2020

Cas9-Primed Adaptive Immunity During the CRISPR-Cas Response

Philip M. Nussenzweig

CAS9-PRIMED ADAPTIVE IMMUNITY DURING THE CRISPR-CAS RESPONSE

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Philip M. Nussenzweig

June 2020

CAS9-PRIMED ADAPTIVE IMMUNITY DURING THE CRISPR-CAS RESPONSE

Philip M. Nussenzweig, Ph.D.
The Rockefeller University 2020

Prokaryotes have developed numerous defense strategies to combat the constant threat of viruses (bacteriophages) that endanger them. Clustered, regularly interspaced short palindromic repeats (CRISPR) loci provide archaea and bacteria with adaptive immune systems that allow them to counteract these rapidly evolving genetic parasites. These diverse systems all generally contain two components: a set of CRISPR-associated (*cas*) genes and a series of repetitive DNA elements intercalated with variable sequences known as spacers. Following viral infection, these sequences are acquired from the viral genome and integrated in the CRISPR array as new spacers. Spacers are then transcribed into CRISPR RNAs (crRNAs) that direct the Cas nucleases to destroy the invader following sequence-specific recognition of either DNA or RNA. Thus, spacers function as a form of immunologic memory that can be called upon again and again to defend the cell from reinfection.

In type II CRISPR-Cas systems, spacer sequences direct the Cas9 nuclease to target infecting bacteriophages and cleave their double-stranded (ds)DNA genomes. Whether and how pre-exiting anti-viral spacers in type II systems affect memory generation and the acquisition of new spacers is unknown. Here, in my thesis work, I demonstrate that previously acquired spacers promote additional spacer capture from the vicinity of the Cas9 cut site at an enhanced rate. I go on to show that Cas9-mediated dsDNA break (DSB) formation is required for spacer-mediated spacer acquisition and that the rate of spacer acquisition is correlated with the efficiency of Cas9 cleavage. As a result of this mechanism, cells with preexisting viral immunity can utilize their spacer-derived crRNAs to direct the acquisition of additional spacers in a new phase of immune response known as primed spacer acquisition or priming.

A consequence of priming is that immune cells can acquire additional spacers as Cas9 destroys the infecting virus. I go on to show that spacers acquired during Cas9-mediated priming endow potent benefits to bacterial communities faced with virulent bacteriophages. In particular, priming suppresses the emergence of CRISPR escaper and related viruses that emerge during Cas9 targeting. I show that this anti-viral immunity is achieved in three ways. Firstly, priming expands the hosts immune repertoire, thereby improving the existing anti-phage immunity. In addition, I show that primed spacer acquisition allows the host to contain the propagation escapers that have mutations in their target sequence that abrogate Cas9 targeting. Finally, by preemptively immunizing the host with additional spacers during the initial Cas9 targeting response, priming allows the host to anticipate secondary infections by escaper and related viruses. This “prophylactic” immunity is a unique feature in CRISPR systems that allows type II systems to overcome future threats from viruses that would have overcome the defense provided by the initial anti-viral spacer.

CRISPR-Cas immune systems allow their host to rapidly adapt to the viruses that challenge them. Collectively, my thesis work has revealed a new phases of the type II-A CRISPR-Cas9 immune response that is fundamental to how these systems defend their hosts against bacteriophages.

For my mother, father, and all my other teachers

ACKNOWLEDGMENTS

I am profoundly grateful for the many opportunities and resources that I have been privileged enough to receive throughout my education and to the many people that facilitated them. The experience of carrying out my graduate education at The Rockefeller University has in particular been an enormous privilege. The science performed at The Rockefeller and the scientists preforming it have been a continual source of inspiration that has unquestionably elevated my own work. I would like to thank the Dean's Office for their assistance as well as all the staff in the cleaning services, food services, mailroom, maintenance department, and student housing for making my time at The Rockefeller so comfortable and enjoyable. In addition, I would also like to thank my friends, family, colleagues, and mentors who made this work possible.

First and foremost, I would like to thank my advisor, Dr. Luicano A. Marraffini for his mentorship and constant support. Luciano took on the risk of allowing me to join his laboratory as his first MD-PhD student in the summer of 2014 and for that I will be forever grateful to him. The subsequent nearly 6 years have been some years of the best of my life and Luciano, as well as the environment he fostered in the lab, are directly responsible for that. Luciano is a remarkable scientist and teacher whose passion, creativity, and ability to think critically about experiments has been hugely inspirational to me. From Luciano I have learned how to think about and execute experiments as a scientist and I hope to carry those lessons with me throughout my career. In addition, Luciano is an exemplary mentor and person whose qualities I aspire to. His compassion, level-headedness, patience, and steadfast belief in his trainees has lifted me during the times I needed it the most.

I am also indebted to the members of my thesis committee, Drs. Carl F. Nathan and Charles M. Rice, and the chair of my committee, Dr. Sohail Tavazoie for their guidance, suggestions, and support throughout the years. It has been an honor to have such exceptional scientists oversee my development and Ph.D. thesis.

I would also like to thank my external examiner, Dr. Aaron Whiteley, for graciously agreeing to serve in this role. Dr. Whiteley has been enormously generous with his time during the current COVID-19 pandemic and I very grateful that he was able to participate in my defense and exam via Zoom. Dr. Whiteley's work on bacterial immune system that respond to nucleotide-based signals has been thrilling to follow during my graduate career.

The Weill Cornell Medicine/Memorial Sloan-Kettering/Rockefeller University Tri-Institutional MD-PhD program, of which I am student in, has also facilitated my development and allowed me to carry out my research at Rockefeller University. In particular I would like to thank Dr. Olaf Anderson for his leadership of the MD-PhD program as well as his personal mentorship and support during the most difficult moments of medical school. Renee Horton and the other administrators at the "Tri-I" have also been very helpful and made my time at Weill Cornell Medicine more congenial.

The studies presented here would not be possible without tremendous assistance from Dr. Connie Zhao and the members of the Rockefeller University Genomics Resource Center. Specifically, Bin Zhang, Dr. (Helen) Hong Duan, and Christine Lai have all provided invaluable expertise and instruction throughout my training.

The Marraffini Lab is an extraordinary place to be a graduate student and scientist, and I am very thankful to have worked with and enjoyed the company of so many great colleagues and friends. Collectively the group has created a wonderful, intellectually stimulating, and above all else fun environment that permeates all our lab meetings, celebrations, and extracurricular actives. My life has forever been enriched by my time working beside you all.

I would like to thank Dr. David Bikard who initially served as my rotation mentor and who taught me many of the fundamental techniques that I have used throughout my Ph.D. During my time working alongside him, David was a fantastic, encouraging mentor and friend who introduced me to CRISPR biology and whose love of science helped convince me to join the lab as a graduate student. I learned a tremendous amount from Dr. Joshua Modell whose scientific advice and ever-present humor were invaluable to the work presented here. I also have had the enormous fortune of calling Dr. Andrew Varble my "baymate" for the entirety of my time in the lab. Andrew is an exemplarily scientist who has provided me with support and guidance at every stage of my development as a scientist and I am very thankful for his friendship. The post-doctoral associates and fellows, Drs. Poulami Samai, Alexander Meeske, Charlie Mo, and Naama Aviram, have all given me advice and served as role models. In particular I would like to thank Poulami for her baking, Charlie for watching soccer with me, Naama for lifting me with her optimism even in the face of my pessimism and Alex for his inspirational passion for microbiology and occasionally allowing me to spend time with his dog.

To my fellow students in the lab, Dr. Wenyan Jiang, Dr. Gregory Goldberg, Dr. Robert Heler, Dr. Jon McGinn, Dr. Nora Pyenson, Jakob Rostøl, Pascal Maguin, Amer Hossain, Claire Kenney, Dalton Banh and Amanda Shilton, together you have served as the life blood of the lab. You have all contributed immeasurably to my happiness and bettered me countlessly both personally and professionally. Dr. Wenyan Jiang greatly elevated the lab with his calming presence, encouraging advice, useful reagents, and inspirational dedication. I would like to thank Dr. Gregory Goldberg for his always thought-provoking discussions. Dr. Robert Heler was an always extremely encouraging and helpful colleagues and I would like to thank him for teaching me how to study spacer acquisition. I would like to thank Dr. Jon McGinn for the many productive scientific discussions I had with him about CRISPR-Cas biology. I sat across from Dr. Nora Pyenson for 5 years and during that time she enriched my life with her art and joy for life and repeatedly challenged me to leave my comfort zone for the better. I would like thank Jakob Rostøl for highly insightful scientific discussions, his experimental suggestions as well as making me smile or laugh almost every day we were in lab together. Pascal Maguin has served as a sympathetic compatriot as we both worked (and often struggled) to study type II-A spacer acquisition and I am thankful for his generosity and thoughtfulness. I would like to thank Amer Hossain for his patience and accompanying me on needed trips to the gym, musical performances, and vending machines. I never thought there would be other MD-PhD students during my time in

the lab, but I am very grateful they were Claire and Dalton. Claire's exuberance, selflessness, and warmth has greatly improved my time in the lab especially during long nights at the bench and the computer. I have also greatly enjoyed spending time with Dalton as well as our conversations about microbiology, food, and of course housing. Finally, I would like to thank our most recent student Amanda Shilton whose kindness has brightened the lab.

In the lab, I would also like to thank our technicians, Phone Ko, Albina Kozlova, Jacob Mathai, Ashely Thornal, and Jessica Fyodorova, as well as our former lab manager Rahul Bhosle, who each provided me with reagents that facilitated my experiments and fed me during lab meeting.

I would additionally like to acknowledge my friend and MD-PhD classmate Margaret M. Fabiszak who has accompanied me through every stage of my graduate training. Whether it be studying for exams in medical school, kayaking during unending program repeats, or discussing life and science, Meg manages to make everything around her better. I am very grateful for her friendship.

On a more personal note, I want to express my deepest gratitude to my family. I would like to thank my partner Helen S. Tian, for her support, encouragement and love. Helen has improved my life tremendously and made me a better scientist, writer, cook, and person in the process. Helen's confidence in me strengthened and protected me on my most difficult days. I would also like to thank my sister Maya for being an unwavering pillar of comfort and being my most trusted confidant. Finally, I would like to acknowledge my parents Drs. Michel C. Nussenzweig and Svetlana Mojsov. Their unconditional love, support, and understanding has made me the person I am and I dedicate this thesis to them.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AAP……abridged anchor primer

acr……anti-CRISPR

bp……base pair(s)

CARF domain……CRISPR-associated Rossman fold domain

Cas……CRISPR-associated

Cascade…….CRISPR-associated complex for antiviral defense

CFU…….colony forming units

Cm^R……chloramphenicol-resistant

cOAs...……cyclic oligoadenylates

cA_3…….cyclic triadeylates

cA_4…….……cyclic tetra-adeylates

CRISPR……...…Clustered Regular Spaced Short Palindromic Repeats

pCRISPR……plasmid-based CRISPR system

pCRISPR-BsaI…a pCRISPR containg a single set of repealing flanking a spacer with two oppositely oriented BsaI restriction sties…….

crRNA……..CRISPR RNA

Csn……CRISPR-associated endonuclease

pre-crRNA……precursor CRISPR RNA

dCas1……catalytically dead Cas1

dCas9……catalytically dead Cas9

DNA……deoxyribonucleic acid

DSB……double-stranded DNA break

dsDNA……double-stranded DNA

ssDNA……single-stranded DNA

ssDNase……single-stranded DNAse

Erm^R……erythromycin-resistant

HEPN domain……higher eukaryotes and prokaryotes nucleotide-binding domain

IHF……integration host factor

kbp……kilobase pair(s)

LAS……leader-anchoring sequence

MOI……multiplicity of infection

NGS…… next generation sequencing

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

PAC……primed acquisition complex

PAM……protospacer adjacent motif

PCR……polymerase chain reaction

PFU……plaque forming units

Phage……bacteriophage

Pre-spacer……substrate(s) for spacer integration

Priming……primed spacer acquisition

R–M……Restriction–Modification

RNA……ribonucleic acid

RPM……Reads per million of phage

RPM_{phage}……Reads per million of phage reads

RPM_{total}……Reads per million of total reads

RT……reverse transcriptase

SNP……single nucleotide polymorphism

SR…… single repeat

TdT……terminal deoxynucleotidyl transferase

tracrRNA……..……trans-activating CRISPR RNA

E. coli……..……...*Escherichia coli*

S. epidermidis...……..……*Staphylococcus epidermidis*

S. aureus……...…..…….Staphylococcus aureus

S. pyogenes……..…….*Streptococcus pyogenes*

S. thermophilus…….……*Streptococcus thermophilus*

CHAPTER 1: INTRODUCTION

1.1. Bacteriophages: central actors in biology

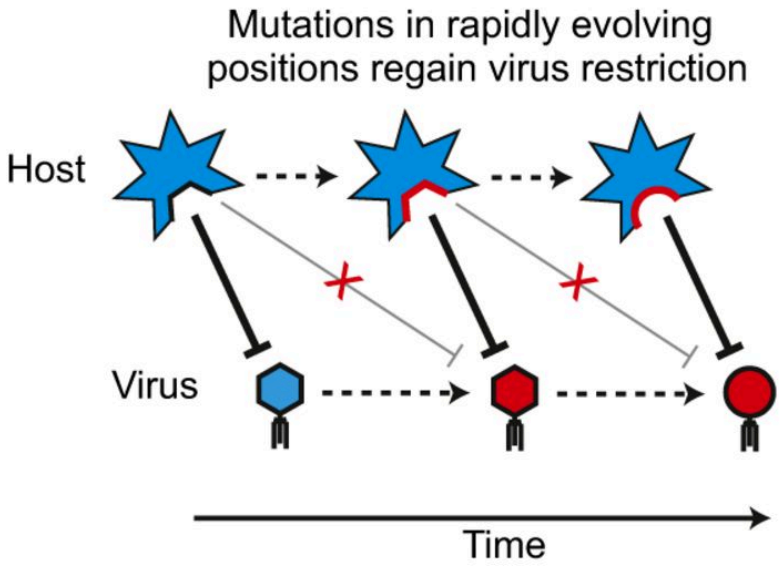


Figure 1.1. Red Queen interactions between host and viruses.
A schematic outlining the Red Queen hypothesis. Host defense systems restrict the replication of viruses and in response viral escaper variants are selected over time that evade host restriction (red hexagon and circle). As a consequence of viral evolution, mutations in the host that restrict escaper viruses are adapted and selected for (red surface on host) giving rise to an ongoing genetic arms race. Adapted from (Daugherty and Malik, 2012).

Bacteriophages (phages) are the viruses that infect and parasitize bacterial cells. Phages have been identified in every biome where bacteria are found and they often vastly outnumber their hosts (Dion et al., 2020). In some marine environments, for instance, the viral particle-to-bacterial cell ratio has been measured to be as high as 100:1 (Brum et al., 2013). As a result, bacteriophages subject their bacterial hosts to continuous selective pressure and serve as major drivers of microbial evolution (Fernandez et al., 2018). To counteract this constant threat, bacteria have developed a diverse set of defense systems that restrict viral propagation and protect bacterial populations from extinction (Rostol and Marraffini, 2019a). Bacteriophages are however highly mutable and continually acquire means of escaping host immunity (Kupczok et al., 2018). Bacteria and their phages therefore engage in a classic evolutionary paradigm known as the "Red Queen hypothesis" (Daugherty and Malik, 2012). This hypothesis, whose name derives from the Red Queen in Lewis Carroll's *Through the Looking Glass*, states an organism must continually evolve in order to maintain its fitness when confronted with a predator. This struggle for fitness forces both the bacterium and the phage into a perpetual cycle of adaptation and counter-adaptation in order to survive (**Figure 1.1)**. The consequence of this ongoing genetic arms race can be felt across biology. In the oceans, aquatic phage infections turnover 20% of the microbial biomass and influence nutrient cycles

(Suttle, 2005, 2007). In the gut microbiome, the competition between bacteria and their phages modulates the composition of the microbial community and the metabolites they produce (Hsu et al., 2019). Finally, bacteriophages that have evaded anti-phage immunity are well documented contributors to human disease due to their role as vectors for the horizontal transmission of virulence and antibiotic resistance genes to bacterial pathogens (Brüssow et al., 2004). Phages have also been widely adopted in the laboratory as model organisms and have confirmed some of the most fundamental principles in molecular biology including the random nature of mutation(Luria and Delbruck, 1943), the identity of DNA as the genetic material (Hershey and Chase, 1952), and the basis of gene expression (Jacob and Monod, 1961). More recently, bacteriophages have been repurposed as novel therapeutic agents capable of treating bacterial infections refractory to conventional antimicrobial therapies (Bikard et al., 2014; Nobrega et al., 2015). The continued understanding of the coevolution of these viruses and their bacterial hosts therefore remains enshrined as a central focus in biology.

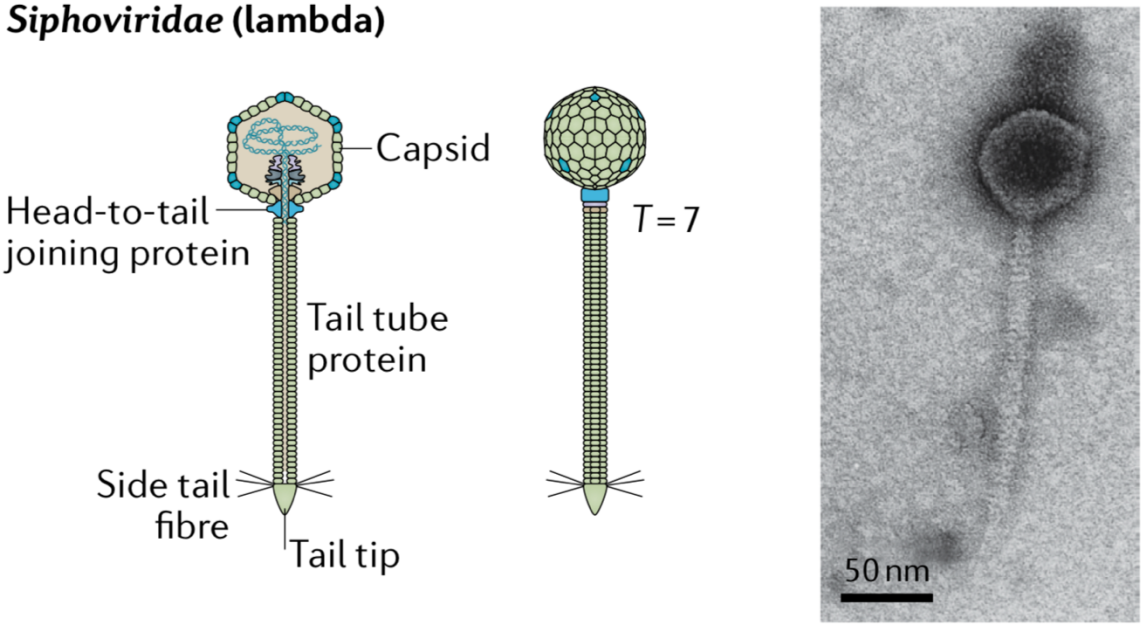


Figure 1.2. The morphology of a phage lambda.
A schematic and electromicrograph of phage lambda; a prototypic, temperate, tailed dsDNA phage of *Siphoviridae* family. Adapted from (Dion et al., 2020).

While tremendously diverse in their genomic composition and structure, all bacteriophages typically contain protein-based capsids that encases their DNA- or RNA-based genomes (**Figure 1.2)** (Dion et al., 2020). In addition to these structures, a large majority of isolated phages have a hollow tail structure (Brum et al., 2013) that allows the virus to penetrate the bacterial cell wall and membrane and then inject their genomes into the host cytosol. Following injection, the phage begins either their lytic or lysogenic lifecycle (Herskowitz and Hagen, 1980). Lytic phages immediately begin replicating multiple copies of their genomes and producing the components necessary for packaging and lysis of the host cells. Lysogenic phages by contrast integrate their genomes into the

host chromosome as “prophages” and have their lytic genes repressed. As a result, these integrated prophages can remain in the host as genetic parasites that are replicated together with the bacterial chromosome. Temperate phages, such as the classic genetic model phage lambda (**Figure 1.2**), can choose to engage in either lifecycle (Oppenheim et al., 2005) (**Figure 1.3**). This lysis-lysogen decision is controlled by a transcriptional repressor that acts as a regulatory ‘genetic switch’ that prevents the transcription of lytic genes necessary for replication (Oppenheim et al., 2005). Functional temperate prophages however retain the ability to reverse this genetic switch and initiate the lytic cycle. To do so, the “induced” prophage must excise from the bacterial chromosome and derepress its lytic genes following proteolytic cleavage of the repressor. As a consequence of Red Queen interactions, anti-phage defense systems have developed mechanisms that inhibit every stage of the lifecycle of both lytic and lysogenic phages (Rostol and Marraffini, 2019a). One of the most versatile and diverse of these defenses is the CRISPR-Cas system.

