To coordinate this specific integration, Cas1 in the Cas1-Cas2 complex recognizes a short region at the beginning of the repeat but importantly, it also recognizes a short DNA sequence at the end of the leader (the leader anchoring-sequence or LAS) (**Fig. 1.9**)[129,130,134,135]. Recognizing both the LAS and the beginning of the repeat favors the first nucleophilic attack at that site[134,136]. Indeed, mutating the LAS sequence resulted in spacer acquisition at other locations in the array[134,135]. While in some type I systems, the host factor IHF is required for DNA bending of the leader sequence to specifically integrate new spacers at the leader-repeat junction[137], no known host factor has been implicated for type II-A systems so far. Following the first integration at the leader-repeat junction, Cas1-Cas2 recognizes a region at the end of the repeat sequence, which acts as a molecular ruler to guide the second integration at the end of the repeat[129–131]. This ensures that the repeats are correctly duplicated. These two integration events split the first repeat of

the array into two single-stranded repeats with a dsDNA spacer in the middle (**Fig. 1.9**). The mechanisms by which the ssDNA repeats are converted to dsDNA and ligated to resolve the array post-integration remain poorly understood. A recent single-molecule study using Cas1-Cas2 from the *E. faecalis* type II-A CRISPR system showed that a novel transcription-assisted DNA-repair mechanism could resolve the array post-integration[138]. However, additional *in vivo* validation is needed.

Cas1-Cas2 site-specific integration is important to avoid genome instability and prevent DNA from being integrated randomly throughout the genome. Additionally, polarized acquisition has also been shown to be important for immunity. It was shown in the *S. pyogenes* type II-A CRISPR-Cas system that spacers at the beginning of the array provided better defense against bacteriophage than spacers located further downstream[134]. This may explain why polarized acquisition is conserved among CRISPR-Cas systems[96]. In the same study, the level of crRNA transcripts was lower for spacers located downstream in the array than for spacers at the beginning of the array, which may explain the difference in immunity[134]. This difference in transcription was also observed in other CRISPR-Cas systems[113,139–141]. This may result in a larger pool of Cas9 loaded with crRNAs matching the first spacers of the array in one cell. As such, polarized acquisition coupled with higher transcription of spacers at the beginning of the array could represent a strategy by CRISPR-Cas systems to deploy

a stronger immunity against recently encountered viruses that are more likely to be in the environment and to re-infect the bacteria.



1.3.8 Selection of prespacers with a PAM sequence

Figure 1.9 . Prespacer integration by Cas1-Cas2.

The Cas1-Cas2 integrase binds a short DNA duplexed with short 3' overhangs, the prespacer (blue). Cas1 in the complex recognizes the LAS (yellow) in the leader (grey), and the beginning the repeat sequence. This recruits the complex at the leader-repeat junction to catalyze the first 3' -OH nucleophilic attack resulting in the ligation of one strand of the prespacer. Then, the complex mediates the second nucleophilic attack at the repeat-spacer junction by recognizing the end of the repeat sequence, ligating the other prespacer strand. Post-integration, the structure is resolved by unknown host DNA repair mechanisms, resulting in the formation of a new spacer (blue) at the beginning the array.

Before integration into the array, a suitable prespacer sequence must be selected and processed for integration. In type II-A CRISPR-Cas systems, the molecular events orchestrating this selection are some of the least understood. In these systems, the selected prespacer must be flanked by a PAM to become a functional spacer[8,91,116]. While *in vitro* Cas1 and Cas2 are sufficient for spacer integration[130,131], *in vivo* every Cas protein (Cas9, Cas1, Cas2, and Csn2) is required for functional spacer acquisition[91,92]. This suggests that Cas9 and Csn2 play a role before integration during prespacer selection. Indeed, Cas9 was shown to be responsible for the selection of prespacers with a correct PAM *in vivo*[91,92]. Mutating the PAM-interacting domain of Cas9 resulted in PAM-less spacers being acquired[91]. As such, Cas9 uses the same PAM-interacting domain for DNA cleavage during interference and spacer selection during acquisition. On the other hand, Cas9's nuclease activity is not required for prespacer selection[91,92]. Cas9 and Csn2 have been shown to form a supercomplex with the Cas1-Cas2 integrase[91,142,143], likely representing the link between prespacer selection and integration. However, after selection of a prespacer, Cas9 and Csn2 may have to disassemble from the Cas1-Cas2 integrase, as their presence inhibited prespacer integration *in vitro*[130,142]. Csn2's exact role in spacer selection remains unknown. *In vitro*, Csn2 assembles into a tetrameric ring that binds dsDNA ends and can slide along the DNA[144–147]. A recent structural study of the supercomplex indicates that Csn2 interacts directly with the Cas1-Cas2 complex and that it forms a ring with DNA in its center[142]. Additionally, the study shows that upon treatment of the supercomplex with DNase, Cas9 was no longer

associated with the complex suggesting that DNA may tether Cas9 to the complex[142]. As such, the authors put forward a speculative model in which a Csn2-Cas1-Cas2 complex can engage onto a dsDNA end using Csn2 and slides along the DNA until encountering a Cas9 molecule bound to a PAM sequence. This would trigger a chain of events resulting in the hand-off of the Cas9 selected prespacer to the Cas1-Cas2 complex for integration. While it is an interesting model that fits with the limited data available on prespacer selection, a significant amount of work will be required to confirm its validity and to uncover the molecular details governing it.

Apart from PAM selection, the processing of the selected DNA into a prespacer of the correct spacer size for integration in the CRISPR array is currently not understood for type II-A systems. One study found that *in vitro* Cas9 in the supercomplex can process long prespacer DNA, using the HNH nuclease domain, into prespacers of the correct length (30 bp) for integration[148]. However, the same study showed that *in vivo*, mutating the HNH domain of Cas9 still resulted in the integration of spacers of the correct size[148]. This suggests that perhaps *in vivo* additional host factors can perform prespacer trimming, similarly to some type I CRISPR-Cas systems[149,150]. However, this possibility remains to be demonstrated.

1.3.9 Selection of prespacers from foreign DNA

In addition to the PAM requirement, the prespacer must also be selected from foreign DNA. Indeed, autoimmunity by the acquisition of self-targeting spacers (derived

from the host genome) by type II-A CRISPR-Cas systems is toxic to the cells[92,151]. While RM systems can innately recognize invaders by the modification state of the DNA, naïve CRISPR-Cas systems (systems without a pre-existing spacer matching the invader DNA) cannot discern host from foreign DNA. As such, strategies must be in place to bias acquisition toward foreign DNA. From recent studies, it is becoming evident that CRISPR-Cas systems across multiple types (I, II, and III) can use free DNA ends for spacer acquisition[152–154]. Looking at self-acquisition events, independent groups have noticed that the acquired spacer sequences preferentially matched regions of the bacterial chromosome often prone to dsDNA breaks[153–155]. This was confirmed by restriction of the host chromosome by the yeast meganuclease I-sceI, which resulted in a "hotspot" in self-acquisition at the site of the DNA break[153–155]. Additionally, these DNA ends can be degraded by the host dsDNA repair complex (RecBCD or AddAB, in Gram-negative or Gram-positive bacteria, respectively), to create additional substrates enhancing spacer acquisition[153–155]. In their hosts, the RecBCD and AddAB complexes mediate the processing of dsDNA ends for repair by homologous recombination[156]. These complexes recognize DNA ends and degrade both strands into ssDNA fragments until encountering a specific short sequence known as a *chi* site, upon which DNA degradation stops and repair is performed[156]. As such, spacer acquisition at dsDNA breaks is limited by the nearest chi sites[153–155]. Consequently, it is thought that *chi* sites, which are enriched in host genomes compared to bacteriophage genomes, limit spacer acquisition from the host and result in biased acquisition from foreign DNA ends[155]. With the observation that the

substrates for the type II-A Cas1-Cas2 integrase are dsDNA fragments[130,131], it remains unknown how the processing of free DNA ends by RecBCD and AddAB enhances spacer acquisition because their degradation products are ssDNA fragments[156]. An alternate explanation is that their degradation products are not substrates for spacer acquisition, but instead, RecBCD and AddAB may generate additional free DNA ends by sporadically falling off the DNA. In turn, these could be the point of entry of the supercomplex on the DNA through Csn2, which *in vitro* binds dsDNA ends and can slide along the DNA[144,145]. However, this has not been validated experimentally and remains to be demonstrated.

Harnessing dsDNA ends for acquisition can also favor prespacer selection toward bacteriophages as most of them have linear dsDNA genomes[11], while bacterial genomes are circular. Indeed, our laboratory demonstrated that the *S. pyogenes* type II-A CRISPR-Cas system preferentially acquires new spacers from the DNA end of a linear viral genome entering the cell upon infection[154]. Although not strictly required, AddAB degradation of the viral end enhanced acquisition during this process[154]. However, acquisition of spacers from bacteriophage in this system is rare under laboratory conditions, estimated to occur in 1 out of 10^7 cells[91]. This suggests that spacer acquisition from the DNA end of the viral genome happens in just a small subset of bacteria. This is consistent with the fact that bacteriophages have evolved ways to protect their viral DNA ends by encoding host exonuclease inhibitors and by circularizing their genomes quickly after entering the cells[157–159]. As such, it is likely that type II-A CRISPR-Cas systems also rely on dsDNA ends generated by other

processes. One known mechanism is utilized by CRISPR-Cas systems themselves using pre-existing spacers to cleave the phage DNA. This type of acquisition is referred to as primed acquisition, and it has mostly been studied in type I-E and I-F CRISPR-Cas systems[90,160]. However, recently it was demonstrated that type II-A CRISPR-Cas systems could also perform primed acquisition[127,161,162]. During this process, a pre-existing spacer is transcribed into a crRNA that guides Cas9 to the phage genome resulting in DNA cleavage and the generation of two dsDNA ends. These DNA ends are then used for the acquisition of new spacers[161]. As such, the spacers acquired after Cas9 cleavage matches regions of DNA directly next to the cut site[161]. This may represent a way for type II-A systems to pre-emptively acquire additional spacers to defend themselves against the rise of bacteriophages that possess a mutated target sequence. However, this type of acquisition relies on the cells harboring pre-existing spacers matching the foreign DNA. Therefore, this does not explain how naïve cells may recognize and acquire spacers without relying on the free dsDNA end of an incoming linear viral genome[154]. One possibility is that naïve type II-A CRISPR-Cas systems rely on other DNA cleaving defense systems to generate free DNA ends for spacer acquisition. This may also provide some explanation regarding the timing of spacer acquisition against bacteriophage. It is hard to imagine that during one infection cycle, the bacterium has the time to select and integrate a new spacer to transcribe it into a crRNA and guide the effector complex to that same invader for destruction. It seems more likely that a bacteriophage is first disarmed by some other mechanisms,

providing an opportunity for CRISPR to acquire a spacer to use in a subsequent infection by that invader.

In this thesis we explore this idea by investigating the interplay between RM and CRISPR-Cas systems. We observed that RM systems can enhance the type II-A CRISPR-Cas immune response and we characterized the molecular events governing this synergy. Interestingly, this interaction is reminiscent of eukaryotic immunity, where the innate response (by RM systems) offers a first-line temporary defense, and also activates a second, more robust adaptive response (by CRISPR-Cas systems).

CHAPTER 2. THE RESTRICTION ACTIVITY OF RM SYSTEMS ENHANCE THE TYPE II-A CRISPR-CAS IMMUNE RESPONSE.

2.1 Background

Restriction-Modification (RM) systems (reviewed in Chapter 1.2) are bacterial defense systems providing innate immunity against bacteriophages. Typically, these systems encode endonuclease and methyltransferase activities that “restrict” and “modify”, respectively, the same short DNA sequence[35]. The methylation of chromosomal target sites inhibits cleavage by the restriction enzyme to prevent attack of “self” DNA. On the other hand, unmodified sites on the phage, or “foreign”, genome are recognized and cleaved by the restriction endonuclease to provide anti-viral defense. This discrimination strategy enables an innate defense because RM systems can restrict viruses never encountered before so long as they harbor RM recognition sites. However, this innate immunity can easily be overcome by phages, as methyltransferases can erroneously modify incoming viral genomes before restriction occurs (See Chapter 1.2.3 and **Fig. 1.3**)[57,58,69]. This allows the completion of the phage’s lytic cycle and the release of viral progeny with modified genomes, which will go on to infect and kill the rest of the bacterial population.

Clustered regularly interspaced short palindromic repeats (CRISPR) loci and their associated genes (*cas*) constitute another defense system that cleaves foreign DNA (reviewed in more detail in Chapter 1.3). CRISPR-Cas systems, unlike RM systems,

provide adaptive immunity. The CRISPR locus contains short DNA repeats (30-40 nucleotides), separated by equally short unique sequences of viral or plasmid origin, called spacers[75,85,86], which are acquired during infection[77]. These are transcribed and processed into short RNA guides, the CRISPR RNAs (crRNAs) that mediate, through base pair complementarity, target recognition, and cleavage by Cas nucleases[79,163,164]. Depending on the *cas* gene content, CRISPR-Cas systems can be classified into six types, with type II being one of the most stdued.[84]. Type II-A CRISPR systems are thought to use free double-stranded DNA (dsDNA) ends as substrates for new spacers[154,161], which are recognized and incorporated into the CRISPR locus by the Cas9-Cas1-Cas2-Csn2 supercomplex[91,131,142,148]. Finally, immunity is achieved by the introduction of dsDNA breaks (DSBs) on the phage genome by the RNA-guided nuclease Cas9[80].

Despite being two of the most well-studied defense systems, that often cohabit in the same host[36], only two studies explored the possibility of interactions between RM and CRISPR-Cas systems. The first found that in *Streptococcus thermophilus* with a type II-A CRISPR-Cas locus carrying spacers against phage Φ2972, the propagation of unmodified viruses was further reduced by three orders of magnitude during heterologous expression of the *Lactococcus lactis* type II RM system LlaDCHI[165]. Southern blots showed cleavage of the viral genome by both systems, further demonstrating that these defense mechanisms can work together to increase immunity. Using the same experimental system, the second study revealed that infection of naïve bacteria; i.e., without phage-targeting CRISPR spacers, with a mixed

population of methylated and unmethylated phage increased the number of CRISPR-resistant colonies, with a proportional correlation to the amount of unmodified Φ2972[166]. This result suggested that the inactivation of the viral lytic cycle through restriction prevents the irreversible damage of the host cell and at the same time enables the acquisition of new spacers from the inactivated phage. However, the molecular mechanisms connecting CRISPR-Cas and RM systems remain unknown. For example, it is not known whether RM systems enhance CRISPR spacer acquisition simply because they prevent the completion of the phage's life cycle. This could allow the CRISPR system time to sample the viral DNA to select and integrate a new spacer without succumbing to the infection. However, another possibility could be that on top of preventing the lysis of the host, RM systems may create, through their restriction activities, the DNA substrates needed by the CRISPR acquisition machinery to generate new phage-targeting spacers.

Our laboratory has previously established a heterologous system in *Staphylococcus aureus* RN4220 to study spacer acquisition by the *Streptococcus pyogenes* SF370 type II-A CRISPR-Cas system[91]. The gram-positive bacterium *S. aureus* was selected because it offers several advantages. First, many bacteriophages have been isolated for *S. aureus*, which can be used to probe phage defense systems[167]. Additionally, multiple plasmids are available enabling easy expression and study of specific genes. Finally, our laboratory chose the laboratory strain RN4220 because it is easily transformed with plasmids, but most importantly, this strain does not harbor any known active phage defense systems[168,169]. This makes this strain an

ideal host to study bacterial immune systems in isolation. The native type II-A CRISPR-Cas system from SF370 *S. pyogenes* was cloned on a plasmid vector (pCRISPR) by a previous graduate student, Dr. Wenyan Jiang. Additionally, ΦNM4γ4, a lytic only mutant of the temperate staphylococcal phage ΦNM4 (member of the *Siphoviridae* family) was generated by Dr. Gregory Goldberg. Using this plasmid and bacteriophage system, the laboratory has previously uncovered multiple aspects of spacer acquisition by *S. pyogenes* Type II-A CRISPR-Cas system[91,96,127,151,154,170]. As such, we decided to use this system to investigate the interplay between RM and CRISPR-Cas systems.

2.2 The SauI RM system provides temporary defense against phage

S. aureus RN4220 harbors a type I RM system, SauI, encoding *hsdM* (methyltransferase), *hsdS* (sequence specificity factor), and *hsdR* (restriction endonuclease) genes[171]. Two complexes with different subunit stoichiometries and activities are formed: M_2S_1, which acts as a methyltransferase capable of modifying a specific sequence defined by its S subunit (5'- $CCAYN_6TGT$-3' and 5'-$ATCN_5CCT$-3')[42], and $R_2M_2S_1$, which initiates DNA translocation upon binding to unmethylated sites[41], cleaving the DNA at a variable distance from the recognition site. However, *S. aureus* RN4220 contains a single nucleotide mutation in the *hsdR* gene that results in a premature stop codon[168], thus preventing type I DNA restriction in this strain.

To investigate the relationship between RM and CRISPR at the molecular level, we restored type I restriction by cloning an intact copy of the *hsdR* gene into the

plasmid pLZ12[172], and introduced the resulting plasmid, pSauI, into RN4220 cells. We tested restriction of the staphylococcal phage ΦNM4γ4[99] (which contains 26 SauI recognition sites) after propagating it on strain sPM02, a *hsdS1*/*hsdM1* and *hsdS2/hsdM2* double mutant incapable of methylating the phage DNA (**Fig. 2.1-AB**).

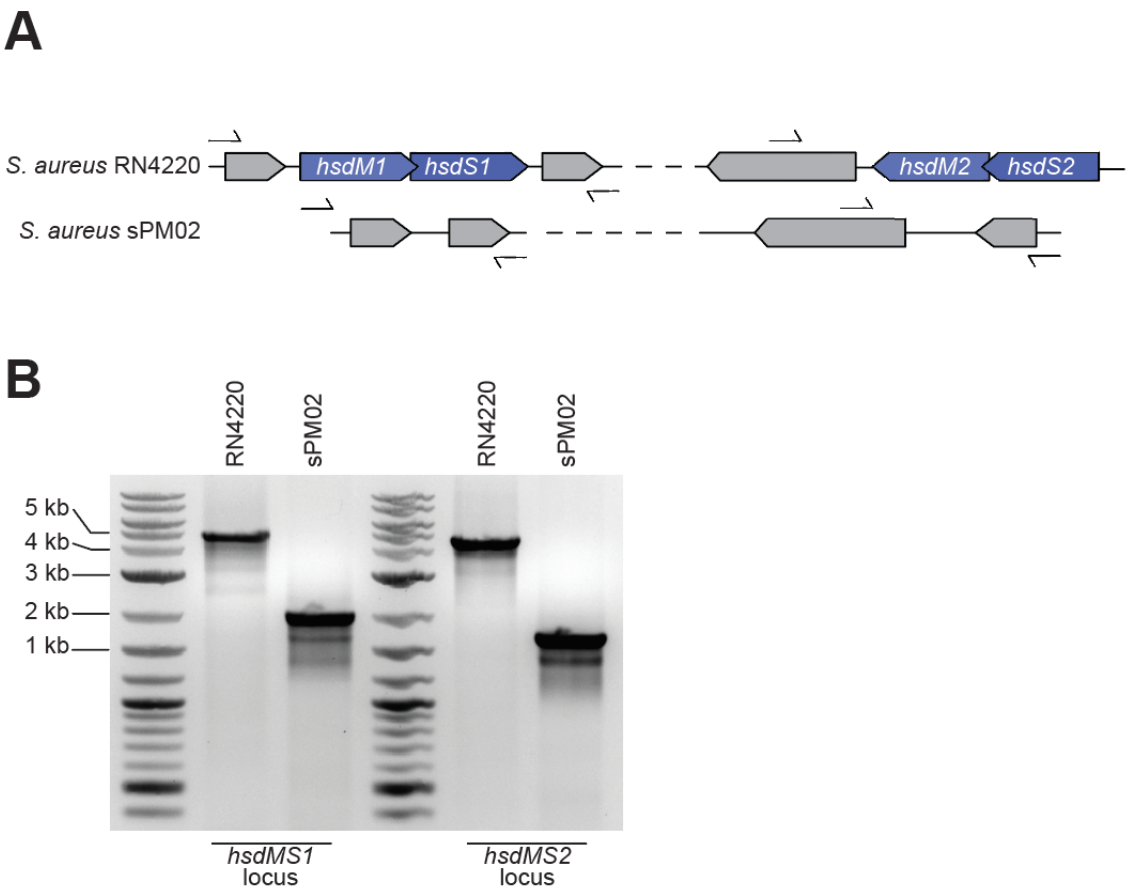


Figure 2.1. Design of strain sPM02.

(A) The two *saul-hsdMS* operons present in *S. aureus* RN4220 were removed (in-frame deletion) to generate strain sPM02. Arrows; primers used to check for the presence of the deletion. **(B)** Agarose gel electrophoresis of PCR products obtained after amplification of the *hsdMS1* and *hsdMS2* loci using template DNA from RN4220 and sPM02 strains, and the primers shown in (**A).**

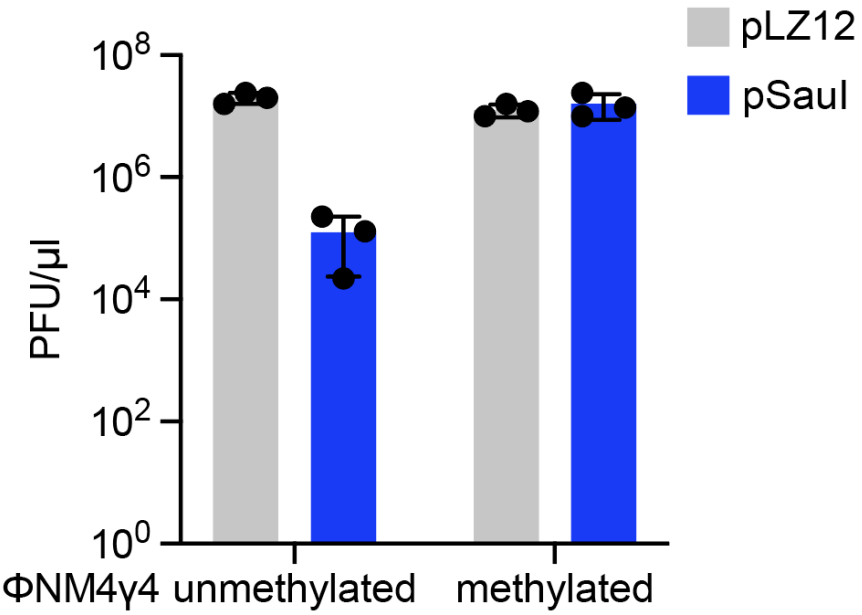


Figure 2.2. SauI RM provides some protection against ΦNM4γ4.

Enumeration of ΦNM4γ4 PFU on lawns of staphylococci expressing SauI or carrying a vector control. Mean of three biological replicates ± SD are reported.

Enumeration of plaque-forming units (PFUs) revealed that phage propagation was reduced by two orders of magnitude on RN4220/pSauI cells when compared to propagation on RN4220/pLZ12 staphylococci or to the propagation of the methylated phage. (**Fig. 2.2**). This relatively limited defense is usually attributed to the rise of phages with modified genomes that can avoid DNA cleavage and lyse the host[173]. To test whether this is the case for the SauI RM system, we treated *S. aureus* RN4220/pSauI and RN4220/pLZ12 with unmodified ΦNM4γ4 at a multiplicity of

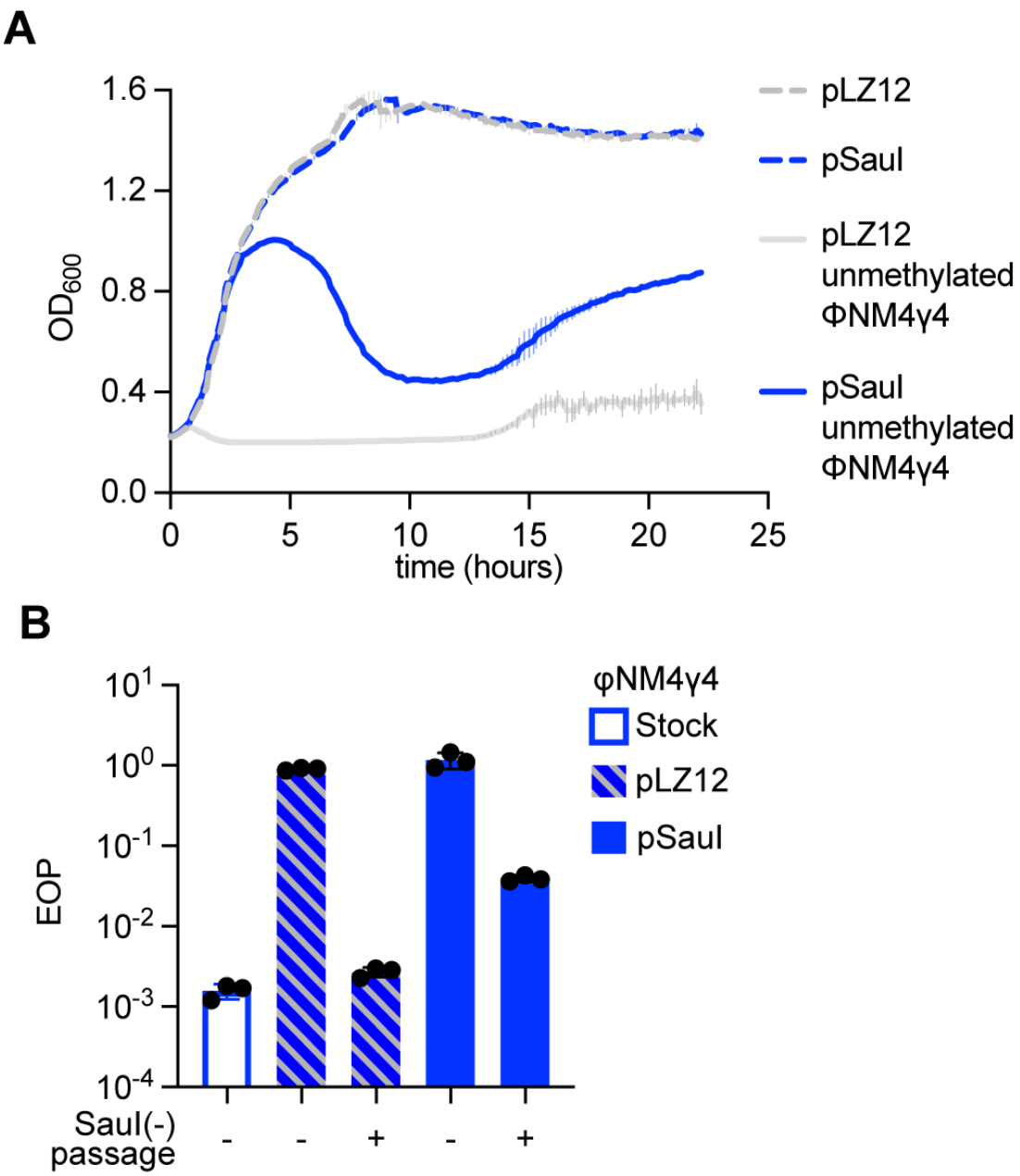


Figure 2.3 SauI RM provides a short-lived protection against □NM4□4 at MOI 10.

(**A**) Growth of staphylococci expressing SauI or carrying an empty vector control in the presence or absence of ΦNM4γ4 infection, measured as the OD_{600} of the cultures over time. MOI ~10. Mean of three biological replicates ± SD are reported. (**B**) EOP of a phage stock, or of phages obtained at the end of the growth curve shown in (**A**), amplified or not through the non-methylating strain sPM02, after plating on lawns of staphylococci expressing SauI, relative to PFUs obtained after plating on lawns of cells carrying an empty vector control. Mean of three biological replicates ± SD are reported.

infection (MOI) equal to 10 and measured the OD_{600} of the cultures to monitor their growth. Staphylococci lacking restriction (RN4220/pLZ12, with functional methylation) succumbed to the virus and were unable to grow. In contrast, cells equipped with SauI (RN4220/pSauI) were initially resistant to infection and displayed similar growth to uninfected cultures, but eventually lysed at about four hours after the addition of phage (**Fig. 2.3-A**) To determine how phages escaped restriction, we collected the supernatants and confirmed that they contained viral particles fully resistant to restriction (**Fig. 2.3-B**). However, when these phages were grown in strain sPM02, incapable of modifying DNA, they became fully sensitive to restriction again (**Fig. 2.3-B**), a result that implicates DNA modification (an epigenetic but not genetic change) in the rapid escape of ΦNM4γ4 from restriction. We observed growth at the end of the experiment for the pSauI cultures and therefore decided to test whether the cells carried mutations that confer phage resistance, such as receptor mutations[167]. We plated the culture, selected 10 colonies, and tested their susceptibility to infection with unmethylated phage (**Fig. 2.4**). ΦNM4γ4 was able to form plaques on lawns of all ten cultures derived from the colonies, with very similar efficiency of plaquing to that observed on a lawn of RN4220/pSauI. These results demonstrate that the recovery is not due to receptor or any other inheritable mutations. Moreover, we repeated the entire with a higher MOI (250) and no regrowth was observed (**Fig. 2.5**). As such the result suggests the presence of a small fraction of uninfected cells at MOI 10 that were able to regrow at the end of the experiment. Altogether, our experiments show that the

SauI RM system provides only temporary protection to staphylococci due to the rise of modified phage progeny that can overcome restriction and lyse the cultures.

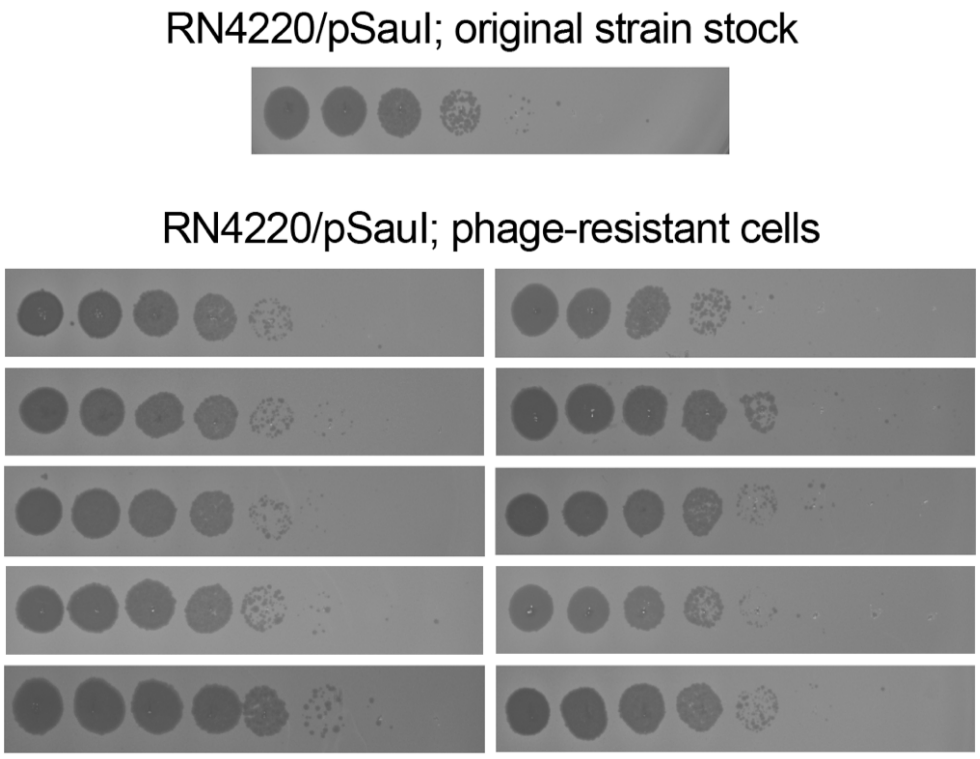


Figure 2.4. Culture recovery is not due to bacteriophage resistant mutants.

Plaque assays using bacterial lawns seeded with cells from 10 single colonies recovered from phage-resistant cells at the end of the growth curve shown in Fig. 2.3.A, in which staphylococci harboring pSauI were infected with unmethylated ΦNM4γ4. 10-fold dilutions of unmethylated phage were spotted on the lawns. A lawn of the original stock strain carrying pSauI was used as control.

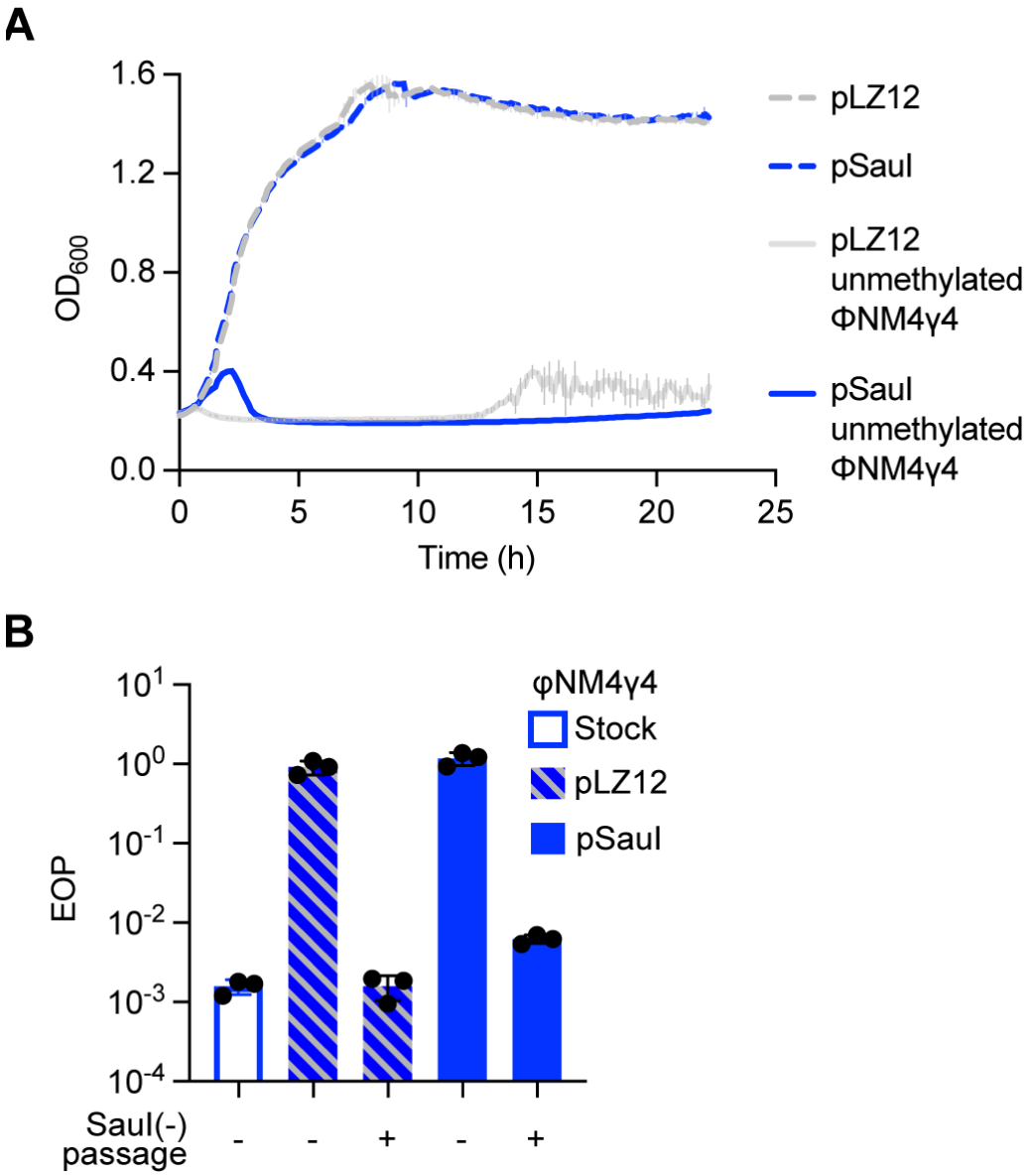


Figure 2.5 SauI RM provides a short-lived protection against ΦNM4γ4 at MOI 250.

(**A**) Growth of staphylococci expressing SauI or carrying an empty vector control in the presence or absence of ΦNM4γ4 infection, measured as the OD_{600} of the cultures over time. MOI ~250. Mean of three biological replicates ± SD are reported. (**B**) EOP of a phage stock, or of phages obtained at the end of the growth curve shown in (**A**), amplified or not through the non-methylating strain sPM02, after plating on lawns of staphylococci expressing SauI, relative to PFUs obtained after plating on lawns of cells carrying an empty vector control. Mean of three biological replicates ± SD are reported.

2.3 SauI restriction increases spacer acquisition during the type II-A CRISPR-Cas response

Previous work showed that phage restriction increased the number of colonies that survive infection via type II-A CRISPR immunity in *S. thermophilus*[166]. We decided to determine whether this is also the case in our *S. aureus* system. To do this, we eliminated all sequences but a single repeat of the type II-A locus CRISPR array in the pCRISPR plasmid (**Fig. 2.6**), for two reasons. First, we wanted to prevent any confounding effects of the phenomenon known as priming, in which previously acquired spacers enhance the acquisition of new ones[161]. Second, the lack of pre-existing spacers produces a stronger transcription repression of the type II locus[174], leading to a low rate of acquisition that would make more evident any enhancement of the process. We infected *S. aureus* RN4220/pCRISPR/pSauI as well as *S. aureus* RN4220/pCRISPR/pLZ12 cultures (five independent replicates) with unmodified ΦNM4γ4, at high MOI and we monitored their growth over time (**Fig. 2.7-A**). Neither

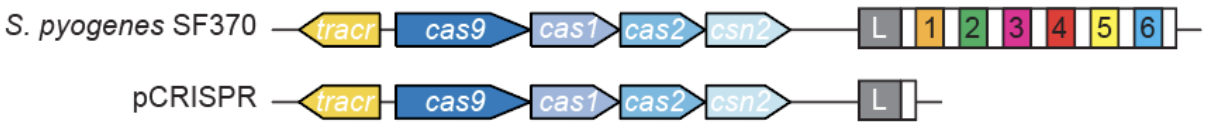
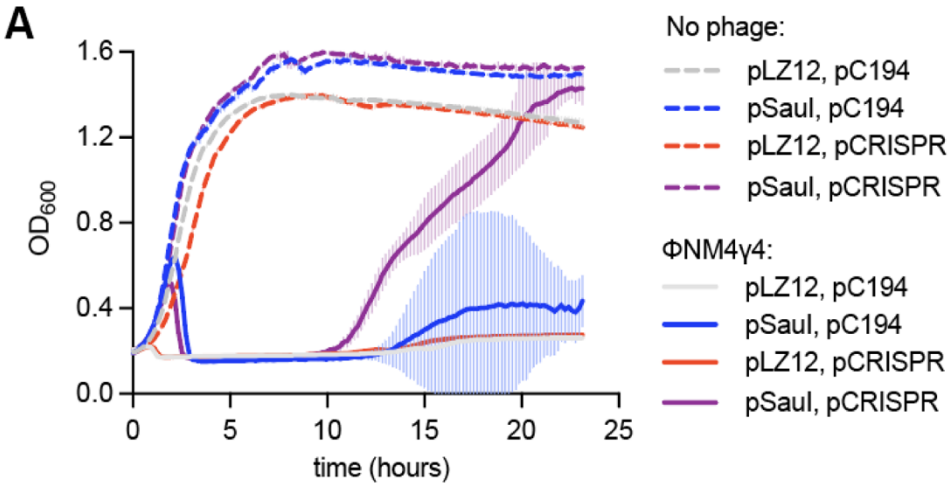
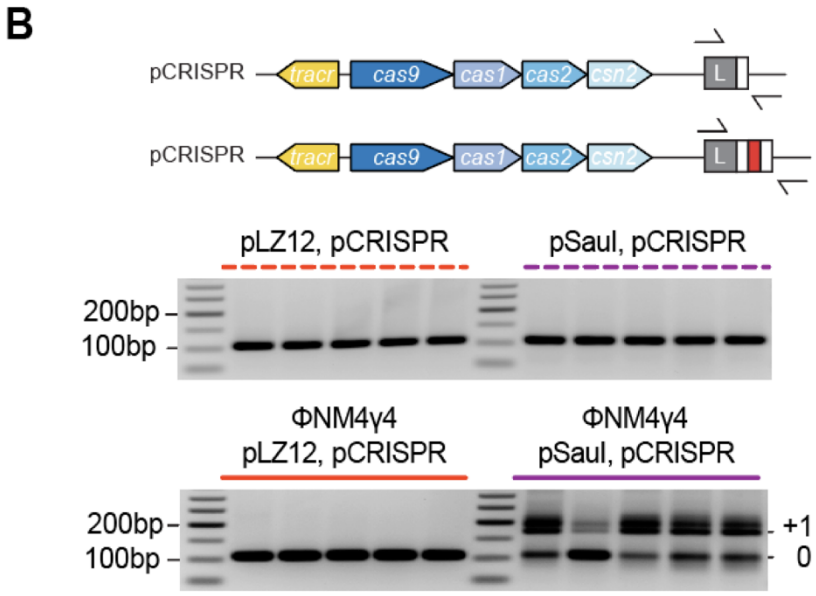


Figure 2.6 Schematic of the *S. pyogenes* type II-A CRISPR-*cas* locus

Grey rectangle, leader sequence ('L"); white rectangle, repeat; colored, numbered rectangles, spacers. It was cloned into the staphylococcal vector pC194 without spacers, just a single repeat.


A
1.6
1.2
0.8
0.4
0.0
OD600
0
5
10
15
20
25
time (hours)
No phage:
pLZ12, pC194
pSaul, pC194
pLZ12, pCRISPR
pSaul, pCRISPR
ΦNM4γ4:
pLZ12, pC194
pSaul, pC194
pLZ12, pCRISPR
pSaul, pCRISPR


B
pCRISPR
tracr
cas9
cas1
cas2
csn2
L
pCRISPR
tracr
cas9
cas1
cas2
csn2
L
pLZ12, pCRISPR
pSaul, pCRISPR
200bp
100bp
ΦNM4γ4
pLZ12, pCRISPR
ΦNM4γ4
pSaul, pCRISPR
200bp
100bp
+1
0

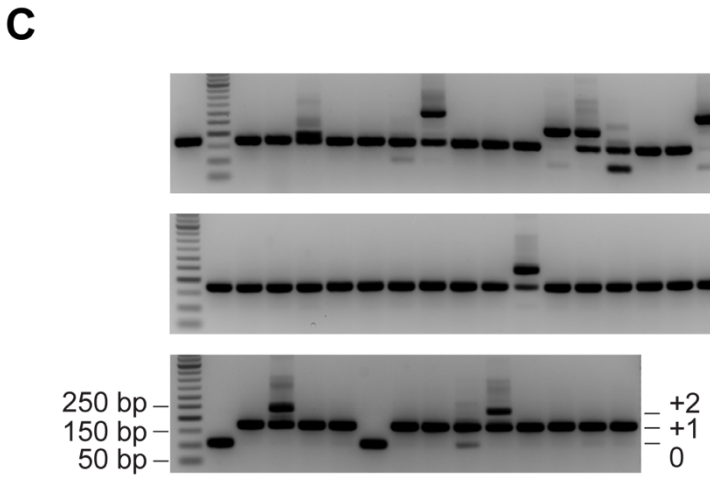

C
250 bp
150 bp
50 bp
+2
+1
0

Figure 2.7 SauI restriction of ΦNM4γ4 promotes the type II CRISPR-Cas response in staphylococci.

(**A**) Growth of staphylococci expressing pSauI or carrying an empty vector control together with pCRISPR or an empty vector in the presence or absence of ΦNM4γ4 infection, measured as the OD_{600} of the cultures over time. MOI ~250. Mean of five biological replicates ± SD are reported. (**B**) Top, representation of the primer binding sites for the amplification of the pCRISPR array. Bottom, agarose gel electrophoresis of PCR products obtained after amplification of the CRISPR array using DNA obtained from the cultures shown in (**A**). (**B**)Agarose gel electrophoresis of PCR products obtained after the amplification of the CRISPR array from 50 individual colonies recovered from one of the pSauI/pCRISPR cultures infected with ΦNM4γ4 at the end of experiment in (**A**)

type II-A CRISPR immunity nor SauI alone were sufficient to enable survival and the cultures rapidly succumbed to infection. In contrast, the combination of CRISPR and SauI restriction allowed an initial growth that was followed by the collapse of the cultures and the eventual recovery of the infected staphylococci. To test whether CRISPR immunity mediated this recovery, we amplified the CRISPR array of each replicate culture (**Fig. 2.7-B**). All the cultures harboring pSauI, but not those carrying the pLZ12 control, showed integration of one or two new spacers into the CRISPR locus. Amplification of pCRISPR isolated from 50 individual colonies followed by Sanger sequencing of the PCR products showed that 48/50 harbored an expanded array (**Fig. 2.7-C**), and that the new spacer sequences matched the ΦNM4γ4 genome (**Table 2.1**). Similar results were obtained after infection at a lower MOI, 10 (**Fig. 2.8 and Table 2.2**). Our data not only corroborate previous results in *S. thermophilus*, but also suggest a dynamic in which restriction provides a first, short-lived, line of defense that is quickly bypassed by the modification of phage DNA and that also stimulates

spacer acquisition by the type II-A CRISPR-Cas immune response to enable the survival of the infected population.

Table 2.1 Spacer sequences from 50 individual colonies from the PCRs shown in Fig. 2.7-C

Colony #	Spacer sequence	PAM
1	AGTATTGGAATCTGATGAATATTCATCTCT	CGG
2	TCATGAAAAAGTGAATTGCTAGTAGTGTGT	TGG
3	CAATAGAGATACTTTATCTAACATGATACAC	GGG
4	Adapted but non-specific sequencing over the spacer region.	N/A
5	TTTTCAAAGAATAAAAAAACTGCTACTTGT	TGG
6	ACATACTCCAAACAATTGATGGATTTGTGT	AGG
7	TTTAGCGATATTAATTATGCTCGTAAGAAT	CGG
8	(ATTTCTGGACTGTTCCATGCTTTTTCAATT) 2X	TGG
9	CAGGACAGACTAATTAACTTAGTCATGAAA	TGG
10	TCGCAATGTGTAGAGATATAGAACTTCACT	GGG
11	AGTATTGGAATCTGATGAATATTCATCTCT	CGG
12	AGCTTGCGTCATATAAATAATTTTCGTTCT, AGTAGCTACTACTAAGACATCAATTTTAGT	AGG and TGG
13	Adapted but non-specific sequencing over the spacer region.	N/A
14	Adapted but non-specific sequencing over the spacer region.	N/A
15	AGTATTGGAATCTGATGAATATTCATCTCT	CGG
16	ACATACTCCAAACAATTGATGGATTTGTGT	AGG
17	GTTTTTAAAATCCGATAAAATAACATTGCC, GCTAAGACTGTGAAGCATAATACTGCTACT, ATAAATAAAAAAGTTACTACTCACACACTA	TGG, AGG, AGG
18	TTTAGCGATATTAATTATGCTCGTAAGAAT	CGG
19	CGAATAACTCACGTTCCATTGAATACTGTGT	AGG
20	ACATACTCCAAACAATTGATGGATTTGTGT	AGG
21	ACATACTCCAAACAATTGATGGATTTGTGT	AGG
22	TGATGTAGCTAAACATGTTGCGATGATGTC	AGG
23	CAATAGAGATACTTTATCTAACATGATACAC	GGG
24	CAATAGAGATACTTTATCTAACATGATACAC	GGG
25	TGATGTAGCTAAACATGTTGCGATGATGTC	AGG
26	CGAATAACTCACGTTCCATTGAATACTGTGT	AGG
27	AGTATTGGAATCTGATGAATATTCATCTCT	CGG

28	CGAATAACTCACGTTCCATTGAATACTGTGT	AGG
29	Adapted but non-specific sequencing over the spacer region.	N/A
30	TGATGTAGCTAAACATGTTGCGATGATGTC	AGG
31	ATTTCTGGACTGTTCCATGCTTTTTCAATT	TGG
32	AGTATTGGAATCTGATGAATATTCATCTCT	CGG
33	CCCAATGATCTTATTGGTAAGTTTTGTCACT	TGG
34	TTTGGAGTATGTAGAAGTACAGTATACAAC	TGG
35	AGTATTGGAATCTGATGAATATTCATCTCT	CGG
36	AGTATTGGAATCTGATGAATATTCATCTCT	CGG
37	unadapted	N/A
38	TTTAGCGATATTAATTATGCTCGTAAGAAT	CGG
39	AGGAATTGAGACACCTCAATATATACTTGC	TGG
40	CAGGACAGACTAATTAACTTAGTCATGAAA	TGG
41	TGATGTAGCTAAACATGTTGCGATGATGTC	AGG
42	unadapted	N/A
43	CGAATAACTCACGTTCCATTGAATACTGTGT	AGG
44	CAATAGAGATACTTTATCTAACATGATACAC	GGG
45	ATAAATAAAAAANTTACTACTCACACACTA	AGG
46	ATTTCTGGACTGTTCCATGCTTTTTCAATT.	TGG
47	CAGGACAGACTAATTAACTTAGTCATGAAA	TGG
48	TTTTCAAAGAATAAAAAAACTGCTACTTGT	TGG
49	TGATGTAGCTAAACATGTTGCGATGATGTC	AGG
50	TGATGTAGCTAAACATGTTGCGATGATGTC	AGG

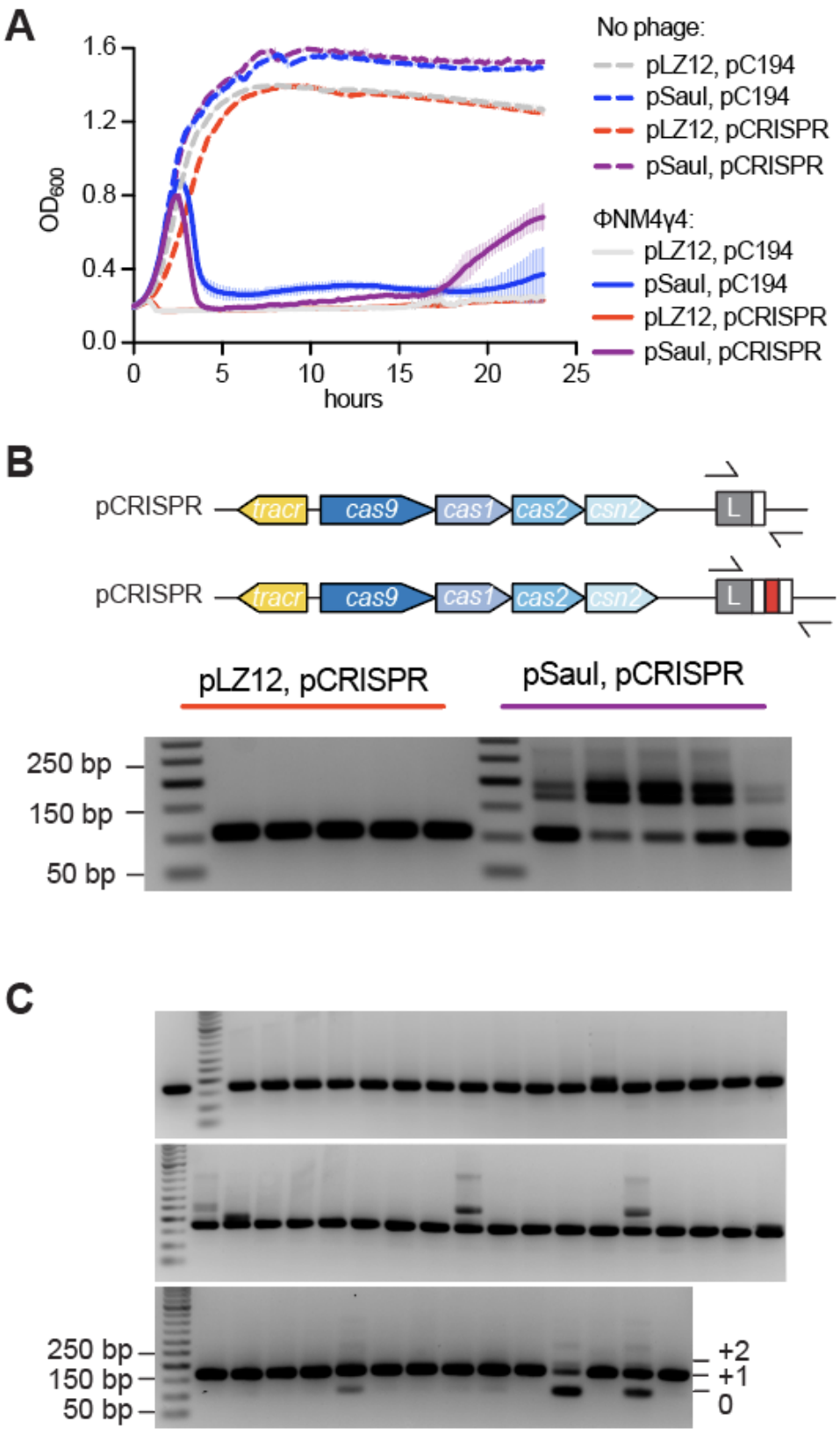

A
1.6
1.2
0.8
0.4
0.0
OD600
0
5
10
15
20
25
hours
No phage:
pLZ12, pC194
pSaul, pC194
pLZ12, pCRISPR
pSaul, pCRISPR
ΦNM4γ4:
pLZ12, pC194
pSaul, pC194
pLZ12, pCRISPR
pSaul, pCRISPR
B
pCRISPR
tracr
cas9
cas1
cas2
csn2
L
pCRISPR
tracr
cas9
cas1
cas2
csn2
L
pLZ12, pCRISPR
pSaul, pCRISPR
250 bp
150 bp
50 bp
C
250 bp
150 bp
50 bp
+2
+1
0

Figure 2.8. SauI restriction of ΦNM4γ4 promotes the type II CRISPR-Cas response in staphylococci at MOI 10.

(**A**) Growth of staphylococci expressing pSauI or carrying an empty vector control together with pCRISPR or an empty vector in the presence or absence of ΦNM4γ4 infection, measured as the OD_{600} of the cultures over time. MOI ~10. Mean of five biological replicates ± SD are reported. (**B**) Top, representation of the primer binding sites for the amplification of the pCRISPR array. Bottom, agarose gel electrophoresis of PCR products obtained after amplification of the CRISPR array using DNA obtained from the cultures shown in (**A**). (**C**)Agarose gel electrophoresis of PCR products obtained after the amplification of the CRISPR array from 50 individual colonies recovered from one of the pSauI/pCRISPR cultures infected with ΦNM4γ4 at the end of experiment in (**A**).

Table 2.2. Spacer sequences from 50 individual colonies from the PCRs shown in Fig. 2.9-C

Colony #	Spacer sequence	PAM
1	ATGATAAGATTACTTACGTAATGCGAAAGG	TGG
2	TTGCTAATAGTTTGTTCGGCGAACTTCAAA	AGG
3	GAATAACTCACGTTCCATTGAATACTGTGT	AGG
4	TATCTCTATTGACACCAATTTCTTCAGAAA	GGG
5	GAATAACTCACGTTCCATTGAATACTGTGT	AGG
6	TATCTCTATTGACACCAATTTCTTCAGAAA	GGG
7	AAAACAATTGATTGAATTAGTTACTCGATT	AGG
8	TATCTCTATTGACACCAATTTCTTCAGAAA	GGG
9	GAATAACTCACGTTCCATTGAATACTGTGT	AGG
10	CTAACGACGGTACTTATTCCGTCGTTGCTAC	TGG
11	AAAACAATTGATTGAATTAGTTACTCGATT	AGG
12	TATCTCTATTGACACCAATTTCTTCAGAAA	GGG
13	Adapted but non-specific sequencing over the spacer region.	NA
14	TTGCTAATAGTTTGTTCGGCGAACTTCAAA	AGG
15	TATCTCTATTGACACCAATTTCTTCAGAAA	GGG
16	TATCTCTATTGACACCAATTTCTTCAGAAA	GGG
17	TTGCTAATAGTTTGTTCGGCGAACTTCAAA	AGG
18	GAATAACTCACGTTCCATTGAATACTGTGT	AGG
19	TATCTCTATTGACACCAATTTCTTCAGAAA	GGG
20	TATCTCTATTGACACCAATTTCTTCAGAAA	GGG
21	TATCTCTATTGACACCAATTTCTTCAGAAA	GGG
22	TGATGTAGCTAAACATGTTGCGATGATGTC	AGG
23	TATCTCTATTGACACCAATTTCTTCAGAAA	GGG

24	AAAAAGAATGAAACAATCAAGAGAAAAACA	AGG
25	TGATGTAGCTAAACATGTTGCGATGATGTC	AGG
26	AAAAAGAATGAAACAATCAAGAGAAAAACA	AGG
27	TATCTCTATTGACACCAATTTCTTCAGAAA	GGG
28	TATCTCTATTGACACCAATTTCTTCAGAAA	GGG
29	GAATAACTCACGTTCCATTGAATACTGTGT	AGG
30	CTAACGACGGTACTTATTCCGTCGTTGCTAC	TGG
31	TATCTCTATTGACACCAATTTCTTCAGAAA	GGG
32	TATCTCTATTGACACCAATTTCTTCAGAAA	GGG
33	TGATACACGGGAGAACAAAACCATCCTACC	CGG
34	TATCTCTATTGACACCAATTTCTTCAGAAA	GGG
35	GAATAACTCACGTTCCATTGAATACTGTGT	AGG
36	TGATGTAGCTAAACATGTTGCGATGATGTC	AGG
37	TATCTCTATTGACACCAATTTCTTCAGAAA	GGG
38	TGATGTAGCTAAACATGTTGCGATGATGTC	AGG
39	TATCTCTATTGACACCAATTTCTTCAGAAA	GGG
40	TATCTCTATTGACACCAATTTCTTCAGAAA	GGG
41	TGATGTAGCTAAACATGTTGCGATGATGTC	AGG
42	TGATGTAGCTAAACATGTTGCGATGATGTC	AGG
43	GAATAACTCACGTTCCATTGAATACTGTGT	AGG
44	TGATGTAGCTAAACATGTTGCGATGATGTC	AGG
45	TCAAAAAATACAACCAACTGGCACGGATCAAT	TGG
46	TATCTCTATTGACACCAATTTCTTCAGAAA	GGG
47	Adapted but non-specific sequencing over the spacer region.	NA
48	TATCTCTATTGACACCAATTTCTTCAGAAA	GGG
49	Adapted but non-specific sequencing over the spacer region.	NA
50	TGATGTAGCTAAACATGTTGCGATGATGTC	AGG

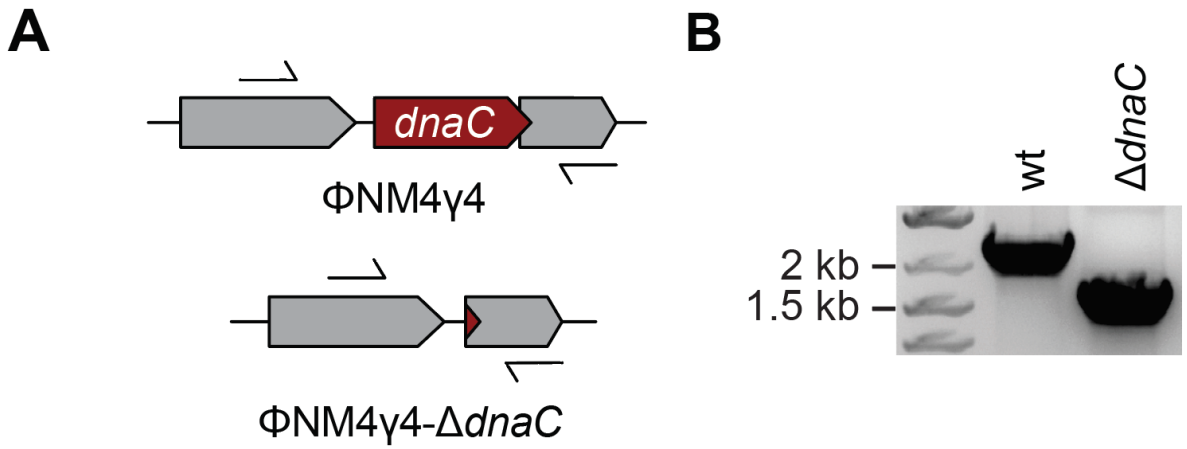


Figure 2.9. Deleting dnaC from ΦNM4γ4

(**A**) The *dnaC* gene of ΦNM4γ4 was removed (in-frame deletion) to generate ΦNM4γ4-D*dnaC*. Arrows; primers used to check for the presence of the deletion. (**B**) Agarose gel electrophoresis of PCR products obtained after amplification of the *dnaC* locus using template DNA from ΦNM4γ4 and ΦNM4γ4-D*dnaC*, and the primers shown in (**A**).