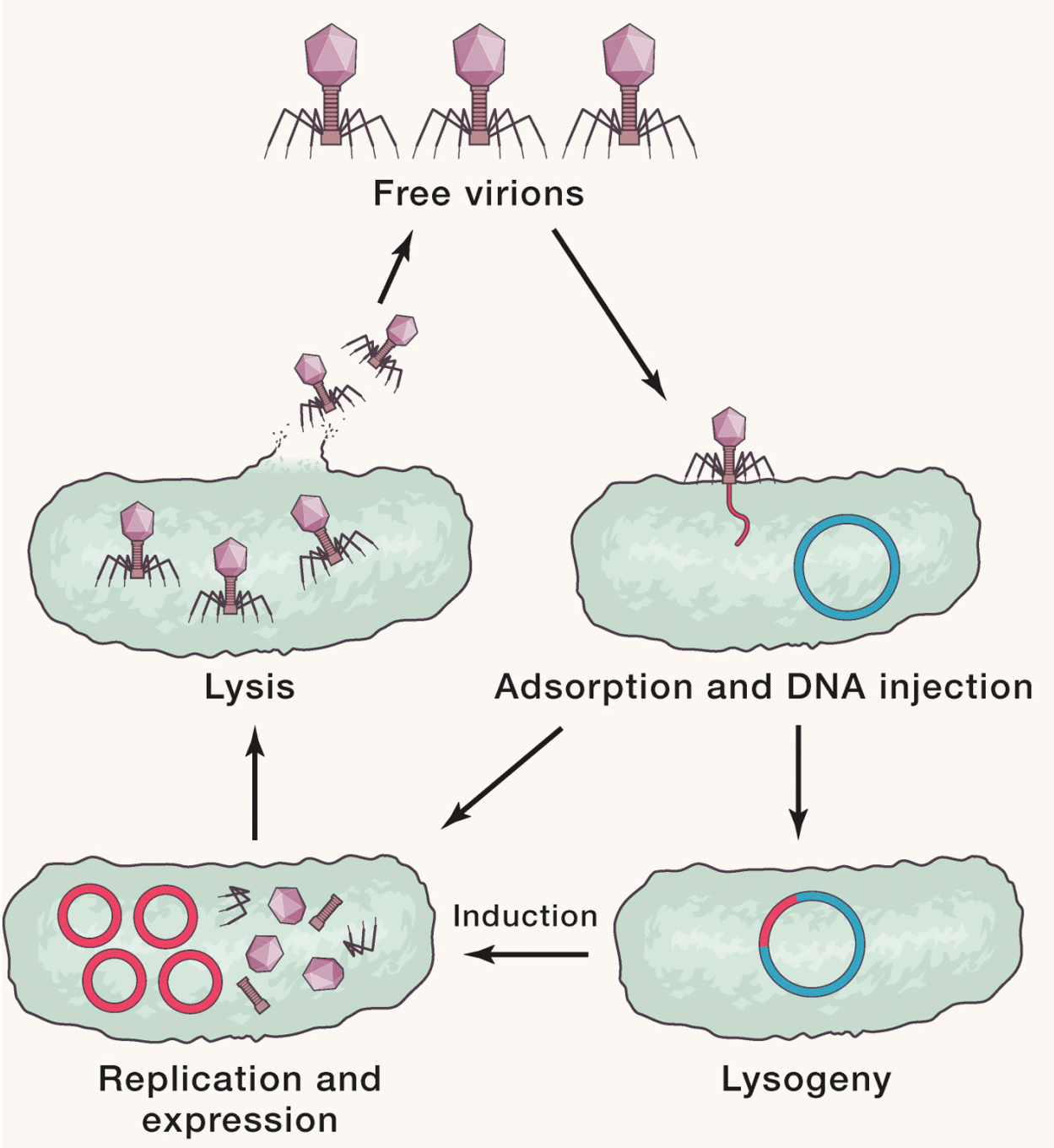


Figure 1.3. The lifecycles of temperate phages.
A schematic depicting the potential lifecycles of a temperature phage. Free virions are adsorbed onto the bacterial cell wall and then inject their DNA into the host cytosol. The phage can choose to either to initiate lysogeny (bottom right cell) and integrate into the bacterial chromosome (blue segment) as a prophage (red segment) or immediately begin replicating (red circles) and expressing genes required for packaging and lysis (bottom left cell). Prophages can however be induced to excise from the bacterial chromosome and enter the lytic cycle of replication. Adapted from (Ofir and Sorek, 2018).

1.2 CRISPR-CAS SYSTEMS: RNA-GUIDED PROKARYOTIC ADAPTIVE IMMUNE SYSTEMS

Prokaryotic organisms have evolved multiple defense systems to protect themselves against parasitic nucleic acids such as plasmids and bacteriophages (Rostol and Marraffini, 2019a). Clustered regularly interspaced short palindromic repeats (CRISPRs) are unique among these systems in that they are able to continually update and tailor their immune responses to match the specific nucleic acid sequence of their invaders. CRISPR loci are widely distributed and are present in 42% bacteria and 85% of archaea (Makarova et al., 2019), and have been reported to exist within bacteriophages (Al-Shayeb et al., 2020; Seed et al., 2013) as well as other mobile genetic elements (McDonald et al., 2019). Research over the last decade has revealed a remarkable diversity in the architecture and mechanisms employed by these systems. Currently, CRISPRs are organized into two classes, comprised of six types (I–VI) and a total of 50 subtypes based on their sequences (Makarova et al., 2019). Despite this diversity between systems, CRISPR loci all contain two basic core anatomical features (**Figure 1.4**). The first is a series of CRISPR-associated (*cas*) genes dedicated to nucleic acid manipulation (Jansen et al., 2002), and the second is an array of short partially palindromic, repetitive non-coding DNA sequences interlaced by equally short, variable sequences known as 'spacers' (Barrangou et al., 2007; Marraffini and Sontheimer, 2008). These spacers specify the sequence of the targets of Cas nucleases and lie at the center of the CRISPR-Cas immune response, which consists of three phases. In the first phase, known as spacer acquisition or immunization, new spacers are captured from the invader's nucleic acids and are integrated into the CRISPR array (**Figure 1.4A**). In the second phase, the newly acquired spacers are transcribed into small CRISPR RNAs (crRNAs) that combine with RNA-guided Cas nucleases to recognize and destroy foreign nucleic acids (**Figure 1.4B**). Class 1 systems, including types I, III, and IV, are defined by the use of multisubunit RNA-guided Cas nucleases, whereas Class II systems, including types II, V, and VI, use a single crRNA:Cas ribonucleoprotein complex. Due to the sequence-specific nature of the CRISPR-Cas response, immunity can be overcome when parasites mutate the target sequence in their genomes, thus preventing immune recognition (Andersson and Banfield, 2008; Deveau et al., 2008). To counteract these 'escapers', CRISPR systems have developed a third phase in their mechanism of immunity: primed spacer acquisition (**Figure 1.4C**). During CRISPR priming, preexisting spacers direct the targeting machinery to catalyze a second, more rapid round of spacer acquisition (Datsenko et al., 2012; Nussenzweig et al., 2019; Swarts et al., 2012). Here we summarize the molecular mechanisms used by different CRISPR-Cas systems to execute these three phases of immunity.

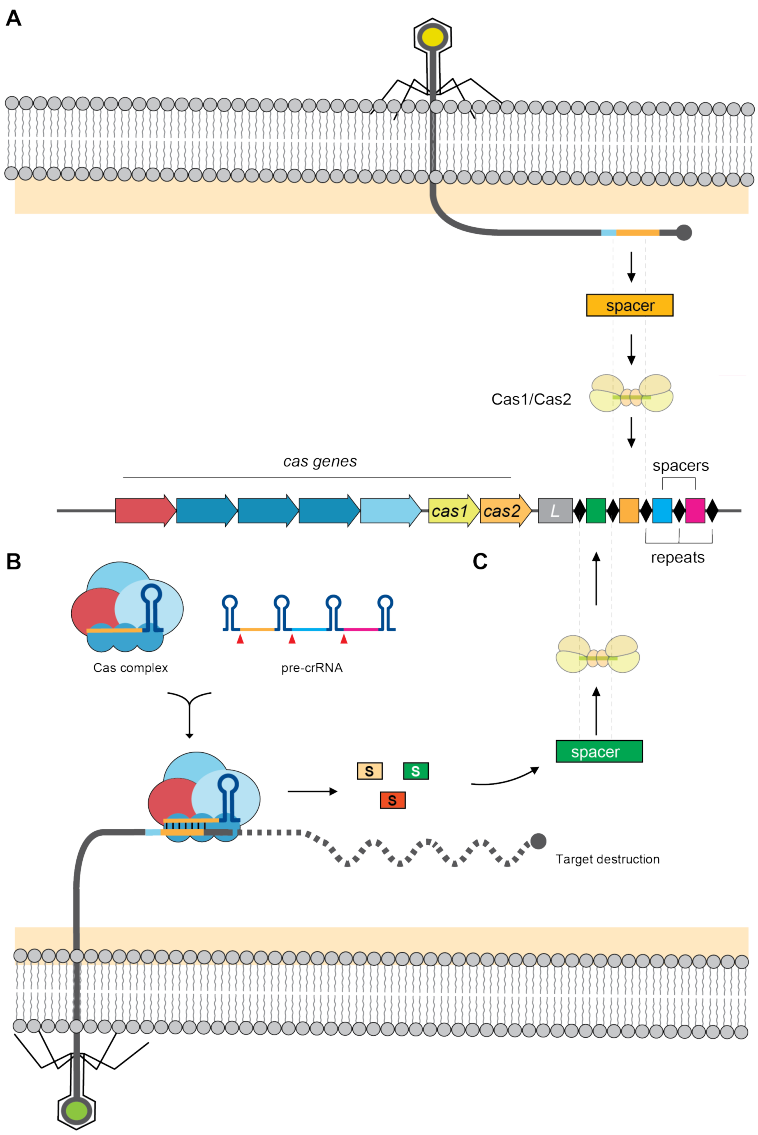


Figure 1.4. The anatomy of CRISPR locus and phases of the CRISPR-Cas immune response.

The three phases of CRISPR-Cas immunity against bacteriophages. CRISPR loci comprise of a set of CRISPR-associated (*cas*) genes, encoding the machinery for nucleic acid targeting (red and blue arrows) as well as the Cas1-Cas2 spacer acquisition machinery (yellow and orange arrows). The CRISPR array comprises of semi-palindromic repeats (black diamonds) interlaced with short 'spacer' sequences (colored rectangles). (**A**) In the first phase, naive spacer acquisition or immunization, a spacer is selected and integrated by the Cas1-Cas2 complex into the CRISPR array (orange spacer). (**B**) In the second 'targeting' phase, the newly immunized CRISPR array is transcribed into pre-crRNAs that are processed into mature CRISPR RNAs (crRNAs). The mature crRNAs combine with the Cas targeting complex and together seek out and destroy foreign nucleic acid in a sequence-specific manner. (**C**) In the primed spacer acquisition phase, new substrates for spacer acquisition are generated concurrently with target destruction and integrated into the CRISPR array (green spacer).

1.3 crRNA biogenesis: the forging of specificity

CRISPR RNAs (crRNAs) have two features: a constant palindromic sequence derived from the CRISPR array's repeats and a variable sequence derived from the array's spacers that are complementary to the targets of the immune system(Bolotin et al., 2005; Brouns et al., 2008; Haft et al., 2005; Pourcel et al., 2005). Fundamentally, all CRISPR systems must distinguish between these RNAs and other transcripts to execute immunity. To solve this, the constant region is used as an identifier to distinguish it from other cellular RNAs and specify its recruitment for RNA processing (Hale et al., 2008; Haurwitz et al., 2010). In addition, recognition of the constant region of the crRNA by the CRISPR immunosurveillance complex is also often coupled to the complex's activation (Jackson et al., 2014; Jinek et al., 2014; Liu et al., 2017b). crRNA maturation therefore potentially represents a central point of regulation before the targeting phase of immunity is initiated. Production of crRNAs usually initiates with the transcription of a long-precursor crRNA (pre-crRNA), containing several repeat spacer units (**Figure 1.4B**), typically from a promoter located in the leader region upstream of the CRISPR array. How subsequent maturation of the pre-crRNA is mechanistically coordinated is CRISPR class specific, with Class 1 systems (types I and III) employing a dedicated endoribonuclease (Carte et al., 2008; Haurwitz et al., 2010) and Class 2 systems (types II, V, and VI) executing maturation with the same machinery that performs target destruction (Deltcheva et al., 2011; East-Seletsky et al., 2016; Fonfara et al., 2016).

1.3.1 Class 1 crRNA maturation

Both type I and III systems utilize the endoribonucleases Cas6 or Cas5d to process their pre-crRNAs (Carte et al., 2008; Haurwitz et al., 2010; Nam et al., 2012). In both systems, the palindromic sequences within the constant region of the pre-crRNA allow for the formation of stem-loop structures that are recognized by Cas6 (Carte et al., 2008; Haurwitz et al., 2010). These stem-loops of type I and III cRNAs are in fact so similar that the archaeon *Sulobobus islandicus* utilizes the same Cas6 enzyme to process the crRNAs from both its type I-A and its two type III-B systems (Deng et al., 2013). The importance of the stem-loop structure is highlighted by studies on the type I-B system, whose repeats are non-palindromic and therefore cannot produce hairpin structures autonomously. Crystal structures of *Methanococcus maripaludis* Cas6b in complex with crRNA reveal that upon RNA binding, Cas6 utilizes a tyrosine aromatic ring as an RNA base-pair mimic to help form a hybrid stem-loop structure necessary for crRNA maturation (Shao et al., 2016). Generally, Cas6 recognition of the hairpin stimulates a conformational rearrangement that licenses cleavage immediately downstream of the hairpin, thus liberating the individual mature crRNA from the longer precursor (Sashital et al., 2011). For the majority of type I systems, Cas6 remains bound to the crRNA post-processing, protecting it from cellular RNases, and the Cas6:crRNA complex is incorporated together into the larger multiprotein surveillance complex (Jore et al., 2011). Type III crRNAs, however,

undergo a second trimming step that removes the hairpin and releases the crRNA from Cas6 (Hale et al., 2008). How these unprotected mature crRNAs produced by type III systems are trafficked into their respective complexes is currently unanswered (Sokolowski et al., 2014).

1.3.2 Class 2 crRNA maturation

Class 2 systems lack a dedicated crRNA maturation factor and instead rely on the same machinery that performs nucleic acid defense to handle crRNA processing (Deltcheva et al., 2011; East-Seletsky et al., 2017; Fonfara et al., 2016). The problem of correctly sorting crRNAs from the total pool of cellular nucleic acid is therefore particularly important. Class 2 systems have developed two strategies to address this challenge. The first strategy, carried out by type V-A and VI systems, is to recognize 5′-hairpins or 'pseudoknots' in pre-crRNA and cleave upstream of them, like in Class 1 Cas5/6 RNases (Dong et al., 2016; Liu et al., 2017b; Yamano et al., 2016). Alternatively, type II and V-B systems require a second *trans*-activating CRISPR (tracr) RNA to anneal to the constant region of the crRNA to act as an intermediate between the nascent crRNA and the effector (Cas9 and Cas12b, respectively) (Deltcheva et al., 2011; Shmakov et al., 2017). Stem-loop structures on the tracr side of the duplex are then used to recruit Cas9 or Cas12. Intriguingly, the original study that identified the tracrRNA in *Streptococcus pyogenes* showed that it is initially transcribed in two forms *in vivo*, as a 181- and an 89-nucleotide species, from two distinct promoters (Deltcheva et al., 2011). Both of these isoforms are capable of being processed into functional, mature tracrRNAs, but how these isoforms are regulated and if they have different effects on immunity are yet unknown. In addition, the *S. pyogenes* CRISPR-Cas9 system is distinct in that, following pre-crRNA:tracr binding, Cas9 outsources its processing to host ribonucleases; first RNase III cleaves the 5′-end of the crRNA:tracrRNA duplex, thus separating individual crRNAs from the long transcript, and then an unidentified factor executes a second 5′-trimming reaction (Deltcheva et al., 2011). This requirement is not universal in type II CRISPRs: crRNA processing and anti-plasmid immunity occur in the *Neisseria meningitidis* type II-C in the absence of RNase III *in vivo* (Zhang et al., 2013).

1.4 CRISPR targeting: executing sequence-specific immunity

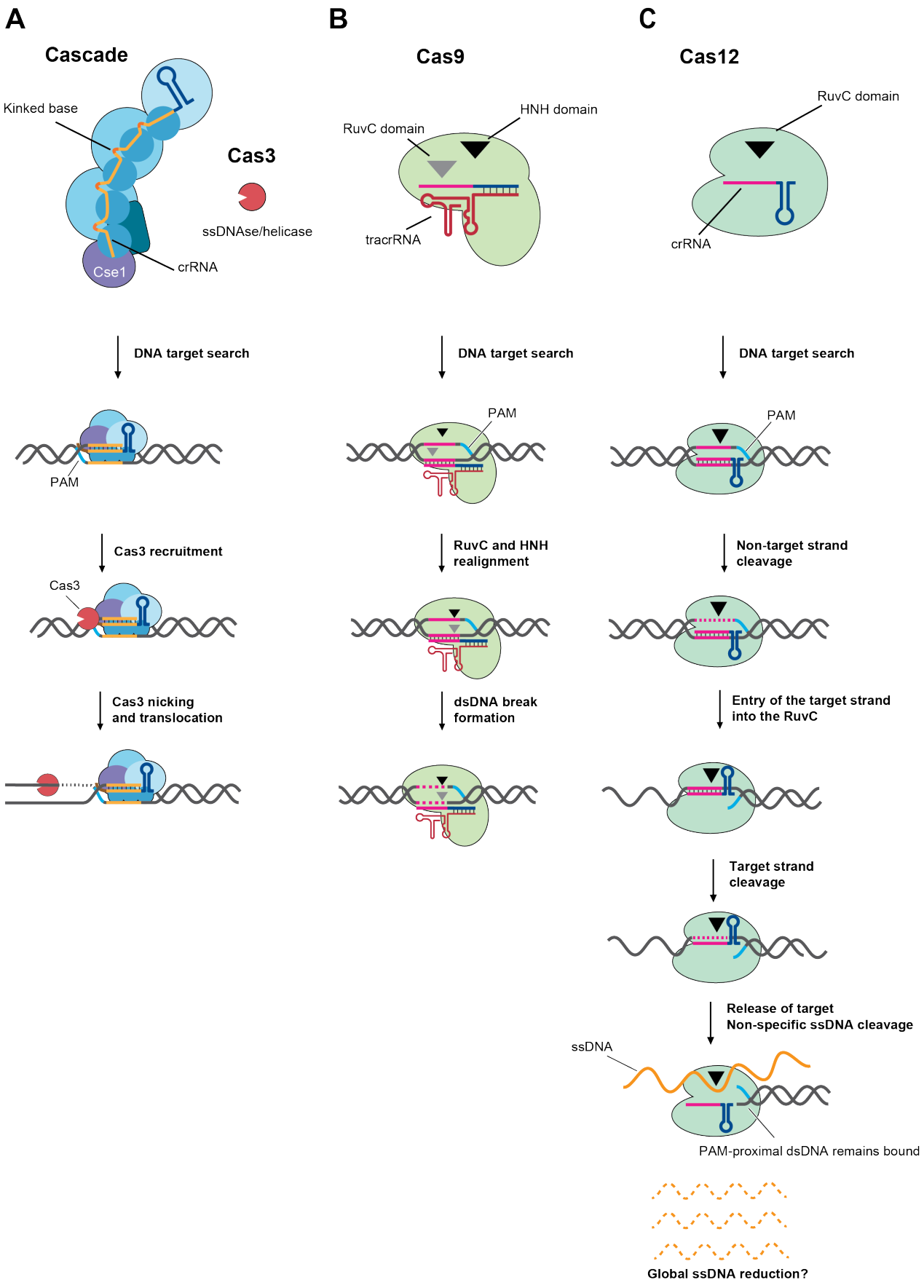
Once mature crRNAs are loaded into their cognate RNA-guided Cas nucleases, they form a ribonucleoprotein complex that surveys the host cytosol for the presence of foreign invaders (Brouns et al., 2008). The subsequent targeting phase is defined by a highly accurate and programmable sequence-specific cleavage of the target nucleic acid (Gasiunas et al., 2012; Jinek et al., 2012). This unique feature has allowed many CRISPR systems to be repurposed as genome editing tools (Doudna and Charpentier, 2014) and has generated a wealth of mechanistic insight into how Cas effectors execute nucleic acid cutting. Traditionally, CRISPR systems have been organized according to their class (ie use of a single protein or multiprotein effector complex) (Makarova et al., 2019).

While this class-based organization is highly useful for examining other facets of CRISPR biology, here we have chosen to organize CRISPR systems based on the identity of the nucleic acid that activates them: DNA- versus RNA-activated CRISPR systems.