2.4 Saul inactivation of the bacteriophage lytic cycle is not sufficient to enhance spacer acquisition

A previous work in *S. thermophilus* suggested that defective phages could drive spacer acquisition by type II-A CRISPR systems[166]. It was hypothesized that restriction of the phage genome would prevent the completion of the lytic cycle and the death of the host, allowing for the process of spacer acquisition to occur and thus leading to the observed increase in phage-resistant colonies. We decided to test this hypothesis using our experimental system through the engineering of a phage that, similar to the conditions of infection in the presence of restriction, could inject its genome but fail at mounting a lytic cycle. To do this, we deleted the *dnaC* gene from phage ΦNM4γ4, which was shown previously to be essential for DNA replication and lysis in the related phage 80α[175](**Fig. 2.9**). ΦNM4γ4-Δ*dnaC* was unfit to propagate in *S. aureus*, but

otherwise able to form plaques on (**Fig. 2.10-A**) and limit the growth of (**Fig. 2.10-B**) cultures expressing *dnaC* from a plasmid, pDnaC. To quantify viral replication directly, we performed quantitative PCR (qPCR) to measure the relative amounts of phage DNA at 10 and 30 minutes after infection of RN4220/pLZ12 cells with unmethylated ΦNM4γ4 or ΦNM4γ4-Δ*dnaC* (**Fig. 2.10-C**). While the qPCR value for the wild-type genome showed a 15-fold increase from 10 to 30 minutes, the value for the mutant genome remained low, confirming its inability to carry DNA replication. To determine if the effect of the Δ*dnaC* mutation is comparable to that of restriction, we performed the same experiment using RN4220/pSauI cells. We found that in the presence of SauI activity, the levels of phage DNA were equivalent to those for the ΦNM4γ4-Δ*dnaC* phage in the absence of restriction (**Fig. 2.10-D**, compare to **Fig. 2.10-C**), suggesting that the replication of the mutant virus is similar to that of the wild-type phage in the presence of restriction.

Finally, we tested whether infection with the defective ΦNM4γ4-Δ*dnaC* phage could lead to an increase in spacer acquisition like the one observed for SauI restriction of the wild-type phage (**Fig. 2.10-E**). To do this we infected RN4220/pCRISPR/pSauI and RN4220/pCRISPR/pLZ12 cultures with unmethylated ΦNM4γ4 or ΦNM4γ4-Δ*dnaC* and collected cells after 30 minutes to extract plasmid DNA. CRISPR loci within the pCRISPR plasmids were amplified and the PCR products were subjected to next-generation sequencing (NGS) to obtain the sequences and relative abundance of the

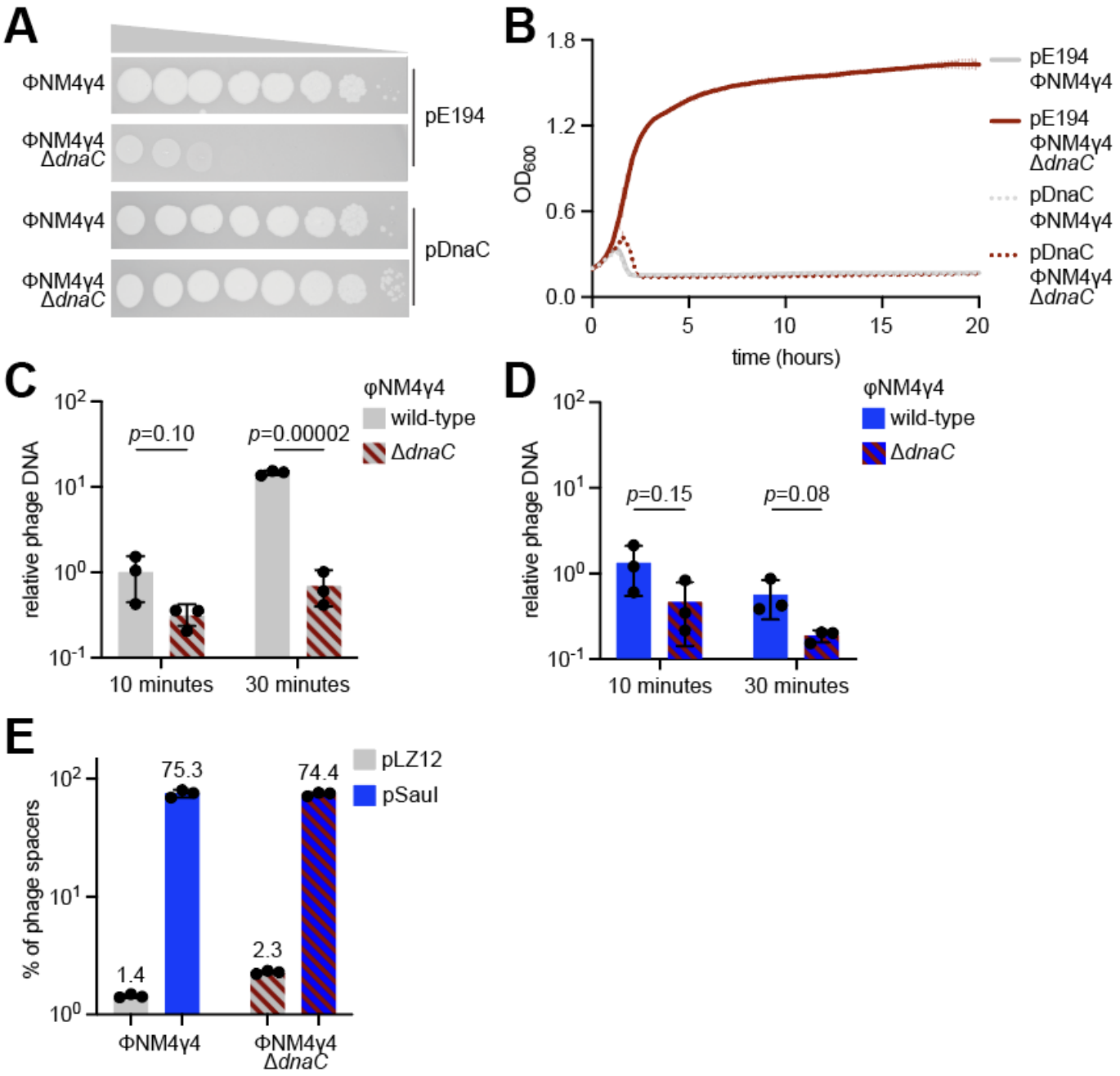

A
ΦNM4γ4
ΦNM4γ4 ΔdnaC
pE194
ΦNM4γ4
ΦNM4γ4 ΔdnaC
pDnaC
B
OD600
1.8
1.2
0.6
0.0
0
5
10
15
20
time (hours)
pE194 ΦNM4γ4
pE194 ΦNM4γ4 ΔdnaC
pDnaC ΦNM4γ4
pDnaC ΦNM4γ4 ΔdnaC
C
relative phage DNA
φNM4γ4
wild-type
ΔdnaC
p=0.10
p=0.00002
10 minutes
30 minutes
D
relative phage DNA
φNM4γ4
wild-type
ΔdnaC
p=0.15
p=0.08
10 minutes
30 minutes
E
% of phage spacers
pLZ12
pSaul
75.3
74.4
1.4
2.3
ΦNM4γ4
ΦNM4γ4 ΔdnaC

Figure 2.10. Inactivation of the viral lytic cycle is not sufficient to promote CRISPR immunity.

(**A**) Detection of plaque formation after spotting 10-fold serial dilutions of ΦNM4γ4 or ΦNM4γ4-Δ*dnaC* phages on top agar plates seeded with *S. aureus* expressing DnaC or carrying an empty vector control. (**B**) Growth of staphylococci expressing DnaC or carrying a vector control after infection with ΦNM4γ4 or ΦNM4γ4-Δ*dnaC* phages, measured as the OD_{600} of the cultures over time. MOI ~1. Mean of three biological replicates ± SD are reported. (**C**) Phage DNA quantification via qPCR, relative to host DNA, 10 and 30 minutes post-infection of *S. aureus* harboring pCRISPR in the absence of restriction with ΦNM4γ4 and ΦNM4γ4-Δ*dnaC*. MOI ~0.1. Mean of three biological replicates ± SD are reported. (**D**) Same as (**C**) but following infection of staphylococci expressing SauI. (**E**) Quantification of phage-derived spacers, relative to total new spacers, acquired 30 minutes after infection of staphylococci harboring pCRISPR and expressing SauI or carrying a vector control with ΦNM4γ4 or ΦNM4γ4-Δ*dnaC* phages, via NGS of the CRISPR locus. MOI ~250. Mean of three biological replicates ± SD are reported**.**

acquired spacers (**Fig. 2.10.-E**). Since the lytic cycle of ΦNM4γ4 takes 40-50 minutes (**Fig. 2.11-A**), sample collection at 30 minutes prevents both the depletion of cells that acquire spacers that mediate poor CRISPR immunity and succumb to phage infection, as well as the positive selection of cells containing spacers with high efficiency of targeting. As expected from our previous results, in the presence of SauI activity >70% of the new spacers matched the viral genome, after infection with both phages (**Fig. 2.1.0-E** and **Fig. 2.11-B**). In the absence of restriction, this fraction decreased to ~1%. In this case, the great majority were derived from the pCRISPR plasmid (however these spacers do not significantly affect plasmid stability over time[151]), not only after infection with the wild-type phage, but also during infection with the non-replicating ΦNM4γ4-Δ*dnaC*. This result demonstrates that not any defective phage can drive

spacer acquisition and suggests that specific events that occur during SauI restriction, beyond halting the progression of the viral lytic cycle, are important for the generation of new spacers.

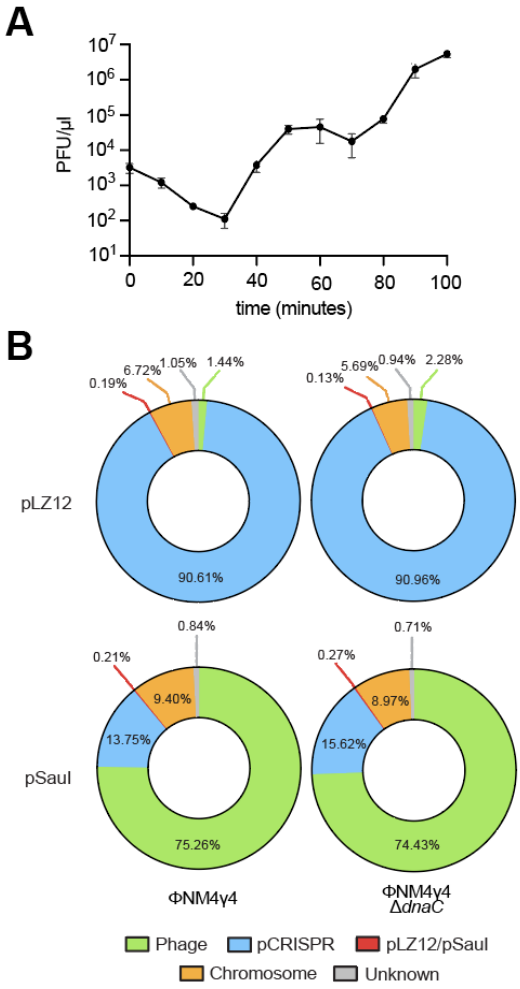


Figure 2.11. ΦNM4γ4 life cycle and spacer origin in the presence or absence of SauI restriction.

(**A**) Quantification of the number of phage in the supernatant after 10-minute intervals following infection of *S. aureus* RN4220 with ΦNM4γ4 at MOI ~0.1. Mean of three biological replicates ± SD are reported. (**B**) Quantification, using NGS reads, of spacers derived from the DNA substrates indicated with different colors, relative to total new spacers, acquired 30 minutes after infection of *S. aureus* harboring pCRISPR and expressing SauI or carrying a vector control, pLZ12, with ΦNM4γ4 or ΦNM4γ4-Δ*dnaC* phages. MOI ~250. Mean of three biological replicates are reported.

2.5 Summary

In this chapter, we described our initial investigation of the interplay between RM and CRISPR-Cas systems. As previously suggested by others [57,69,176], we found that an RM system alone provided only an innate temporary defense against bacteriophage. In our experiments, upon infection of *S. aureus* harboring its native type I RM system, SauI, with unmethylated phages, we observed initial protection against the phage followed by rapid lysis of the cultures due to the emergence of methylated escaper phages. Additionally, the *S. pyogenes* type II-A CRISPR-Cas system harboring no spacers was also insufficient on its own to provide immunity and guarantee survival of the cultures upon treatment with phages. This was unsurprising, as it was previously observed that this system has a very low spacer acquisition rate[91,174]. Interestingly, when we combined both the RM and the CRISPR-Cas system, we first observed initial protection by the RM system, which was quickly overcome by the emergence of methylated phages. However, the cultures were able to survive and rapidly regrow to high optical densities. Upon inspection of the CRISPR array, we noticed that the bacteria had acquired spacers targeting the bacteriophage genome. In agreement with a previous study[166], our results revealed that RM systems stimulate spacer acquisition by a naïve type II-A CRISPR-Cas system. The previous study[166] had hypothesized that the inactivation of phage by RM systems would prevent it from damaging the host cell beyond recovery and allow time for the process of spacer acquisition to occur[166]. We performed experiments to test this hypothesis in our system. We reasoned that two events take place during the RM immune response:

(1) the viral DNA is restricted, and (2) the bacteriophage cannot complete its life cycle. To disentangle these events from each other, we engineered a defective phage capable of injecting its genome but incapable of initiating DNA replication and host cell lysis. Interestingly, we found no enhancement in spacer acquisition using this defective phage in the absence of restriction. In turn, restriction of the defective phage by the RM system resulted in an increase in spacer acquisition by the *S. pyogenes* type II-A CRISPR-Cas system. As such, our results reveal that simply preventing a bacteriophage from completing its life cycle is not sufficient to stimulate spacer acquisition, suggesting that the restriction activity of an RM system might be critical in generating the DNA substrates for this process.

CHAPTER 3. NEW SPACERS ARE ACQUIRED AT THE SITE OF CLEAVAGE BY RESTRICTION ENZYMES

3.1 Background

A growing body of work across multiple CRISPR types suggests that free DNA ends are substrates for the acquisition of new spacers[153–155,161]. Due to their DNA packaging mechanism, *cos* phages have linear dsDNA genomes with fixed DNA sequences on each end[177,178]. During viral DNA injection, the free DNA end entering the host always has the same sequence, known as a *cos* site. Our laboratory previously described that naïve type II-A CRISPR-Cas systems preferentially acquire spacers at or near the *cos* site upon injection of the viral genome in the host[154]. Additionally, we also reported that Cas9 cleavage of the viral DNA triggers the acquisition of additional spacers directly at the Cas9 cleavage site[161]. Together, these earlier reports suggest that free DNA ends are preferential substrates for spacer acquisition.

In Chapter 2, we reported that a type I RM system promoted the acquisition of new spacers by the *S. pyogenes* type II-A CRISPR-Cas system. Additionally, our results suggest that the restriction activity of the RM system is critical to stimulate spacer acquisition. In this chapter, we investigate the hypothesis that the free DNA ends generated by restriction enzymes during the RM immune response are used for the acquisition of new spacers by the *S. pyogenes* type II-A CRISPR-Cas system.

3.2 Restriction sites are hotspots of spacer acquisition

To determine if new spacers originate from the DNA ends generated by SauI cleavage sites, we mapped the spacer sequences obtained from the NGS experiments of Figure 2.10-E from Chapter 2 to the ΦNM4γ4 genome (**Fig. 3.1-B**). However, even after infection at a very high MOI of 250, which should markedly increase the frequency of spacer acquisition[154], the graphs did not reveal any notable correlation between the new spacer sequences and the SauI recognition sites. We hypothesized that the difficulty to interpret these data was due to both the large number of SauI sites (twenty-six, **Fig. 3.1-A**) as well as the random cleavage location of this type I restriction enzyme[30]. Therefore, we decided to use a type II RM, which would cleave the viral genome at a few and defined sequences. We cloned the BglII RM system from *B. globigii* on a modified pLZ12 vector (carrying mutations that eliminate a BglII site, see Methods), obtaining pBglII. This system is composed of three genes, *bglIIC, bglIIM, and bglIIR*, coding for a controller protein (C.BglII), a methyltransferase (M.BglII), and a restriction enzyme (R.BglII), respectively, and specifically recognizes and cleaves 5'-A↓GATCT-3' sequences that contain non-methylated cytosines[179]. The ΦNM4γ4 genome contains a single BglII site (**Fig. 3.2-A**), which we named site "A", and it is susceptible to cleavage in vitro (**Fig. 3.2-AB**). In vivo, we looked at the propagation of ΦNM4γ4 in lawns of RN4220/pBglII or RN4220/pLZ12, and found that BglII restriction reduced the PFU count by approximately one order of magnitude (**Fig. 3.2-C**). In liquid

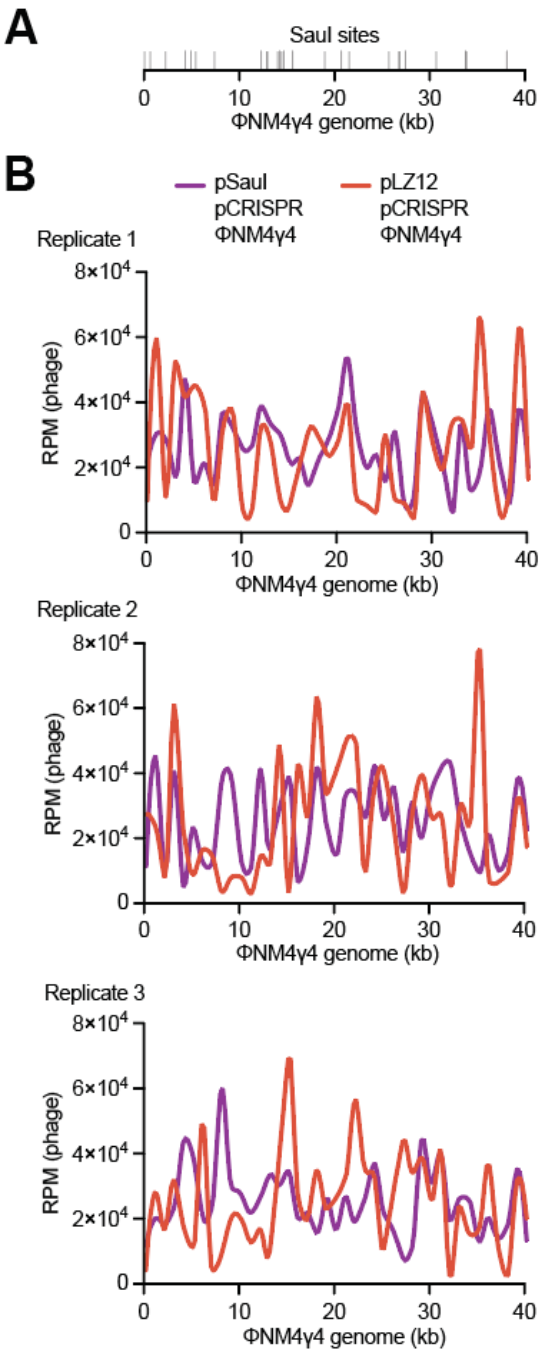


Figure 3.1. Pattern of spacer acquisition in the presence of SauI restriction.

(**A**) Schematic representation of the ΦNM4γ4 genome showing its 26 SauI sites (grey lines). (**B**) Distribution of spacer abundance (measured as RPM of phage-matching reads) obtained in Fig. 2.8.B across the ΦNM4γ4 genome. Maps for three independent replicates are shown.

cultures treated with ΦNM4γ4, cells expressing BglII continued growing for ~ 30 minutes longer than those harboring the vector control, but ultimately succumbed to infection (**Fig.3.2-D**), presumably due to the rise of methylated phage. As a control, we engineered a mutant virus in which the "A" site was eliminated through the introduction of silent mutations (to TGACTT), ΦNM4γ4-Δ*BglII* (**Fig. 3.2-A**). The genomic DNA of this phage was not cleaved in vitro (**Fig. 3.2-B**) and the PFU count was not reduced in lawns of RN4220/pBglII when compared to RN4220/pLZ12 (**Fig. 3.2-C**), and the expression of BglII did not prolong the growth of liquid cultures (**Fig. 3.2-D**).

After establishing a functional BglII RM system in staphylococci, we tested whether it could synergize with the type II-A CRISPR-Cas response as was the case for SauI. First, we infected RN4220/pCRISPR/pBglII or RN4220/pCRISPR/pLZ12 cultures with ΦNM4γ4 or ΦNM4γ4-Δ*BglII* and followed their growth over time (**Fig. 3.3-A**). CRISPR immunity was able to promote the regrowth of cells infected with wild-type phage, but not ΦNM4γ4-Δ*BglII*, only when they expressed BglII. PCR amplification of the CRISPR array at the end of the experiment detected acquisition of new spacers for all the replicates of the cultures carrying pBglII but not for those harboring the pLZ12 control or infected with the unrestricted ΦNM4γ4-Δ*BglII* phage (**Fig. 3.3-B**). These results confirmed that, similar to SauI restriction, the heterologous expression of the BglII RM system can promote the type II-A CRISPR-Cas immune response in staphylococci.

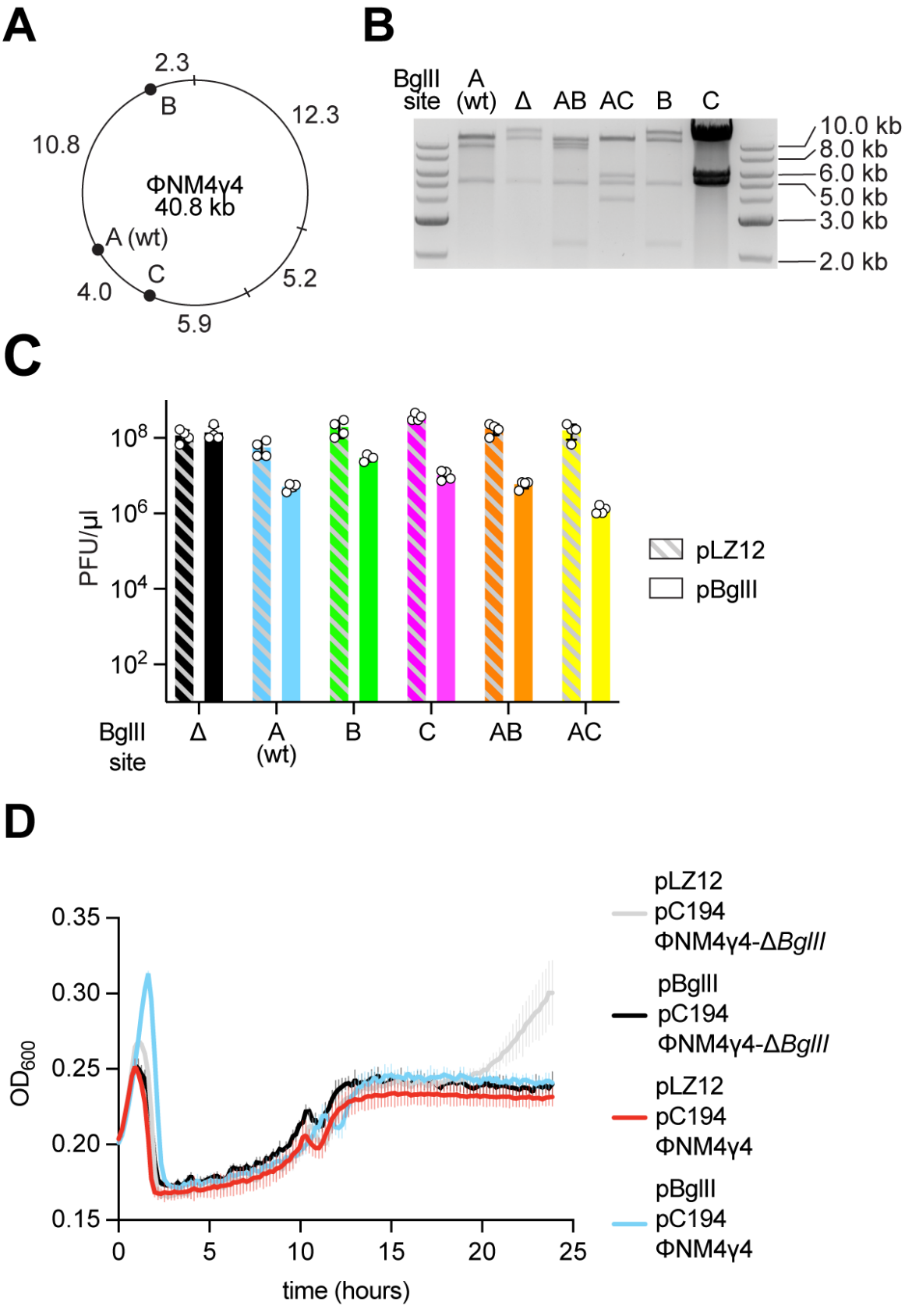

A
2.3
B
12.3
10.8
ΦNM4γ4
40.8 kb
A (wt)
C
4.0
5.2
5.9
B
BglII site
A (wt)
Δ
AB
AC
B
C
10.0 kb
8.0 kb
6.0 kb
5.0 kb
3.0 kb
2.0 kb
C
PFU/µl
10^8
10^6
10^4
10^2
BglII site
Δ
A (wt)
B
C
AB
AC
pLZ12
pBglII
D
OD600
0.35
0.30
0.25
0.20
0.15
0
5
10
15
20
25
time (hours)
pLZ12 pC194 ΦNM4γ4-ΔBglII
pBglII pC194 ΦNM4γ4-ΔBglII
pLZ12 pC194 ΦNM4γ4
pBglII pC194 ΦNM4γ4

Figure 3.2. BglII provides some protection against ΦNM4γ4

(**A**) Map of the ΦNM4γ4 genome showing the different BglII sites analyzed in this study (black circles, "A", "B", "C"). The three BssHII sites used for restriction mapping are shown as well (black lines). (**B**) Agarose gel electrophoresis of restriction fragments of different ΦNM4γ4 phages, after digestion with BssHII and BglII. The sizes of molecular weight markers are shown. (**C**) Enumeration of PFU generated by different ΦNM4γ4 phage stocks on lawns of staphylococci expressing BglII or carrying an empty vector control. Mean of four biological replicates ± SD are reported. (**D**) Growth of staphylococci expressing BglII or carrying a vector control after infection with ΦNM4γ4 or ΦNM4γ4-Δ*BglII* phages, measured as the OD_{600} of the cultures over time. MOI ~10. Mean of five biological replicates ± SD are reported.

Next, we used the heterologous BglII RM system to test our hypothesis that the free DNA ends generated by restriction are substrates for spacer acquisition. To do this, we infected RN4220/pCRISPR/pBglII cells with ΦNM4γ4 or ΦNM4γ4-Δ*BglII* and extracted plasmid DNA for the amplification of the CRISPR array, 30 minutes post-infection. The PCR products were subjected to NGS to capture the sequence and relative abundance of the new spacers. Using this data, we first determined the fraction of reads corresponding to spacers acquired from the viral DNA (**Fig. 3.4-B**). Consistent with our previous results, BglII restriction increased this value by ~ 10-fold, from 0.5 % during ΦNM4γ4-Δ*BglII* infections, to 6.7 % in experiments using wild-type ΦNM4γ4. We also mapped these reads along the phage genome to obtain the pattern of spacer acquisition (**Fig. 3.4-C**). We detected a hotspot of spacer acquisition at the BglII cleavage site A in ΦNM4γ4, which was absent for ΦNM4γ4-Δ*BglII*.

To determine if this result was specific to this BglII site, we engineered phages lacking site A but having sites B and C (**Fig. 3.4-A**). Both engineered sites were cleaved in vitro by the commercial restriction enzyme (**Fig. 3.2-B**), reduced plaque formation to a similar extent as site A in the wild-type phage (**Fig. 3.2-C**), and increased

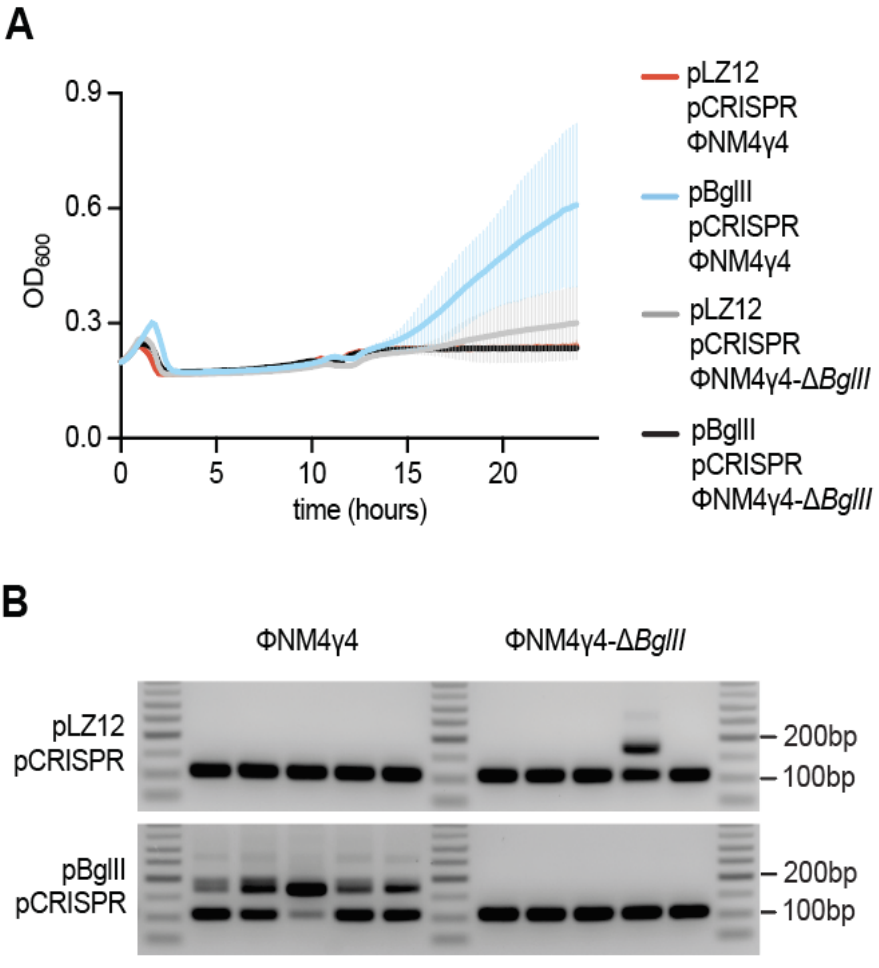


Figure 3.3. BglII restriction stimulates CRISPR immunity.

(**A**) Growth of staphylococci harboring pCRISPR and expressing BglII or carrying a vector control after infection with ΦNM4γ4 or ΦNM4γ4-Δ*BglII* phages, measured as the OD_{600} of the cultures over time. MOI ~10. Mean of five biological replicates ± SD are reported. (**B**) Agarose gel electrophoresis of PCR products obtained after amplification of the CRISPR array using DNA obtained from the cultures used in (**A**).

the fraction of spacers acquired from the viral genome during the type II-A CRISPR response by approximately an order of magnitude (**Fig. 3.4-B**). Most importantly, analysis of the spacer acquisition map of these phages showed hotspots centered at the new sites, as well as the elimination of the peak in site A (**Fig. 3.4-D**). Finally, we wanted to investigate the effect of multiple BglII sites and therefore we created phages with sites A and B, or A and C (**Fig. 3.4-A**). Both viral genomes were restricted in vitro at both sites (**Fig. 3.2-AB**). In vivo, the additional BglII cleavage did not reduce plaque formation further than any of the single sites (**Fig. 3.2-C**). In contrast, the fraction of spacers acquired from the viral genome during type II-A CRISPR immunity was significantly increased to ~15% (**Fig. 3.4-B**). Note that this value is still considerably lower than what we detected for the SauI RM system (~70%, **Fig. 2.12-E**) and is most likely due to the presence of a much larger number of SauI target sites (twenty-six, **Fig. 3.1-A**) in the ΦNM4γ4 genome. More important, the spacer acquisition pattern of the AB and AC phages showed hotspots centered at both BglII sites (**Fig. 3.4-B**). Altogether, these experiments demonstrate that restriction of ΦNM4γ4 by BglII promotes the acquisition of new spacers at the cleavage site, and suggest that the dsDNA ends generated by this RM system are substrates for the spacer integration complex of type II-A CRISPR-Cas systems.

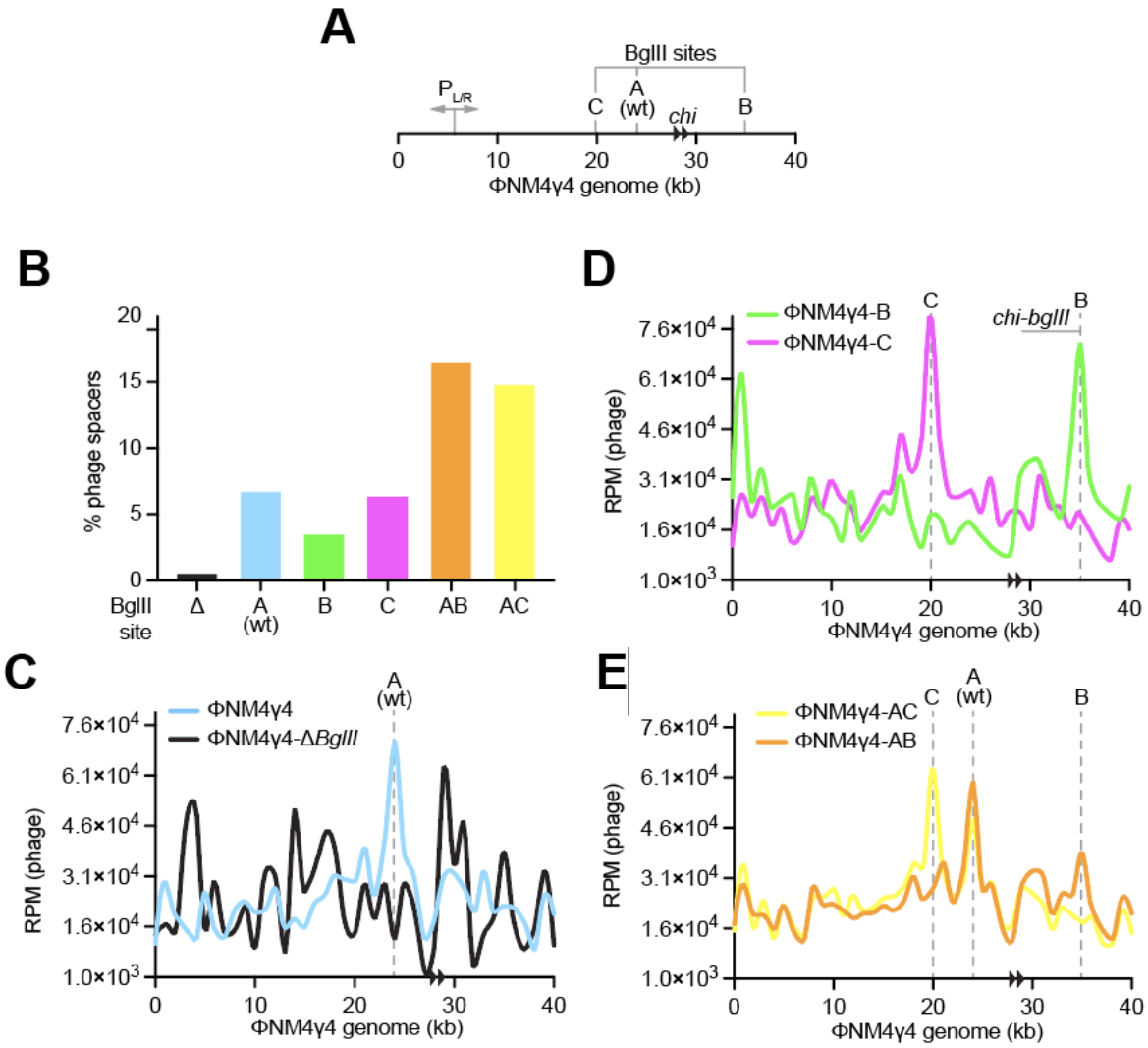


Figure 3.4. BglII promotes spacer acquisition at the restricition sites.

(**A**) Schematic representation of the ΦNM4γ4 genome showing the BglII sites analyzed in this study, its two *chi* sites (black arrowheads) and the $P_{L/R}$ bidirectional promoter. (**B**) Quantification of phage-derived spacers, relative to total new spacers, acquired 30 minutes after infection of staphylococci harboring pCRISPR and expressing BglII or carrying a vector control with ΦNM4γ4 phages containing different BglII sites, via NGS of the CRISPR locus. MOI ~25. (**C**) Distribution of spacer abundance (measured as RPM of phage-matching reads) obtained in (**B**) across the ΦNM4γ4 genome, using data from wild-type and Δ*BglII* infections. (**D**) Same as (**B**) using data from ΦNM4γ4-B and ΦNM4γ4-C infections. (**E**) Same as (**B**) using data from ΦNM4γ4-AB and ΦNM4γ4-AC infections.

3.3 AddAB amplifies spacer acquisition from BglII restriction sites

Previously, we showed that the DNA repair complex AddAB, which is the main nuclease that processes free dsDNA ends in Gram-positive bacteria[156], is responsible for the amplification of spacer acquisition from the *cos* site of the staphylococcal phage Φ12γ3[154]. *Cos* phages translocate a specific free dsDNA end during every infection[177], which is thought to be processed by AddAB to generate the DNA substrates for the acquired spacers. This processing is limited by *chi* sites, a 7 base pair sequence (5'-GAAGCGG-3' for *S. aureus*) that stops DNA degradation by AddAB[180], and therefore it creates a hotspot of spacer acquisition between the *cos* site and the first upstream *chi* site[154]. In contrast to Φ12γ3, ΦNM4γ4 is a *pac* phage that, due to its headful packaging mechanism[181], injects a variable dsDNA end of its genome during each infection event. Therefore, a *cos-chi* hotspot is not detected for this phage[154]. To determine whether and how AddAB affects spacer acquisition from the restriction site, we decided to analyze the spacers acquired by the type II-A CRISPR-Cas system upon BglII restriction of the *cos* phage Φ12γ3, and take advantage of the *chi-cos* hotspot of spacer acquisition observed after infection with this phage as an internal control. To do this, we removed from Φ12γ3 (through the introduction of silent mutations) three BglII recognition sequences residing within the *chi-cos* hotspot, and three *chi* sites located in close proximity upstream of the remaining BglII restriction sequence,

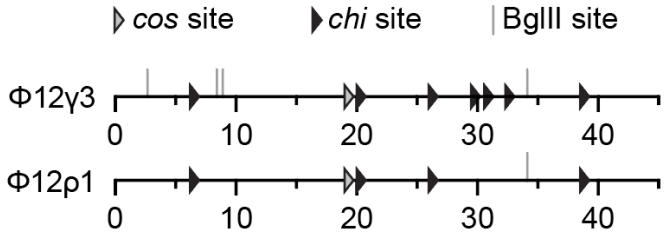


Figure 3.5. Schematic representation of the Φ12γ3 and Φ12ρ1 genomes.

Schematic representation of the Φ12γ3 and Φ12ρ1 genomes showing the BglII sites analyzed in this study (grey lines), their *chi* sites (black arrowheads) and the *cos* site (white arrowhead).

generating phage Φ12ρ1 (**Fig. 3.5**). BglII was able to cleave the genomic DNA of this phage in vitro (**Fig. 3.6-AB)**, and also mediated a reduction in plaque formation (**Fig. 3.6-C**).

To determine the pattern of spacer acquisition, we infected RN4220/pCRISPR or RN4220/pCRISPR/pBglII cultures with Φ12ρ1 and performed NGS experiments. As was the case for infections with ΦNM4γ4, restriction increased the fraction of viral-derived spacers approximately 4-fold (**Fig. 3.7-A**). After mapping the reads obtained from the control cells (without BglII restriction) to the Φ12ρ1 genome, we observed the pattern that we described previously for Φ12γ3[154], with the *cos-chi* spacer acquisition hotspot (**Fig. 3.7-B**). As reported before[154], a strong peak around 10 kb, immediately downstream of the *chi* site, was observed. Although we do not understand the factors that lead to the generation of this peak, given that it is the result of one spacer with an

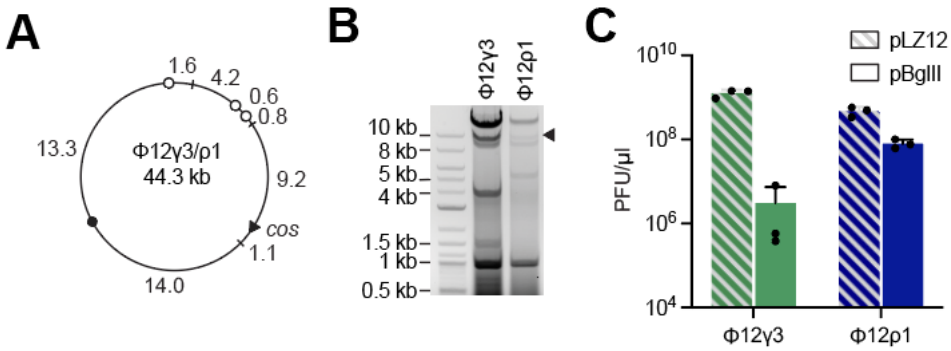


Figure 3.6. BglII provides some protection against Φ12γ3 and Φ12ρ1.

(**A**) Map of the Φ12γ3 and Φ12ρ1 genomes showing the BglII present in both phages (black circles), as well as those removed from Φ12γ3 to generate Φ12ρ1 (white circles). The three BsrBI sites used for restriction mapping (black lines), as well as the *cos* site (black arrowhead), are also shown. (**B**) Agarose gel electrophoresis of restriction fragments of Φ12γ3 and Φ12ρ1 phage DNA, after digestion with BsrBI and BglII. The sizes of molecular weight markers are shown. The black arrowhead marks a restriction fragment produced after the annealing of the *cos* site. (**C**) Enumeration of PFU generated by Φ12γ3 or Φ12ρ1 on lawns of staphylococci expressing BglII or carrying an empty vector control. Mean of three biological replicates ± SD are reported.

unusually high number of reads (5'-TCGCCGTATGTGTAGTGCGC-3', 10 to 100 fold higher than the rest of the spacer reads in that bin), we suspect that this sequence is preferred for binding and/or integration by the Cas1-Cas2 integrase complex[131]. In the presence of BglII, a new hotspot of spacer acquisition appeared, encompassing the restriction site and the region immediately upstream limited by the nearest *chi* sequence (**Fig. 3.7-C**). The marked reduction of acquisition beyond (upstream of) the *chi* site suggests the involvement of AddAB in this process; i.e., degrading the free dsDNA ends generated after BglII restriction. This was confirmed after repeating the experiment with *S. aureus* JW418 hosts, a derivative of RN4220 carrying a mutation

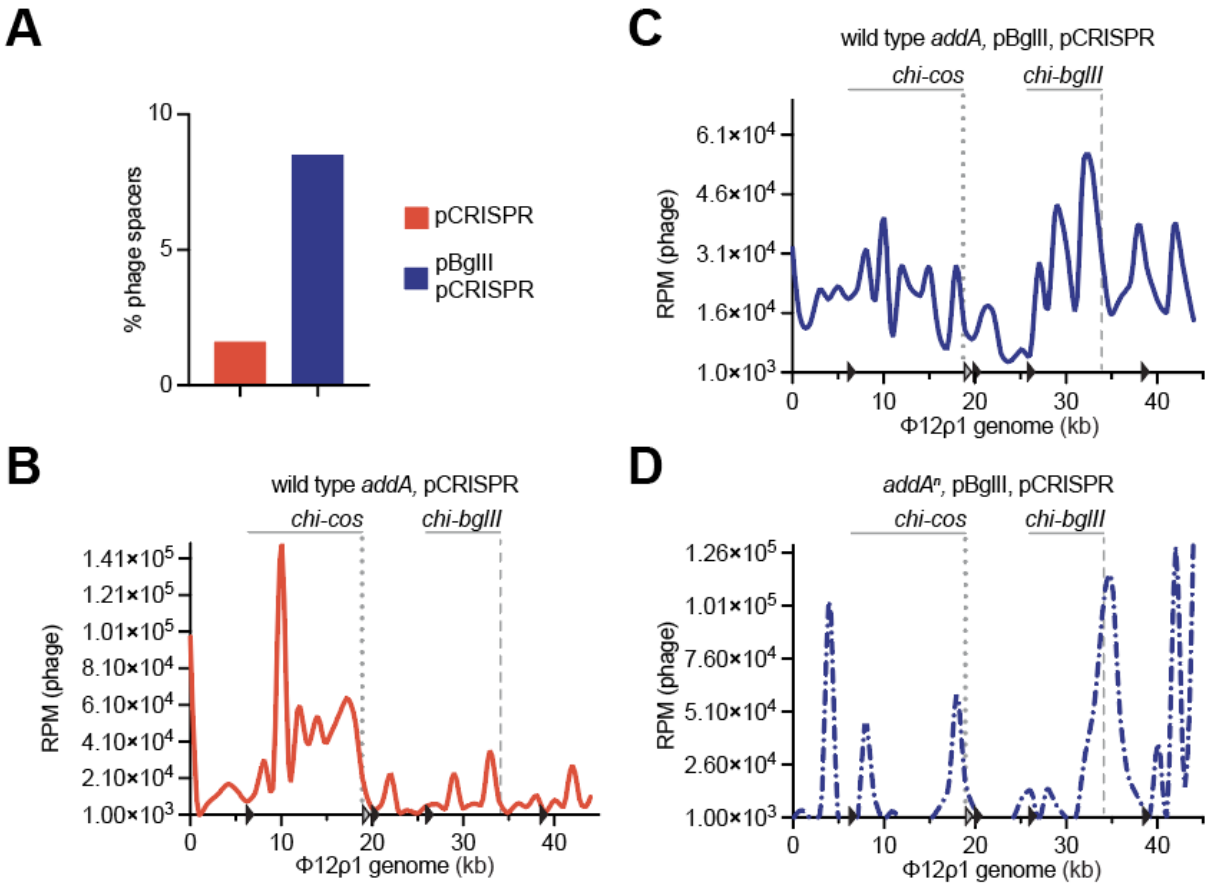


Figure 3.7. AddAB nuclease activity amplifies the region of spacer acquisition.

(**A**) Quantification of phage-derived spacers, relative to total new spacers, acquired 30 minutes after infection of staphylococci harboring pCRISPR in the presence or absence of BglII expression with Φ12ρ1, via NGS of the CRISPR locus. MOI ~25. (**B**) Distribution of spacer abundance (measured as RPM of phage-matching reads) obtained in (**A**) across the Φ12ρ1 genome, using data from infection of cells not expressing BglII. (**C**) Same as (**B**) using data from infection of cells carrying pBglII. (**D**) Same as (**C**) using data from infection of *addA*^{n} mutant staphylococci.

in the *addA* gene that inactivates the nuclease activity of the AddAB complex[154]. In this genetic background, the pattern of spacer acquisition displayed sharp peaks; i.e., not extended, at the *cos* and BglII sites (**Fig. 3.7-D**). In addition, the distribution of new spacers displayed a very low level of acquisition from the rest of the genome outside

of the main peaks, as well as peaks at ~4, 8, and 42 kb of unknown origin. Interestingly, during infection of wild-type hosts, spacer acquisition from the *chi-cos* area seemed to be diminished at the expense of the *chi-bglII* hotspot. We believe this to be the consequence of the rapid circularization of the Φ12ρ1 genome at the *cos* site, which eliminates the dsDNA ends for the spacer acquisition machinery, making this process infrequent. In contrast, approximately 90 % of the phages are cut by BglII, an estimation based on the decrease in PFUs caused by this restriction nuclease (**Fig. 3.2-B**), greatly elevating the chances of spacer acquisition from the cleavage site.

While ΦNM4γ4 does not have a *cos* sequence, it contains two contiguous *chi* sites (**Fig. 3.4-A**). Both of them are downstream of BglII sites A and C, and in an orientation that will be recognized by an AddAB nuclease complex starting degradation at the upstream restriction site, approximately 35 kb away (once the phage's genome circularize). We believe that because of this distance, we cannot discern whether we have amplification of acquisition from site A and C extending to the *chi* sites. BglII site B, on the other hand, is located only ~10 kb downstream of the *chi* sites, which should limit the AddAB upstream activity that originates at the DSB generated at this restriction site. This short distance creates a hotspot of spacer acquisition in the ΦNM4γ4-B genome delimited by the BglII-B site and these two chi sites (**Fig. 3.4-D**). AddAB degradation of the DNA downstream of the DSB generated by restriction is not

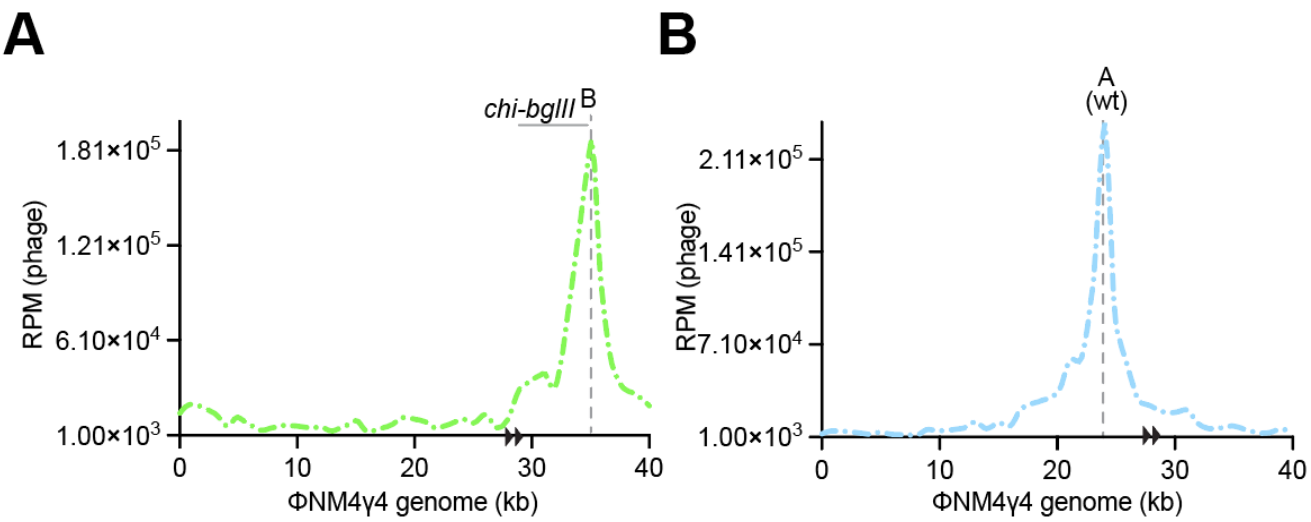


Figure 3.8. Pattern of spacer acquisition for ΦNM4γ4 in an *addA*^{n} background.

(**A**) Distribution of spacer abundance (measured as RPM of phage-matching reads) across the ΦNM4γ4-B genome, using NGS data collected after infection of *addA*^{n} mutant staphylococci cells carrying pCRISPR and pBglII. (**B**) Same as (**A**) using data from ΦNM4γ4-A(wt) infection experiments.

antagonized by the *chi* sites and therefore can produce only a general increase of spacer acquisition, but not generate hotspots. To directly test the involvement of AddAB in spacer acquisition, we repeated the experiment using the *addA* mutant host *S. aureus* JW418. We found that the pattern of spacer acquisition after infection of JW418/pCRISPR/pBglII cells with ΦNM4γ4-B contained a single peak at the BglII recognition site, which was narrow and not as extended to the upstream *chi* sites (**Fig. 3.8-A**), with low levels of acquisition from the rest of the viral DNA. A similar result was obtained for the BglII site A when we infected AddA mutant cells with wild-type ΦNM4γ4 (**Fig. 3.8-B**). Therefore, absence of AddAB activity eliminated both background as well as the minor hotspots of spacer acquisition and revealed only one strong and sharp

hotspot at the BglII cut site. Altogether, these experiments demonstrate that AddAB is involved in the amplification of spacer acquisition that starts at the BglII restriction site, allowing it to expand to other regions of the viral genome.

3.4 Summary

In this chapter, we investigated whether the DNA ends generated by restriction enzymes during viral DNA cleavage are the preferred sites for the selection of new spacers by the *S. pyogenes* type II-A CRISPR system. As previously observed for a type I RM system in Chapter 2, restriction of the viral genome by a type II RM system, BglII, also increased spacer acquisition. Given that type II RM systems cleave at specific DNA sequences, we mapped the sequences of the newly acquired spacers onto the phage genome and observed a hotspot of spacer acquisition centered at the BglII restriction cleavage site. By engineering phages with BglII recognition sites at different locations in the phage genome, we observed new spacer acquisition hotspots centered around those sites. Furthermore, increasing the number of restriction sites in the bacteriophage genome resulted in higher total levels of spacer acquisition by the type II-A CRISPR system. In addition, we also demonstrated that the host DNA repair complex AddAB can further enhance spacer acquisition at the restriction site by degrading the DNA starting at the site of RM cleavage until the nearest *chi* site, which halts AddAB degradation. As such, our experiments reveal that during the RM immune response, spacers are preferentially extracted from the site of DNA cleavage by

restriction enzymes, suggesting that the free DNA ends generated by restriction cleavage are substrates for the acquisition of new spacers by type II-A CRISPR-Cas systems. Our laboratory has previously demonstrated that the free DNA end from a phage's injecting linear genome as well as free DNA ends generated by Cas9 cleavage are also preferred sites for spacer acquisition by type II-A CRISPR systems[154,161]. The results described in this chapter support and broaden this claim by demonstrating that RM cleavage sites are another way in which DNA ends can be generated and harnessed by the type II-A CRISPR-Cas acquisition machinery.

CHAPTER 4. RM SYSTEMS PROVIDE A SHORT-LIVED INNATE IMMUNITY THAT STIMULATES A ROBUST ADAPTIVE IMMUNE RESPONSE

4.1 Background

Our results from the previous chapters suggest that restriction is a first, vulnerable, line of defense against phage infection that can stimulate CRISPR immunity as a second, more reliable, response. Interestingly, cultures that combine restriction (either SauI or BglII) and type II-A CRISPR immunity to fight against phage infection at a high multiplicity of infection (MOI), collapse 2-5 hours after infection and recover at 10-15 hours (**Fig. 2.7-A** and **3.3-A**), while integrating new spacers as early as 30 minutes post-infection (**Figs. 2.12-E** and **3.4-B**). As such, in this chapter we wanted to investigate the timing and the dynamics of the interaction between these two systems.