1.4.1 DNA-activated CRISPR targeting (Types I, II, and V)

Type I systems are the most widely distributed CRISPR-Cas systems and historically the first to be characterized in depth (Makarova et al., 2019). The type I-E system, originally identified in *Escherichia coli*, serves as the model locus and its targeting machinery comprises of two core components: (1) the crRNA-containing complex known as the CRISPR-associated complex for antiviral defense (Cascade) and (2) the signature ATP-dependent helicase/nuclease Cas3 that defines the locus (Brouns et al., 2008; Sinkunas et al., 2011) (**Figure 1.5A**). Following crRNA binding, Cascade employs its Cse1 (Cas8) subunit to survey the cell for a three nucleotide DNA sequence known as the protospacers adjacent motif (PAM) (Sashital et al., 2012). Correct PAM recognition triggers Cascade-induced DNA strand separation and invasion of the crRNA into target DNA (Hayes et al., 2016; Sinkunas et al., 2013). The crRNA then directionally zips with the complementary DNA and expands toward the end of the protospacer, forming an R-loop that stably locks the complex onto the DNA (Blosser et al., 2015; Rutkauskas et al., 2015; Xiao et al., 2017a). A particularity of Cascade complex is that its Cas7 backbone introduces kinks in the crRNA at every 6th base, leading to their outward exposure and precludes their participation in immunity *in vivo* (Fineran et al., 2014; Jackson et al., 2014; Mulepati et al., 2014). Following full base-pairing and R-loop induced locking, Cascade undergoes a conformational change that is sufficient to recruit Cas3 by the Cse1 subunit of the complex (Hochstrasser et al., 2014; Huo et al., 2014; Mulepati and Bailey, 2013; Westra et al., 2012).The first Cas3 to arrive cleaves the non-complementary strand and generates a 200-300 nucleotide single-stranded DNA (ssDNA) gap, thereby destroying the R-loop and unlocking Cascade from its target (Mulepati and Bailey, 2013; Sinkunas et al., 2013). Cas3 then utilizes its helicase activity to translocate 3'->5' along the non-complimentary strand (Mulepati and Bailey, 2013; Redding et al., 2015). Data from plasmid and phage infections suggests that Cas3-mediated immunity ultimately generates full double-stranded DNA (dsDNA) degradation, but how this is achieved *in vivo* remains unclear. Early studies of the type I system demonstrated that Cas3 helicase activity is essential for plasmid immunity (Westra et al., 2012), suggesting that Cas3 may remove secondary structures or proteins as it moves along the ssDNA; this could allow the binding of host nucleases or additional Cas3s responsible for further DNA degradation, dsDNA formation, and the completion of targeting (Mulepati and Bailey, 2013; Redding et al., 2015).

Figure 1.5. DNA-activated CRISPR targeting.
The mechanism of action of three prototypical DNA-activated CRISPR-Cas targeting complexes. (**A**) The type I-E targeting machinery comprises two components: (1) the CRISPR-associated complex for antiviral defense (Cascade) and (2) the helicase/nuclease Cas3. Cascade identifies DNA targets by searching for the protospacer adjacent motif (PAM) and then recruiting Cas3 via Cse1. Cas3 then unidirectionally translocates along one strand away from Cascade and performs DNA nicking. (**B**) The type II-A targeting machinery is composed of Cas9 and the crRNA:tracrRNA duplex. Cas9 identifies DNA targets by searching for the PAM. Target recognition simultaneously realigns the HNH and RuvC domains, leading to blunted double-stranded DNA (dsDNA) breaks. (**C**) The type V-A targeting machinery contains Cas12 and the crRNA guide. DNA targets are identified by searching for the PAM. After target binding, the non-target strand is the first to cleaved by the RuvC domain and is then released. The target strand then moves into the RuvC domain for its cleavage. Following staggered dsDNA break formation and release of the target strand, the PAM-proximal dsDNA remains bound, allowing the RuvC to remain active and capable of cleaving single-stranded DNA (ssDNA) substrates.


A
Cascade
Kinked base
Cas3
ssDNAse/helicase
Cse1
crRNA
DNA target search
PAM
Cas3 recruitment
Cas3
Cas3 nicking
and translocation
B
Cas9
HNH domain
RuvC domain
tracrRNA
DNA target search
PAM
RuvC and HNH
realignment
dsDNA break
formation
C
Cas12
RuvC domain
crRNA
DNA target search
PAM
Non-target strand
cleavage
Entry of the target strand
into the RuvC
Target strand
cleavage
Release of target
Non-specific ssDNA cleavage
ssDNA
PAM-proximal dsDNA remains bound
Global ssDNA reduction?

The type II and V effectors Cas9 and Cas12 (**Figure 1.5B-C**), famous for their use in genome engineering (Cong et al., 2013; Zetsche et al., 2015), are remarkably similar in their structure and execution of immunity but differ in how they generate dsDNA breaks. While Cas9 contains two nuclease domains (an HNH and a RuvC domain) that cleave each strand of the target DNA and generate blunt breaks (Sapranauskas et al., 2011), Cas12 uses a single RuvC nuclease domain to generate staggered dsDNA breaks (Swarts et al., 2017; Zetsche et al., 2015). For both enzymes, as with Cascade, the loading of a mature crRNA initiates a search for compatible PAMs that are obligate for DNA recognition and subsequent strand invasion (Jeon et al., 2018; Redding et al., 2015; Rollins et al., 2017; Singh et al., 2018; Sternberg et al., 2014). Directional DNA unwinding dependent on crRNA-DNA complementarity then drives the rearrangement of the nuclease domains allowing them to engage with their cleavage sites (Jeon et al., 2018; Singh et al., 2018; Sternberg et al., 2015; Sternberg et al., 2014; Szczelkun et al., 2014). For Cas9, movement of the HNH domain simultaneously triggers activation of the RuvC domain at the non-complementary strand, ensuring the cleavage of both strands (Sternberg et al., 2015). How Cas12's single RuvC domain could generate a staggered double-stranded break remained a mystery until recently. Using DNA substrates containing phosphorothioate to inhibit nuclease activity on one strand or the other, biochemical analysis revealed that the non-targeting strand is cleaved first, thereby releasing the target strand and making it accessible to the RuvC for a second cleavage event (Swarts and Jinek, 2019).

The shared strategy adopted by DNA-targeting systems (**Figure 1.5**) for executing target search and recognition brings several advantages. The initial search of a PAM allows these enzymes to first interact with a small portion of the overall binding site before deciding to probe flanking sequences with the crRNA, and as a result these enzymes dramatically reduce the total number of sequences they sample (Jeon et al., 2018; Redding et al., 2015; Rollins et al., 2015; Singh et al., 2018; Sternberg et al., 2014). The effectiveness of this simplified target search is illustrated by a single-molecule study that demonstrated that the crRNA-Cascade complex can avoid 90% of the lambda phage genome while searching for its cognate target (Redding et al., 2015). PAM recognition has added benefits given that Cascade, Cas9, and Cas12 all lack ATP-dependent helicase activity and therefore rely on the energy released from PAM strand separation to initiate R-loop formation (Hayes et al., 2016; Jeon et al., 2018; Singh et al., 2018; Sternberg et al., 2014). The dependence on the PAM, however, makes DNA-activated systems highly vulnerable to CRISPR escaper viruses that can break CRISPR immunity by introducing a single point mutation at this sequence (Deveau et al., 2008; Pyenson et al., 2017; Semenova et al., 2011). An additional strategy adopted by DNA-activated CRISPR-Cas systems is the use of crRNA to proofread the PAM-flanking sequence for mismatches during complex activation (Jeon et al., 2018; Rollins et al., 2015; Singh et al., 2018; Sternberg et al., 2014). This successive base-pairing step between the crRNA and the target allows for a second round of quality control to avoid incorrect, "off" targets (Rouillon et al., 2013; Sternberg et al., 2015; Szczelkun et al., 2014). The initial base pairing with the 7-12 nucleotides that flank the PAM, known as the seed sequence, is critical for target cleavage

(Jinek et al., 2012; Semenova et al., 2011). Mismatches between the seed region and the crRNA will prematurely abort the stable heteroduplex formation required for complex activation(Rutkauskas et al., 2015; Singh et al., 2018; Sternberg et al., 2014). Indeed, phages that mutate their target seed sequence have been shown to be able to escape DNA-activated CRISPR targeting (Nussenzweig et al., 2019; Semenova et al., 2011).

Fundamental differences between the systems emerge when comparing mechanistic requirements for DNA cleavage. Cascade is far more promiscuous in terms of the number of PAMs it can recognize because Cse1's uses of the minor rather than the major groove of DNA to establish contacts (Hayes et al., 2016). In addition, a full R-loop must form before Cascade can recruit Cas3 and initiate DNA destruction (Rutkauskas et al., 2015). By contrast, Cas9's and Cas12's nucleases are gradually activated in concert with successive base-pairing with the target, meaning that they probably sample partially-activated conformations prior to full R-loop formation (Jiang et al., 2016a; Jiang et al., 2015; Singh et al., 2018). These differences likely have profound effects on how these systems interact with targets that have only partial homology to the crRNA and on how the subsequent primed spacer acquisition response is stimulated (**Chapter 1.6**).

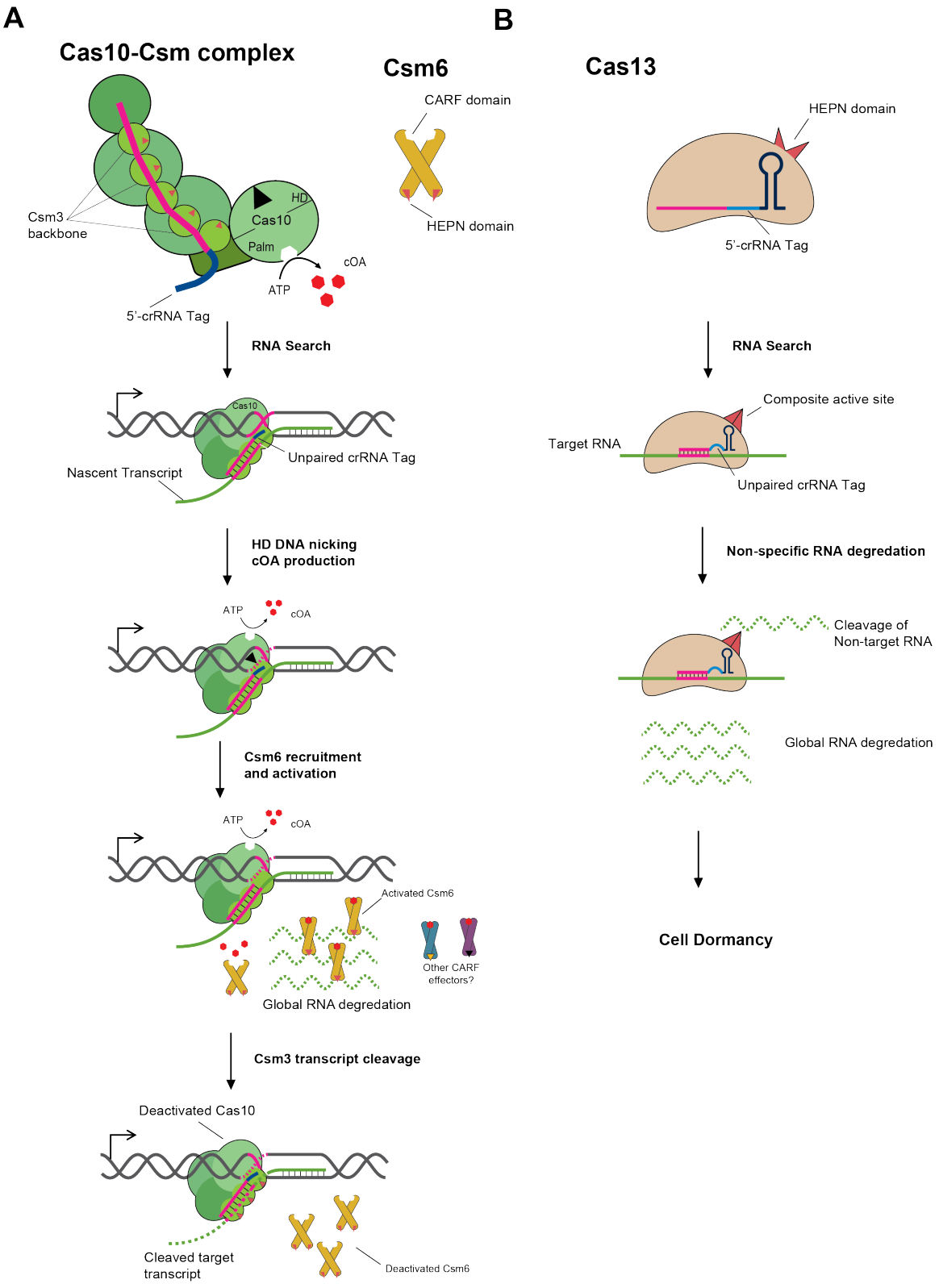
In addition to the sequence-specific dsDNA cleavage activity of Cas12, studies have shown that Cas12 can also perform sequence-independent single-stranded DNA cleavage in *trans* across multiple V subtypes (Chen et al., 2018) (**Figure 1.5C**). A recent structural study of the *Francisella novicida* Cas12a has revealed how this *trans* nuclease activity is triggered following annealing of the crRNA with its target sequence (Swarts and Jinek, 2019). Following *cis*-cleavage of the target strand, the PAM-proximal end of the DNA is released, freeing the RuvC domain to interact with new substrates. Crucially, the crRNA and target DNA strand remain paired, allowing Cas12a to stay active and cleave incoming non-target ssDNAs. Thus, type V systems possess the *cis*-cleavage properties observed in type I and II DNA-activated CRISPR-Cas systems while also possessing the collateral, non-specific *trans* activity of RNA-activated type III and VI systems. What role Cas12 *trans* cleavage plays in its immune response is unknown but ssDNA destruction may be important for interfering with the lifecycle of mobile genetic elements such as phages and conjugative plasmids that generate multiple ssDNA intermediates during rolling circle replication (Novick, 1998). Interestingly, like Cas9, Cas12a remains firmly bound to its target DNA after double-stranded *cis*-cleavage occurs (Jeon et al., 2018; Singh et al., 2018), implying that Cas12a stays active in the *trans*-cleavage competent state until it is actively dislodged. This suggests that a prolonged state of *trans*-activity may be induced leading to degradation of host ssDNA and entry into a dormancy state similar to that observed in the Csm6 and Cas13 responses. Finally, another recent study reports the existence of a type V-G locus encoding a Cas12 capable of recognizing and cleaving ssRNA *in cis* (Yan et al., 2018), further suggesting that some type V loci may use elements of both DNA- and RNA-activated systems. Uncovering the nature of immune responses guided by these new CRISPR-Cas12 systems should prove to be exciting new avenue of future research

Finally, the events that follow target DNA cleavage by DNA-activated systems, which are ultimately responsible for target destruction *in vivo*, remain uncharacterized. Neither Cas9 nor Cas12 dissociates from its cleaved DNA targets *in vitro* except under extremely harsh conditions (Jeon et al., 2018; Singh et al., 2018; Sternberg et al., 2014), suggesting that other cellular factors are required for their dislodgment. In agreement with this observation, a study demonstrated that collisions between RNA polymerase and Cas9 are sufficient to dislodge Cas9 and convert it into a multi-turnover enzyme with enhanced anti-phage immunity *in vivo* (Clarke et al., 2018). Alternatively, tight post-cleavage binding may be needed to ensure that DNA repair enzymes do not immediately mend the nascent dsDNA break. Cascade's status after Cas3 nicking remains less clear. One single-molecule study on the *E. coli* type I-E complex demonstrated that, like Cas9 and Cas12, Cascade remains tightly bound (Redding et al., 2015), while a second study on the type I-E complex from *Thermobifida fusca* showed that Cascade translocates along ssDNA in complex with Cas3 (Dillard et al., 2018). A better understanding of these discrepancies will yield further mechanistic insights into how DNA-activated immunity is ultimately achieved.

1.4.2 RNA-activated CRISPR targeting (Types III and VI)

Type III CRISPR loci are defined by the presence of the signature Cas10 nuclease which organizes with Csm2-5 (type III-A) (Rouillon et al., 2013) or Cmr1,3,4, and 5 (type III-B) (Zhang et al., 2012) to form a surveillance complex guided by the crRNA (Makarova et al., 2019) (**Figure 1.6A**). The surveillance complexes of these type III systems are unique in that they can degrade both DNA and RNA (Deng et al., 2013; Goldberg et al., 2014; Hale et al., 2009; Peng et al., 2015; Samai et al., 2015; Staals et al., 2014; Tamulaitis et al., 2014; Zebec et al., 2014). This is achieved by base pairing between the crRNA and nascent RNA transcripts, which allows for non-specific Cas10 DNA binding and activates the single-stranded DNase (ssDNase) activity of its HD domain (Kazlauskiene et al., 2016). Type III immunity thus requires active transcription across the target DNA in order to create RNA targets capable of activating the complex and tethering it to DNA(Deng et al., 2013; Goldberg et al., 2014; Liu et al., 2019). The ssDNase activity of Cas10 is, however, turned off by the sequence-specific RNase activity of Csm3 or Cmr4 that destroys the target RNA bound to the complex, leading to Cas10's deactivation and dissociation from DNA (Kazlauskiene et al., 2016; Samai et al., 2015; Staals et al., 2014; Tamulaitis et al., 2014). It is believed that the cleavage of the activating transcript both limits the region of DNase activity to the area of target transcription and also prevents "off-target" *trans* cleavage of ssDNA that might result in host DNA damage.

Figure 1.6. RNA-activated CRISPR targeting.
The mechanism of action of two prototypical RNA-activated CRISPR-Cas targeting complexes. (**A**) The type III-A targeting machinery has two components: the Cas10-Csm complex containing the crRNA, and the HEPN-domain-containing RNase Csm6. Cas10 has two activities: (1) DNA nicking mediated by the HD domain and (2) conversion of ATP to cyclic oligoadenylate (cOA) (red diamonds) mediated by the Palm domain. Targeting begins with the search for crRNA-matching target RNAs lacking the sequence complementary to the crRNA tag (the anti-tag). Binding to the nascent RNA transcripts tethers the complex to DNA and activates Cas10's dual activity. Csm6 is activated by newly produced cOA leading to non-specific, global RNA depletion. Csm6 and Cas10 are deactivated following Csm3 cleavage of the target transcript. (**B**) The type VI targeting machinery is composed of Cas13 and tag-containing crRNA. Target identification leads to HEPN domain allosteric activation, creating a composite RNase active-site that leads to non-specific RNA degradation and cell dormancy.


A
Cas10-Csm complex
Csm6
CARF domain
HEPN domain
Csm3
backbone
HD
Cas10
Palm
cOA
ATP
5'-crRNA Tag
RNA Search
Cas10
Unpaired crRNA Tag
Nascent Transcript
HD DNA nicking
cOA production
ATP
cOA
Csm6 recruitment
and activation
ATP
cOA
Activated Csm6
Other CARF
effectors?
Global RNA degredation
Csm3 transcript cleavage
Deactivated Cas10
Cleaved target
transcript
Deactivated Csm6
B
Cas13
HEPN domain
5'-crRNA Tag
RNA Search
Composite active site
Target RNA
Unpaired crRNA Tag
Non-specific RNA degredation
Cleavage of
Non-target RNA
Global RNA degredation
Cell Dormancy

Cas10 has a recently discovered secondary function as an oligonucleotide synthetase that converts cytosolic ATP into cyclic oligoadenylates (cOAs) (**Figure 1.6A**). This activity is performed by Cas10's Palm domain following target RNA binding to the complex (Kazlauskiene et al., 2017; Niewoehner et al., 2017). cOA acts as a second messenger that induces a secondary immune response directed by the RNase Csm6 (Kazlauskiene et al., 2017; Niewoehner et al., 2017). The presence of cOA is sensed by Csm6's CRISPR-associated Rossman fold (CARF) domain triggering the allosteric activation of its two Higher Eukaryotes and Prokaryotes Nucleotide binding (HEPN) domains, thereby creating a composite ribonuclease active site that executes non-specific RNA degradation (Jia et al., 2019a). This results in a global depletion of both host and foreign transcripts, causing a growth arrest (Rostol and Marraffini, 2019b) that prevents the propagation of the invader (Jiang et al., 2016b) in a manner that resembles the activity of toxins in abortive infection (Abi) systems. The potentially toxic effects of Csm6 are restricted by three events: (1) Csm3/Cmr4-mediated deactivation of Cas10 following cleavage of the activating RNA transcript, which prevents further cOA synthesis (Rouillon et al., 2018) (2) degradation of existing cOA by host ring nucleases or the Csm6 CARF domain itself (Athukoralage et al., 2018) and (3) the ultimate destruction of the invading DNA by Cas10's ssDNase activity, preventing *de novo* transcription and Cas10 reactivation (Rostol and Marraffini, 2019b).

The Csm6 RNase response is essential when the DNase activity of Cas10 is compromised or weakened during the initial type III response. This can occur when the target is transcribed at low levels, leading to slower complex recruitment (Rostol and Marraffini, 2019b), when the Cas10-Csm complex is activated by transcripts expressed late during infection and the phage has already replicated too many copies of its genome (Jiang et al., 2016b), or when the target contains mismatches with the crRNA that inhibits HD activation (Jiang et al., 2016b). It is hypothesized that during these scenarios, the secondary Csm6 response is compensatory and arrests replication of the invader via the global reduction of the host's and the invader's transcripts (Rostol and Marraffini, 2019b). The ensuing growth arrest gives the Cas10-Csm complex the added time needed to destroy the invader.

Type VI systems, defined by the signature Cas13 nucleases, are the most recent CRISPR loci to have their immune response characterized (Makarova et al., 2019; Meeske et al., 2019) (**Figure 1.6B**). Like other class 2 effectors, Cas13 forms a bilobed structure whose ability to interact with targets is triggered following crRNA binding to its central channel (Liu et al., 2017a). Unlike Cas9 and Cas12, Cas13 lacks a DNase domain and instead contains two conserved HEPN domains (Liu et al., 2017a). Following base-pairing with the crRNA, target RNA functions as an activator that triggers a conformational change that, similar to Csm6/Csx1, induces the allosteric activation of the HEPN domain and creates a composite catalytic site capable of non-specific RNase activity (Liu et al., 2017a). The location of this active site on the outer surface further distinguishes Cas13 from Cas9 and Cas12 and facilities *trans* cleavage of non-target RNAs (East-Seletsky et al., 2016; Liu et al., 2017b). *In vitro* studies have also shown that Cas13 is capable of degrading the target RNA responsible for activating it (East-

Seletsky et al., 2016). These *cis*-cleavage events of the target RNA are, however, not confined to the region of crRNA-target RNA complementarity and their mechanistic basis remains unclear.