4.2 BglII defense is rapidly overcome by phage DNA methylation

To investigate the dynamics of the interplay between restriction and spacer acquisition by the type II-A CRISPR system, we infected RN4220/pCRISPR/pBglII or RN4220/pCRISPR/pLZ12 cultures with ΦNM4γ4 at an MOI of 1. We decreased the MOI to induce a smoother transition in the timing of phage modification and of the expansion of the CRISPR array, in order to avoid abrupt changes caused by a very high MOI that could prevent us from observing the dynamics of the process we want

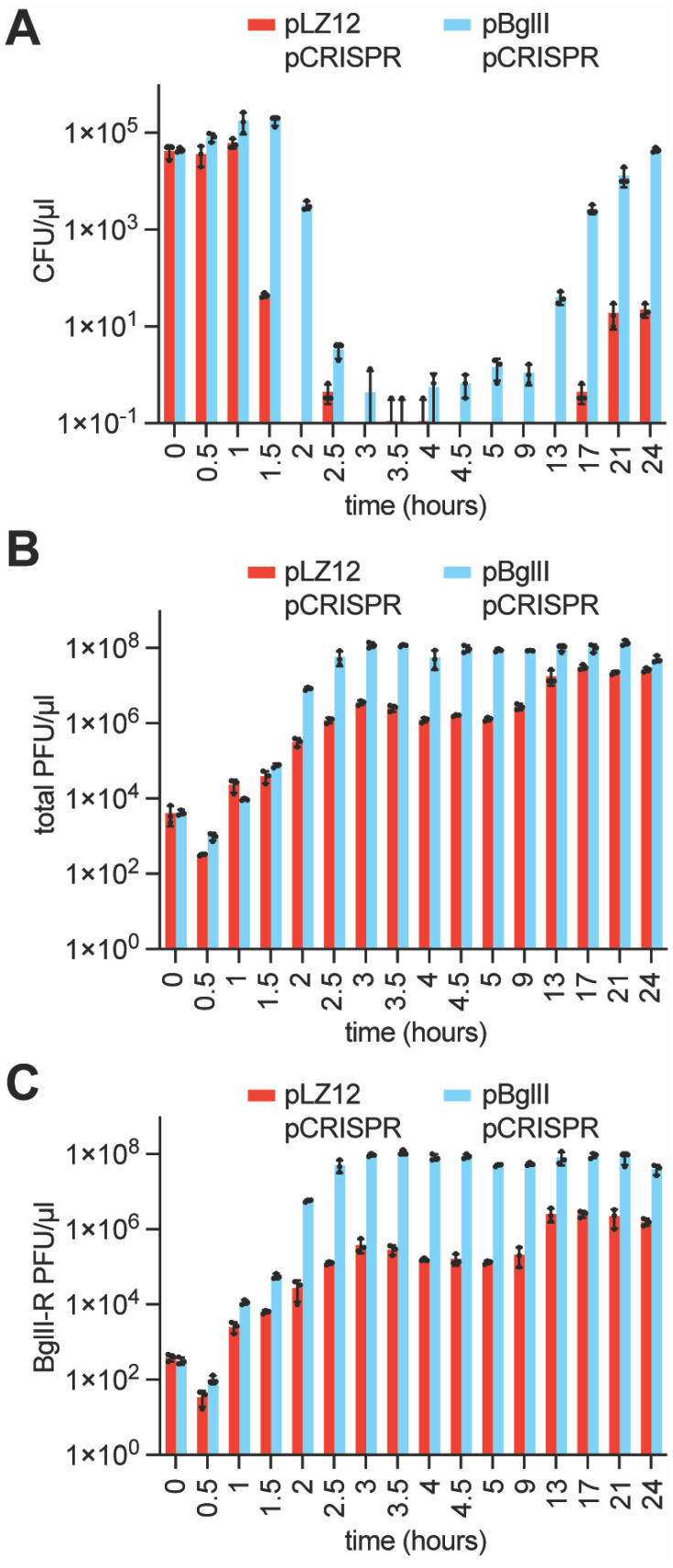

A
pLZ12 pCRISPR
pBglII pCRISPR
CFU/μl
1×10^5
1×10^3
1×10^1
1×10^-1
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 9 13 17 21 24
time (hours)
B
pLZ12 pCRISPR
pBglII pCRISPR
total PFU/μl
1×10^8
1×10^6
1×10^4
1×10^2
1×10^0
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 9 13 17 21 24
time (hours)
C
pLZ12 pCRISPR
pBglII pCRISPR
BglII-R PFU/μl
1×10^8
1×10^6
1×10^4
1×10^2
1×10^0
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 9 13 17 21 24
time (hours)

Figure 4.1 Dynamics of ΦNM4γ4 restriction over time.

A) Enumeration of CFUs present at different times following infection of *S. aureus* cells harboring pCRISPR and expressing BglII or carrying a vector control with ΦNM4γ4 (MOI ~1). Mean of three technical replicates ± SD are reported. (**B**) Enumeration of total PFUs present in the cultures of the experiment described in (**A**), after plaquing on the CRISPR(-), RM(-) strain RN4220. Mean of three technical replicates ± SD are reported. (**C**) Same as (**B**) but plaquing on staphylococci expressing BglII to enumerate PFUs resistant to restriction. Mean of three technical replicates ± SD are reported.

to study. We took 16 samples for a period of 24 hours and measured the CFU, PFU, DNA methylation, and spacer acquisition. The CFU count reflected the growth curves we previously observed, upon infection at a higher MOI (**Fig. 4.1-A**), with a rapid collapse and later recovery in the absence of restriction, and delayed lysis and earlier regrowth of cells expressing BglII. Expectedly, the total number of phage in the culture supernatants followed the opposite trend (**Fig. 4.1-B**), with a decrease in PFU at the beginning, followed by a steady accumulation of phage. Viral titers are approximately two orders of magnitude higher in cultures carrying BglII, most likely a result of the partial survival of staphylococci due to restriction earlier during infection, which increased the number of hosts for the subsequent infection by methylated phages. As noted before, we hypothesized that this higher PFU number could be attributed to the evasion of BglII restriction through methylation of the phage DNA. To test this, we plated the culture supernatants on lawns of staphylococci expressing the BglII RM system (**Fig. 4.1-C**) and calculated the fraction of total phage resistant to restriction (**Fig. 4.2-A**). We observed that approximately 90 % of the phage from RN4220/pCRISPR/pLZ12 cultures remained sensitive to BglII restriction over the 24-

hour period. In contrast, ΦNM4γ4 taken from RN4220/pCRISPR/pBglII cultures gained resistance to restriction within an hour of infection. To determine if this is due to DNA

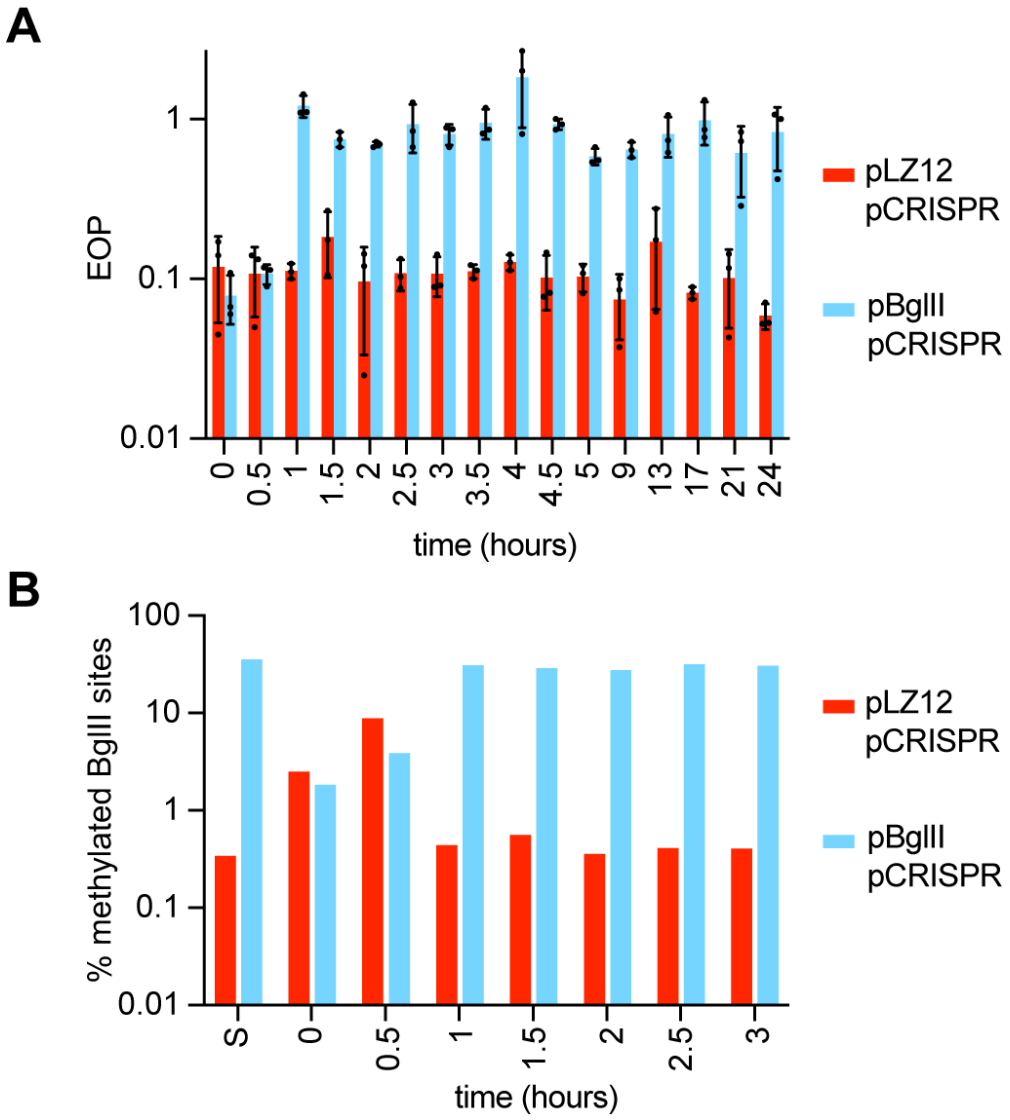


Figure 4.2. Phage DNA methylation occurs shortly after phage infection.

(**A**) EOP of ΦNM4γ4 phages obtained at different time points following infection (MOI ~1) of *S. aureus* cells harboring pCRISPR and expressing BglII or carrying a vector control, after plating on lawns of staphylococci expressing BglII, relative to PFUs obtained with cells carrying a vector control. Mean of three technical replicates ± SD are reported. (**B**) Percent of methylated BglII sites on a unmethylated or methylated ΦNM4γ4 stocks ("S") and on phages obtained in (**A**), measured using bisulfite NGS.

methylation, we performed bisulfite sequencing, a method that uses this chemical to convert unmodified, but not methylated, cytosines into uracil residues, which are recognized as thymine in subsequent PCR amplification and NGS[182]. We found that whereas the phage DNA isolated from RN4220/pCRISPR/pLZ12 supernatants contained background levels of converted uracil residues, the phage refractory to BglII restriction contained maximum levels of genomic modification after one hour of infection (**Fig. 4.2-B**). These results demonstrate that ΦNM4γ4 can rapidly evade RM systems through DNA methylation.

4.3 The type II-A CRISPR immune response initially targets the BglII recognition site and expands to attack other regions of the viral genome

Next, we investigated the details of the type II-A CRISPR-Cas immune response during restriction. Amplification of the CRISPR array showed that, in the presence of BglII activity, by 13 hours after infection with ΦNM4γ4, a small fraction of the population harboring new spacers was detected by PCR (**Fig. 4.3-A**, compare the intensity of the PCR products with and without a new spacer). Accordingly, this is also the time when we observed a rebound in CFU in the culture (**Fig. 4.1-A**). This fraction increased over the next 8 hours, and at 21 hours post-infection, the great majority of the cells had expanded CRISPR arrays. In contrast, in the absence of restriction, the first sign of spacer acquisition appeared at our last data point, 24 hours post-infection, for only a small proportion of the staphylococci in the RN4220/pCRISPR/pLZ12 culture. This

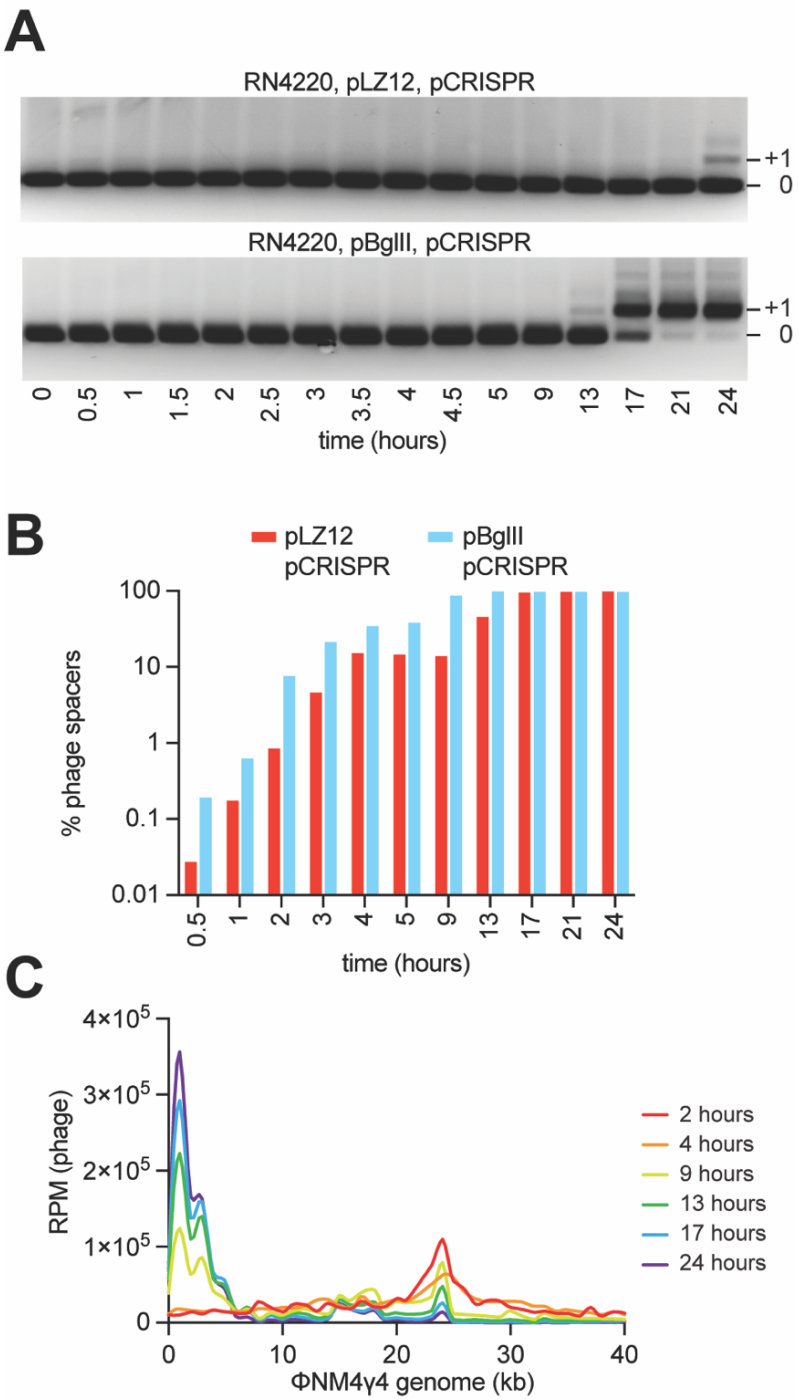


Figure 4.3 Spacer acquisition occurs shortly after phage infection.

(**A**) Agarose gel electrophoresis of PCR products obtained after amplification of the CRISPR array using DNA obtained from the cultures used in Fig. 4.1-A. (**B**) Quantification of phage-derived spacers, relative to total new spacers, acquired at the different time points of the experiment shown in Fig. 4.1-A (**C**) Distribution of spacer abundance (measured as RPM of phage-matching reads) obtained after NGS of the CRISPR locus present in the cultures used in Fig. 4.1-A across the ΦNM4γ4 genome, using data from the indicated time points.

result shows a much less efficient CRISPR-Cas immune response in the absence of restriction, which leads to a limited recovery of the cells in this culture (**Fig. 4.1-A**). Next, we performed NGS analysis to determine more accurately the timing and levels of spacer acquisition (Fig. **Fig. 4.3-B**). As previously observed in 2, as early as 30-60 minutes after infection with ΦNM4γ4, the levels of viral spacers acquired in the presence of BglII restriction were an order of magnitude higher than those acquired in the absence of the nuclease. Given that the duration of the ΦNM4γ4 lytic cycle is 40-50 minutes, the NGS data after the first two time points reflects not only spacer acquisition, but also the enrichment of staphylococci harboring anti-phage spacers. Therefore, the fraction of ΦNM4γ4-derived spacers increases steadily, and at 17 hours post-infection all spacers in both populations matched the phage genome. However,

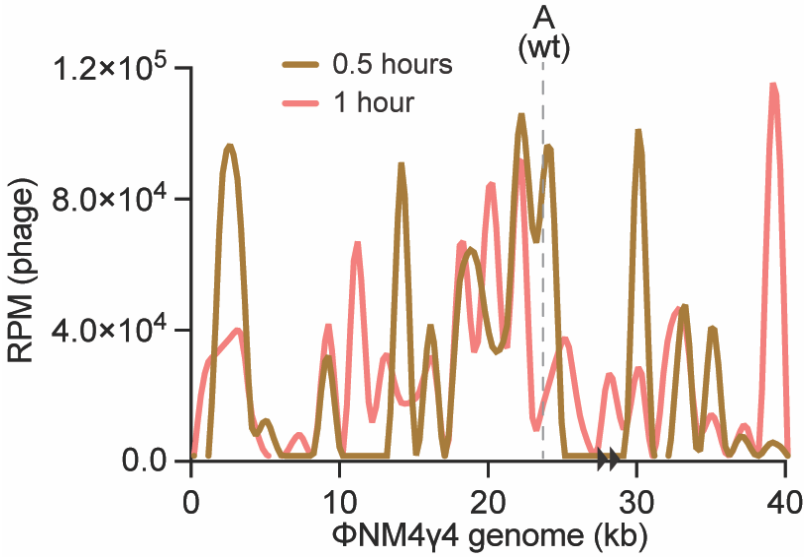


Figure 4.4. Pattern of spacer acquisition at 0.5 and 1 hour.

Distribution of spacer abundance (measured as RPM of phage-matching reads) obtained after NGS of the CRISPR locus present in the cultures used in Fig. 4.2-A across the ΦNM4γ4 genome, using data from the 0.5- and 1-hour time points

the CFU count at this time point is too low for the pLZ12/pCRISPR sample (**Fig. 4.1-A**) and therefore the expanded CRISPR array cannot be amplified via PCR. We also mapped the new spacers to obtain their distribution across the phage genome over time. Compared to the previous experiment performed at a much higher MOI (250), at MOI 1 the number of spacers acquired 30 and 60 minutes after infection were not sufficient to generate a map (**Fig. 4.4**). At two hours, however, a distinct peak was detected at the BglII site (**Fig. 4.3-C**). Interestingly, this peak decreased with time, seemingly at the expense of a growing area of spacer acquisition at the 5' end of the viral genome (**Fig. 4.3-C**), a region which has been shown before to be a preferred source of new spacers against ΦNM4γ4[154,170]. This result suggests that, after the initial stimulation of the type II-A CRISPR-Cas response by BglII restriction, the subsequent interplay between additional spacer acquisition and phage selection of the most efficient crRNA guides further shape the distribution of spacer sequences in the bacterial population.

CHAPTER 5. DISCUSSION

Previous work that studied the relationship between type II-A CRISPR immunity and RM systems found that infections with a mix of modified and unmodified phage, resistant and susceptible to restriction, respectively, increased the levels of CRISPR-resistant survivors in a manner proportional to the levels of restriction-sensitive virus[166]. It was hypothesized that the inactivation of the phage would prevent lysis and allow time for the process of spacer acquisition to occur. In this thesis, we describe the molecular mechanisms that explain and expand these results. We found that upon infection with unmodified phage, the cleavage of the viral DNA by restriction nucleases provides only a temporary protection that is rapidly overcome by DNA methylation, causing the death of most bacteria in the culture (**Fig. 5.1**). Although not efficient enough to ensure survival, RM activity stimulates the type II-A CRISPR-Cas immune response. As early as 30 minutes after infection, restriction promotes spacer acquisition in a small fraction of the cells of the bacterial population. New spacers are extracted predominantly from both ends of the cleavage site, likely due to the creation of the dsDNA ends used by the type II supercomplex as substrates for integration into the CRISPR array[148,154,161]. The host's DNA repair machinery (RecBCD in *E. coli*; AddAB in *S. aureus*) can further degrade the restricted DNA at the cleavage site[183]. Assuming that these complexes fall off the substrate DNA at some rate, their activity would generate additional free dsDNA ends. As a result, a hotspot of spacer acquisition is generated in the region spanning the restriction site and the first *chi* site with the appropriate orientation to stop AddAB activity. In turn, the newly acquired spacers direct Cas9 to cleave the viral DNA, an event that not only provides immunity but also

generates more free dsDNA ends in other areas of the viral genome that are also used by the spacer acquisition machinery, a process known as priming[161]. We believe that over time, priming expands the repertoire of new spacers from an original set targeting the vicinity of the restriction site to a final spacer population matching many other zones of the phage DNA. The fraction of CRISPR-immunized cells (each containing a different spacer) is able to grow in the presence of modified phage, allowing the recovery and survival of the infected culture. Ultimately, escaper phages with target mutations that abrogate Cas9 recognition and/or cleavage will rise. However, these phages do not take over the bacterial population for two reasons. First, the frequency of escape mutations is low, ~10^{-5} in our experimental system (much lower than the frequency of escape from RM systems)[184]. Second, the population is immunized with many different spacers, ensuring the neutralization of phages that can escape the defense provided by a single spacer[126,127]. As a result of this, the type II-A CRISPR-Cas immune response provides a more stable defense than restriction. The synergy between RM and CRISPR systems is reminiscent of mammalian immunity, where the activation of pattern recognition receptors (PRRs) during the innate immune response triggers an immediate defense that also activates a second, more robust, highly specific, but temporally delayed, adaptive immunity[185].

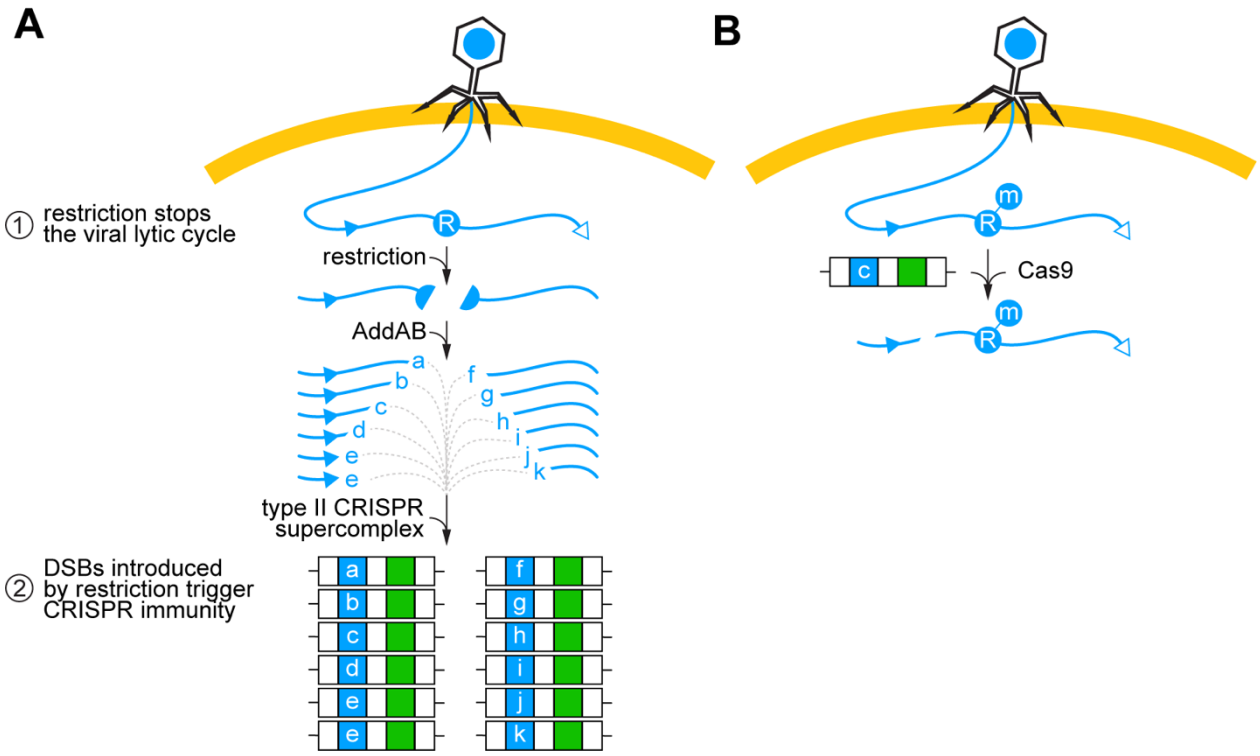


Figure 5.1. Restriction prevents the death of the host and at the same time provides the substrates for new spacers

(**A**) In our model for the synergistic effect of RM and type II-A CRISPR systems, cleavage of the viral DNA at the restriction site (blue circle, "R") shortly after infection has a dual effect: (i) it prevents the completion of the lytic cycle and death of the host, and (ii) generates free dsDNA ends that are processed by AddAB and then used by the type II supercomplex to acquire new spacers (blue squares, a, b,…, k). *Chi* sites (blue triangles), if positioned in the correct direction to inhibit AddAB, limit the region of spacer acquisition to a hotspot between the restriction and the *chi* site. Acquisition from the free dsDNA at the injected *cos* site is less frequent. (**B**) When methylation of the restriction site (blue circle, "m") prevents this first line of defense, a subpopulation of bacterial hosts are already immunized with the spacers acquired during restriction, enabling the Cas9 endonuclease to cleave the viral DNA and prevent infection.

While, as previously hypothesized[166], the inactivation of the phage via restriction is probably required to prevent its lytic cycle and allow the events required for spacer acquisition to occur, using a non-replicating phage we demonstrated that this inactivation is not sufficient for the increased levels of CRISPR immunization. Instead, *how* restriction inactivates the phage; i.e., generating free dsDNA ends that are the substrates for the spacer acquisition machinery, is critical for the synergistic relationship between RM and CRISPR systems. Therefore, our results add to a body of work that highlights the requirement of a two-pronged attack on the viral genome, made by different nucleases, for efficient spacer acquisition during the type II CRISPR-Cas immune response (**Fig. 5.2**). First, nucleases have to halt the phage lytic cycle. This can be mediated by AddAB degradation of the injected dsDNA end of the viral genome[154], by Cas9 cleavage during priming[161], or by restriction endonucleases (this thesis). Second, the product of nuclease activity has to generate free DNA ends, the preferred substrates for the Cas9-Cas1-Cas2-Csn2 spacer integration supercomplex[148]. In the case of AddAB, while degrading the injected viral DNA end, the complex could occasionally disengage from its substrate before reaching a *chi* site, leaving free DNA ends for the spacer acquisition machinery[154]. In the case of Cas9 cleavage, it generates blunt DNA ends[81,186] that also serve as spacer substrates[161]. Finally, in the case of restriction endonucleases, DNA ends generated after cleavage, including those with 3' and 5' ssDNA overhangs, are not only used as sources of spacers, but also further degraded by AddAB to expand the region of the phage genome from which spacers are acquired. All three mechanisms also contribute to the

discrimination against acquisition of “self” spacers, from the host genome, that would lead to lethal cleavage of the bacterial chromosome[115], a form of type II-A autoimmunity. Injection of free DNA ends is an obligate step of the infection cycle of most dsDNA phages, Cas9 is usually programmed with crRNAs that target invaders,

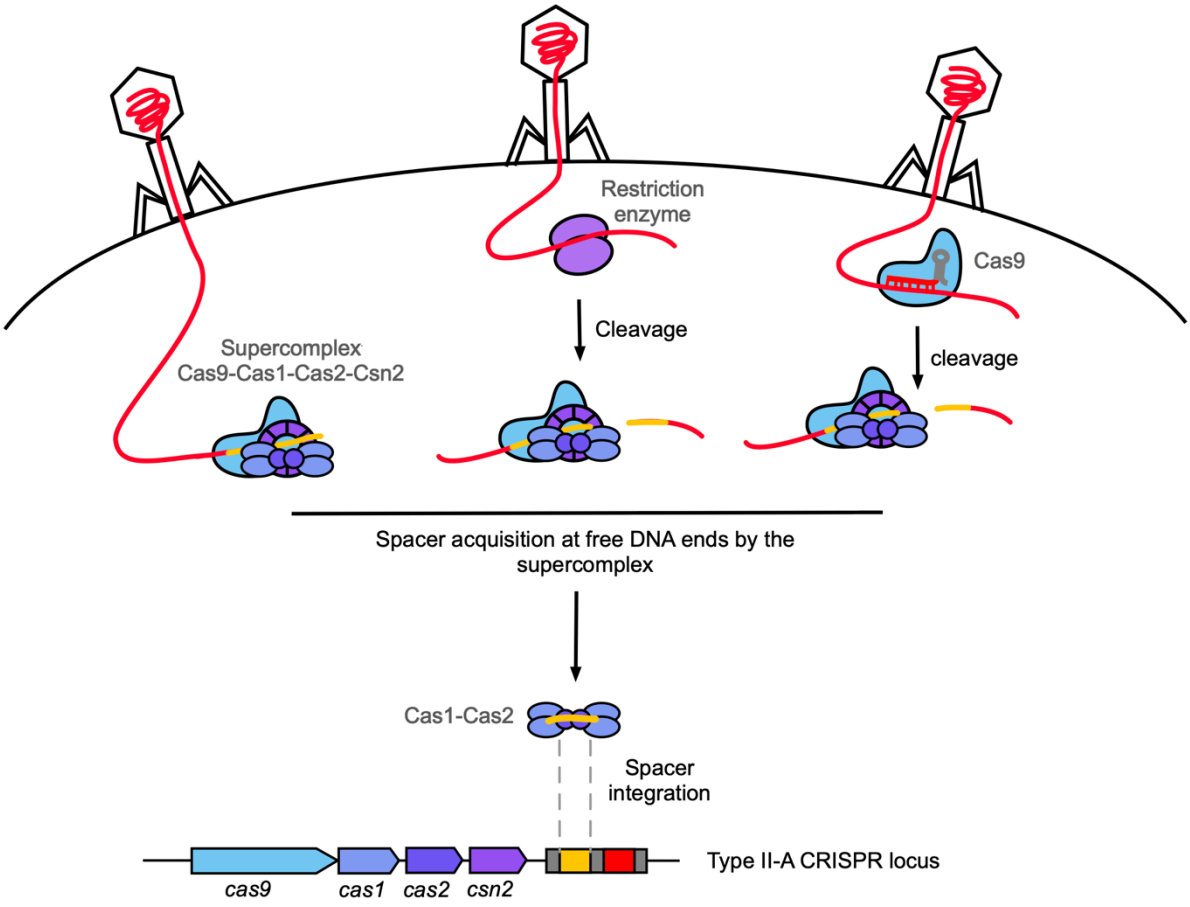


Figure 5.2. Spacer acquisition by type II-A CRISPR-Cas systems stimulated by DNA ends generated by three different processes.