Given its crRNA-guided RNase activity, Cas13 was initially hypothesized to defend against RNA phages. However, analysis of its associated type VI loci revealed that they contained spacers matching dsDNA phages. This raised the question of how type VI immunity against DNA parasites could be achieved using the *cis*- and/or *trans*-RNase activity of Cas13. The answer came in a study that examined the consequences of Cas13 *trans*-cleavage during DNA phage infection (Meeske et al., 2019). As with Csm6 in type III systems, Cas13 non-specific degradation of host and invader transcripts is sufficient to arrest phage replication (Meeske et al., 2019), but unlike Csm6, which is deactivated by the regulatory RNase Csm3/Cmr4, Cas13 activity is not shut-off (presumably because the viral DNA is not degraded and continues to produce target transcripts), and the host cell enters a state of dormancy following transcript depletion (Meeske et al., 2019). Immunity is therefore passively achieved at the population level as infected cells harboring activated Cas13 trap viruses in their cytosol, thus preventing viral spread to uninfected sister cells. In addition, by inducing prolonged global RNA destruction, type VI systems provide broad cross-protection against phage viruses not recognized by the crRNA, including CRISPR escapers, that infect concurrently or after Cas13 activation (Meeske et al., 2019).

The requirement for target RNAs to stimulate type III and VI systems provides them with the distinct advantage of being able to coexist with foreign DNAs under the condition that they remain transcriptionally silent. The best example of this conditional tolerance comes from type III-A systems that target the lytic genes of temperate phages (Goldberg et al., 2014). As part of their lifecycle, temperate phages integrate into the host genome as quiescent prophages that actively repress their own transcription. The type III CRISPR system cannot recognize a prophage that does not transcribe its target and therefore allows the prophage to remain in the host genome. In turn, the CRISPR-immune host cell can receive the beneficial phenotypes that accompany stable prophage symbiosis (Brüssow et al., 2004).

Interestingly, bioinformatic searches have identified many different CARF-domain-containing genes associated with type III CRISPR loci (Burroughs et al., 2015; Makarova et al., 2014; Shah et al., 2018; Shmakov et al., 2018). These genes are predicted to encode a diverse set of proteins, including nucleases, kinases, transmembrane domains, and transcription factors, that could be involved in foreign nucleic acid defense (Anantharaman et al., 2013; Shah et al., 2018). This suggests that cOAs produced during CRISPR-Cas10 targeting is a signal to activate wide range of uncharacterized immune responses. Indeed, two studies have shown that CARF-domain containing nucleases associated with type III loci can be activated by either cyclic triadenylates or tetra-adenylates (cA_3 or cA_4) (Lau et al., 2020; McMahon et al., 2020). Intriguingly, Cas10 can produce multiple cOA isoforms with different ring sizes (Kazlauskiene et al., 2017), suggesting that individual, isoform specific effectors may be activated during a coordinated immune response. Future structural and *in vivo* studies will be

needed to elucidate the complexity of these novel CRISPR-associated genes and the immune response they execute.

1.4.3 Self or Non-self?: avoiding autoimmunity

The success of any immune system requires a mechanism to distinguish between invasive foreign (non-self) elements and its own intrinsic (self) components. For CRISPR-Cas systems this represents an especially difficult challenge because the "antigenic" sequences they recognize are by definition already present within their own loci as spacers in the CRISPR array. How each systems avoids self-targeting is based on the identity of the nucleic acids that activates it, with DNA-activated systems (types I, II, V) prioritizing the recognition of non-self (Mojica et al., 2009; Semenova et al., 2011; Zetsche et al., 2015) and RNA-activated systems (types III and VI) prioritizing the recognition of self (Marraffini and Sontheimer, 2010; Meeske and Marraffini, 2018).

DNA-activated Cascade, Cas9, and Cas12 (**Figure 1.5**) all rely on unique PAMs required for initiating targeting binding during their DNA targeting responses (**Chapter 1.4.1**). As a result these complexes only stably interact with DNAs that contains both a PAM and a crRNA-matching sequences. Because repeat sequences lack PAMs, the spacers within the CRISPR array are protected against self-targeting (Mojica et al., 2009). This conservative approach of solely interacting with "non-self" DNAs, while well-suited to avoid off-targets, makes these immune systems susceptible to rapidly evolving phages that can mutate their PAMs and thereby no longer be recognized (Deveau et al., 2010; Pyenson et al., 2017; Semenova et al., 2011). Indeed, it may be the selective pressure posed by these viral escapers that has driven both type I and type II systems to evolve primed spacer acquisition mechanisms to counteract them (**Chapter 1.6** and **Chapter 3**).

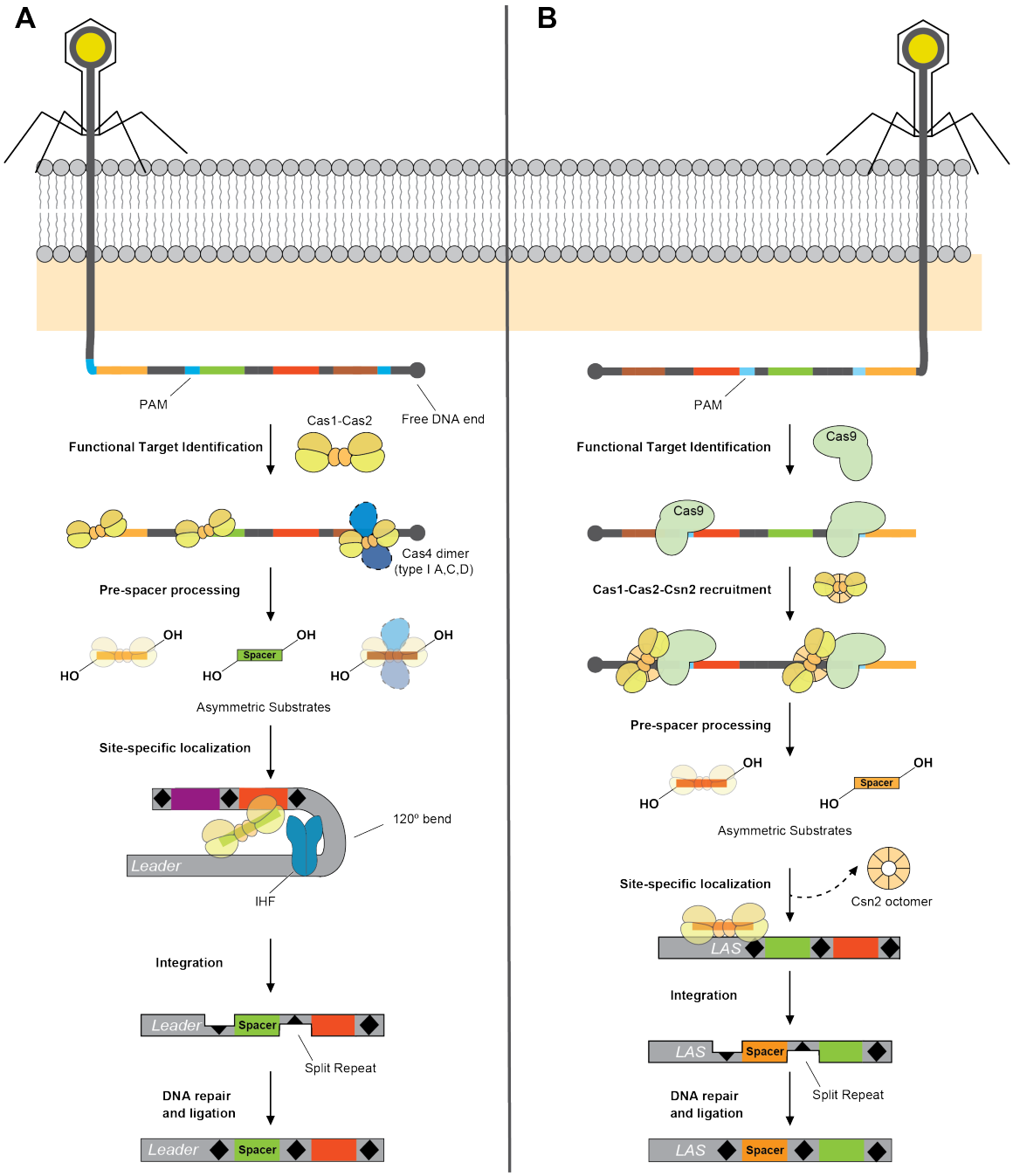
The type III and VI RNA-activating CRISPR systems, by contrast, avoid autoimmunity by identifying self (Marraffini and Sontheimer, 2010; Meeske and Marraffini, 2018) (**Figure 1.6**). Because transcription of the CRISPR array generates RNAs that have the same sequence as, rather than being complementary to, the crRNA, auto-activation of these systems can only occur after antisense transcription of the CRISPR array: a common phenomenon in prokaryotes (Lillestøl et al., 2009). Sensing of self-transcripts is achieved via an extended, invariant 5'-region of the crRNA derived from the CRISPR repeat known as the "tag" (Marraffini and Sontheimer, 2010; Meeske and Marraffini, 2018). Base pairing between the tag and its complementary sequence on the 3' end of the target RNA (i.e., the "anti-tag") is sufficient to identify the RNA as a "self" and abort both type III and VI targeting (Marraffini and Sontheimer, 2010; Meeske and Marraffini, 2018). Three recent studies on the type III-A systems have provided mechanistic insights into this process (Jia et al., 2019b; Wang et al., 2019; You et al., 2019). Collectively they present a model in which base-paring between the tag and anti-tag locks the Csm complex preventing the activation of the HD and Palm domains of Cas10, thereby inhibiting both ssDNA degradation and Csm6/Csx1 activation during the type III response. Csm3-mediated *cis*-cleavage of the anti-tag containing RNA is however permitted thereby releasing

the “lock” on the complex and allowing subsequent RNA binding (Kazlauskiene et al., 2016; Samai et al., 2015; Staals et al., 2014; Tamulaitis et al., 2014). Structures of Cas13 in complex with the anti-tag have yet to be solved but structures of Cas13 bound to an activating RNA suggest that base-pairing with the tag may similarly interfere with the allosteric activation of its HEPN1 domain (Liu et al., 2017a). Together these data suggest that RNA-activating systems avoid autoimmunity by deliberately inhibiting their own activation when encountering a “self-antigen".

1.5 Spacer acquisition: the creation of immunologic memory

The ability of CRISPR-Cas loci to provide an adaptive immune response, capable of being continuously updated to target new invader sequences, is unique among prokaryotes. CRISPR immunization is achieved by adding new fragments of foreign DNA to the CRISPR array as spacers, therein protecting the cell from subsequent infection (Barrangou et al., 2007). The CRISPR array can therefore be seen as a molecular memory system that holds records of its previous encounters with foreign genetic elements. Unlike higher order eukaryotes, the immunologic memories generated by CRISPRs are heritable and passed down to daughter cells, thereby rapidly transmitting immunity across the microbial community. The basis for this powerful and lasting immunization hinges on a two-part cleavage-ligation reaction that integrates new spacers into the CRISPR array (Arslan et al., 2014; Nunez et al., 2014; Nunez et al., 2015b) (**Figure 1.7**). Processed precursors (pre-spacers) containing two 3′-OH groups on each strand are used for a two-step nucleophilic attack on the first repeat of the array catalyzed by Cas1 within the Cas1/Cas2 complex. The first nucleophilic reaction likely uses the 5′-OH of the pre-spacer to attack the leader-repeat junction, generating a half-integrated intermediate. A second nucleophilic attack resolves this intermediate by ligating the 3′ end of the pre-spacer to a leader-distal site within the repeat. At the end of this reaction, the newly integrated, double-stranded spacer is ligated to single-stranded repeat DNA, which are then "filled-in" by host DNA repair enzymes and polymerases (Arslan et al., 2013; Ivancic-Bace et al., 2015; Nunez et al., 2014; Nunez et al., 2015b; Wright and Doudna, 2016; Wright et al., 2017; Xiao et al., 2017b; Yosef et al., 2012). For this integration reaction to successfully yield immunity the acquisition machinery must first perform three quality-control steps: (1) functional DNA sequences that meet the requirements for targeting (pre-spacers) must be captured, (2) the 5′-3′ orientation pre-spacer must be controlled to ensure the transcription of an targeting crRNA, and finally (3) pre-spacers must be localized to the CRISPR array for site-specific integration (**Figure 1.7**). An almost universal participant in spacer acquisition across CRISPR classes and types is the conserved Cas1-Cas2 complex which participates during all three aspects of spacer acquisition (Arslan et al., 2014; Makarova et al., 2020; Nunez et al., 2014; Nunez et al., 2015b; Wang et al., 2015; Wright et al., 2017). The molecular mechanisms used to solve these challenges are among the least thoroughly characterized aspects CRISPR immunity. Although considerable improvements in our understanding have been made in recent years, spacer acquisition in the laboratory setting has yet to be reported for type V, type VI, or canonical type III systems, despite the fact that all of these loci encode the required *cas1* and *cas2* genes. The sections below describe the known molecular mechanisms and tools used to acquire spacers

Figure 1.7. Naïve spacer acquisition.
The type I and type II naïve spacer acquisition mechanisms against a bacteriophage. (**A**) During type I spacer acquisition functional viral targets (pre-spacers) are identified by Cas1-Cas2 binding the PAM in type I-E system and Cas1-Cas2-Cas4 in type I-A, I-A, -C, -D, and -U. Pre-spacers are then processed, creating asymmetric substrates with 3′-OH groups on each strand. The pre-spacers are localized to the CRISPR array by integration host factor (IHF) which introduces a 120° bend in the DNA. (**B**) In type II spacer acquisition, pre-spacers are identified by Cas9 binding to the PAM. The spacer acquisition complex (Cas1-Cas2-Csn2) is recruited and pre-spacers are processed creating asymmetric substrates with 3′-OH groups on each strand. Csn2 dissociates from the acquisition complex and Cas1-Cas2 are localized to the CRISPR array by binding the leader-anchoring sequence (LAS). In both types, Cas1 mediates spacer integration by carrying out a two-step transesterification reaction that separates the first repeat (black diamond) into single-strands. The split repeat is repaired and ligated, completing the integration reaction.


A
PAM
Free DNA end
Functional Target Identification
Cas1-Cas2
Cas4 dimer
(type I A,C,D)
Pre-spacer processing
OH
HO
Spacer
Asymmetric Substrates
Site-specific localization
120° bend
Leader
IHF
Integration
Split Repeat
DNA repair
and ligation
B
PAM
Functional Target Identification
Cas9
Cas1-Cas2-Csn2 recruitment
Pre-spacer processing
Asymmetric Substrates
Site-specific localization
Csn2 octomer
LAS
Integration
Split Repeat
DNA repair
and ligation

In order to generate immunity, acquired spacers must match sequences that can be recognized and cleaved by the effector complex. An early question in the field was whether CRISPR systems contain an active quality-control mechanism that guarantees the selection of these "functional" spacers. Alternatively, without any pre-spacer discrimination, cells that acquired non-functional spacers would be killed by pathogens and disappear from the population. A hallmark of the characterized spacer acquisition mechanisms (types I and II) is the development of strategies to prioritize the capture of PAM-flanked pre-spacers. In type I systems, the Cas1-Cas2 complex is composed of two Cas1 dimers sandwiching a single Cas2 dimer in a butterfly shaped arrangement (Nunez et al., 2014; Wang et al., 2015). In the type I-E subtype, the Cas1-Cas2 complex selects functional spacers by directly interacting with the PAM sequence via Cas1 (Wang et al., 2015) (**Figure 1.7A**). Substrate binding rearranges the complex, allowing the Cas1 homodimer to trim the respective ends of the pre-spacer and generate the 3'-OH groups needed for integration (Arslan et al., 2013; Nunez et al., 2014). Cas1-mediated PAM selection is however not universal to all type I subtypes. In the *Bacillus halodurans* type I-C system, Cas4 forms a symmetric complex with Cas1-Cas2 ($Cas4_2$-$Cas1_4$-$Cas2_2$) with a Cas4 monomer pairing with a Cas1 on each side of the butterfly structure (Lee et al., 2019) (**Figure 1.7A**). In this system, as well in type I-A and -D, Cas1-Cas2 alone are sufficient to mediate spacer acquisition but, in the absence of Cas4, the selection of functional spacers is lost (Lee et al., 2018; Shiimori et al., 2018). Cas4 ensures spacer functionality by recognizing PAM-containing pre-spacers bound by Cas1-Cas2 (Almendros et al., 2019; Lee et al., 2018; Shiimori et al., 2018). These PAM-containing pre-spacers are then trimmed at their ends producing ssDNA overhangs with 3'-OH groups. In all these systems, quality control is ensured by coupling the identification of PAM to pre-spacer processing, thereby allowing for the preferential acquisition of functional spacers (Lee et al., 2018; Shiimori et al., 2018). In addition to type I systems, Cas4 is also found in type II, III, and V loci, raising the question of whether Cas4 performs a similar quality-control function in spacer acquisition for these systems.

In type II-A systems, the spacer acquisition machinery comprises of Cas1, Cas2, Csn2, and the Cas9:tracr:crRNA complex, all of which are required for spacer acquisition *in vivo* (Heler et al., 2015; Wei et al., 2015b) (**Figure 1.7B**). Here, selection of PAM-containing pre-spacers is performed by Cas9 with the same PAM-interacting residues used to identify the PAM during target DNA cleavage (Heler et al., 2015). The pre-spacer is then presumably handed off to the Cas1-Cas2 sub-complex, where it is processed prior to integration, although mechanistic details of this transition have yet to be elucidated. Notably, spacer acquisition can occur when Cas9's dsDNase activity of Cas9 is eliminated (dCas9), and therefore Cas9 cleavage is not essential for function during pre-spacer capture (Heler et al., 2015; Wei et al., 2015b). How the third member of the type II spacer acquisition complex, Csn2, fits into this model has remained mysterious. *In vitro* integration reactions with *S. pyogenes* Cas1-Cas2 are inhibited in the presence of Csn2 (Wright and Doudna, 2016), suggesting that it is not part of the complex when pre-spacers are integrated into the CRISPR array. This, combined with genetic and structural data showing that two Csn2 tetramers

form a channel that threads dsDNA ends (Arslan et al., 2013; Bernheim et al., 2017; Wilkinson et al., 2019), suggests that Csn2 may traffic Cas1-Cas2 to PAM-containing pre-spacers identified by Cas9.

The type III system requires its targets to be transcribed in order to provide full immunity. This requirement therefore presents a unique problem for the type III spacer acquisition machinery: how to prioritize the acquisition of the spacers from actively transcribed regions? The *Marinomonas mediterranea* type III-B system has ingeniously solved this problem by fusing a reverse transcriptase (RT) domain to Cas1, allowing it to capture transcripts, integrate them into the array first as ssRNA spacers, and then convert them to cDNAs (Silas et al., 2016). Bioinformatic studies have recently uncovered several type VI loci that contain RT-Cas1 fusions (Toro et al., 2019) suggesting that this strategy may also have been adopted by these systems. The vast majority of conventional type III systems however lack RT-Cas1 fusions (Makarova et al., 2019). This raises the question of whether these systems bias their spacer acquisition toward transcriptionally active sites or simply rely on negative selection to remove non-functional spacers from the population.

All Cas1-Cas2 complexes seem to use free DNA ends as the substrates for pre-spacers. These free ends originate by dsDNA breaks that occur at the origin of replication of plasmids and phages (Levy et al., 2015) or at the ends of the phage genome injected during infection (Modell et al., 2017). For type I and II systems, this bias has been shown to help avoid the acquisition of self-targeting spacers derived from the host chromosome, since these are circular DNA molecules that replicate relatively infrequently compared with multi-copy plasmids and phages (Levy et al., 2015). As a consequence of this, deleterious self-acquisition from host chromosome events preferentially occur at the terminus and at highly transcribed genes such as ribosomal RNAs (Shiimori et al., 2017), regions that often accumulate dsDNA breaks and therefore contain free DNA ends.

A striking feature of CRISPR arrays across types is the relatively constant length of spacers, implying that it is fixed prior to integration. Structural studies on the *E. coli* Cas1-Cas2 complex have demonstrated that the length of the spacer is intrinsically set by the distance between the two Cas1 homodimers that flank the complex (Nunez et al., 2015a; Wang et al., 2015). Whether the Cas1-Cas2 complex acts as the molecular ruler for other systems' spacers is unknown but uncovering the basis of spacer length will undoubtedly provide further insights into how pre-spacers are processed.

In order to ensure the transcription of functional crRNAs, newly acquired spacers must be oriented properly so that they match their targets. For all DNA-activated CRISPR-Cas systems, spacer orientation is likely determined in part by the presence of the PAM in pre-spacers captured by Cas1-Cas2. In the Cas4 containing type I systems, differential single-stranded trimming of the PAM-containing and PAM-lacking ends of the pre-spacer creates an asymmetric substrate that can be used to determine the spacer orientation (Lee et al., 2018; Shiimori et al., 2018) (**Figure 1.7A**). In the *Pyrococcus furiosus* type I-A system, this asymmetry is additionally reinforced by using two distinct Cas4 genes (*cas4-1* and *cas4-2*), each providing unique sequence specificity (Shiimori et al., 2018).