Based on this work and past work from the laboratory, we postulate that type II-A CRISPR-Cas system can utilize three different DNA ends for the acquisition of new spacers upon infection: the viral DNA end entering the host upon injection (left)[154], the DNA ends generated by restriction enzymes (middle), and the DNA ends produced by Cas9 cleavage[161]. We speculate that these DNA ends (yellow) are recognized by the Cas9-Cas1-Cas2-Csn2 supercomplex, resulting in the selection of a prespacer for integration in the CRISPR locus by the Cas1-Cas2 integrase.

and restriction only cleaves foreign, unmodified, DNA. Therefore, all these strategies will bias spacer acquisition towards DNA invaders and therefore contribute to the fundamental goal of all immune systems: to recognize and eliminate invading pathogens with maximal efficacy and minimal damage to self.

CHAPTER 6. OUTLOOK

CRISPR-Cas systems provide adaptive immunity in prokaryotes by acquiring and storing in the CRISPR locus short fragments of DNA, known as spacers, from invading viruses[77,88]. During later infection by these viruses, the spacers are transcribed into CRISPR RNAs (crRNAs), which guide CRISPR-associated (Cas) nucleases to matching viral DNA sequences for destruction[96]. As such, in order to provide defense, CRISPR-Cas systems require an initial exposure to a virus to become "immunized" with new spacers. In contrast, Restriction-Modification (RM) systems provide innate immunity. Typically, these systems provide defense by encoding a methyltransferase and a restriction endonuclease that recognize the same short DNA sequence[35]. The methyltransferase modifies the host DNA at these sequences, preventing autoimmunity caused by the recognition and self-targeting of these sequences by the endonuclease[35]. In contrast, viral DNA remains unmethylated and is subsequently recognized and destroyed by the restriction endonuclease. Unlike CRISPR systems, RM systems can provide defense against many diverse kinds of viruses without prior exposure to the foreign DNA. Due to the extensive distribution of both CRISPR-Cas and RM systems (present in about 40% and 90% of bacterial genomes, respectively)[32,187], there is a wide range of organisms that contain both systems[36].

In my thesis work, I explored the interactions between RM and CRISPR-Cas systems in providing immunity against bacteriophages. By studying the *S. pyogenes* type II-A CRISPR-Cas system together with either a type I or a type II RM system, I initially observed that RM stimulates a stronger CRISPR adaptive immune response during infection by bacteriophages. Upon further investigation, I uncovered that cleavage of viral DNA by type I and II restriction enzymes generated dsDNA ends that serve as substrates for spacer acquisition by the CRISPR system. Additionally, I found that by further degrading the dsDNA ends generated by RM cleavage, the host DNA repair complex AddAB can further enhance spacer acquisition by producing additional substrates. Altogether, my work revealed an interplay between RM and CRISPR-Cas systems that can be summarized into two stages that together provide optimal anti-viral defense. First, the RM system provides a first-line defense against bacteriophage by restricting viral DNA, which also generates free dsDNA end substrates for CRISPR spacer acquisition. The initial immunity provided by the RM system is short-lived and ultimately overcome by the rapid rise of methylated phages. However, now the CRISPR-Cas system has acquired new spacers, which enable robust targeting of these methylated phages and allow for the survival of the population. Interestingly, this interplay involving a first innate immune response stimulating a second, more robust adaptive immunity was thus far thought to be reserved for eukaryotic immunity. Here, I show that prokaryotes also have their own version of linking innate and adaptive immunity by utilizing RM systems as an initial layer of defense that stimulates a stronger defensive response by CRISPR-Cas systems.

In my work, I specifically explored the effects of type I and II restriction nucleases on type II-A CRISPR spacer acquisition. Although it is not known what type of DNA ends are generated by SauI, BglII cleavage results in 4-base 5' overhangs[179]. Many other type II restriction nucleases, however, generate 3' overhangs. The laboratory has previously shown that chromosomal cleavage by the yeast homing endonuclease I-SceI, which produces in 4-base 3' overhangs[188], also leads to AddAB-enhanced hotspots of type II-A CRISPR spacer acquisition. Therefore, it appears that the type II-A spacer acquisition machinery can utilize both types of overhangs usually generated by restriction enzymes. Some type III RM systems have also been shown to produce DNA ends with small overhangs,[35] suggesting that they could also facilitate the type II-A CRISPR-Cas response. Additional experimental work will be required to determine whether these RM systems, as well as the least-studied systems belonging to type IV, can generate substrates for spacer acquisition. Additionally, our laboratory has previously shown that blunt DNA ends produced by Cas9 cleavage also serve as substrates for spacer acquisition[161]. As such, restriction enzymes producing blunt DNA ends are also likely to also enhance spacer acquisition. In type II-A CRISPR-Cas systems, all the Cas proteins are required for spacer acquisition and form a supercomplex, Cas9-Cas1-Cas2-Csn2, which is thought to select prespacer before integration by the Cas1-Cas2 integrase[91,142,143,148] (see Chapter 1.3.2 and 1.3.3). Interestingly, in vitro, Csn2 forms a ring that can bind DNA ends and slide along the DNA[144]. Furthermore, DNA ends generated by different restriction enzymes producing

double-stranded breaks (DSBs) with blunt ends or with short 5' or 3' overhangs were shown to be recognized by Csn2[144]. As such, perhaps the DNA ends generated during the RM immune response can be directly recognized by the Cas9-Cas1-Cas2-Csn2 complex through Csn2 binding. While my work and the ones from previous lab members show that DNA ends are utilized for spacer acquisition[154,161], this process represents only the very beginning of spacer acquisition (see Chapter 1.3.2 and 1.3.3). However, how these DNA ends are converted into short DNA duplexes for integration in the CRISPR locus by the Cas1-Cas2 integrase remains unknown. Future studies are required to provide the link connecting these DNA ends substrates and prespacer integration.

My work underscored the advantages of having an RM system and a CRISPR-Cas system together during phage infection. However, there exist costs to carrying defense systems. For instance, it was previously reported in *E. coli* that the type II RM system, EcoRI, can at low frequency restrict the host chromosome, resulting in autoimmunity[189]. Additionally, CRISPR-Cas systems can cause autoimmunity through the acquisition of self-targeting spacers. Perhaps the potential risks of carrying both systems are minimal compared to the advantages that they offer against phage infection. Future studies could examine the possible fitness trade-offs of carrying both systems and whether they are due to the link between restriction-mediated cleavage and CRISPR spacer acquisition described in my work. Furthermore, a recent study in *Pectobacterium carotovorum* showed that one RM system could silence another RM

system by methylating its promoter[190], providing the first-known example of an RM system using its methyltransferase to regulate another defense system. This raises the interesting possibility of whether the expression of CRISPR-Cas systems could also be modulated by RM systems, perhaps to maximize immunity against foreign invaders while dampening the likelihood of autoimmunity.

With regards to the different CRISPR types, both the previous work in *S. themophilus*[166] and my work examined the interactions between RM systems and type II-A CRISPR-Cas; therefore whether restriction can enhance spacer acquisition in other CRISPR types is unknown. In addition to type II, the type I-E system of *Escherichia coli*[155] and type III-A system of *Staphylococcus epidermidis*[153] can use dsDNA ends as preferred substrates for new spacers, with the RecBCD (*E. coli*) and AddAB (*S. epidermidis*) complexes also being involved in the expansion of the hotspot of spacer acquisition at the dsDNA break. These similarities with the spacer acquisition mechanism of type II-A systems suggest that restriction would also enhance the type I and III CRISPR-Cas responses through the incorporation of new spacers from restricted phage DNA. In contrast, the importance of DSBs for spacer acquisition in the other CRISPR types (IV, V, and VI) is not currently known. Therefore, whether and how restriction of viral DNA would affect these CRISPR systems remains to be determined. While it was shown that bacterial genomes encoding RM systems are more likely to also harbor CRISPR-Cas systems[36], no study to-date has looked at the co-occurrence of RM and CRISPR systems based on their types. A bioinformatic

investigation to look at whether specific types of RM and CRISPR-Cas systems are preferentially found together could potentially reveal new ways in which these systems interact with each other.

Recent work revealed that prokaryotic Argonautes (pAgo), a family of guided-dependent nucleases, can be loaded with small DNA guides to mediate anti-phage defense and plasmid targeting[191,192]. The mechanisms by which these small DNA guides are created, as well as the exact sources from which they are acquired, remain unclear. Interestingly, the small DNA guides appear to be generated at DSBs and this process involves the DNA repair complex RecBCD[192]. This is strikingly similar to the generation of prespacer substrates for CRISPR acquisition that our laboratory and others have described[153–155,161]. Similar to the interplay between RM and CRISPR that was elucidated through my thesis work, there exists the possibility of a dynamic between RM and pAgo, where restriction creates DSBs that may trigger a pAgo defense against methylated phages that arise later during infection. Additionally, Cas9 cleavage also produces DSBs that can serve as substrates for spacer acquisition[161], suggesting that CRISPR targeting could also generate small DNA guides for pAgo. If true, all these events in concert could represent a way for different defense systems to unify and amplify the overall immune response against a specific invader.

Finally, anti-phage defense systems were found to frequently cluster together on bacterial genomes forming “defense islands”[193,194]. Exploiting this powerful

observation, within the last ten years, a multitude of novel anti-phage defense systems have been uncovered using a combination of bioinformatic approaches in concert with experimental validation[191,195–199]. Even though these systems co-exist together on defense islands, most have been studied in isolation to precisely decipher their mechanisms of action. My work illustrated the cooperation between RM and CRISPR systems, but the study of the interactions between different prokaryotic immune systems is still only in its infancy. As such, I predict that the synergy between RM and CRISPR-Cas is but one example out of many. Future studies will likely reveal many more instances of such interplay and it will be fascinating to unravel the complex dynamics that exist between diverse defense systems.

CHAPTER 7. MATERIALS AND METHODS

7.1 Bacterial strains and growth conditions

Growth of *S. aureus* was carried out in brain-heart infusion broth (BHI) at 37 °C with agitation at 220 RPM. Whenever needed, the media was supplemented with chloramphenicol at 10 µg/ml, erythromycin at 10 µg/ml, or spectinomycin at 250 µg/ml for maintenance of pC194[200], pE194[201], and pLZ12-derived[172] plasmids, respectively. For bacteriophage infection, the media was supplemented with 5 mM $CaCl_2$ to facilitate adsorption. The bacterial strains and phage used in this study can be found in Table 3.1.

7.2 Bacteriophage propagation

Overnight cultures of *S. aureus* were diluted 1:100 in fresh BHI supplemented with 5 mM $CaCl_2$ and the appropriate antibiotic, if needed, and grown for 1 h 15 min at 37 °C. A small volume of bacteriophage was added, and the cultures were grown for an additional 4 hours. Then, the cultures were spun down for 5 minutes at 4300 RPM and the lysates were filtered through 0.45 µm syringe filters (Acrodisc). Plaque formation assays were conducted to assess the number of infectious particles in the resulting stocks.

7.3 Plasmid construction

All the plasmids used in this study can be found in Table 3.2 along with their cloning strategies and the oligos used to construct them. All the constructed plasmids were electroporated in *S. aureus* as previously described elsewhere[202].

7.4 Strains construction

To make the *S. aureus* RN4220 *hsdM1/hsdS1* and *hsdM2/hsdS2* double knockout, sPM02, we used a method previously described[154]. Briefly, RN4220 was electroporated with pPM48 and integrants were checked by PCR (Phusion High-Fidelity, ThermoFischer) using the primer PM96/PM37 and then isolated. Selection for plasmid excision was performed using a temperature-sensitive *cat* targeting Cas9 phagemid, pJW326. Deletion of *hsdM1/hsdS1* was confirmed by PCR using primers PM37/PM174. The resulting strain was electroporated with pPM49 and integrants were isolated and confirmed by PCR with the primer pair PM181/PM185. Selection for plasmid excision was performed using a temperature-sensitive *cat* targeting Cas9 phagemid, pJW326. Deletion of *hsdM2/hsdS2* was confirmed by PCR using the primer pair PM181/PM184 resulting in strain sPM02.

7.5 Phage construction

To make phage ΦNM4γ4-*BglII*$_{AB}$ and ΦNM4γ4-*BglII*$_{AC}$, ΦNM4γ4[91] was spotted on a layer of *S. aureus* harboring pPM96 or pPM98, respectively. pPM96 and pPM98 harbor the *S. pyogenes* type II-A CRISPR system with a unique spacer targeting

ΦNM4γ4. Each spacer was chosen carefully to have its PAM sequence (5′-NGG-3′) in a region of DNA that resembles a BglII recognition site except for one bp. CRISPR Phage escapers for pPM96 and pPM98 were selected on soft agar and checked by PCR (Phusion High-Fidelity, ThermoFischer) using primers PM438/PM439 and PM454/PM455, respectively. Phage escapers that acquired a mutation in the PAM sequence creating a new BglII recognition site were spotted again on a lawn of RN4220 harboring either pPM96 or pPM98 to further purify the mutant phage from wild type phage. Then, a single plaque for each was isolated and propagated on *S. aureus* sPM02.

To make ΦNM4γ4-Δ*BglII*, ΦNM4γ4-*BglII*$_B$ and ΦNM4γ4-*BglII*$_C$, phage ΦNM4γ4, ΦNM4γ4-*BglII*$_{AB}$ and ΦNM4γ4-*BglII*$_{AC}$ were propagated in liquid cultures of *S. aureus* RN4220 harboring pPM117, a plasmid with roughly a 1000 bp of homology to the ΦNM4γ4 genome centered on a mutated BglII recognition sequence (site A). To select for phage with a mutated BglII site A, each resulting lysate was used in a plaque formation assay with *S. aureus* RN4220 harboring a plasmid with the *S. pyogenes* type II-A CRISPR and a spacer targeting wild type ΦNM4γ4 (pPM116) but not the recombined phage. Correct editing was confirmed by PCR using the primer pair PM493/PM472. A single plaque for each was propagated again on pPM117 to further purify edited phage from wild type. Then, a single plaque for each was isolated and propagated on *S. aureus* sPM02.

To make ΦNM4γ4-Δ*dnaC*, ΦNM4γ4 was propagated in a liquid culture of *S. aureus* RN4220 harboring a plasmid containing 700 bp arms with downstream and upstream homology to the *dnaC* gene of ΦNM4γ4, pPM135. The resulting lysate was spotted on *S. aureus* harboring pPM235 and pPM134. pPM134 provided the *dnaC* gene on a plasmid to allow phage with the deleted *dnaC* gene to form plaques and pPM235 allowed for selection of the recombinant phage by having a spacer targeting ΦNM4γ4 but not the *dnaC* deletion mutant. Correct deletion of *dnaC* was verified by PCR amplification with the primer pair PM592/PM593. A single plaque was propagated again on pPM235/pPM134, and one of the resulting plaques was used to lyse a liquid culture of pPM134 to obtain a final stock of ΦNM4γ4-Δ*dnaC*.

To make Φ12ρ1, Φ12γ3^{154} was propagated on *S. aureu*s RN4220 with pPM166 to mutate one of its BglII recognition sites through silent mutations. A plaquing assay with *S. aureus* harboring pPM167 was performed with the resulting lysate to select for recombinant phage. PCR amplification with PM824/PM825 was performed to check for the mutated BglII recognition site. One plaque was purified further on pPM167 and then propagated on RN4220. The resulting lysate was propagated on *S. aureu*s RN4220 with pPM168 to mutate a second BglII site through silent mutations. A plaquing assay with *S. aureus* harboring pPM169 was performed to select for recombinant phage. PCR amplification with PM822/PM823 was performed to verify that the second BglII recognition site was mutated. One plaque was purified further on pPM169 and then propagated on RN4220. The resulting lysate was propagated on *S. aureu*s with pPM170 to mutate a third BglII site through silent mutations. A plaquing

assay with *S. aureus* RN4220 harboring pPM171 was performed to select for recombinant phage. PCR amplification with PM826/PM827 was performed to verify that the third BglII recognition site was mutated. One plaque was purified further on pPM171 and then propagated on RN4220. The resulting phage was propagated on S. aureus pPM176 to mutate two *chi* sites through silent mutations. A plaquing assay with a 1:1 mixed culture of *S. aureus* harboring pPM177 or pPM178 was performed to select for phage with both *chi* sites mutated. Correct editing was verified by PCR amplification with PM902/PM903. Then, a single plaque was purified further on the mixed culture and then propagated on RN4220. The resulting phage stock was propagated on *S. aureus* harboring pPM179 to mutate a third *chi* site. The resulting lysate was used in a plaque assay with *S. aureus* harboring pPM180 to select for recombinant phage. Plaques were PCR amplified with the primer pair PM904/PM905 to check for mutation of the *chi* site. A single plaque was then spotted on a lawn of pPM180 to further purify the phage and a single plaque was then amplified on sPM02. This resulted in phage Φ12ρ1. The final stock of Φ12ρ1was again amplified by PCR to check that all the mutations created successively were still present.

To make ΦNM4γ4$_{\text{I-sceI}}$, ΦNM4γ4 was propagated in a liquid culture of *S. aureus* harboring a plasmid, pJW241, containing about 2000 bp of homology to the ΦNM4γ4 genome with an I-sceI recognition sequence in the middle. The resulting lysate was spotted on a lawn of *S. aureus* harboring a plasmid with the *S. pyogenes* type II-A CRISPR system with a spacer targeting ΦNM4γ4 but not the edited phage, pJW237.

Plaques were PCR amplified to check for proper editing and a single edited plaque was amplified on sPM02 to create a stock of ΦNM4γ4$_{I\text{-}sceI}$.

7.6 Colony formation assay

Ten-fold dilutions of *S. aureus* were spotted on BHI agar plate supplemented with the appropriate antibiotic, if needed. The plates were incubated overnight at 37 °C and colony-forming units (CFU) were enumerated the next day.

7.7 Plaque formation assay

Ten-fold dilutions of bacteriophage were spotted on a layer of *S. aureus* cells suspended in 50 % heart infusion agar (HIA) supplemented with 5 mM $CaCl_2$ and the appropriate antibiotic, if needed. The plates were incubated overnight at 37 °C and plaque-forming units (PFU) were enumerated the next day.

7.8 ΦNM4γ4 growth curve

Overnight cultures of *S. aureus* RN4220 were diluted 1:100 in 50 ml of fresh BHI supplemented with 5 mM $CaCl_2$ and grown for 1 h 15 min at 37 °C. Following incubation, the cultures were infected with ΦNM4γ4 at MOI = 0.1. Immediately, 1 ml was removed from each culture and filtered through a 0.45 µm syringe filter (Acrodisc) to remove the bacteria. Then every 10 minutes, 1 ml was removed and filtered from

each culture. All the filtered lysates were used in plaque formation assays to enumerate the number of phage particles at each time point.

7.9 SauI escaper assay

Overnight cultures of *S. aureus* RN4220 harboring either pLZ12 or pSauI (pPM61) were launched from single colonies. The following day, the cultures were diluted 1:100 in fresh BHI supplemented with spectinomycin at 250 µg/mL and 5 mM $CaCl_2$ and outgrown for about 1 hour 15min and normalized for optical density. The cultures were then aliquoted to a 96-well plate (Cellstar) and half were infected with unmodified ΦNM4γ4 (obtained by lysing *S. aureus* sPM02) at an MOI of 10. Absorbance at 600 nm was measured every 10 minutes for 22 hours using a microplate reader (TECAN Infinite 200 PRO). At the end of the experiment, the bacteriophage obtained from the infected wells were collected by briefly centrifuging the cultures at high speed and collecting the lysates. The lysates were then propagated on *S. aureus* sPM02. The original lysates and the ones passaged on sPM02 were used to perform plaque formation assays with RN4220/pLZ12 and RN4220/pPM61 to assess the sensitivity of each phage stock to the SauI R-M system.

7.10 CRISPR and RM synergy growth curves

Overnight cultures were launched from single colonies. The next day, the cultures were diluted 1:100 in fresh BHI and 5 mM $CaCl_2$. The cultures were outgrown for about 1 hour 15 minutes and then normalized for optical density and 150uL of cultures were

seeded in a flat-bottom 96-well plate (Cellstar). Half of the cultures were infected with phage and the absorbance at 600 nm was recorded every 10 minutes for 24 hours in a microplate reader (TECAN Infinite 200 PRO). After 24 hours, 2 µl from the uninfected and infected pCRISPR cultures were resuspended in 30 µl of colony lysis buffer (250 mM KCl, 5 mM $MgCl_2$, 50 mM Tris-HCl at pH 9.0 and 0.5% Triton X-100) supplemented with 200 ng/µl of lysostaphin. The reactions were incubated for 20 minutes at 37° C and then for 10 minutes at 98° C in a thermocycler. 0.5 µl of each reaction was used to PCR amplify the CRISPR locus using the primer pair PM223/PM225 with TopTaq© master mix (Qiagen). The PCR products were analyzed on a 2 % agarose gel stained with ethidium bromide and imaged with FluorChem HD2 (Protein simple). Additionally, one well for pSauI-pCRISPR cultures infected with ΦNM4γ4 was streaked on a BHI agar plate at the end of the experiment and incubated overnight at 37 °C. The next day, 50 single colonies were resuspended in 30 µl of colony lysis buffer (see above) supplemented 200 ng/µl of lysostaphin. 0.5 µl of each reaction was PCR amplified as described just above with the primer pair PM223/PM225. The PCR products were analyzed on a 2 % agarose gel stained with ethidium bromide and imaged with FluorChem HD2 (Protein simple). Also, each PCR product was sent for Sanger sequencing to obtain the sequences of the newly acquired spacers.

7.11 ΦNM4γ4-ΔdnaC infectivity assay

Plaque formation assays (see above) were performed using ΦNM4γ4 and ΦNM4γ4-ΔdnaC on lawns of *S. aureus* RN4220 harboring either an empty pE194 plasmid or pE194 with a copy of the dnaC gene, pDnaC. Liquid infection was also tested by recording the absorbance of infected culture over time. Briefly, overnight of *S. aureus* RN4220 harboring either pE194 or pDnaC were launched from single colonies. The next day, the cultures were diluted 1:100 in fresh BHI supplemented with 10 µg/ml erythromycin and 5 mM $CaCl_2$, outgrown for 1 hour 15 minutes and normalized for optical density. Cultures were seeded to wells of a flat bottom 96-well plate (Cellstar) and each culture was either uninfected or infected with ΦNM4γ4 or ΦNM4γ4-ΔdnaC at an MOI equal to 1. Growth curves were obtained by measuring the absorbance at 600 nm every 10 minutes for 20 hours in a microplate reader (TECAN Infinite 200 PRO).

7.12 DNA extraction and qPCR for phage DNA replication assay

Overnight cultures of pLZ12 and pPM61 (pSaul) were launched from single colonies. The next day, the cultures were diluted 1:100 in 20 ml of fresh BHI supplemented with 250 µg/ml spectinomycin and 5 µM $CaCl_2$ and outgrown until OD_{600} reached 0.4. The cultures were then infected with either ΦNM4γ4 and ΦNM4γ4-ΔdnaC at an MOI equal to 0.1. At 10 minutes and 30 minutes post-infection, 10 ml from each culture was collected and spun down at 10000 RPM for 3 minutes and the pellets were flash-frozen in liquid nitrogen and stored at -80 °C until DNA extraction. For DNA

extraction, the pellets were resuspended in 500uL of P1 buffer (Qiagen) supplemented with 2 µg/µl lysozyme (Ambi Products) and 200 ng/µL^{-1} of lysostaphin (Sigma-Aldrich) and incubated for 10 minutes at 37° C. Then, 60 µl of 10 % N-lauroylsarcosine (Sigma-Aldrich) was mixed-in and 600 µl of phenol/chloroform/isoamyl alcohol (Fischer Scientific) was added. The samples were vortexed at high force until the mixtures were white and opaque and then spun down at 13000 RPM for 5 minutes. 550 µl of the top layer was collected and 60 µl of 3M sodium acetate (Fisher Scientific) was added and briefly vortexed. 1 ml of 100% ethanol was added to precipitate the DNA and then spun down at 13000 RPM for 1 minute. The DNA pellets were washed with 200 µl of 70 % ethanol and spun down again for one minute. The ethanol was removed, and the DNA pellets were air-dried at room temperature. Once dried, they were resuspended in 500 µl of water. 100 ng of DNA was used from each sample for qPCR using Fast SYBR Green Master Mix (Life Technologies) and QuantStudio® 3 Real-Time PCR System (Applied Biosystems). The relative phage DNA content was calculated by the ΔΔCt method. Ct values were measured for a phage-specific primer set (PM1250/PM1251), then normalized to Ct values for a host-specific primer set (PM1248/1249) to control for total DNA content. Then phage DNA content values were normalized to the 10 min pLZ12 sample infected with ΦNM4γ4, which was set to 1.

7.13 30 minutes post-infection spacer acquisition assay

Spacer acquisition was performed as previously described[161] with slight modifications. Briefly, overnight cultures launched from single colonies were diluted 1:100 in 10 ml of BHI with 5mM $CaCl_2$ and outgrown for about 1 hour 15 minutes at

37° C with agitation. For the addA nuclease mutant, strain JW418 was used instead of RN4220. The optical densities (OD_{600}) were measured for all the cultures and each was diluted to OD_{600} = 0.3. Cultures were infected with the appropriate bacteriophage at MOI = 250 for experiments with pSauI and MOI= 25 for pBIII. 30 minutes post-infection, the cells were spun down at 10000 RPM for 3 minutes and flash-frozen in liquid nitrogen. The frozen pellets were stored at -80 °C until plasmid extraction for next-generation sequencing.

7.14 Spacer acquisition time course experiment

Overnight cultures of S. aureus RN4220/pLZ12/pGG32 and RN4220/pPM120/pPM118 were launched from single colonies. The next day, the cultures were diluted 1:100 in 1 liter of fresh BHI supplemented with 5mM $CaCl_2$, outgrown for 1 hour 15 minutes. Each culture was infected with ΦNM4γ4 at an MOI equal to 1. From each culture, 25 ml was retrieved at each time point (every 0.5 hours for the first 5 hours post-infection and then every 4 hours). From the 25ml sample, 1 ml aliquot was removed, spun down, and resuspended in 1ml of fresh BHI and used in colony formation assays. The rest of the 25 ml was spun down at 10000 RPM for 3 minutes at 4 °C. The supernatant was collected and filtered through a 0.45 µm syringe filter (Acrodisc) and used to perform plaque formation assays on both *S. aureus* RN4220/pLZ12 and RN4220/pPM120. The pellet was flash-frozen in liquid nitrogen and stored at -80° c until plasmid extraction for next-generation sequencing.

7.15 Phage DNA extraction and bisulfite sequencing on the time course samples

Phage DNA was extracted from the spacer acquisition time course samples collected during the first 3 hours post-infection. Additionally, DNA was also extracted from control stocks of unmethylated ΦNM4γ4 (propagated on *S. aureus* sPM02) and BglII methylated ΦNM4γ4 (propagated on *S. aureus* sPM02/pPM212). The phage supernatants were concentrated using Ultra-4 100K centrifugal 50-ml spin columns (Amicon) to about 500 µl. The concentrates were resuspended in 15 ml of DNase I buffer (DNase I buffer, 20 mM Tris-HCl, pH 8.0, and 2 mM $MgCl_2$) and concentrated again. This was done three times to buffer exchange the lysate with DNAse I buffer. 450 µL of each concentrate was treated with 25 units of DNase I (Sigma) for 1 h at 37° C. DNAse I was inactivated by adding 25 µl of stop solution (Sigma) and by heating the reactions for 10 min at 70 °C. The samples were then treated with 8 units of proteinase K (NEB) and 12 µl of 20 % SDS for 1 h at 37 °C. Finally, the DNA was extracted using phenol /chloroform/isoamyl alcohol extraction (Fisher) and resuspended in 100 µl of water. 500 ng of each DNA sample was bisulfite treated using the epiTect kit (Qiagen). Following bisulfite treatment, the region of DNA encompassing the single BglII site of ΦNM4γ4 was PCR amplified using pyroMARK PCR kit (Qiagen) and a set of top strand-specific primers (PM1033/PM1035). Correct PCR amplification was checked on a 2 % agarose gel stained with ethidium bromide. The amplicons were cleaned up using a QIAquick PCR purification kit (Qiagen) and prepared for high-throughput sequencing exactly as previously described[161]. Then, the samples were sequenced on a MiSeq instrument (Illumina). A custom python script

was used to compute the number of reads containing an intact BglII site 5'-AGATCT-3' and the number of reads containing a bisulfite modified BglII site 5'-AGATTT-3'.

7.16 Spacer acquisition with I-SceI cleavage

Overnight cultures launched from single colonies of RN4220/pC194/pWJ259/ and RN4220/pWJ250/pWJ259 were diluted 1:100 in 10 ml of BHI supplemented 5mM $CaCl_2$ and 1 mM IPTG. The cultures were outgrown for about 1 hour and 15 minutes at 37° C with agitation and then normalized for optical density. Following the outgrowth, the cultures were infected with $\Phi NM4\gamma4_{I\text{-}SceI}$ at an MOI equal to 1. The cultures were collected 5 hours post-infection and quickly spun down at 10000 RPM for 3 minutes. The resulting pellets were flash-frozen in liquid nitrogen and store at -80 °C until plasmid extraction for next-generation sequencing.

7.17 CRISPR plasmid extraction and amplification for next-generation sequencing

For all the deep sequencing spacer acquisition experiments, the CRISPR plasmids from the frozen pellets were extracted using a modified QIAprep Spin Miniprep Kit (Qiagen) protocol described previously (Modell *et al.*, 2017). The CRISPR loci were amplified using 250 ng of plasmid DNA with Phusion High-Fidelity enzyme (Thermo Fischer). For the acquisition time course experiment, each PCR reaction was performed using modified PM900 and PM901 primers containing 5 random bp and a unique 3-6bp barcode at their 5' ends to keep track of each sample in the downstream NGS data. Successful amplification was checked on a 2 % agarose gel stained with

ethidium bromide. Then, the PCR products were cleaned using a MinElute PCR purification kit (Qiagen). The PCR amplicons corresponding to expanded CRISPR loci were extracted using the PippinHT instrument (Sage Science) set on the range mode (136bp to 450bp) with a 2 % agarose gel cassette. For all the other acquisition experiments, PCR amplifications of the CRISPR loci were performed as described elsewhere with some modifications[148]. In short, the plasmids were PCR amplified with a cocktail of three reverse primers mixed 1:1 (PM375, PM376, and PM377) and one forward primer PM168 to preferentially amplify expanded CRISPR loci[91]. For each sample, a modified pPM168 primer with 5 random nucleotides and a unique 3-6 bp barcode at its 5' end was used to track each sample in the resulting next-generation sequencing data. The PCR products were analyzed on a 2 % agarose gel stained with ethidium bromide and locations corresponding to expanded CRISPR loci were gel extracted using QIAquick gel extraction kit (Qiagen). For all the acquisition experiments, the PCR amplicons were prepared for high-throughput sequencing exactly as previously described[161]. Samples from the acquisition time course were sequenced on a NovaSeq instrument (Illumina) and all the other experimental samples were sequenced on a MiSeq instrument (Illumina).

7.18 Statistical analysis

7.18.1 High Throughput Sequencing Data Analysis

Using python, the spacer sequences were extracted and the number of reads for each spacer was recorded from the Illumina raw FASTQ files. For all the deep

sequencing experiments (except for the time course acquisition experiment), the number of reads for each spacer was corrected to account for the PCR biased introduced by the reverse primer cocktail (PM375, PM376, PM377), as previously described[154]. Each spacer sequence was aligned to the phage, plasmids, and bacterial genome. If an exact match was found (except for Figure 2.10-E), the origin of the spacer and its PAM sequence was recorded. Additionally, for spacer sequences matching the phage genome, the position of the spacer sequence on the genome was recorded. For the quantification of spacer acquisition in the presence of the SauI RM system (Figure 2E), the spacers were matched to each possible DNA source using Bowtie[203,204] on usegalaxy.org[205] to allow for mismatches. To create the pattern of acquisition for each phage, the number of reads for each phage spacer sequence with a correct PAM (5′-NGG-3′) was binned in roughly 1 kb bins along the phage genome (985 bp for ΦNM4γ4 and its mutants and 993 bp for Φ12ρ1). Following binning, the number of reads in each bin was normalized to the number of PAM sequences in the phage genome within that bin. Finally, reads per million were calculated as RPM_{phage}, as previously described[154]. The sequence 5'-AGACAAAAATAGTCTACGAG-3' was removed from our NovaSeq data as it corresponds to the leader sequence but is sometimes perceived by our script as a spacer in some reads representing a DNA recombination event. The error bars in the quantification of spacer acquisition in Figure 2.10-E represent the SD of 3 biological replicates. Statistical analysis was carried using Prism 9.2.0 (GraphPad).

7.18.2 Growth curves

The error bars for growth curves in Figures 2.7-A, 2.8-A, 3.2-D, and 3.3-A represent the SD of 5 biological replicates. In Figures 2.3-A, 2.5-A, and2.10-B, the error bars represent the SD of 3 biological replicates. Statistical analysis was carried using Prism 9.2.0 (GraphPad).

7.18.3 Plaque assays

The error bars for the quantification of plaque assays in Figures 2.2, 2.3-B, 2.5-B, and 2.11-A represent the SD of 3 biological replicates. In Figure 3.2-C the error bars represent the SD of 4 biological replicates. The error bars in Figures 3.6-C, 4.1-B, and 4.1-C represent the SD of 3 technical replicates. Statistical analysis was carried using Prism 9.2.0 (GraphPad).

7.18.4 Colony formation assay

The error bars in Figure 4.1-A represent the SD of 3 technical replicates. Statistical analysis was carried using Prism 9.2.0 (GraphPad).

7.18.5 qPCR quantification

The error bars in Figures 2.10-C and 2.10-D represent the SD of 3 biological replicates. Statistical analysis was carried using Prism 9.2.0 (GraphPad).

Table 7.1 Bacterial strains and phages used in this study

Name	Description	Made in this study?
***S. aureus* RN4220**	Laboratory strain of *S. aureus*	No[206]
***S. aureus* sPM02**	RN4220 *HsdS/HsdM* double mutant	Yes
***S. aureus* TB4**	*S. aureus* Newman cured of all prophages	No[167]
ΦNM4γ4	Lytic only ΦNM4 mutant	No[91]
ΦNM4γ4-Δ*dnaC*	ΦNM4γ4 mutant with gene *DnaC* deleted	Yes
ΦNM4γ4-Δ*BglII*	ΦNM4γ4 mutant with mutation in its single BglII site (site A)	Yes
ΦNM4γ4-BglII$_{AB}$	ΦNM4γ4 mutant with a second BglII site (site B)	Yes
ΦNM4γ4-BglII$_{AC}$	ΦNM4γ4 mutant with a second BglII site (site C)	Yes
ΦNM4γ4-BglII$_{B}$	ΦNM4γ4-BglII$_{AB}$ with a mutated BglII site (site A)	Yes
ΦNM4γ4-BglII$_{C}$	ΦNM4γ4-BglII$_{AC}$ with a mutated BglII site (site A)	Yes
ΦNM4γ4-I-sceI	ΦNM4γ4 mutant with an I-sceI site	Yes
Φ12γ3	Lytic only Φ12 mutant	No[154]
Φ12ρ1	Φ12γ3 mutant with 4 mutated BglII sites and 3 mutated chi sites	Yes

Table 7.2 Plasmids used in this study

Plasmid name	Plasmid contents	Made in this study?
pAV268	Cloning vector	Yes
pC194	Empty cloning vector	No[200]
pDB114	Type II-A, BsaI spacer for cloning new spacers	No[207]
pE194	Empty cloning vector	No[201]
pGG32	S. pyogenes Type II-A CRISPR-Cas with a single repeat (pCRISPR)	No[91]
pWJ215	*S. aureus* codon optimized *I-sceI* under an IPTG inducible promoter	No[154]
pWJ237	Type II-A, ΦNM4γ4 targeting spacer	Yes
pWJ241	Homology to ΦNM4γ4 with an I-sceI recognition sequence for allelic exchange	Yes
pJW250	Same as pWJ215 but with a stronger promoter	Yes
pJW259	*S. pyogenes* Type II-A CRISPR-Cas with a single repeat and hyper-*cas9*	No[154]
pLM9B	Empty cloning vector	No[208]

pLZ12	Empty cloning vector	No[172]
pPM48	Upstream and downstream homology to *S. aureus* RN4220 *hsdM1/hsdS1* for allelic exchange	Yes
pPM49	Upstream and downstream homology to *S. aureus* RN4220 *hsdM2/hsdS2* for allelic exchange	Yes
pPM61	Fixed SauI *hsdR* gene (pSauI)	Yes
pPM82	Full BglII R-M system (pBglII)	Yes
pPM96	Type II-A, ΦNM4γ4 targeting spacer	Yes
pPM98	Type II-A, ΦNM4γ4 targeting spacer	Yes
pPM116	Type II-A, ΦNM4γ4 targeting spacer	Yes
pPM117	Homology to ΦNM4γ4 with a mutated BglII site for allelic exchange	Yes
pPM118	Same as pGG32 but with a mutated BglII site in the pC194 plasmid backbone	Yes
pPM120	Full BglII R-M system (pBglII) on pLZ12 (same as pPM82 without BglII site on the pLZ12 backbone)	Yes
pPM134	*dnaC* gene from ΦNM4γ4 (pDnaC) on pE194	Yes
pPM135	Upstream and downstream homology to ΦNM4γ4 *dnaC* for allelic exchange	Yes
pPM166	Homology to Φ12γ3 for allelic exchange	Yes
pPM167	Type II-A, Φ12γ3 targeting spacer	Yes
pPM168	Homology to Φ12γ3 for allelic exchange	Yes
pPM169	Type II-A, Φ12γ3 targeting spacer	Yes
pPM170	Homology to Φ12γ3 for allelic exchange	Yes
pPM171	Type II-A, Φ12γ3 targeting spacer	Yes
pPM176	Homology to Φ12γ3 for allelic exchange	Yes
pPM177	Type II-A, Φ12γ3 targeting spacer	Yes
pPM179	Homology to Φ12γ3 for allelic exchange	Yes
pPM180	Type II-A, Φ12γ3 targeting spacer	Yes
pPM212	*bglIIM* gene	Yes
pPM235	Type II-A, ΦNM4γ4 *dnaC* targeting spacer	Yes
pT181	Cloning vector	No[209]
pWJ244	Suicide vector for allelic replacement in *S. aureus*	No[154]
pWJ326	Phagemid vector	No[154]

Table 7.3 Primers used in this study

Name	Sequence	Purpose
AV108	GCAAAAACAGGTTTAAGCCTCGC	Cloning
AV109	AATGAGTGGCAAAATGCTAGCC	Cloning
AV186	AATCGATAACCACATAACAGTCATAAAAC	Cloning
AV744	GGAGATATACATATGATGAAGGGTAAAATTGCACTTTATTC	Cloning
AV745	TTAGAATAGGCGCGCCGTCACTCTAAAAAAGTGGTAAAAGTG	Cloning
AV746	CATATGTATATCTCCAATTGTTATCCGCTCACAATTCC	Cloning
AV759	TTATGTGGTTATCGATTCATGTTCATATTTATCAGAGCTCGTGC	Cloning
AV760	CGGCGCGCCTATTCTAAGAGTGATCGTTAAATTTATACTGCAA TCGG	Cloning
JW554	AAACGTATTTAAAGGAGGTGATTACCATGCTTAAG	Cloning
JW555	AAAACTTAAGCATGGTAATCACCTCCTTTAAATAC	Cloning
JW556	CTATGAGTGGCTAGCATTTTGCCACTCATTGTGTACTTTGATT TAGGTAAAACATCAGG	Cloning
JW557	TAGGGATAACAGGGTAATTTTAAGCATGGTAATCACCTCC	Cloning
JW558	ATTACCCTGTTATCCCTAATTTTAGGATATAGCTTCTGGGCG	Cloning
JW567	GTGCTCTGCGAGGCTTAAACCTGTTTTTGCCCAAACACCACG TATTAACGC	Cloning
JW582	CTTTATCTACAAGGTGTGGCATAATGTGTGTAATTGTGAGCGG ATAACAATTAAGC	Cloning
JW584	ACACACATTATGCCACACCTTG	Cloning
PM37	GAATATTAACTACAGGTGTGTTTACAGCAG	Cloning
PM96	AAGCTGTATTACCAGATCTATCAG	Cloning
PM160	CATAAATATATATTTTAAAAATATCCCACATGAAAGACGAAAAG AAGCATGA	Cloning
PM161	CCTAAATTTAATGTTCGTTTTCCTCCGCCTTTGAATAAGTAA	Cloning
PM164	CATAAATATATATTTTAAAAATATCCCACTCGCAGTATATCGTG AAGCTAA	Cloning
PM165	TGAAAGCATTTGAAAGCTACCCCTGATAACAACCTATTAATTG AAATG	Cloning
PM168	AGTGCGATTACAAAATTTTTTAGAC	Specific PCR amplification of expanded CRISPR loci
PM174	TCAGCACCAACTTCTTCAGTTG	Cloning
PM181	CTTTGCGTATTTTGATTTCATCTTATC	Cloning
PM184	GTAAAAATACAATTGTTACCAACAAGCATG	Cloning
PM185	TTTCACCAACTCAATTGTTGATTGG	Cloning
PM210	GATTAATGGAGGGACTTGAATCTGTGCCAGTTCGTAATGTC	Cloning

PM211	CGGTGTGTAATGATTCAGCCGGTCATAACCTGAAGGAAGATC TG	Cloning
PM212	CATTACGAACTGGCACAGATTCAAGTCCCTCCATTAATCGTAG	Cloning
PM213	TCTTCCTTCAGGTTATGACCGGCTGAATCATTACACACCG	Cloning
PM326	TGGCACAGATCCACACACCATCTTCCTTAAC	Cloning
PM327	GGTTATGACCGCCATTTTATTGTTCATTTTGAACTTTTTTATAT AC	Cloning
PM328	TGGTGTGTGGATCTGTGCCAGTTCGTAATG	Cloning
PM329	ATAAAATGGCGGTCATAACCTGAAGGAAGATCTG	Cloning
PM438	CCAAAACTGGGTTAATTCTAATAGTTGG	Cloning
PM439	GCGTATACAGTTCGTCTACAG	Cloning
PM375	AAAACAGCATAGCTCTAAAACG	Specific PCR amplification of expanded CRISPR loci
PM376	AAAACAGCATAGCTCTAAAACA	Specific PCR amplification of expanded CRISPR loci
PM377	AAAACAGCATAGCTCTAAAACT	Specific PCR amplification of expanded CRISPR loci
PM384	AAACACTAGGTTTGCTAGTATCTTTTGAATATGGG	Cloning
PM385	AAAACCCATATTCAAAAGATACTAGCAAACCTAGT	Cloning
PM388	AAACTCATCTTCAATAGTAGCCATTAATGCCGAAG	Cloning
PM389	AAAACTTCGGCATTAATGGCTACTATTGAAGATGA	Cloning
PM454	TTTGACCAAGCTGTTATCTTTGG	Cloning
PM455	GCATACTCAATACCTTGTAAGATACC	Cloning
PM463	AAACGGGTTTAAAGAAGATTTAATAGATCTTAGCG	Cloning
PM464	AAAACGCTAAGATCTATTAAATCTTCTTTAAACCC	Cloning
PM465	TCTACTTAAGTCAATTAAATCTTCTTTAAACCCTTTAACCGC	Cloning
PM466	ATTTAATTGACTTAAGTAGACATAGTTTTGATATTGATTCCAGC AG	Cloning
PM467	TTTTTAAAATTTCTAAGAGGCTAACTAGTATGATTCG	Cloning
PM468	ATAAATATATCACTTGTAACACCTCTGATACC	Cloning
PM469	GTTACAAGTGATATATTTATGTTACAGTAATATTGACTTTTAAA AAAGG	Cloning
PM470	CCTCTTAGAAATTTTAAAAATATCCCACTTTATCCAATTTTCG	Cloning
PM472	AATTCTTCCGAAAGAACGTTTAACTG	Cloning

PM476	CAAGAGCGTTATGACTGTTATGTGGTTATCGATTATAGG	Cloning
PM477	ACATAACAGTCATAACGCTCTTGCATAATTCACGCTGAC	Cloning
PM478	CTGAAGGATTAATGGGATCCTAATGAATTCATCTGC	Cloning
PM479	CATTAGGATCCCATTAATCCTTCAGGTTATGACCGCC	Cloning
PM493	CGTGATTGGCTTATTGGTGG	Cloning
PM495	ATAAATATATGTCTAATACTGTTTTAATTAAGTTATCGATATCCG	Cloning
PM496	TTTTTAAAATTGCTAAAGCTAAAGGGATATTATTGTATTTCCTAAGTC	Cloning
PM498	AGCTTTAGCAATTTTAAAAATATCCCACTTTATCCAATTTTCG	Cloning
PM499	AGTATTAGACATATATTTATGTTACAGTAATATTGACTTTTAAAAAAGG	Cloning
PM522	GATAACAATTGGGATTAGGGGGATATTATGAAACC	Cloning
PM523	ACGATCACTCTTAGTTTTTAAAAATTCTTTGGTTACCATGCATC	Cloning
PM553	AAGCTTGTACTTAGGAGGATGATTATTTATGAAACCACTATTCAGCGAAAAG	Cloning
PM554	AAATAATCATCCTCCTAAGTACAAGCTTAATTGTTATCCGCTCACAATTCC	Cloning
PM555	GCATCTCGCTCCCTGAAATCGTCAATATCCCCCTAATCCCAATAACTTTC	Cloning
PM556	GACGATTTCAGGGAGCGAGATGCATGGTAACCAAAGAATTTTTAAAAACTAAAC	Cloning
PM592	TAATTATTCACCCCCAATCTAACGC	Cloning
PM593	TCAATCGTTGCGTTTATATATGCTTGC	Cloning
PM783	ATTTTAAAAATATCCCACTTTATCCAATTTTCG	Cloning
PM783	ATTTTAAAAATATCCCACTTTATCCAATTTTCG	Cloning
PM784	ATATATTTATGTTACAGTAATATTGACTTTTAAAAAAGG	Cloning
PM784	ATATATTTATGTTACAGTAATATTGACTTTTAAAAAAGG	Cloning
PM785	TTACTGTAACATAAATATATGTTGTTTTTACAGGTAATTTTGACACTG	Cloning
PM786	TACTACGGTCAAGATATATAAATTTCCCTGATTTAGAATTAGTCTTTTTATTC	Cloning
PM787	TTATATATCTTGACCGTAGTAAAAGAATCGTTAGACAATCTGATATCAC	Cloning
PM788	AAGTGGGATATTTTTAAAATGAAATAAATTCCAAGTATTTACGCGC	Cloning
PM789	AAACGATTGTCTAACGATTCTTTTAGATCTGTCAG	Cloning
PM790	AAAACTGACAGATCTAAAAGAATCGTTAGACAATC	Cloning
PM797	AAACCATCCTCAAGACTTATTAAGTCAATTAGTTG	Cloning
PM798	AAAACAACTAATTGACTTAATAAGTCTTGAGGATG	Cloning
PM803	AAGTGGGATATTTTTAAAATATTTTTCTGCAATAGAGGCAAG	Cloning
PM804	TTTACGGTTTAAGTCCAACCCCATAAATTCAACAACTTGACTAC	Cloning

PM805	GGTTGGACTTAAACCGTAAATGGTCAAAGTTAAAAACG	Cloning
PM806	TTACTGTAACATAAATATATAACTTTTCTTCTAGCCATCATTCC	Cloning
PM807	GCAACAGCTCGTGATTTACTTGCGATTTCTATAGCAAG	Cloning
PM808	AGTAAATCACGAGCTGTTGCTTGAACAATATTCTCGACTAAC	Cloning
PM809	AAACTTAGTCGAGAATATTGTTCAAGCAACTGCAG	Cloning
PM810	AAAACTGCAGTTGCTTGAACAATATTCTCGACTAA	Cloning
PM816	TTTTTAAAATGACTTTCCAGTAACTGCAATTG	Cloning
PM817	TCTTTAATTAAATCACGTGCTATTTCTAATTCAGTATCTGAAAT ATAATGC	Cloning
PM818	GCACGTGATTTAATTAAAGAACAACGTTTTGATGATTTAGATTT ATTAC	Cloning
PM819	ATAAATATATGCGTCTTTCGGGATTTTACAG	Cloning
PM820	CGAAAGACGCATATATTTATGTTACAGTAATATTGACTTTTAAA AAAGG	Cloning
PM821	CTGGAAAGTCATTTTAAAAATATCCCACTTTATCCAATTTTCG	Cloning
PM822	CCTGAACATGATCTTGAAAAATGGC	Cloning
PM823	GCTCTTCCACTAGGCAGTTC	Cloning
PM824	CCATTCAACGTCTACACTAGTAGG	Cloning
PM825	CGAAATTTTCGTATTGTCAACATTAAATACG	Cloning
PM826	CAATGTTCATCCTCAAGACTTATTAAGTC	Cloning
PM827	CACACAATAATTCATCGCCGC	Cloning
PM848	AAAATTGAGGCTGATTTATCTAATAACCTTATAGCTGAAATAGA AAAAAG	Cloning
PM849	ATAAATATATCAGTCTTACTCAATTCTTTAACAGTG	Cloning
PM850	TCTGGACCAGCCTCAATCAAGGCTAAAAATCCAC	Cloning
PM851	TTTTTAAAATTCAATAGAGGACACTTCTGCAG	Cloning
PM852	CCTCTATTGAATTTTAAAAATATCCCACTTTATCCAATTTTCG	Cloning
PM853	AGTAAGACTGATATATTTATGTTACAGTAATATTGACTTTTAAA AAAGG	Cloning
PM854	AAACACAAAGCAAAAAAACTTTTGAAAATTGAAGG	Cloning
PM855	AAAACCTTCAATTTTCAAAAGTTTTTTTGCTTTGT	Cloning
PM858	TCGCCTGCAGCCTCAACCATTTCTTCTGCAC	Cloning
PM859	TTTTTAAAATAATGTTGCTAATAAAGCTATTGGCG	Cloning
PM860	ATGGTTGAGGCTGCAGGCGATAAAATCAAAG	Cloning
PM861	ATAAATATATTTTAATACGTTGGCCTTGTCTTG	Cloning
PM862	ACGTATTAAAATATATTTATGTTACAGTAATATTGACTTTTAAAA AAGG	Cloning
PM863	TAGCAACATTATTTTAAAAATATCCCACTTTATCCAATTTTCG	Cloning
PM864	AAACTCAAAAAAGGTGCAGAAGAAATGGTTGAAGG	Cloning
PM865	AAAACCTTCAACCATTTCTTCTGCACCTTTTTTGA	Cloning

PM900	CAAAATTTTTTAGACAAAAATAGTC	PCR amplification of CRISPR loci
PM901	TAACCCTCTTTCTCAAGTTATC	PCR amplification of CRISPR loci
PM902	ATAGTGCTGCAGATATTCAATATATGGG	Cloning
PM903	CTACAGCATCTTTTTTAGATTTTGCC	Cloning
PM904	GATGCAATTTCAAGTTCGGC	Cloning
PM905	CAATTGGAAATGTAAGTGTGCTCC	Cloning
PM1206	AAACAGAAAAGTAAGAGTAATCGGAGACGATTTCG	Cloning
PM1207	AAAACGAAATCGTCTCCGATTACTCTTACTTTTCT	Cloning
PM1248	TGTCAAATGTGCGCCTTCACG	qPCR
PM1249	CCGTTCTGGTTCGAGTTTGGTTC	qPCR
PM1015	TGGCACAGATGTAGGGTTAATAGAAGACTTTGAAACTGATATTG	Cloning
PM1016	TTAACCCTACATCTGTGCCAGTTCGTAATGTCTGGTC	Cloning
PM1250	TAGGAGCTATACGTGGTATGACATCG	qPCR
PM1251	CACCAGTTCCACTACAACGTGAC	qPCR
W1005	GTGAAGACGAAAGGGCCTCGTG	Cloning
W1055	GTGGGATATTTTTAAAATATATATTTATG	Cloning

Table 7.4 Cloning strategies

Name	Cloning strategy
pAV268	PCR amplification of pLM-9B using AV746/AV186, *S. aureus* ST398 with AV744/AV745 and pE194 with AV759/AV760, followed by Gibson assembly of the three PCR fragments[210].
pWJ237	Linearization of pDB114 with BsaI enzyme, followed by oligos annealing and ligation of JW554/JW555 into the linearized vector.
pJW241	PCR amplification of pT181 using AV108/AV109, and of ΦNM4γ4 with JW556/JW557 and JW558/JW567, followed by Gibson assembly of the three PCR fragments.
pWJ250	PCR amplification of pJW215 using JW582/JW584, followed by Gibson assembly of the PCR fragment.
pPM48	PCR amplification of pWJ244 using W1005/W1055, and of *S. aureus* RN4220 with PM160/PM161 and PM162/PM163, followed by Gibson assembly of the three PCR fragments.
pPM49	PCR amplification of pWJ244 using W1005/W1055, and of *S. aureus* RN4220 with PM164/PM165 and PM166/PM167, followed by Gibson assembly of the three PCR fragments.
pPM61	PCR amplification of pLZ12 using PM210/PM211, and of *S. aureus* TB4 with PM212/PM213, followed by Gibson assembly of the two PCR fragments.
pPM82	PCR amplification of pLZ12 using PM328/PM329, and of *B. subtilis subsp. Globigii* with PM326/PM327, followed by Gibson assembly of the two PCR fragments.
pPM96	Linearization of pDB114 with BsaI enzyme, followed by oligos annealing and ligation of PM384/PM385 into the linearized vector.
pPM98	Linearization of pDB114 with BsaI enzyme, followed by oligos annealing and ligation of PM388/PM389 into the linearized vector.
pPM116	Linearization of pDB114 with BsaI enzyme, followed by oligos annealing and ligation of PM463/PM464 into the linearized vector.
pPM117	PCR amplification of pC194 using PM469/PM470, and of ΦNM4γ4 with PM465/PM467 and PM466/PM468, followed by Gibson assembly of the three PCR fragments.
pPM118	PCR amplification of pGG32 using PM476/PM477, followed by circularization of the PCR product by Gibson assembly.
pPM120	PCR amplification of pPM82 using PM478/PM479, followed by circularization of the PCR product by Gibson assembly.
pPM134	PCR amplification of pAV268 using PM522/PM554, and of ΦNM4γ4 with PM523/PM553, followed by Gibson assembly of the two PCR fragments.
pPM135	PCR amplification of pC194 using PM499/PM498, and of ΦNM4γ4 with PM496/PM555 and PM495/PM556, followed by Gibson assembly of the three PCR fragments.
pPM166	PCR amplification of pC194 using PM783/PM784, and of Φ12γ3 with PM785/PM786 and PM787/PM788, followed by Gibson assembly of the three PCR fragments.
pPM167	Linearization of pDB114 with BsaI enzyme, followed by oligos annealing and ligation of PM789/PM790 into the linearized vector.

pPM168	PCR amplification of pC194 using PM820/PM821, and of Φ12γ3 with PM816/PM817 and PM818/PM819, followed by Gibson assembly of the three PCR fragments.
pPM169	Linearization of pDB114 with BsaI enzyme, followed by oligos annealing and ligation of PM797/PM798 into the linearized vector.
pPM170	PCR amplification of pC194 using PM783/PM784, and of Φ12γ3 with PM803/PM804 and PM805/PM806, followed by Gibson assembly of the two PCR fragment. The resulting plasmid was then amplified with PM807/PM808, followed by circularization of the PCR fragment by Gibson assembly.
pPM171	Linearization of pDB114 with BsaI enzyme, followed by oligos annealing and ligation of PM809/PM810 into the linearized vector.
pPM176	PCR amplification of pC194 using PM852/PM853, and of Φ12γ3 with PM850/PM851, PM846/PM847, and PM848/849, followed by Gibson assembly of the four PCR fragments.
pPM177	Linearization of pDB114 with BsaI enzyme, followed by oligos annealing and ligation of PM854/PM855 into the linearized vector.
pPM179	PCR amplification of pC194 using PM862/PM863, and of Φ12γ3 with PM858/PM859 and PM860/PM861, followed by Gibson assembly of the three PCR fragments.
pPM180	Linearization of pDB114 with BsaI enzyme, followed by oligos annealing and ligation of PM864/PM865 into the linearized vector.
pPM212	PCR amplification of pPM120 using PM1015/PM1016, followed by circularization of the PCR product by Gibson assembly.
pPM235	Linearization of pDB114 with BsaI enzyme, followed by oligos annealing and ligation of PM1206/PM1207 into the linearized vector.

CHAPTER 8. REFERENCES

1. Twort, F. W. An Investigation On The Nature Of Ultra-Micropscopic Viruses. *Lancet* 186, 1241–1243 (1915).

2. D'Herelle, F. Sur un microbe invisible antagoniste des bacilles dysentériques. *C.R. Acad. Sci.* (1917).

3. HERSHEY, A. D. & CHASE, M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *The Journal of general physiology* 36, 39–56 (1952).

4. Luria, S. E. & Delbrück, M. MUTATIONS OF BACTERIA FROM VIRUS SENSITIVITY TO VIRUS RESISTANCE. *Genetics* 28, 491–511 (1943).

5. Salmond, G. P. C. & Fineran, P. C. A century of the phage: past, present and future. *Nat Rev Microbiol* 13, 777–786 (2015).

6. Sauer, B. Functional expression of the cre-lox site-specific recombination system in the yeast Saccharomyces cerevisiae. *Mol Cell Biol* 7, 2087–2096 (1987).

7. Weiss, B. & Richardson, C. C. Enzymatic breakage and joining of deoxyribonucleic acid, I. Repair of single-strand breaks in DNA by an enzyme system from Escherichia coli infected with T4 bacteriophage. *Proc National Acad Sci* 57, 1021–1028 (1967).

8. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821 (2012).

9. Suttle, C. A. Viruses in the sea. *Nature* 437, 356–361 (2005).

10. Dion, M. B., Oechslin, F. & Moineau, S. Phage diversity, genomics and phylogeny. *Nature Publishing Group* 18, 125–138 (2020).

11. Ackermann, H.-W. 5500 Phages examined in the electron microscope. *Arch Virol* 152, 227–243 (2007).

12. Herskowitz, I. & Hagen, D. The Lysis-Lysogeny Decision of Phage lambda: Explicit Programming and Responsiveness. *Annu Rev Genet* 14, 399–445 (1980).

13. Ptashne, M. *A genetic switch: phage lambda revisited*. vol. 3 (Cold Spring Harbor Laboratory Press Cold Spring Harbor, NY, 2004).

14. Benzer, S. & Jacob, F. ETUDE DU DEVELOPPEMENT DU BACTERIOPHAGE AU MOYEN DIRRADIATIONS PAR LA LUMIERE ULTRA-VIOLETTE. in *ANNALES DE L INSTITUT PASTEUR* vol. 84 186–204.