Whereas Cas4-1 recognizes the PAM, Cas4-2 recognizes a NW motif on the opposite side of the pre-spacer. Each of these sequence motifs are then used as molecular rulers to determine the length of 3′-overhangs generated by both Cas4 proteins. In both type I and II systems that lack Cas4, PAM sequences must also be removed prior to spacer integration, implying the existence of similar PAM-processing steps. For type I-E systems, recent *in vitro* studies have suggested that asymmetry is generated due to Cas1 binding and protecting the PAM proximal end of the prespacer while host 3′-5′ exonucleases trim the PAM-distal end (Kim et al., 2020; Ramachandran et al., 2019). This is in agreement with the finding that in *Streptococcus thermophilus* Cas2 is fused with an exonuclease DnaQ-like domain that aids in the maturation of prespacers (Drabavicius et al., 2018). Additionally, how spacer orientation is coordinated in RNA-activated CRISPR systems that lack PAMs is completely unknown. Notably, when the *M. mediterranea* type III-B system that harbors RT-Cas1 is heterologously expressed in *E. coli* or used for *in vitro* integration reactions, it integrates spacers in both directions (Silas et al., 2016). This suggests that additional factors biasing spacer orientation, yet undiscovered, are present in the natural host.

Once pre-spacers have been captured and oriented according to the location of the PAM, they must be brought to the CRISPR array for integration. Specific integration is particularly important to prevent the genomic instability that would be generated from non-specific integration of viral sequences elsewhere in the host chromosome (Wright and Doudna, 2016). Cas1 is believed to have evolved from an ancestral transposase (Casposon) (Krupovic et al., 2017), and therefore its domestication from a non-specific to an array-specific integrase represents a pivotal moment in CRISPR-Cas evolution. An early clue for how this specificity is achieved came from the observation that spacer acquisition is polarized; with new spacers successively added to the leader end of the array (Barrangou et al., 2007). The explanation for both of these phenomena comes from the intrinsic specificity of Cas1 for sequences at the leader-repeat junction. In type II systems this sequence, known as the leader-anchoring sequence (LAS), is essential for polarized acquisition (McGinn and Marraffini, 2016; Wei et al., 2015a). Absence of a LAS in either engineered bacteria (McGinn and Marraffini, 2016) or naturally occurring mutants (Wright and Doudna, 2016), results in ectopic integration of spacers at LAS-like sites in the CRISPR array. In the type I-E system, a leader sequence also specifies polarized integration, but here the host integration factor (IHF) protein is also required (Nunez et al., 2016). IHF binds the leader sequence and generates a 120° bend in the DNA, inducing Cas1-Cas2 binding and spacer integration (Yoganand et al., 2017). It has been hypothesized that the bent leader-repeat site may create an ideal integration target for the Cas1 integrase but exactly how this topology promotes spacer acquisition is unclear. Notably bacteriophages, transposases, and mammalian retroviruses all preferentially integrate at bent target sites (Muller and Varmus, 1994). IHF is, however, not present in gram-positive bacteria or many archaea, suggesting that these organisms have repurposed other DNA-bending factors for integration or depend solely on Cas1-Cas2 intrinsic specificity. Interestingly, polarized acquisition has potential long-term consequences for immunity, as the first spacer in the CRISPR array provides the strongest defense (McGinn and Marraffini, 2016). Transcriptional profiling of several CRISPR loci revealed that

spacers located closer to the leader end are transcribed to higher levels resulting in higher amounts of leader-proximal crRNAs (Deltcheva et al., 2011; Hale et al., 2008), and therefore it has been hypothesized that this results in enhanced immunity (McGinn and Marraffini, 2016).

After the pre-spacer has been localized and undergone its first transesterification reaction at the leader-repeat junction, the second nucleophilic attack has to take place in the repeat. This step is vital for ensuring proper placement of the spacer within the array and to guarantee that the repeats are duplicated correctly following integration. Multiple studies in both type I and II systems have shown that the decision of where to anchor the spacer is dictated by sequence motifs at the distal end of the repeat (Goren et al., 2016; Wang et al., 2016). These sequences are believed to act as molecular rulers that specify where the second nucleophilic attack occurs. Recent structural studies of the type II-A Cas1-Cas2 complex bound to the repeat during integration indicate that the Cas1-Cas2 complex induces sequence-independent bending that is necessary for positioning the Cas1 active site during spacer placement (Xiao et al., 2017b). Future studies will be needed to resolve the interplay between these "ruler motifs" and the complicated structural gymnastics that guide Cas1-Cas2 spacer insertion. Finally, the characterization of the steps that occur after spacer integration and that lead to complex disassembly and repeat duplication is still needed to complete our understanding of spacer acquisition.

1.6 crRNA-directed "primed" spacer acquisition in the type I CRISPR-Cas systems

The sequence-specificity of CRISPR-Cas immunity makes it vulnerable to escape by rapidly evolving invaders such as bacteriophages that can mutate the target sequences recognized by the crRNA. Indeed, escaper viruses rapidly emerge within hours of infection of monoclonal CRISPR-immune hosts with a single targeting spacer (Pyenson et al., 2017). As a result, many targeting spacers must be present in the population to produce sterilizing immunity that eliminates the existing bacteriophages and prevents the emergence of new escaper viruses (van Houte et al., 2016). In order to rapidly acquire a diverse pool of spacers, CRISPR systems have developed a third phase to their immune response known as primed spacer acquisition or priming. CRISPR priming utilizes preexisting spacers already present in the array to mediate further spacer acquisition (Nussenzweig et al., 2019; Semenova et al., 2011). Thus, priming is able to enhance the rate of spacer acquisition by using the crRNA to direct the acquisition machinery to the invaders' nucleic acids. Furthermore, by selectively biasing spacer acquisition to foreign nucleic acids, priming may limit acquisition of DNA from the host chromosome and avoid self-targeting. Primed spacer acquisition has been reported extensively in type I systems. Here we review the molecular mechanisms of type I priming.

Type I priming (**Figure 1.8**) has been observed in multiple CRISPR systems, with the *E. coli* type I-E and the potato bacterium *Pectobacterium atrosepticum* type I-F systems being the most thoroughly characterized to date. In both systems, mutations that diminish the ability of Cascade to form an R-loop with the target sequence stimulate the priming response (Datsenko et al., 2012; Fineran et al., 2014; Richter et al., 2014). As a result, CRISPR escapers with mutations in either the PAM or target sequence are subject to primed spacer acquisition (Datsenko et al., 2012). The mechanistic basis for the primed spacer response against these targets is dependent on recruitment of the Cas1-Cas2 complex to Cascade (Redding et al., 2015). If R-loop formation is impaired by mutations at the PAM or target sequence, Cas3 recruitment is abrogated, allowing Cas1-Cas2 to arrive before Cas3 and act as a negative regulator of Cas3 nuclease activity (Redding et al., 2015; Rollins et al., 2015). Cas1-Cas2 and Cas3 then translocate together along the non-targeted strand as a "primed acquisition complex" (PAC) (Dillard et al., 2018; Redding et al., 2015). Cas3-dependent translocation of Cas1-Cas2 along ssDNA therefore provides an elegant explanation for the strand bias observed during type I priming (Datsenko et al., 2012; Swarts et al., 2012). In agreement with this, in the type I-B system, in which Cas3 can be loaded onto either strand, priming yields spacers from both strands (Li et al., 2014). Notably, primed spacer acquisition has also been observed in the presence of perfectly matching targets with canonical PAMs but at lower rates than priming against imperfect targets (Semenova et al., 2016; Staals et al., 2016; Xue et al., 2015). This may be explained by the fact that when Cascade is bound to perfect targets, Cas3 is likely recruited before Cas1-Cas2 and rapidly destroys the target before robust spacer acquisition can occur. In support of this model, inhibition of Cas3 nuclease activity with a phage derived CRISPR inhibitor can increase the rate of primed spacer acquisition against a perfect target (Staals et al., 2016). Interestingly, in type I-F, Cas2 and Cas3 are fused, further suggesting a

functional link between DNA destruction and spacer acquisition (Fagerlund et al., 2017; Rollins et al., 2017). A fundamental question left unanswered by this model is how pre-spacers are generated following Cas1-Cas2-Cas3 translocation away from Cascade. Crucially, Cas3-mediated target degradation is capable of producing DNA fragments with the characteristic asymmetric structure needed for functional spacer acquisition (see **Chapter 1.5**), implying that Cas3 directly participates in pre-spacer production (Künne et al., 2016; Shiriaeva et al., 2019). If Cas3 activity is required for spacer acquisition, how is the nuclease reactivated following Cas1-Cas2 inhibition? One possibility is that collisions between the PAC and DNA-binding proteins stall its translocation and allow sufficient time for Cas3 to reactivate and produce DNA breaks (Dillard et al., 2018). Alternatively, the inherent ability of Cas1-Cas2 to recognize PAMs may cause the PAC to stall at PAMs and allow for Cas3-mediated pre-spacer generation.

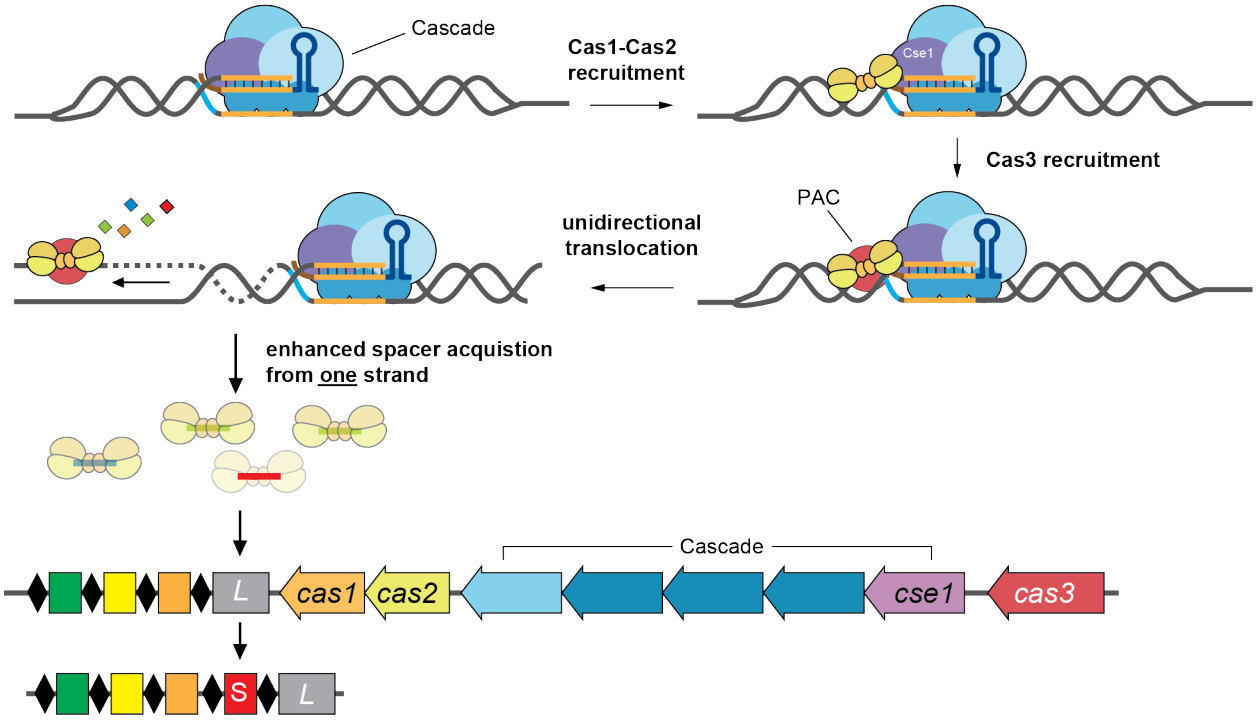


Figure 1-8. Primed spacer acquisition in the type I CRISPR-Cas system. The mechanism of spacer acquisition in type I CRISPR-Cas systems. In the type I-E system, Cascade binds PAM-containing targets and Cas3 associates with the complex prior to Cas1-Cas2 recruitment. The primed spacer acquisition complex (PAC) forms and translocates unidirectionally producing substrates (pre-spacers) (colored squares) for spacer acquisition from one strand.

A central question in type I priming is how Cascade chooses to either recruit Cas3 and engage in DNA targeting or recruit Cas1-Cas2 and engage in spacer acquisition. One "conformation controlled" model based on structural and single-molecule studies proposes that Cascade enters into a priming-specific state during primed spacer acquisition (Blosser et al., 2015; Redding et al., 2015).

In this model, the Cse1 subunit of Cascade has an "open" conformation in the presence of mismatched targets or mutant PAMs which promotes Cas1-Cas2-dependent recruitment of Cas3 for priming (Xue et al., 2015; Xue et al., 2016). In addition, a distinct, "locked", state recruits Cas3 to perfect targets for their destruction (Blosser et al., 2015; Redding et al., 2015). Alternatively, a model that does not rely on alternative conformations has been proposed in which targeting and priming are the results of the same DNA destroying pathway (Semenova et al., 2016). This second model is based on two findings. First, when priming is stimulated against perfect and imperfect targets the same pattern of spacer acquisition is generated (Semenova et al., 2016), suggesting that there is no functional distinction in how Cascade interacts with these targets. Second, target mutations that differentially affect Cascade's ability to form R-loops, and presumably enter the "locked" confirmation, do not correlate with differences in the efficiency of priming against these targets (Krivoy et al., 2018; Musharova et al., 2019). In this alternative, "kinetically controlled", model of priming, imperfect targets produce more potent priming solely because they are destroyed less rapidly and are therefore more readily available to the acquisition machinery (Semenova et al., 2016). Future work will be needed to distinguish between these two competing models.

1.7 Studying type II-A spacer acquisition in a *Staphylococcus aureus* heterologous system

When this work began the use of the gram-positive bacterium *Staphylococcus aureus* as a model organism to study CRISPR-Cas immune systems had been well established by our laboratory (Goldberg et al., 2014; Heler et al., 2015; Marraffini and Sontheimer, 2008). In these studies, the native *Streptococcus pyogenes* SF370 type II-A CRISPR-Cas locus and the native RP62a type III-A CRISPR-Cas locus were individually cloned into plasmid vectors (pCRISPRs) (Goldberg et al., 2014). In addition, Drs. David Bikard and Wenyan Jiang had together developed parental pCRISPR vectors that had their endogenous CRISPR arrays replaced with simplified arrays containing a single set of repeats flanking a placeholder spacer with two oppositely oriented BsaI restriction sties (pCRISPR-BsaI) (Jiang et al., 2013). This allowed for the programing of pCRISPR systems with a spacer of interest following digestion of pCRISPR-BsaI vectors with BsaI and ligation with annealed oligonucleotides containing BsaI-compatible overhangs.

The laboratory strain derived from NCTC8325, RN4220 (Nair et al., 2011) was selected to carry these pCRISPRs. The rationale for the use of RN4220 is based on three features of this cell line: (1) it lacks functional restriction-modification (R-M) systems allowing it to be easily transformed with plasmids and genetically manipulated, (2) it has been cured of all its prophages making it highly susceptible to bacteriophage infection, and (3) it lacks endogenous CRISPR-Cas systems allowing the immunity provided by heterologously expressed pCRISPRs to be examined in isolation (Nair et al., 2011). Genetic manipulation and phage infection of *S. pyogenes* SF370 by contrast was significantly more difficult and, at the time, our lab lacked the genetic tools and phages to carry out such work.

An equally important component for this model system was the isolation of phages capable of infecting these pCRISPR-containing RN4220 derivates and activating the CRISPR-Cas immune response. In this respect, I was very fortunate for the work of Dr. Gregory Goldberg who had, shortly before I had joined the lab, isolated strictly lytic variants of the staphylococcal temperate phages ΦNM1 and ΦNM4 (known as ΦNM1γ6 and ΦNM4γ4 respectively) (Goldberg et al., 2014) that were originally present in the *S. aureus* Newman chromosome as prophages (Bae et al., 2006). Finally, it was experimentally validated by Dr. Robert Heler that these lytic phages could elicit naive spacer acquisition in RN4220 containing pCRISPRs derived from the wild-type SF370 locus that lacked any pre-exiting immunity to these phages (Heler et al., 2015). Robert also developed an assay to assess the acquisition of new spacers into the SF370 type II-A locus via PCR amplification of the pCRISPR array (Heler et al., 2015). This previous work established an ideal model system to investigate type II-A spacer acquisition *in vivo*. In brief, RN4220 containing pCRISPRs were infected in liquid culture or soft agar with ΦNM1γ6 and/or ΦNM4γ4 and then assayed for *de novo* spacer acquisition via PCR (**Figure 1.9**) (Heler et al., 2015). Thus, I was given an ideal platform from which to explore the central question of my thesis: how CRISPR-Cas9 targeting affects *de novo* spacer acquisition.

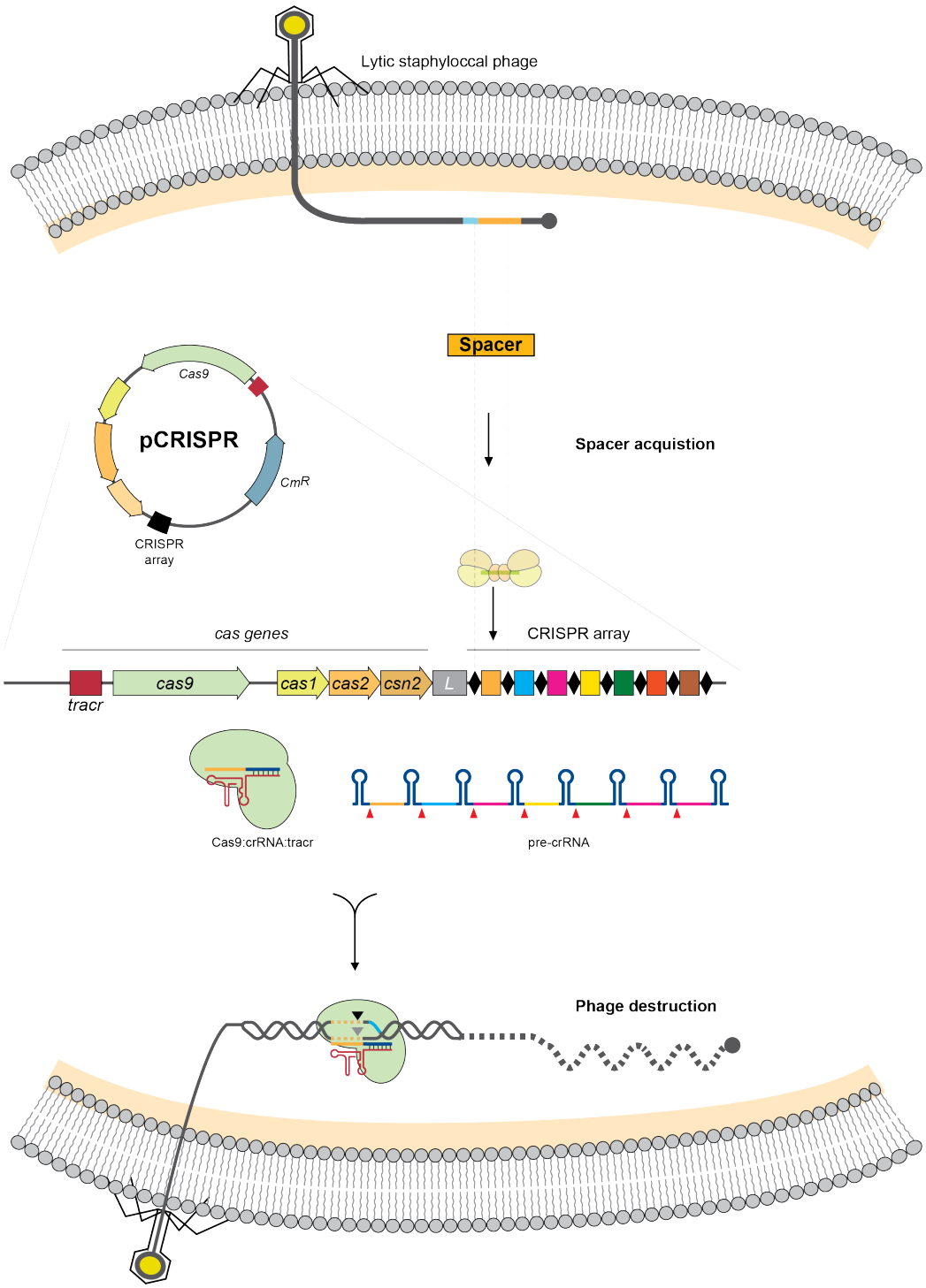


Figure 1.9. The *S. aureus* model system for studying type II-A CRISPR-Cas immunity

A schematic of showing the experimental model system used in this work to study type II CRISPR-Cas immunity. The SF370 *S. pyogenes* type II-A CRISPR-Cas locus is located on a plasmid that expresses the *cas* genes and crRNA. This system is capable of both spacer acquisition and DNA targeting against lytic staphylococcal phages leading to the destruction of the phage.

CHAPTER 2: CAS9 CLEAVAGE OF VIRAL GENOMES GENERATES IMMUNOLOGICAL MEMORIES

2.1 CRISPR immune cells acquire additional spacers

Recently we discovered that spacer acquisition in this system has a marked preference for linear dsDNA ends (Modell et al., 2017). Since it is well established that these are also the products of Cas9 cleavage both in vivo (Garneau et al., 2010) and in vitro (Gasiunas et al., 2012; Jinek et al., 2012), we decided to investigate if type II-A CRISPR-Cas targeting of phage invaders could lead to spacer acquisition in cultures that are already immune. To test this, the native CRISPR array of pCRISPR was replaced with a single-spacer array containing spacer 174 [pCRISPR(*spc174*); the sequences of all spacers used in this study are compiled in **Table 4.5**], a spacer commonly acquired by this heterologous system following infection with the staphylococcal phage ΦNM4γ4 (**Figure 2.1A**) (Heler et al., 2019). As a control, we utilized a pCRISPR(SR) harboring a single-repeat CRISPR locus; i.e. lacking a pre-existing spacer (we were concerned that possible off-target effects of a "non-targeting" spacer could invalidate it as a negative control). One caveat of this approach is the possibility of a *cis* effect of the pre-existing spacer on the integration of new ones, which could lead to differences of spacer acquisition due to divergent levels of spacer integration into the "regular" vs. "single-repeat" CRISPR loci.

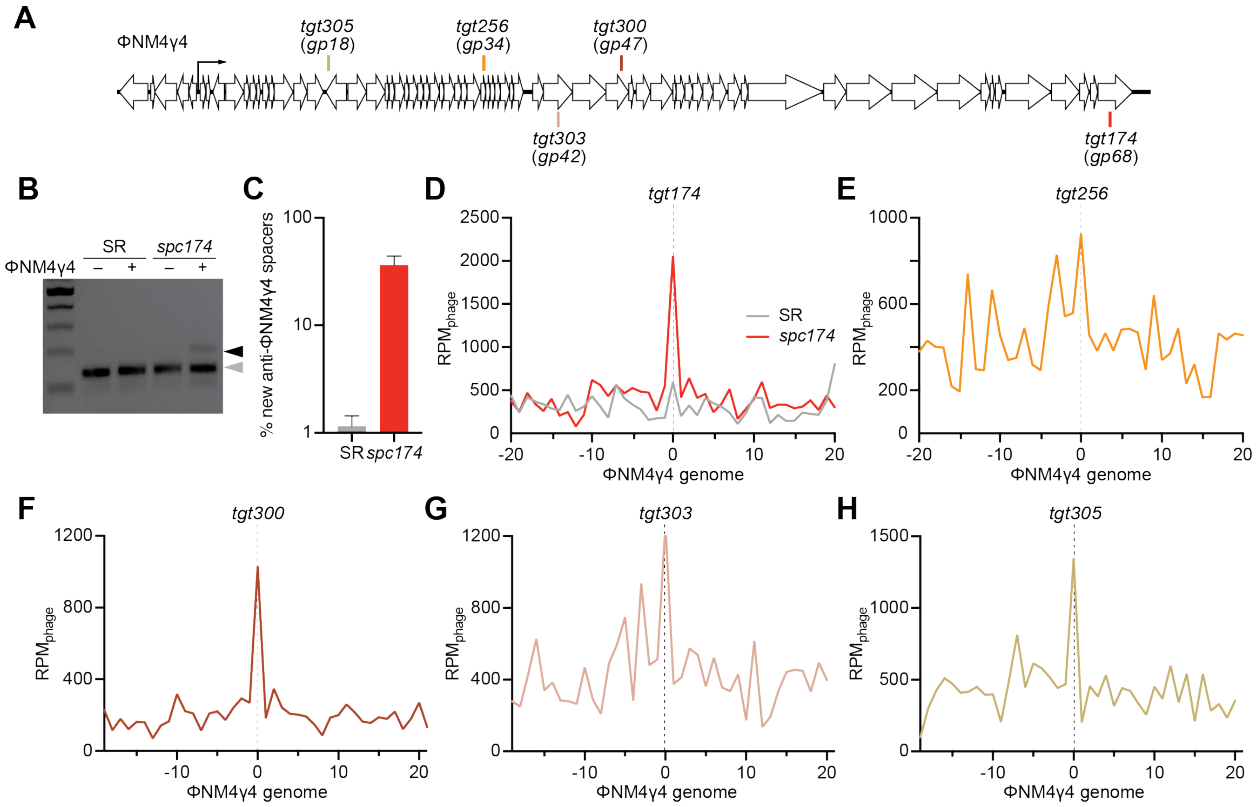
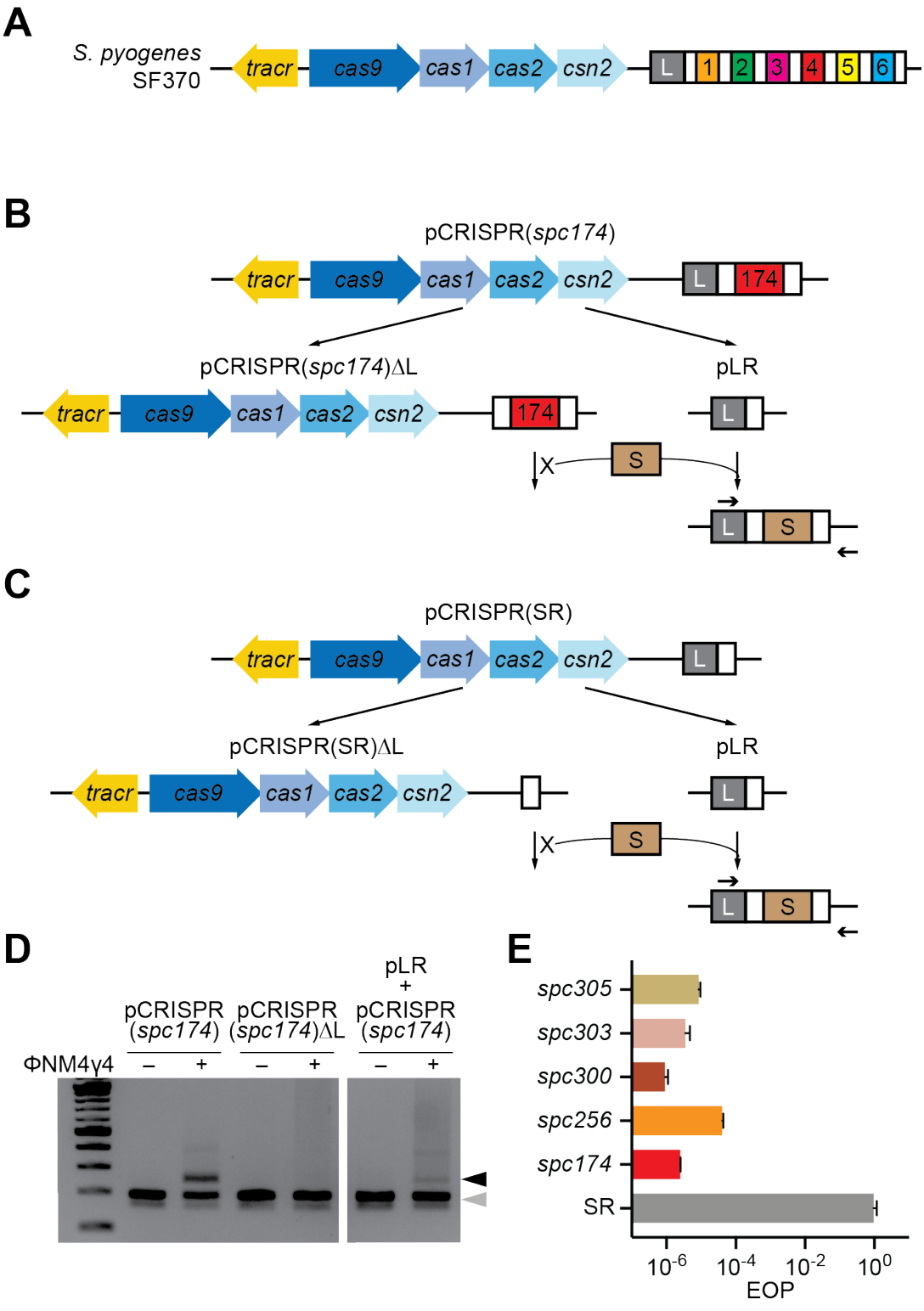


Figure 2.1. Immune cells acquire can additional spacers upon infection. Staphylococcal phage ΦNM4γ4 genome and the Cas9 targets analyzed in this study. (**B**) Staphylococci harboring pCRISPR(*spc174*) that targets the phage ΦNM4γ4 and pCRISPR(SR) were infected (+) or mock-infected (-) for 30 minutes and their DNA purified to amplify the CRISPR locus. PCR products were separated by gel electrophoresis to detect the acquisition of new spacers. Grey and black arrows: non-expanded and expanded, respectively, CRISPR arrays. (**C**) Fraction (%) of spacer sequences matching to the genome of phage ΦNM4γ4 detected in the PCR product of the expanded CRISPR array of staphylococci harboring either pCRISPR(*spc174*) or pCRISPR(SR). Mean ± StDev values of three independent experiments are shown. (**D-H**) Distribution of new spacer sequences detected 30 minutes after infection of staphylococci expressing Cas9 programmed to cleave the targets shown in (**A**). Reads per million of phage reads (RPM_{phage}) are mapped to 1 kb bins of the ΦNM4γ4 genome (shown in linear form, with the specified target in the center). Average curve of three independent experiments is shown.

To avoid spacer integration biases, we deleted part of the leader sequence in both plasmids (generating pCRISPR(*spc174*)ΔL and pCRISPR(SR)ΔL, **Figure 2.2B-C**), a mutation that prevents the integration of new spacers (**Figure 2.2D**) in *cis*. We then transformed the cells with a second plasmid containing only the leader sequence and a single repeat (pLR), which will enable spacer acquisition in *trans* into genetically equivalent CRISPR loci (**Figure 2.2B-D**). We corroborated

efficient targeting by *spc174* in this dual plasmid system (**Figure 2.2E**) and proceeded to test our hypothesis. Cells carrying either sets of plasmids were infected with ΦNM4γ4, collected 30 minutes after infection for plasmid DNA extraction and amplification of its CRISPR locus. In line with previous observations that type II systems acquire spacers preferentially from free DNA ends, we detected an expansion of the CRISPR array (**Figure 2.1B**). This result demonstrated that not only naïve cells, which must acquire new spacers to survive infection, but also immune cells, which a priori do not need extra spacers to destroy the phage, can generate new CRISPR memories.

Figure 2.2: Design of a two-plasmid system for the study of the effect of pre-existing spacers in type II-A CRISPR-Cas spacer acquisition
(**A**) Type II-A CRISPR-*cas* locus of *S. pyogenes* SF370. White rectangles, CRISPR repeats; colored and numbered rectangles, spacers; "L", leader sequence; blue arrows, protein-coding genes; yellow arrow, tracrRNA gene. (**B**) pCRISPR(*spc174*) was generated by cloning the *S. pyogenes* SF370 type II-A CRISPR- *cas* locus into the staphylococcal pC194 vector, and the 6-spacer CRISPR array replaced with a single-spacer array harboring the *spc174* sequence. pCRISPR(*spc174*)DL was generated after the deletion of the leader, which prevents spacer acquisition. To enable the capture of new spacers, a second plasmid (pE194) containing only the leader and a single repeat was added. Arrows indicate priming sites of oligonucleotides used to detect spacer acquisition via PCR. (**C**) To generate a "no pre-existing spacer" control, pCRISPR(SR)DL was constructed, which contains only a single repeat sequence. (**D**) Detection of spacer acquisition by agarose gel electrophoresis of PCR products obtained with primers and plasmid templates shown in (**B**) and (**C**). Grey and black arrows: non-expanded and expanded, respectively, CRISPR arrays. (**E**) Comparison of the efficiency of plaquing (EOP) of ΦNM4γ4 on staphylococci carrying pCRISPR plasmids programmed with spacers *174*, *256*, *300*, *303* and *305*, or without a spacer (SR). Mean ± StDev values of three independent experiments are shown.

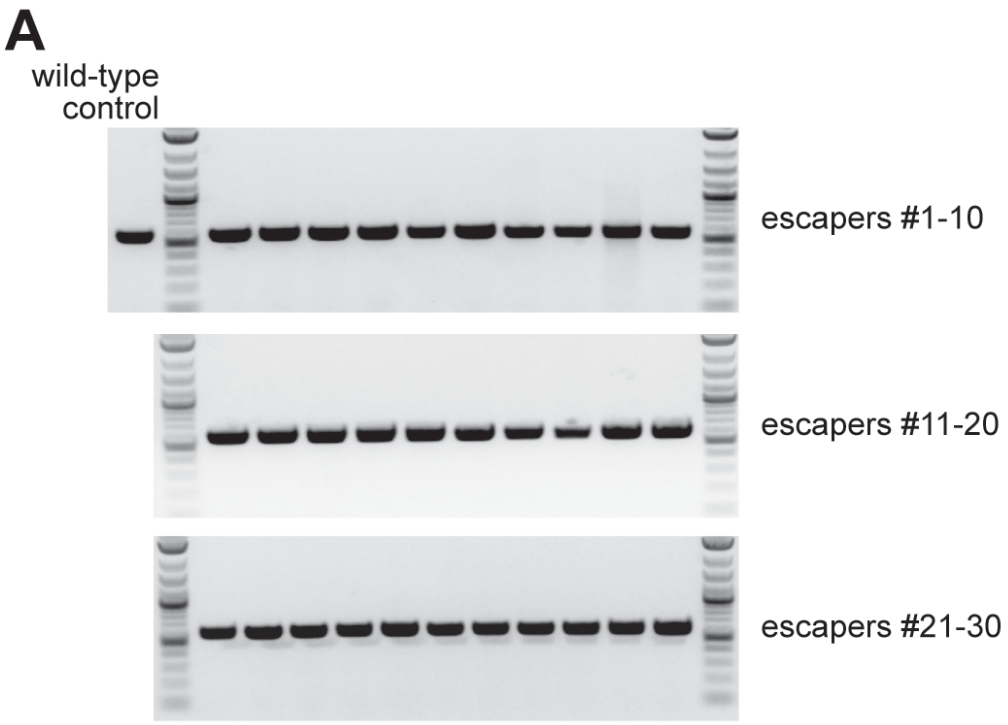

A
S. pyogenes SF370
tracr
cas9
cas1
cas2
csn2
L
1
2
3
4
5
6
B
pCRISPR(spc174)
174
pCRISPR(spc174)ΔL
pLR
S
X
C
pCRISPR(SR)
pCRISPR(SR)ΔL
D
pLR + pCRISPR (spc174)
pCRISPR (spc174)
pCRISPR (spc174)ΔL
ΦNM4γ4
E
spc305
spc303
spc300
spc256
spc174
SR
EOP

2.2 Pre-existing spacers determine the pattern of spacer acquisition

Next, we wanted to determine how the presence of a targeting spacer influences the acquisition of additional ones. To do this, we compared spacer acquisition in cells carrying pCRISPR(*spc174*)ΔL or pCRISPR(SR)ΔL. Both strains were infected with ΦNM4γ4, pLR was extracted, the new spacers amplified and the products subject to next-generation sequencing. Analysis of the data indicated a striking difference in the acquisition of new spacers from the phage genome, with the presence of the pre-existing spacer *spc174* resulting in ~ 30-fold more anti-viral spacer acquisition (**Figure 2.1C**). We then mapped the acquired spacers and their abundances along the ΦNM4γ4 genome. We found that, in the presence of *spc174*, the new spacers clustered around the targeting site (*tgt174*, **Figure 2.1D**). We corroborated this finding by investigating the sequences acquired in the presence of four different targeting spacers, *spc256*, *spc300*, *spc303* and *spc305* (**Figure 2.1A**). All these spacers provided efficient immunity against ΦNM4γ4 (**Figure 2.2E**) and, although not as striking as the peak generated at the *tgt174* site, we observed distinguishable spacer acquisition peaks at the targets of all of them (**Figure 2.1E-H**). Together these results show that the presence of a pre-existing spacer not only increases dramatically the spacer acquisition rate, but also determines the genomic location of the new spacers.

2.2.1 The isolation of *spc174* mutant escaper phages

To investigate the molecular mechanisms behind the enhanced spacer acquisition observed in immune cells that already contain a functional spacer, we decided to test if its targeting capabilities are required. Therefore we looked for phages that were able to escape *spc174* immunity. We analyzed 30 escaper plaques and, as previously described for type II CRISPR-Cas systems (Deveau et al., 2008; Pyenson et al., 2017), we found two types of mutations: within the seed region within the protospacer or within the PAM sequence immediately flanking the protospacer (**Figure 2.3A-B**).



B

frequency	174 protospacer-PAM sequence (5'-3')	mutation
0/30	TGATACATTAACATTTAGTAAATCATTACG-AGG	none
11/30	TGATACATTAACATTTAGTAAATCATTACG-AGA	PAM
06/30	TGATACATTAACATTTAGTAAATCATTACG-AGT	PAM
09/30	TGATACATTAACATTTAGTAAATCATTACG-ATG	PAM
01/30	ACTATGTAATTGTAAATCATTTAGTAATGC-AAG	PAM
02/30	TGATACATTAACATTTAGTAAATCATTGCG-AGG	seed
01/30	ACTATGTAATTGTAAATCATTTAGGAATGC-AGG	seed

Figure 2.3. Analysis of phages that evade *spc174*-mediated CRISPR immunity. (**A**) Agarose gel electrophoresis of PCR products obtained after amplification of the *tgt174* region of 30 different ΦNM4γ4 "escapers". (**B**) Summary of the *tgt174* sequences within the PCR products shown in (**A**), with the mutations in the PAM or seed sequences highlighted.

2.2.2 PAM, but not seed, target mutations abrogate spacer-mediated spacer acquisition

We isolated one mutant phage containing each mutation, ΦNM4γ4$^{174\text{-seed}}$ and ΦNM4γ4$^{174\text{-PAM}}$ (**Figure 2.4A**), and corroborated that indeed they are able to propagate in staphylococci harboring pCRISPR(*spc174*) (**Figure 2.4B**). Therefore these two phages have partially matching sequences with a pre-existing spacer but cannot be targeted. We infected cells containing pCRISPR(*spc174*) and pLR plasmids with these two mutant phages, plated the phage-resistant staphylococci and analyzed ten colonies for the presence of new spacers via PCR of the pLR plasmid. As a control, we infected cells lacking the targeting spacer, harboring pCRISPR(SR). We performed this experiment four times and consistently observed that only a few colonies acquired new spacers after infection with ΦNM4γ4$^{174\text{-PAM}}$ (**Figure 2.5A**), similar to the no-spacer control (Fig. **Figure 2.5B**). In contrast, most of the colonies surviving infection with ΦNM4γ4$^{174\text{-seed}}$ acquired new spacers (Fig. S3C), many more than the control strain (Fig. S3D). Quantification of these data showed a significant increase of the frequency of spacer acquisition from the ΦNM4γ4$^{174\text{-seed}}$ in the presence of *spc174*, but not from the ΦNM4γ4$^{174\text{-PAM}}$ phage (Fig. 2C). Finally, we amplified the expanded pLR loci in bulk and subjected the resulting PCR products to next-generation sequencing (NGS) to determine both the abundance and the distribution of the newly acquired spacers on the phage genome. Similarly to the colony analysis, deep sequencing results showed a marked increase in the proportion of ΦNM4γ4$^{174\text{-seed}}$-derived spacers compared to ΦNM4γ4$^{174\text{-PAM}}$-derived spacers (Fig. 2D). Mapping of the reads showed that, similarly to the new spacers generated from wild-type ΦNM4γ4, acquired spacers from ΦNM4γ4$^{174\text{-seed}}$ cluster around the *spc174* target (**Figure 2.4E**). Although qualitatively similar, spacer acquisition from wild-type ΦNM4γ4 and ΦNM4γ4$^{174\text{-seed}}$ are quantitatively different, with approximately a 10-fold increase in the acquisition frequency during infections with the wild-type phage. The pattern of spacer capture from the ΦNM4γ4$^{174\text{-PAM}}$ phage in the presence of *spc174*, however, is indistinguishable from the pattern of a no spacer control (**Figure 2.4F**); i.e., is not enhanced by the presence of a matching spacer.

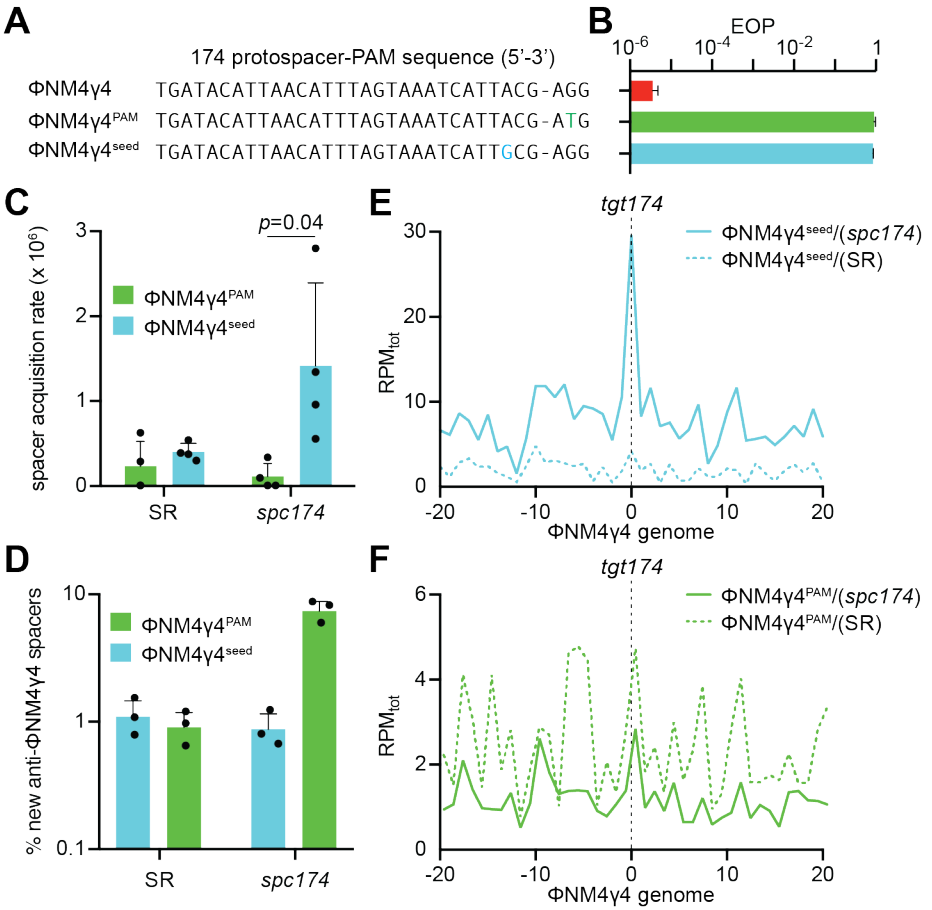


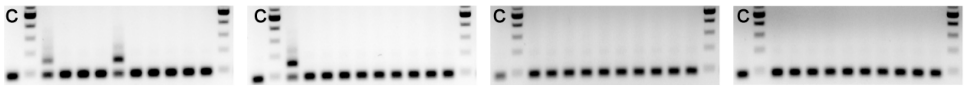
Figure 2.4. PAM, but not seed, target mutations, abrogate new spacer acquisition in immune cells.