15. Rascovan, N., Duraisamy, R. & Desnues, C. Metagenomics and the Human Virome in Asymptomatic Individuals. *Annu Rev Microbiol* 70, 125–141 (2016).

16. Suttle, C. A. Marine viruses — major players in the global ecosystem. *Nat Rev Microbiol* 5, 801–812 (2007).

17. Sokol, N. W. *et al.* Life and death in the soil microbiome: how ecological processes influence biogeochemistry. *Nat Rev Microbiol* 1–16 (2022) doi:10.1038/s41579-022-00695-z.

18. Mayneris-Perxachs, J. *et al.* Caudovirales bacteriophages are associated with improved executive function and memory in flies, mice, and humans. *Cell Host Microbe* (2022) doi:10.1016/j.chom.2022.01.013.

19. Parikka, K. J., Romancer, M. L., Wauters, N. & Jacquet, S. Deciphering the virus-to-prokaryote ratio (VPR): insights into virus–host relationships in a variety of ecosystems. *Biol Rev* 92, 1081–1100 (2017).

20. McLaughlin, R. N., Malik, H. S., Levine, J. D., Kronauer, D. J. C. & Dickinson, M. H. Genetic conflicts: the usual suspects and beyond. *J Exp Biol* 220, 6–17 (2017).

21. Delbrück, M. The Burst Size Distribution in the Growth of Bacterial Viruses (Bacteriophages). *J Bacteriol* 50, 131–135 (1945).

22. Rostøl, J. T. & Marraffini, L. (Ph)ighting Phages: How Bacteria Resist Their Parasites. *Cell Host and Microbe* 25, 184–194 (2019).

23. Samson, J. E., Magadán, A. H., Sabri, M. & Moineau, S. Revenge of the phages: defeating bacterial defences. *Nat Rev Microbiol* 11, 675–687 (2013).

24. Fernández, L., Rodríguez, A. & García, P. Phage or foe: an insight into the impact of viral predation on microbial communities. *Isme J* 12, 1171–1179 (2018).

25. LURIA, S. E. & HUMAN, M. L. A nonhereditary, host-induced variation of bacterial viruses. *Journal of Bacteriology* 64, 557–569 (1952).

26. ANDERSON, E. S. & FELIX, A. Variation in Vi-Phage II of Salmonella typhi. *Nature* 170, 492–494 (1952).

27. BERTANI, G. & WEIGLE, J. J. Host controlled variation in bacterial viruses. *Journal of Bacteriology* 65, 113–121 (1953).

28. Luria, S. E. HOST-INDUCED MODIFICATIONS OF VIRUSES. *Cold Spring Harb Sym* 18, 237–244 (1953).

29. ARBER, W. & DUSSOIX, D. Host specificity of DNA produced by Escherichia coli. I. Host controlled modification of bacteriophage lambda. *Journal of molecular biology* 5, 18–36 (1962).

30. Loenen, W. A. M., Dryden, D. T. F., Raleigh, E. A., Wilson, G. G. & Murray, N. E. Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Research* 42, 3–19 (2014).

31. Maguin, P. & Marraffini, L. A. From the discovery of DNA to current tools for DNA editing. *The Journal of experimental medicine* 218, (2021).

32. Roberts, R. J., Vincze, T., Posfai, J. & Macelis, D. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Research* 43, D298–D299 (2015).

33. Roberts, R. J. *et al.* A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Research* 31, 1805–1812 (2003).

34. Smith, H. O. & Nathans, D. A Suggested nomenclature for bacterial host modification and restriction systems and their enzymes. *J Mol Biol* 81, 419–423 (1973).

35. Tock, M. R. & Dryden, D. T. The biology of restriction and anti-restriction. *Current Opinion in Microbiology* 8, 466–472 (2005).

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

37. Titheradge, A. J. B. Families of restriction enzymes: an analysis prompted by molecular and genetic data for type ID restriction and modification systems. *Nucleic Acids Res* 29, 4195–4205 (2001).

38. Kennaway, C. K. *et al.* Structure and operation of the DNA-translocating type I DNA restriction enzymes. *Gene Dev* 26, 92–104 (2012).

39. Kennaway, C. K. *et al.* The structure of M.EcoKI Type I DNA methyltransferase with a DNA mimic antirestriction protein. *Nucleic Acids Res* 37, 762–770 (2009).

40. Vovis, G. F., Horiuchi, K. & Zinder, N. D. Kinetics of Methylation of DNA by a Restriction Endonuclease from Escherichia coli B. *Proc National Acad Sci* 71, 3810–3813 (1974).

41. Loenen, W. A. M., Dryden, D. T. F., Raleigh, E. A. & Wilson, G. G. Type I restriction enzymes and their relatives. *Nucleic Acids Res* 42, 20–44 (2013).

42. Roberts, G. A. *et al.* Impact of target site distribution for Type I restriction enzymes on the evolution of methicillin-resistant Staphylococcus aureus (MRSA) populations. *Nucleic Acids Research* 41, 7472–7484 (2013).

43. Madhusoodanan, U. K. & Rao, D. N. Diversity of DNA methyltransferases that recognize asymmetric target sequences. *Crit Rev Biochem Mol* 45, 125–145 (2010).

44. Horiuchi, K. & Zinder, N. D. Cleavage of Bacteriophage f1 DNA by the Restriction Enzyme of Escherichia coli B. *Proc National Acad Sci* 69, 3220–3224 (1972).

45. McClelland, S. E. & Szczelkun, M. D. The Type I and III Restriction Endonucleases: Structural Elements in Molecular Motors that Process DNA. *Nucleic Acids Mol Biology* 111–135 (2004) doi:10.1007/978-3-642-18851-0_5.

46. Rosamond, J., Endlich, B. & Linn, S. Electron microscopic studies of the mechanism of action of the restriction endonuclease of Escherichia coli B. *J Mol Biol* 129, 619–635 (1979).

47. Yuan, R., Hamilton, D. L. & Burckhardt, J. DNA translocation by the restriction enzyme from E. coli K. *Cell* 20, 237–244 (1980).

48. Studier, F. W. & Bandyopadhyay, P. K. Model for how type I restriction enzymes select cleavage sites in DNA. *Proc National Acad Sci* 85, 4677–4681 (1988).

49. Jindrova, E., Schmid-Nuoffer, S., Hamburger, F., Janscak, P. & Bickle, T. A. On the DNA cleavage mechanism of Type I restriction enzymes. *Nucleic Acids Res* 33, 1760–1766 (2005).

50. Anton, B. P. *et al.* Cloning and characterization of the BglII restriction–modification system reveals a possible evolutionary footprint. *Gene* 187, 19–27 (1997).

51. Kaminska, K. H., Kawai, M., Boniecki, M., Kobayashi, I. & Bujnicki, J. M. Type II restriction endonuclease R.Hpy188I belongs to the GIY-YIG nuclease superfamily, but exhibits an unusual active site. *Bmc Struct Biol* 8, 48 (2008).

52. Saravanan, M., Bujnicki, J. M., Cymerman, I. A., Rao, D. N. & Nagaraja, V. Type II restriction endonuclease R.KpnI is a member of the HNH nuclease superfamily. *Nucleic Acids Res* 32, 6129–6135 (2004).

53. Pingoud, A., Fuxreiter, M., Pingoud, V. & Wende, W. Type II restriction endonucleases: structure and mechanism. *Cell Mol Life Sci* 62, 685 (2005).

54. Kostiuk, G., Sasnauskas, G., Tamulaitiene, G. & Siksnys, V. Degenerate sequence recognition by the monomeric restriction enzyme: single mutation converts BcnI into a strand-specific nicking endonuclease. *Nucleic Acids Res* 39, 3744–3753 (2011).

55. Kr€ger, D. H., Barcak, G. J., Reuter, M. & Smith, H. O. Eco RII can be activated to cleave refractory DNA recognition sites. *Nucleic Acids Res* 16, 3997–4008 (1988).

56. Zaremba, M. *et al.* DNA synapsis through transient tetramerization triggers cleavage by Ecl18kI restriction enzyme. *Nucleic Acids Res* 38, 7142–7154 (2010).

57. Korona, R. & Levin, B. R. Phage-mediated selection and the evolution and maintenance of restriction-modification. *Evolution* (1993).

58. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nature Reviews Microbiology* 8, 317–327 (2010).

59. Day, R. S. UV-induced alleviation of K-specific restriction of bacteriophage lambda. *J Virol* 21, 1249–51 (1977).

60. Efimova, E. P., Delver, E. P. & Belogurov, A. A. 2-Aminopurine and 5-bromouracil induce alleviation of type I restriction in Escherichia coli: Mismatches function as inducing signals? *Mol Gen Genetics Mgg* 214, 317–320 (1988).

61. Thoms, B. & Wackernagel, W. Genetic control of damage-inducible restriction alleviation in Escherichia coli K12: an SOS function not repressed by lexA. *Mol Gen Genetics Mgg* 197, 297–303 (1984).

62. Blakely, G. W. & Murray, N. E. Control of the endonuclease activity of type I restriction-modification systems is required to maintain chromosome integrity following homologous recombination. *Mol Microbiol* 60, 883–893 (2006).

63. Prakash-Cheng, A. & Ryu, J. Delayed expression of in vivo restriction activity following conjugal transfer of Escherichia coli hsdK (restriction-modification) genes. *J Bacteriol* 175, 4905–4906 (1993).

64. Prakash-Cheng, A., Chung, S. S. & Ryu, J. The expression and regulation of hsdK genes after conjugative transfer. *Mol Gen Genetics Mgg* 241, 491–6 (1993).

65. Kulik, E. M. & Bickle, T. A. Regulation of the Activity of the type ICEcoR124I Restriction Enzyme. *J Mol Biol* 264, 891–906 (1996).

66. Nagornykh, M. O. *et al.* Regulation of gene expression in a type II restriction-modification system. *Russian Journal of Genetics* 44, 523–532 (2008).

67. Mruk, I. & Kobayashi, I. To be or not to be: regulation of restriction–modification systems and other toxin–antitoxin systems. *Nucleic Acids Res* 42, 70–86 (2014).

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

69. Kirillov, A. *et al.* Cells with Stochastically Increased Methyltransferase to Restriction Endonuclease Ratio Provide an Entry for Bacteriophage into Protected Cell Population. *bioRxiv* (2022) doi:10.1101/2022.03.28.486079.

70. Pleška, M. & Guet, C. C. Effects of mutations in phage restriction sites during escape from restriction-modification. *Biology letters* 13, 20170646 (2017).

71. Rusinov, I., Ershova, A., Karyagina, A., Spirin, S. & Alexeevski, A. Lifespan of restriction-modification systems critically affects avoidance of their recognition sites in host genomes. *Bmc Genomics* 16, 1084 (2015).

72. ISHINO, S., MIZUKAMI, T., YAMAGUCHI, K., KATSUMATA, R. & ARAKI, K. Cloning and Sequencing of the meso-Diaminopimelate-D-dehydrogenase (ddh) Gene of Corynebacterium glutamicum. *Agr Biol Chem Tokyo* 52, 2903–2909 (1988).

73. Mojica, F. J., Ferrer, C., Juez, G. & Rodríguez-Valera, F. Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning. *Molecular Microbiology* 17, 85–93 (1995).

74. Jansen, R., Embden, J. D. A. van, Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology* 43, 1565–1575 (2002).

75. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution* 60, 174–182 (2005).

76. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct* 1, 7 (2006).

77. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* 315, 1709–1712 (2007).

78. Marraffini, L. A. & Sontheimer, E. J. CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. *Science* 322, 1843–1845 (2008).

79. Brouns, S. J. J. *et al.* Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science* 321, 960–964 (2008).

80. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468, 67–71 (2010).

81. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc National Acad Sci* 109, E2579–E2586 (2012).

82. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* 339, 1231143–823 (2013).

83. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* 339, 823–826 (2013).

84. Makarova, K. S. *et al.* Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol* 18, 67–83 (2020).

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

86. Pourcel, C., Salvignol, G. & Vergnaud, G. 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 (2005).

87. Marraffini, L. A. CRISPR-Cas immunity in prokaryotes. *Nature* 526, 55–61 (2015).

88. Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *Journal of Bacteriology* 190, 1390–1400 (2008).

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

90. Datsenko, K. A. *et al.* Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature Communications* 3, 945 (2012).

91. Heler, R. *et al.* Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* 519, 199–202 (2015).

92. Wei, Y., Terns, R. M. & Terns, M. P. Cas9 function and host genome sampling in Type II-A CRISPR-Cas adaptation. *Genes & Development* 29, 356–361 (2015).

93. Nuñez, J. K. *et al.* Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nat Struct Mol Biol* 21, 528–534 (2014).

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

95. Hale, C., Kleppe, K., Terns, R. M. & Terns, M. P. Prokaryotic silencing (psi)RNAs in Pyrococcus furiosus. *Rna* 14, 2572–2579 (2008).

96. Nussenzweig, P. M. & Marraffini, L. A. Molecular Mechanisms of CRISPR-Cas Immunity in Bacteria. *Annual review of genetics* 54, 93–120 (2020).

97. Sinkunas, T. *et al.* In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. *Embo J* 32, 385–394 (2013).

98. Sinkunas, T. *et al.* Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system: Cas3 nuclease/helicase. *Embo J* 30, 1335–1342 (2011).

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

100. Elmore, J. R. *et al.* Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR–Cas system. *Genes & Development* 30, 447–459 (2016).

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

102. Samai, P. *et al.* Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell* 161, 1164–1174 (2015).

103. Niewoehner, O. *et al.* Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. *Nature* 548, 543–548 (2017).

104. Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G. & Siksnys, V. A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science* 357, 605–609 (2017).

105. Rostøl, J. T. *et al.* The Card1 nuclease provides defence during type III CRISPR immunity. *Nature* 590, 624–629 (2021).

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

107. McMahon, S. A. *et al.* Structure and mechanism of a Type III CRISPR defence DNA nuclease activated by cyclic oligoadenylate. *Nat Commun* 11, 500 (2020).

108. Pinilla-Redondo, R. *et al.* Type IV CRISPR–Cas systems are highly diverse and involved in competition between plasmids. *Nucleic Acids Res* 48, 2000–2012 (2020).

109. Crowley, V. M. *et al.* A Type IV-A CRISPR-Cas System in Pseudomonas aeruginosa Mediates RNA-Guided Plasmid Interference In Vivo. *Crispr J* 2, 434–440 (2019).

110. Sapranauskas, R. *et al.* The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Res* 39, 9275–9282 (2011).

111. Zetsche, B. *et al.* Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* 163, 1–14 (2015).

112. Abudayyeh, O. O. *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* 353, aaf5573-17 (2016).

113. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471, 602–607 (2011).

114. Chylinski, K., Rhun, A. L. & Charpentier, E. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *Rna Biol* 10, 726–737 (2013).

115. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology* 31, 233–239 (2013).

116. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507, 62–67 (2014).

117. Meeske, A. J. & Marraffini, L. A. RNA Guide Complementarity Prevents Self-Targeting in Type VI CRISPR Systems. *Mol Cell* 71, 791-801.e3 (2018).

118. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nature Publishing Group* 13, 722–736 (2015).

119. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat Commun* 9, 1911 (2018).

120. Liao, C. & Beisel, C. L. The tracrRNA in CRISPR Biology and Technologies. *Annu Rev Genet* 55, 1–21 (2021).

121. Nishimasu, H. *et al.* Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* 156, 935–949 (2014).

122. Jinek, M. *et al.* Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science* 343, 1247997 (2014).

123. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. A Cas9–guide RNA complex preorganized for target DNA recognition. *Science* 348, 1477–1481 (2015).

124. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 513, 569–573 (2014).

125. Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature* 527, 110–113 (2015).

126. Houte, S. van *et al.* The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* 532, 385–388 (2016).

127. Pyenson, N. C. & Marraffini, L. A. Co-evolution within structured bacterial communities results in multiple expansion of CRISPR loci and enhanced immunity. *Elife* 9, e53078 (2020).

128. McGinn, J. & Marraffini, L. A. Molecular mechanisms of CRISPR-Cas spacer acquisition. *Nature Publishing Group* 3, 711–12 (2018).

129. Kim, J. G., Garrett, S., Wei, Y., Graveley, B. R. & Terns, M. P. CRISPR DNA elements controlling site-specific spacer integration and proper repeat length by a Type II CRISPR–Cas system. *Nucleic Acids Res* 47, 8632–8648 (2019).

130. Wright, A. V. & Doudna, J. A. Protecting genome integrity during CRISPR immune adaptation. *Nature structural & molecular biology* 23, 876–883 (2016).

131. Xiao, Y., Ng, S., Nam, K. H. & Ke, A. How type II CRISPR-Cas establish immunity through Cas1-Cas2-mediated spacer integration. *Nature* 550, 137–141 (2017).

132. Wang, J. *et al.* Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell* 163, 840–853 (2015).

133. Nuñez, J. K., Harrington, L. B., Kranzusch, P. J., Engelman, A. N. & Doudna, J. A. Foreign DNA capture during CRISPR–Cas adaptive immunity. *Nature* 527, 535–538 (2015).

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

135. Wei, Y., Chesne, M. T., Terns, R. M. & Terns, M. P. Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in Streptococcus thermophilus. *Nucleic Acids Research* 43, 1749–1758 (2015).

136. Wright, A. V. *et al.* Structures of the CRISPR genome integration complex. *Science* 357, 1113–1118 (2017).

137. Nuñez, J. K., Bai, L., Harrington, L. B., Hinder, T. L. & Doudna, J. A. CRISPR Immunological Memory Requires a Host Factor for Specificity. *Mol Cell* 62, 824–833 (2016).

138. Budhathoki, J. B. *et al.* Real-time Observation of CRISPR spacer acquisition by Cas1–Cas2 integrase. *Nat Struct Mol Biol* 27, 489–499 (2020).

139. Elmore, J. R. *et al.* Programmable plasmid interference by the CRISPR-Cas system in Thermococcus kodakarensis. *Rna Biol* 10, 828–840 (2013).

140. Nickel, L. *et al.* Two CRISPR-Cas systems in Methanosarcina mazei strain Gö1 display common processing features despite belonging to different types I and III. *Rna Biol* 10, 779–791 (2013).

141. Richter, H. *et al.* Characterization of CRISPR RNA processing in Clostridium thermocellum and Methanococcus maripaludis. *Nucleic Acids Res* 40, 9887–9896 (2012).

142. Wilkinson, M. *et al.* Structure of the DNA-Bound Spacer Capture Complex of a Type II CRISPR-Cas System. *Mol Cell* 75, 90-101.e5 (2019).

143. Ka, D., Jang, D. M., Han, B. W. & Bae, E. Molecular organization of the type II-A CRISPR adaptation module and its interaction with Cas9 via Csn2. *Nucleic Acids Research* 46, 9805–9815 (2018).

144. Arslan, Z. *et al.* Double-strand DNA end-binding and sliding of the toroidal CRISPR-associated protein Csn2. *Nucleic Acids Research* 41, 6347–6359 (2013).

145. Koo, Y., Jung, D. & Bae, E. Crystal structure of Streptococcus pyogenes Csn2 reveals calcium-dependent conformational changes in its tertiary and quaternary structure. *PLoS ONE* 7, e33401 (2012).

146. Ellinger, P. *et al.* The crystal structure of the CRISPR-associated protein Csn2 from Streptococcus agalactiae. *Journal of Structural Biology* 178, 350–362 (2012).

147. Nam, K. H., Kurinov, I. & Ke, A. Crystal structure of clustered regularly interspaced short palindromic repeats (CRISPR)-associated Csn2 protein revealed Ca2+-dependent double-stranded DNA binding activity. *Journal of Biological Chemistry* 286, 30759–30768 (2011).

148. Jakhanwal, S. *et al.* A CRISPR-Cas9-integrase complex generates precise DNA fragments for genome integration. *Nucleic Acids Research* (2021) doi:10.1093/nar/gkab123.

149. Ramachandran, A., Summerville, L., Learn, B. A., DeBell, L. & Bailey, S. Processing and integration of functionally oriented prespacers in the Escherichia coli CRISPR system depends on bacterial host exonucleases. *J Biol Chem* 295, 3403–3414 (2020).

150. Kim, S. *et al.* Selective loading and processing of prespacers for precise CRISPR adaptation. *Nature* 579, 141–145 (2020).

151. Heler, R. *et al.* Mutations in Cas9 Enhance the Rate of Acquisition of Viral Spacer Sequences during the CRISPR-Cas Immune Response. *Molecular Cell* 65, 168–175 (2017).

152. Amitai, G. & Sorek, R. CRISPR–Cas adaptation: insights into the mechanism of action. *Nature Reviews Microbiology* 14, 1–10 (2016).

153. Aviram, N., Thornal, A. N., Zeevi, D. & Marraffini, L. A. Different modes of spacer acquisition by the Staphylococcus epidermidis type III-A CRISPR-Cas system. *Nucleic Acids Res* gkab1299- (2022) doi:10.1093/nar/gkab1299.

154. Modell, J. W., Jiang, W. & Marraffini, L. A. CRISPR-Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. *Nature* 544, 101–104 (2017).

155. Levy, A. *et al.* CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* 520, 505–510 (2015).

156. Wigley, D. B. Bacterial DNA repair: recent insights into the mechanism of RecBCD, AddAB and AdnAB. *Nature Publishing Group* 11, 9–13 (2013).

157. Akroyd, J. E., Clayson, E. & Higgins, N. P. Purification of the gam gene-product of bacteriophage Mu and determination of the nucleotide sequence of the gam gene. *Nucleic Acids Research* 14, 6901–6914 (1986).

158. Friedman, S. A. & Hays, J. B. Selective inhibition of Escherichia coli RecBC activities by plasmid-encoded GamS function of phage lambda. *Gene* 43, 255–263 (1986).

159. Murphy, K. C. & Lewis, L. J. Properties of Escherichia coli expressing bacteriophage P22 Abc (anti-RecBCD) proteins, including inhibition of Chi activity. *J Bacteriol* 175, 1756–1766 (1993).

160. Richter, C. *et al.* Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res* 42, 8516–8526 (2014).

161. Nussenzweig, P. M., McGinn, J. & Marraffini, L. A. Cas9 Cleavage of Viral Genomes Primes the Acquisition of New Immunological Memories. *Cell Host and Microbe* 26, 515-526.e6 (2019).

162. Nicholson, T. J. *et al.* Bioinformatic evidence of widespread priming in type I and II CRISPR-Cas systems. *Rna Biol* 16, 566–576 (2018).

163. Hale, C. R. *et al.* RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell* 139, 945–956 (2009).

164. Jore, M. M. *et al.* Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* 18, 529–536 (2011).

165. Dupuis, M.-È., Villion, M., Magadán, A. H. & Moineau, S. CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance. *Nature Communications* 4, 2087 (2013).

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

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

168. Nair, D. *et al.* 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 (2011).

169. Berscheid, A., Sass, P., Weber-Lassalle, K., Cheung, A. L. & Bierbaum, G. Revisiting the genomes of the Staphylococcus aureus strains NCTC 8325 and RN4220. *International Journal of Medical Microbiology* 302, 84–87 (2012).

170. Heler, R., Wright, A. V., Vucelja, M., Doudna, J. A. & Marraffini, L. A. Spacer Acquisition Rates Determine the Immunological Diversity of the Type II CRISPR-Cas Immune Response. *Cell Host and Microbe* 25, 242-249.e3 (2019).

171. Waldron, D. E. & Lindsay, J. A. Sau1: a Novel Lineage-Specific Type I Restriction-Modification System That Blocks Horizontal Gene Transfer into Staphylococcus aureus and between S. aureus Isolates of Different Lineages. *Journal of Bacteriology* 188, 5578–5585 (2006).

172. Perez-Casal, J., Caparon, M. G. & Scott, J. R. Mry, a trans-acting positive regulator of the M protein gene of Streptococcus pyogenes with similarity to the receptor proteins of two-component regulatory systems. *Journal of Bacteriology* 173, 2617–2624 (1991).

173. Bickle, T. A. & Krüger, D. H. Biology of DNA restriction. *Microbiol Rev* 57, 434–50 (1993).

174. Workman, R. E. *et al.* A natural single-guide RNA repurposes Cas9 to autoregulate CRISPR-Cas expression. *Cell* 395, 270-688.e19 (2021).

175. Neamah, M. M. *et al.* Sak and Sak4 recombinases are required for bacteriophage replication in Staphylococcus aureus. *Nucleic Acids Research* 45, 6507–6519 (2017).

176. Levin, B. R. Frequency-dependent selection in bacterial populations. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 319, 459–472 (1988).

177. Hershey, A. D., Burgi, E. & Ingraham, L. COHESION OF DNA MOLECULES ISOLATED FROM PHAGE LAMBDA. *Proc National Acad Sci* 49, 748–755 (1963).

178. Catalano, C. E., Cue, D. & Feiss, M. Virus DNA packaging: the strategy used by phage λ. *Mol Microbiol* 16, 1075–1086 (1995).

179. Duncan, C. H., Wilson, G. A. & Young, F. E. Biochemical and genetic properties of site-specific restriction endonucleases in Bacillus globigii. *Journal of Bacteriology* 134, 338–344 (1978).

180. Halpern, D. *et al.* Identification of DNA motifs implicated in maintenance of bacterial core genomes by predictive modeling. *PLoS Genetics* 3, 1614–1621 (2007).

181. Casjens, S. R. & Gilcrease, E. B. Determining DNA Packaging Strategy by Analysis of the Termini of the Chromosomes in Tailed-Bacteriophage Virions. in vol. 502 91–111 (Humana Press, 2009).

182. Frommer, M. *et al.* A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc National Acad Sci* 89, 1827–1831 (1992).

183. Simmon, V. F. & Lederberg, S. Degradation of bacteriophage lambda deoxyribonucleic acid after restriction by Escherichia coli K-12. *Journal of Bacteriology* 112, 161–169 (1972).

184. Pyenson, N. C., Gayvert, K., Varble, A., Elemento, O. & Marraffini, L. A. Broad Targeting Specificity during Bacterial Type III CRISPR-Cas Immunity Constrains Viral Escape. *Cell Host and Microbe* 22, 343-353.e3 (2017).

185. Palm, N. W. & Medzhitov, R. Pattern recognition receptors and control of adaptive immunity. *Immunol Rev* 227, 221–233 (2009).

186. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* 337, 816–821 (2015).

187. Makarova, K. S. *et al.* Evolution and classification of the CRISPR–Cas systems. *Nature Reviews Microbiology* 9, 467–477 (2011).

188. Colleaux, L., D'Auriol, L., Galibert, F. & Dujon, B. Recognition and cleavage site of the intron-encoded omega transposase. *Proceedings of the National Academy of Sciences* 85, 6022–6026 (1988).

189. Pleška, M. *et al.* Bacterial Autoimmunity Due to a Restriction- Modification System. *Current Biology* 26, 404–409 (2016).

190. Birkholz, N., Jackson, S. A., Fagerlund, R. D. & Fineran, P. C. A mobile restriction-modification system provides phage defence and resolves an epigenetic conflict with an antagonistic endonuclease. *Nucleic Acids Res* (2022) doi:10.1093/nar/gkac147.

191. Swarts, D. C. *et al.* DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* 507, 258–261 (2014).

192. Kuzmenko, A. *et al.* DNA targeting and interference by a bacterial Argonaute nuclease. *Nature* 587, 632–637 (2020).

193. Makarova, K. S., Wolf, Y. I., Snir, S. & Koonin, E. V. Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *Journal of Bacteriology* 193, 6039–6056 (2011).

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

195. Goldfarb, T. *et al.* BREX is a novel phage resistance system widespread in microbial genomes. *The EMBO journal* 34, 169–183 (2015).

196. Ofir, G. *et al.* DISARM is a widespread bacterial defence system with broad anti-phage activities. *Nature Microbiology* 8, 317–1 (2017).

197. Doron, S. *et al.* Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* 359, (2018).

198. Bernheim, A. *et al.* Prokaryotic viperins produce diverse antiviral molecules. *Nature* 589, 120–124 (2021).

199. Cohen, D. *et al.* Cyclic GMP–AMP signalling protects bacteria against viral infection. *Nature* 574, 691–695 (2019).

200. Horinouchi, S. & Weisblum, B. Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. *Journal of Bacteriology* 150, 815–825 (1982).

201. Horinouchi, S. & Weisblum, B. Nucleotide sequence and functional map of pE194, a plasmid that specifies inducible resistance to macrolide, lincosamide, and streptogramin type B antibodies. *Journal of Bacteriology* 150, 804–814 (1982).

202. Goldberg, G. W. *et al.* Incomplete prophage tolerance by type III-A CRISPR-Cas systems reduces the fitness of lysogenic hosts. *Nature Communications* 9, 61 (2018).

203. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357–359 (2012).

204. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25 (2009).

205. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* 44, W3–W10 (2016).

206. Kreiswirth, B. N. *et al.* The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage. *Nature* 305, 709–712 (1983).

207. Bikard, D. *et al.* Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. *Nature Biotechnology* 32, 1146–1150 (2014).

208. Jiang, W. *et al.* Dealing with the evolutionary downside of CRISPR immunity: bacteria and beneficial plasmids. *PLoS Genetics* 9, e1003844 (2013).

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

210. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* 6, 343–345 (2009).

211. Ofir, G. & Sorek, R. Contemporary Phage Biology: From Classic Models to New Insights. *Cell* 172, 1260–1270 (2018).

212. Jiang, F. & Doudna, J. A. CRISPR–Cas9 Structures and Mechanisms. *Annu Rev Biophys* 46, 505–529 (2017).