(**A**) ΦNM4γ4 variants containing different *tgt174* sequences: wild-type, PAM (AGG>ATG) and seed (A_{-3}>G_{-3}). (**B**) Propagation ability of the three ΦNM4γ4 variants described in (**A**) on staphylococci harboring pCRISPR(*spc174*), measured as efficiency of plaquing (EOP). Mean ± StDev values of three independent experiments are shown. (**C**) Spacer acquisition rate after infection with different ΦNM4γ4 phages containing mutations in either the PAM or seed of *tgt174*, measured as the fraction of cells [harboring either pCRISPR(*spc174*) or pCRISPR(SR)] that survive infection through the acquisition of a new spacer. Mean ± StDev values of three independent experiments are shown. (**D**) Fraction (%) of spacer sequences matching to the genome of phage ΦNM4γ4 detected in the PCR product of the expanded CRISPR array of staphylococci harboring either pCRISPR(*spc174*) or pCRISPR(SR) after infection with different ΦNM4γ4 phages containing mutations in either the PAM or seed of *tgt174*. Mean ± StDev values of three independent experiments are shown. (**E**) Distribution of new spacer sequences detected in (**D**) in cells harboring pCRISPR(*spc174*), measured as spacer reads per million of total reads (RPM_{tot}), and mapped to 1 kb bins of the ΦNM4γ4 genome (shown in linear form, with *tgt174* in the center). Average curve of three independent experiments is shown. (**F**) Same as (**E**), but after analysis of bacteria harboring pCRISPR(SR).

Figure 2.5. Analysis of spacer acquisition in bacteriophage-insensitive mutant (BIM) colonies.
S. aureus RN4220 harboring different CRISPR plasmids were infected with different variants of ΦNM4γ4 phage, mixed with top agar on plates and incubated for 24 hours to isolate BIM colonies. Four independent experiments were performed for each host/phage combination. The total number of BIM colonies were counted in each plate and ten were selected to detect the acquisition of new spacers by PCR. The results of the agarose gel electrophoresis are shown. "c" indicates a lane where a PCR product corresponding to the CRISPR array uninfected cells was loaded as a negative control. The lane immediately to the right of the control lane contains molecular markers. The PCR results were used to calculate the fraction of BIM colonies that survived through the acquisition of a new spacer (CRISPR BIMs). This fraction was multiplied by the total number of BIM colonies per plate and divided by the total number of cells infected in the experiment to obtain the spacer acquisition rate. Finally, the values obtained in each experiment where used to calculate the mean and standard deviation of the spacer acquisition rate for each host/phage combination: (**A**) pCRISPR(*spc174*)ΔL/pSR + ΦNM4γ4^{PAM}, (**B**) pCRISPR(SR)ΔL/pSR + ΦNM4γ4^{PAM}, (**C**) pCRISPR(*spc174*)ΔL/pSR + ΦNM4γ4^{seed}, (**D**) pCRISPR(SR)ΔL/pSR + ΦNM4γ4^{seed}.

A

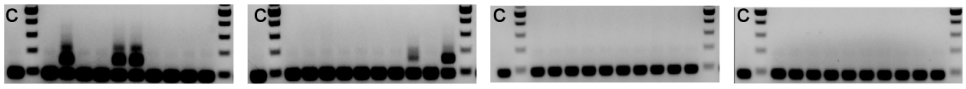
S. aureus RN4220/pCRISPR(*spc174*)ΔL/pSR + ΦNM4γ4^{PAM}



CRISPR BIMs	2/10	1/10	0/10	0/10
	x	x	x	x
total BIMs on plate	365	296	545	536
	÷	÷	÷	÷
total infected cells	2.2E+07	3.2E+07	2.0E+07	1.0E+07
	=	=	=	=
spacer acquisition rate	3.3E-06	9.3E-07	0.0E-00	0.0E-00
mean ± StDev	(1.1±1.6)E-06			

B

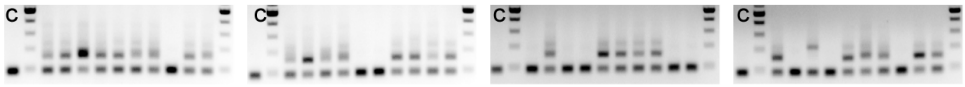
S. aureus RN4220/pCRISPR(SR)ΔL/pSR + ΦNM4γ4^{PAM}



CRISPR BIMs	3/10	2/10	0/10	0/10
	x	x	x	x
total BIMs on plate	290	255	540	571
	÷	÷	÷	÷
total infected cells	1.4E+07	1.8E+07	1.4E+07	6.0E+06
	=	=	=	=
spacer acquisition rate	6.2E-06	2.8E-06	0.0E-00	0.0E-00
mean ± StDev	(2.3±3.0)E-06			

C

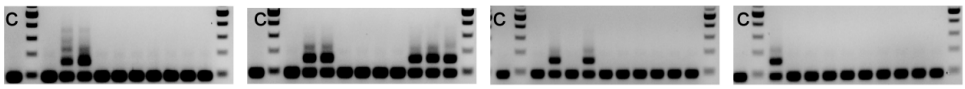
S. aureus RN4220/pCRISPR(*spc174*)ΔL/pSR + ΦNM4γ4^{seed}



CRISPR BIMs	9/10	8/10	5/10	7/10
	x	x	x	x
total BIMs on plate	327	221	381	399
	÷	÷	÷	÷
total infected cells	2.2E+07	3.2E+07	2.0E+07	1.0E+07
	=	=	=	=
spacer acquisition rate	1.3E-05	5.5E-06	9.5E-06	2.8E-05
mean ± StDev	(1.4±0.1)E-05			

D

S. aureus RN4220/pCRISPR(SR)ΔL/pSR + ΦNM4γ4^{seed}



CRISPR BIMs	2/10	5/10	2/10	1/10
	x	x	x	x
total BIMs on plate	207	193	269	223
	÷	÷	÷	÷
total infected cells	1.4E+07	1.8E+07	1.4E+07	6.0E+06
	=	=	=	=
spacer acquisition rate	3.0E-06	5.4E-06	3.8E-06	3.7E-06
mean ± StDev	(4.0±1.0)E-06			

2.2.2 Target orientation does not affect the pattern of spacer acquisition

Our findings resemble the "priming" effect of pre-existing spacers observed during the type I CRISPR-Cas immune response (Datsenko et al., 2012). Similarly to our results, the presence of a pre-existing spacer enhances the rate of acquisition of new spacers, which tend to cluster in the vicinity of the protospacer sequence (Staals et al., 2016). An outstanding feature of type I priming is the orientation bias of the acquired sequences: most of the protospacers that correspond to the new spacers are either located in the same DNA strand as the priming protospacer (defined as the strand that anneals to the crRNA) (Datsenko et al., 2012; Swarts et al., 2012) or in opposite strands at the 5′ or 3′ of the target site (Staals et al., 2016; Westra et al., 2015). Therefore we decided to check whether type II priming displays a similar orientation bias. Analysis of the new spacers acquired as a result of *spc174* priming showed a marked strand bias, both during targeting (infection with wild-type ΦNM4γ4, **Figure 2.6A**) and non-targeting (infection with ΦNM4γ4$^{174\text{-seed}}$, **Figure 2.6B**) conditions.

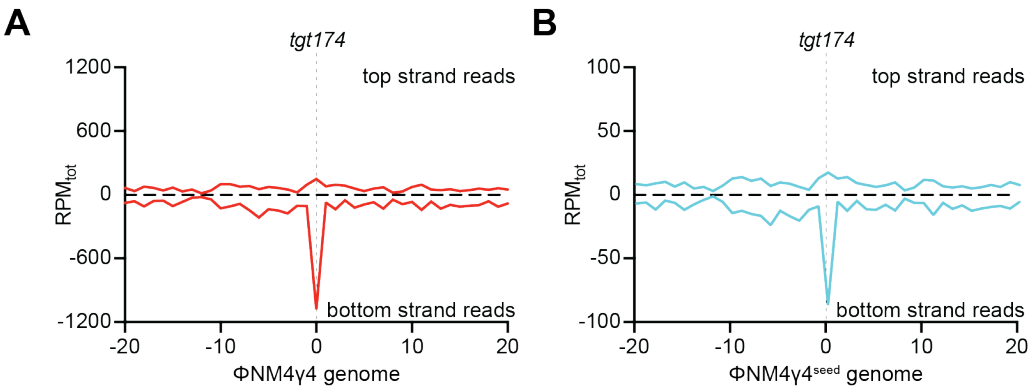


Figure 2.6. Strand bias of *spc174*-mediated spacer acquisition.
(**A**) Abundance of the spacer sequences acquired after infection of staphylococci carrying pCRISPR(*spc174*) with ΦNM4γ4, measured as spacer reads per million of total reads (RPMtot), mapped to 1 kb bins of either the top or bottom strands of the phage genome (shown in linear form, with *tgt174* in the center). Average curve of three independent experiments is shown. (**B**) Same as (**A**) but after infection with ΦNM4γ4^{seed}.

If there is orientation bias as in type I priming, relocating the target of *spc174* on the opposite strand would reverse the observed strand bias. We unsuccessfully attempted to introduce the *spc174* protospacer in the opposite strand of ΦNM4γ4, most likely because the resulting phages are not viable (data not shown). Therefore we decided to investigate the orientation bias during priming by *spc256* (**Figure 2.1E**), which targets a non-essential region of ΦNM4γ4 that can be reversed without loss of viral titers (**Figure 2.7A-B**). This spacer targets a protospacer located in the bottom strand of the ΦNM4γ4 genome. Similarly to

spc174, *spc256* priming showed a strong strand bias, with most new spacers acquired having matching protospacers in the bottom strand (**Figure 2.7C**). When the target was relocated to the top strand in the mutant phage ΦNM4γ4^{256rev}, however, the newly acquired spacers still matched bottom-strand protospacers (**Figure 2.7D**). Therefore, as opposed to priming in type I CRISPR-Cas systems, type II priming lacks an orientation bias.

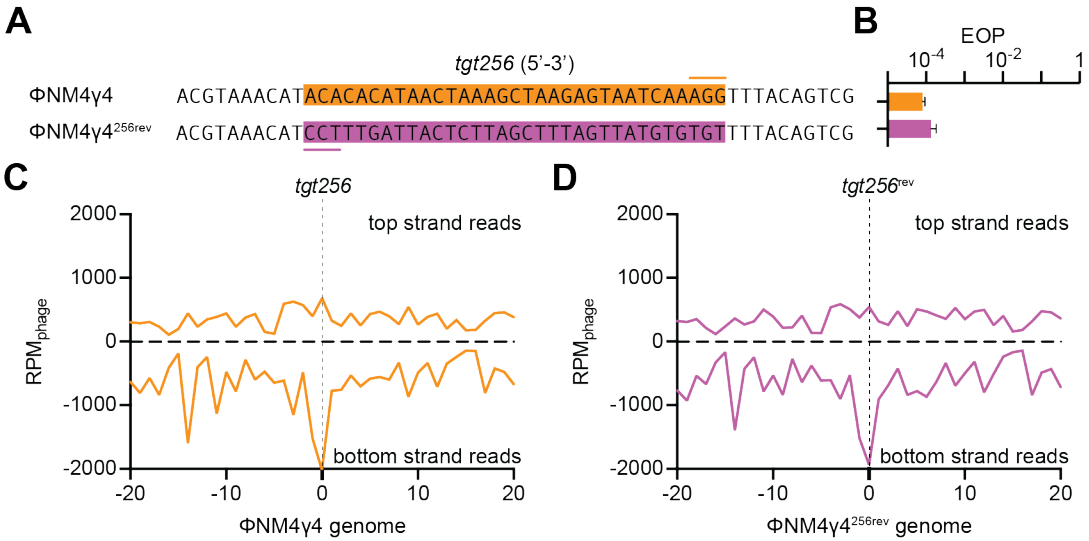


Figure 2.7. Target orientation does not affect the pattern of spacer acquisition. (**A**) The sequence of *tgt256* on the ΦNM4γ4 genome (orange box, the line marks the PAM) was reversed in the ΦNM4γ4^{256rev} mutant virus to allow annealing of the crRNA generated by *spc256* to the other strand of the phage genome. (**B**) Comparison of the efficiency of plaquing (EOP) of ΦNM4γ4 on staphylococci carrying pCRISPR(*spc256*) of both phages shown in (**A**). Mean ± StDev values of three independent experiments are shown. (**C**) Distribution of new spacer sequences acquired after infection of staphylococci carrying pCRISPR(*spc256*) with ΦNM4γ4, measured as spacer reads per million of total phage reads (RPM_{phage}), and mapped to 1 kb bins of either the top or bottom strands of the phage genome (shown in linear form, with *tgt256* in the center). Average curve of three independent experiments is shown. (**D**) Same as (**C**) but after infection with ΦNM4γ4^{256rev}.

2.3 Target DNA cleavage is required for type II primed spacer acquisition

The distribution of the new spacers acquired through type II priming, narrowly centered at the target site of the priming spacer, suggests that Cas9 binding and/or cleavage of the protospacer are important for the process. To investigate this, we first tested the ability of *spc174* to direct the cleavage of the different targets found in ΦNM4γ4, ΦNM4γ4$^{174\text{-seed}}$ and ΦNM4γ4$^{174\text{-PAM}}$ phages. Biochemical characterization of *S. pyogenes* Cas9 has shown that binding of the PAM nucleotides is the fundamental first step in target recognition, which then proceeds to the critical pairing of the protospacer DNA and crRNA seed

sequences (Sternberg et al., 2014). As a result of this mechanism, PAM mutations that prevent its recognition completely abrogate target cleavage, whereas mismatches within the seed region can lead to low levels of nuclease activity. Indeed, when we tested the cleavage of the different *spc174* targets *in vitro* we found strong cleavage of the wild-type target, lower levels of cleavage of the 174-seed target, and no Cas9 nuclease activity against the 174-PAM target DNA (**Figure 2.8A-B**).

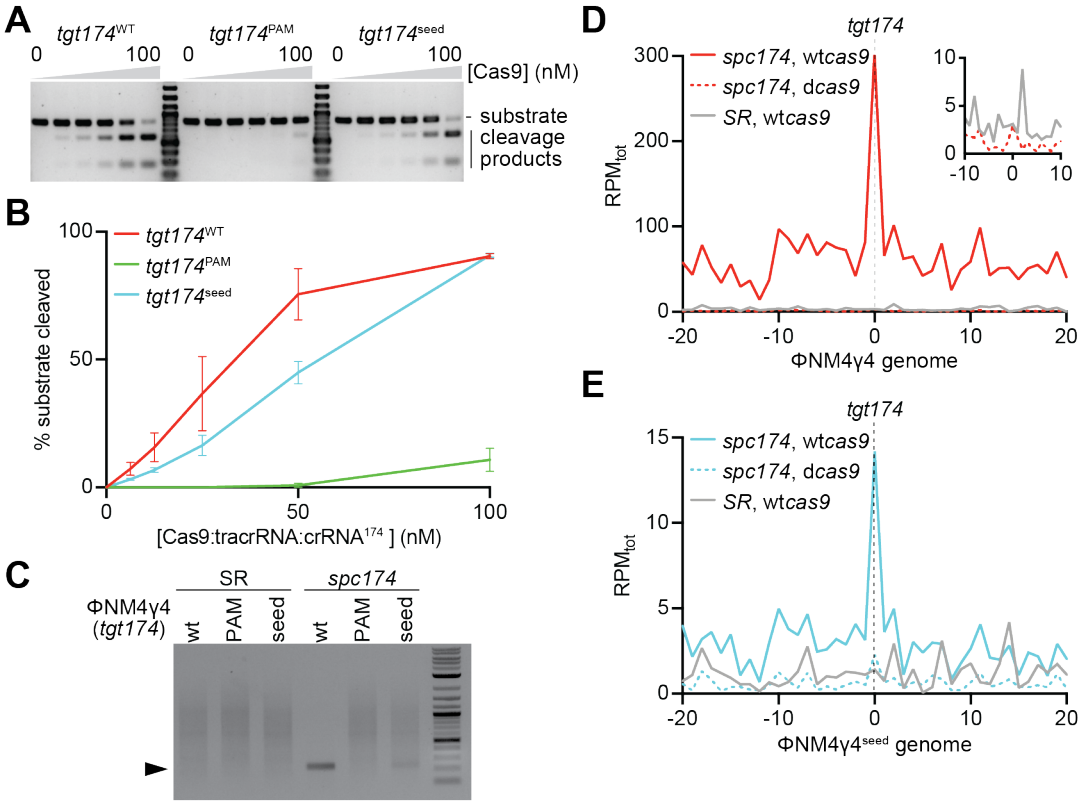


Figure 2.8. Target DNA cleavage is required for type II primed spacer acquisition.

(**A**) In vitro cleavage assay of a dsDNA oligonucleotide containing the different *tgt174* sequences shown in Fig. 2A, incubated with increasing concentrations of a 1:1:1 mix of Cas9:tracrRNA:crRNA^{174}: 0, 6.25, 12.5, 25, 50 and 100 nM. Substrates and cleavage products were separated by agarose gel electrophoresis. A representative image of three replicates is shown. (**B**) The bands corresponding to the cleavage products in (**A**) were quantified and plotted against the concentration of the Cas9 ribonucleoprotein complex. Mean ± StDev values of three independent experiments are shown. (**C**) In vivo cleavage of phage DNA. Cells containing the pCRISPR(*spc174*) or pCRISPR(SR) plasmids where infected with ΦNM4γ4, ΦNM4γ4$^{174\text{-seed}}$ or ΦNM4γ4$^{174\text{-PAM}}$ phages and the cleavage products (marked by the black triangle) of the *tgt174* sequence were amplified and separated by agarose gel electrophoresis. (**D**) Distribution of new spacer sequences acquired after infection of staphylococci carrying either pCRISPR(*spc174*, wt*cas9*), pCRISPR(*spc174*, d*cas9*) or pCRISPR(SR), with ΦNM4γ4, measured as spacer reads per million of total reads (RPM_{tot}), and mapped to 1 kb bins of either the top or bottom strands of the phage genome (shown in linear form, with *tgt174* in the center). Average curve of two independent experiments is shown. The inset shows the data from cells harboring pCRISPR(*spc174*, d*cas9*) or pCRISPR(SR) around the target site, with a different RPM_{tot} scale. (**E**) Same as (**D**) but after infection with ΦNM4γ4^{seed}.

We also tested cleavage of phage DNA in vivo during infection. Cells containing the pCRISPR(*spc174*) or pCRISPR(SR) plasmids where infected with ΦNM4γ4, ΦNM4γ4$^{174\text{-seed}}$ or ΦNM4γ4$^{174\text{-PAM}}$ phages, total DNA extracted after 20 minutes and treated with terminal deoxynucleotidyl transferase (TdT) and deoxycytosine to add poly-dC extensions to the 3' ends of DNA breaks (**Figure 2.9**). The modified DNA was used as template for amplification with a polyG primer and a second primer annealing upstream of the *spc174* target sequence to detect Cas9 cleavage of the phage genome as a PCR product.

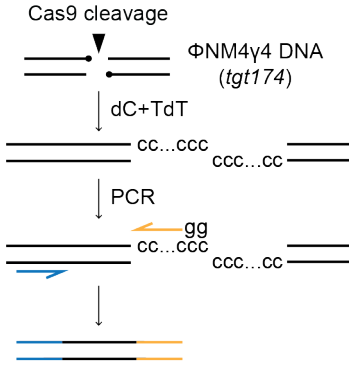


Figure 2.9. Detection of Cas9 cleavage of the phage genome *in vivo*
A schematic depicting assay for detecting cleavage of viral genomes *in vivo*. In brief, total DNA was extracted 30 minutes after infection and treated with terminal deoxynucleotidyl transferase (TdT) and deoxycytosine (dC) to add poly-dC extensions to the 3' ends of DNA breaks (black dot). The modified DNA was used as template for amplification with a polyG primer (orange arrow) and a second specific primer annealing upstream of the *spc174* target sequence (blue arrow) to detect Cas9 cleavage sites on the phage genome as a PCR product.

Whereas no amplification was detected in the non-targeting control samples, strong and weak products were observed for targeting cells infected with wild-type ΦNM4γ4 and ΦNM4γ4$^{174\text{-seed}}$ phages, respectively (**Figure 2.8C**). No PCR product was detected for targeting cells infected with mutant phages lacking a functional PAM in the *spc174* target. Cloning and sequencing of these PCR products confirmed the specificity of the cleavage (data not shown). Therefore both in vitro and in vivo experiments, as expected, showed a strong cleavage of the wild-type *spc174* target, partial cleavage of the target with the seed mutation, and no cleavage of the target with the non-functional PAM. These different levels of Cas9 cleavage are correlated with the level of priming mediated by *spc174*, as spacer acquisition was highest during infection with wild-type ΦNM4γ4, intermediate with ΦNM4γ4$^{174\text{-seed}}$ and very low with ΦNM4γ4$^{174\text{-PAM}}$ (**Figure 2.1C** and **Figure 2.4D**). As mentioned above, this could be due to a requirement for binding and/or cleavage of the target site. To determine if binding alone can induce type II primed spacer acquisition, we performed short-term infection experiments in staphylococci expressing catalytically dead Cas9 (dCas9), which can bind but not cleave the target specified by the crRNA guide (Bikard et al., 2013; Qi et al., 2013), and has been previously shown to be capable of

participating in naïve spacer acquisition (Heler et al., 2015; Wei et al., 2015b). Next-generation sequencing analysis of the acquired spacers in the presence of *spc174* targeting showed the reduction of the peak located in the target site to the same background levels observed in the absence of targeting, during infection with both wild-type ΦNM4γ4 (**Figure 2.8D**) and ΦNM4γ4$^{174\text{-seed}}$ (**Figure 2.8E**) phages. These data demonstrate that target cleavage is fundamental for type II CRISPR-Cas primed spacer acquisition.

2.4 Cleavage-mediated spacer acquisition in Streptococcus thermophilus

In order to confirm our results in a natural experimental system, i.e. as opposed to the heterologous plasmid-based system described above, we explored cleavage-mediated spacer acquisition in *Streptococcus thermophilus*, the first organism shown to employ CRISPR-Cas systems for anti-phage defense (Barrangou et al., 2007). Strain DGCC7710 harbors two type II-A CRISPR loci (CRISPR1 and CRISPR3) that function independently from one another (Carte et al., 2014) and both are capable of spacer acquisition (Barrangou et al., 2007; Wei et al., 2015a). To eliminate the possibility of competition between these two CRISPR loci for the same substrates during spacer acquisition that might complicate interpretation of our results, we used a mutant containing only CRISPR3 (**Figure 2.10A**), strain CRISPR3-naïve (Varble et al., 2019), since this locus encodes the closest relative to the *S. pyogenes* SF370 type II-A system used in our previous experiments (Fonfara et al., 2014). We infected CRISPR3-naïve cells with the phage Φ2972 (**Figure 2.10B**) in soft agar to generate 'bacteriophage-insenstive mutants' (BIMs) colonies as previously described (Hynes et al., 2017). The resulting BIMs were screened for spacer acquisition by PCR and three of these colonies containing a single, unique spacer targeting Φ2972 were saved for subsequent experiments (CRISPR3α, CRISPR3β, and CRISPR3γ, **Figure 2.10C**). First we confirmed the targeting, and thus cleavage, properties of each spacer (*spc-α*, *spc-β*, and *spc-γ*, **Figure 2.10B**) by measuring the ability of Φ2972 to form plaques in lawns of CRISPR-adapted streptococci (**Fig. 2.10D**). We then measured cleavage-mediated spacer acquisition after infection of liquid cultures with Φ2972 at a MOI of 10 for thirty minutes. Cells were collected for genomic DNA extraction, amplification of the CRISPR3 locus and NGS of the obtained PCR product. When analyzed quantitatively, sequencing data showed a markedly higher number of new anti-viral spacers for the three targeting strains relative to the CRISPR3-naïve control (~ 80-fold increase on average, **Figure 2.10D**). Finally, qualitative analysis of the distribution pattern of the new spacers revealed a clustering around their respective Cas9 target sites (**Figure 2.10F**). Altogether, these results demonstrate the existence of cleavage-mediated spacer acquisition in the CRISPR3 locus of *S. thermophilus*, and also show that it shares similar mechanistic features with the *S. pyogenes* SF370 heterologous system.

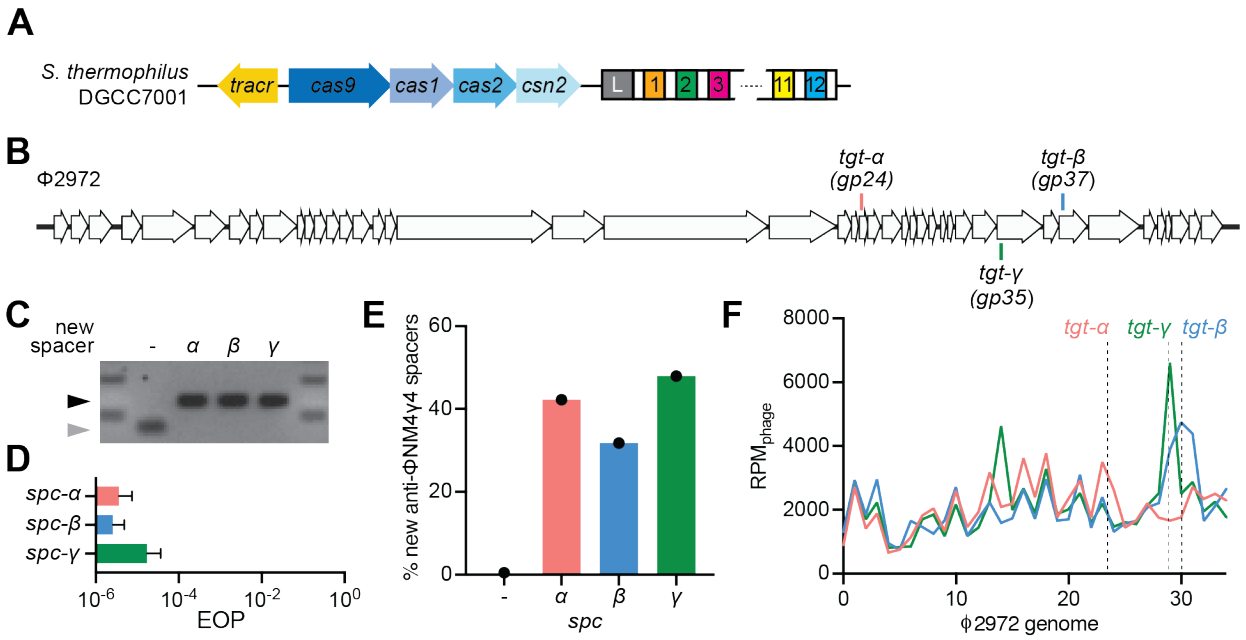


Figure 2.10. Cleavage-dependent spacer acquisition in *Streptococcus thermophilus*.
(**A**) Type II-A CRISPR-*cas* locus of *S. thermophilus* DGCC7710. White rectangles, CRISPR repeats; colored and numbered rectangles, spacers; "L", leader sequence; blue arrows, protein-coding genes; yellow arrow, tracrRNA gene. (**B**) Streptococcal phage Φ2972 genome and the Cas9 targets analyzed in this study. (**C**) Agarose gel electrophoresis of PCR products after amplification of the CRISPR3 array from *S. thermophilus* CRISPR3-naïve that acquired spacers α, β and γ upon infection with Φ2972. Grey and black arrows: non-expanded and expanded, respectively, CRISPR3 arrays. (**D**) Comparison of the efficiency of plaquing (EOP) of phage Φ2972 on CRISPR3-α, -β or -γ streptococci. Mean ± StDev values of three independent experiments are shown. (**E**) Fraction (%) of new spacer sequences matching to the genome of phage Φ2972 detected after infection of CRISPR3-α, -β or -γ streptococci. (-) indicates new spacers acquired in the absence of a targeting spacer by CRISPR-naïve cells. Values for a single experiment are shown. (**F**) Distribution of new spacer sequences detected 30 minutes after infection of streptococci expressing Cas9 programmed to cleave the targets shown in (**B**). Reads per million of phage reads (RPM_{phage}) are mapped to 1 kb bins of the Φ2972 genome. Values for a single experiment are shown.

2.5 Discussion

Our results demonstrate that the presence of pre-existing spacers enhances further spacer acquisition during the type II-A CRISPR-Cas immune response through a mechanism that requires the cleavage of the target DNA and that results in the capture of sequences in the vicinity of the cut site. This phenomenon was observed for both a heterologous *S. pyogenes* SF370 and the native CRISPR3 locus of *S. thermophilus* DGCC7710, suggesting that it is a common feature of type 2 CRISPR systems. As a result of this mechanism, the abundance of the newly acquired spacers is determined by the efficiency of cleavage mediated by the preexisting priming spacer. Spacers with fully functional targets lead to the highest level of spacer acquisition, spacers that target a protospacer with seed mutations promote an intermediate level of priming and those that target a site with a non-functional PAM do not enhance spacer acquisition. We believe that the strict correlation between target cleavage and primed spacer acquisition is a result of the mechanism behind type II-A spacer capture, which displays a marked preference for sequences near free double-stranded DNA (dsDNA) ends (Modell et al., 2017). In this model, the dsDNA ends generated by Cas9 cleavage become substrates for spacer acquisition. The previously reported physical association between Cas9, Csn2 and the Cas1-Cas2 integrase of the type II-A systems (Heler et al., 2015; Wilkinson et al., 2019), could provide an efficient avenue for the recruitment of the spacer acquisition machinery to the free dsDNA ends generated by Cas9. A previous study showed that when the dsDNA end is either the *cos* site injected by a phage or the product of DNA restriction, processing by the AddAB nuclease extends the region of spacer acquisition to the first *chi* site capable of inhibit AddAB (Modell et al., 2017), presumably by generating degradation products with new dsDNA ends for the spacer acquisition machinery. In AddAB mutants, however, the region of acquisition is narrowed and centered around the dsDNA end, lacking the expansion to the *chi* sites. Interestingly, our data shows an acquisition hotspot concentrated at the cleavage site, not limited by *chi* sites, similar to the spacer acquisition pattern in AddAB mutants (Modell et al., 2017). This suggests that the dsDNA ends produced by Cas9 cannot be processed by AddAB, due to either the tight binding of Cas9 to its cleavage products (Garneau et al., 2010; Sternberg et al., 2014) or the expression of an AddAB inhibitor by ΦNM4γ4 (Bobay et al., 2013). Alternatively, as a result of the close association between the acquisition machinery and Cas9, the Csn2/Cas1-Cas2 complex could have preferential access to the dsDNA ends generated after cleavage and prevent the loading of AddAB to the ends.

Although conceptually similar, our results show important differences in the mechanism of primed spacer acquisition carried out by type I and type II CRISPR-Cas systems. In these systems, the crRNA-guided Cascade complex requires both seed and PAM sequences to recognize a protospacer sequence and recruit the Cas3 nuclease to destroy the target genome by nicking one DNA strand at different length intervals (Dillard et al., 2018; Loeff et al., 2018; Mulepati and Bailey, 2013; Sinkunas et al., 2011). The lack of robust priming against perfect targets has been explained by two hypotheses in the context of this targeting

mechanism. Some studies support the idea that, upon interaction with a mutated, but not with a perfect, protospacer, the Cascade complex adopts a "priming" conformation that enables the recruitment of Cas1/Cas2 to initiate spacer acquisition from the target genome (Blosser et al., 2015; Redding et al., 2015). Other work has shown that target cleavage is required for type I priming, but perfect protospacers lead to the rapid degradation of the target genome and prevent Cas1/Cas2 from extracting additional spacers from it (Musharova et al., 2019; Semenova et al., 2016). Indeed, only when an anti-CRISPR Cascade inhibitor is over-expressed in hosts to impair target destruction, high levels of "interference-driven acquisition" can be detected (Staals et al., 2016). Mutated protospacers, on the other hand, lead to the recruitment by Cascade of both Cas3 and Cas1/Cas2, which promote slow degradation of the target genome and primed spacer acquisition, respectively (Musharova et al., 2019; Semenova et al., 2016; Staals et al., 2016). Moreover, in type I-F systems that encode a Cas2-Cas3 fusion protein, Cas1 recruitment to the mutated target site, an essential step for primed acquisition, inhibits the nuclease activity (within Cas3) of the fusion (Rollins et al., 2017), supporting the idea that slow target degradation is important for primed spacer acquisition. In contrast to the results reported for type I priming, where acquisition mediated by targeted protospacers can only be detected when target degradation is artificially decreased, we observed the highest levels of type II primed spacer acquisition in the presence of a perfect protospacer that can be rapidly cleaved. We believe that this difference could be related to the lack of processive degradation after target cleavage by Cas9 compared with Cas3 (Sternberg et al., 2014); i.e. the rapid degradation of DNA containing a perfect target for Cas3 prevents the acquisition of new spacers from sequences that surround the target site. Alternatively, even if the Cas1/Cas2 complex interacts with both the Cas3 and Cas9 nucleases, the integration substrates (also known as pre-spacers) could be more immediately accessible to the integrase complex in the vicinity of the target site after Cas9 cleavage versus Cas3 nicking.

Another difference is the influence of the protospacer location in the orientation of the new spacers. Whereas during type I priming the acquired spacers match the same strand as the priming protospacer (Datsenko et al., 2012; Staals et al., 2016; Swarts et al., 2012), this is not the case for the spacers acquired through type II primed spacer acquisition. This dissimilitude could be related to the disparities in DNA cleavage between Cas3 and Cas9. The recognition of a mutated protospacer by Cascade leads to the recruitment of Cas3 and the Cas1/Cas2 integrase to the target DNA. This priming complex moves unidirectionally along one of the DNA strands (Dillard et al., 2018; Redding et al., 2015), a process that could cause the observed polarity of spacer acquisition during type I priming. In contrast, Cas9 remains bound to the cleaved DNA ends (Sternberg et al., 2014); i.e. the lack of strand-specific or directional movement of this nuclease prevents any influence on the extraction of spacers by Cas1/Cas2.

Finally, the wide use of Cas9 as a tool for genome engineering of the human genome (Doudna and Charpentier, 2014) have revealed the existence of "off-target" sites. These are not simply mutated targets, but sequences with random and imperfect complementarity to the crRNA guide of Cas9 that can nevertheless be recognized and cleaved by the nuclease at low rates (Fu et al.,

2013; Hsu et al., 2013). Limiting Cas9 off-target effects represents a fundamental obstacle for the continued improvement of CRISPR-Cas9 biotechnology. According to our results, these off-target cleavage events should lead to primed spacer acquisition, and therefore it is interesting to speculate that this "defect" of Cas9 targeting has been selected through evolution to boost the otherwise naïve type II CRISPR-Cas immune response and to achieve immunization against phages for which a priori there are no partially matching pre-existing spacers.

CHAPTER 3: TYPE II PRIMED SPACER ACQUISTION MEDIATES A ROBUTS ANTI-VIRAL IMMUNE RESPONSE

3.1 Introduction

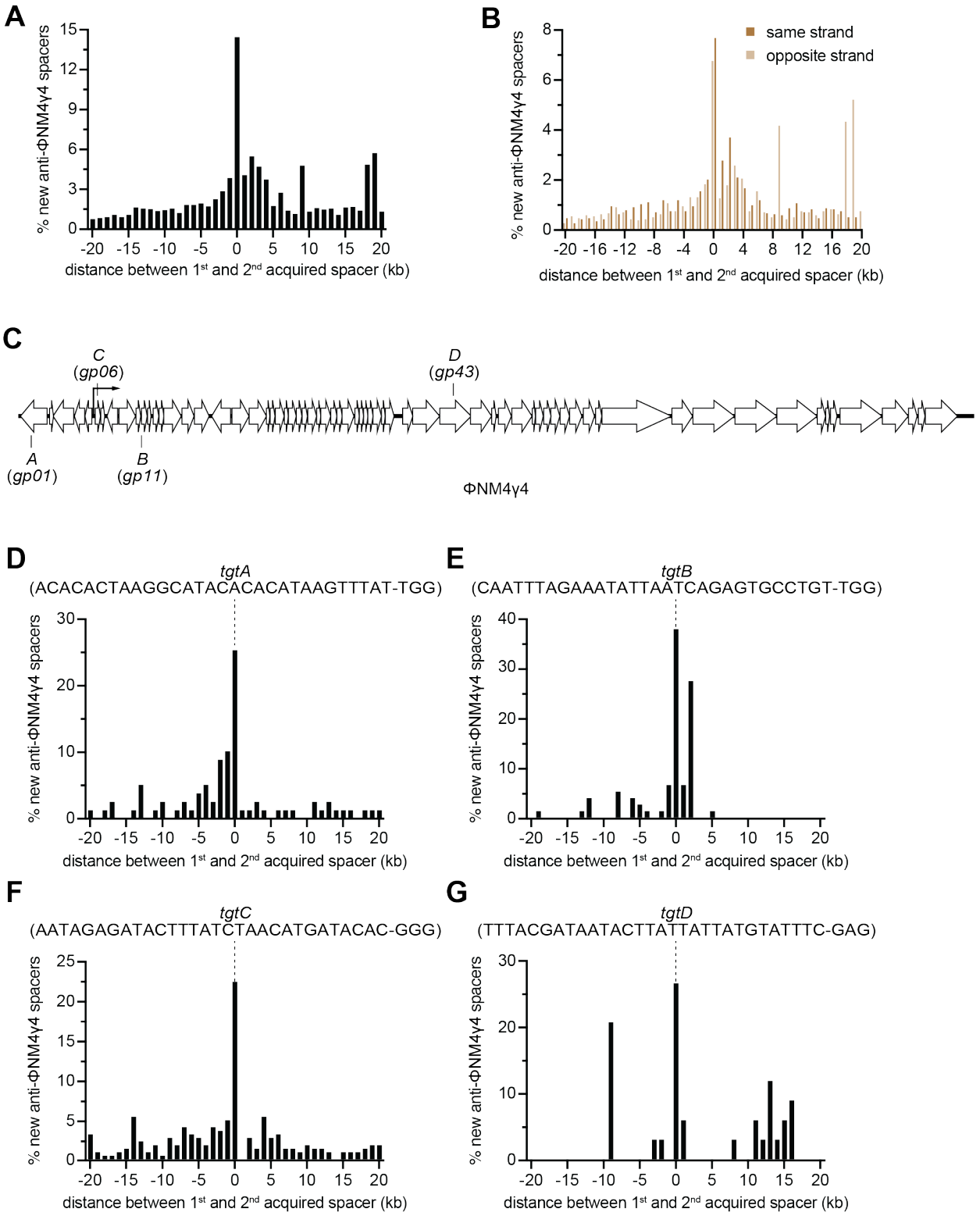
A hallmark of CRISPR immunity is the generation of a heterogenous bacterial population with thousands of different spacer sequences, each of them in a different cell (Barrangou et al., 2007; Heler et al., 2015; Paez-Espino et al., 2013). This prevents the rise of phage escapers containing mutations that can avoid the immunity mediated by one of the acquired spacers (van Houte et al., 2016). These phages can kill the cells in the population harboring that particular spacer but will be eliminated upon infection of other host cells that contain different spacers for which they do not have escape mutations. In addition, the expansion of the spacer repertoire defends the host population from infection with related phages that share target sequences with the previous phage that triggered spacer acquisition. Therefore, the enhancement of spacer acquisition promoted by Cas9 cleavage should mediate a more robust type II-A CRISPR-Cas immune response. On one hand, the cleavage of targets with seed escape mutations would allow the host to *contain* the rise of such escapers after they appear in the population. On the other, the acquisition of additional spacers that occurs after cleavage of perfect targets within immune hosts would enable to *anticipate* the occurrence of phage (seed and PAM) escapers or related phages; i.e. it would equip the host with a different spacer sequence for defense against phages that cannot be cleaved by the first spacer.

3.2 Primed spacer acquisition occurs during long-term phage infections

To investigate these scenarios, we first determined whether primed spacer acquisition occurs during infection of staphylococci harboring a naïve pCRISPR, without spacers matching ΦNM4γ4; i.e whether cleavage-mediated spacer acquisition can happen after naïve cells capture their first spacers. We infected naïve cells with ΦNM4γ4 at a multiplicity of infection (MOI) of 1 and allowed the cultures to grow in the presence of phage for 24 hours before amplifying the CRISPR arrays and submitting the resulting PCR products to next-generation sequencing. We then selected reads that contained two new anti-viral spacers and calculated the genomic distance between the first and second acquired sequences (located at the 5′ end of the array (Barrangou et al., 2007; McGinn and Marraffini, 2016)). Consistent with a previous bioinformatic analysis (Nicholson et al., 2018), the histogram of these distances revealed that the majority of the second spacers map within 1 kb of the protospacer targeted by the first spacer (**Figure 3.1A** and **Figure 3.1C-G**), without any noticeable strand bias (**Figure 3.1B**). This spacer distribution is similar to that obtained in the *spc174* priming experiments **(Figure 2-1C** and **Figure 2.4E**), and thus the data strongly suggests primed spacer acquisition follows the naïve type II-A CRISPR-Cas immune response.

Figure 3.1. Cleavage-dependent spacer acquisition occurs during long-term infection

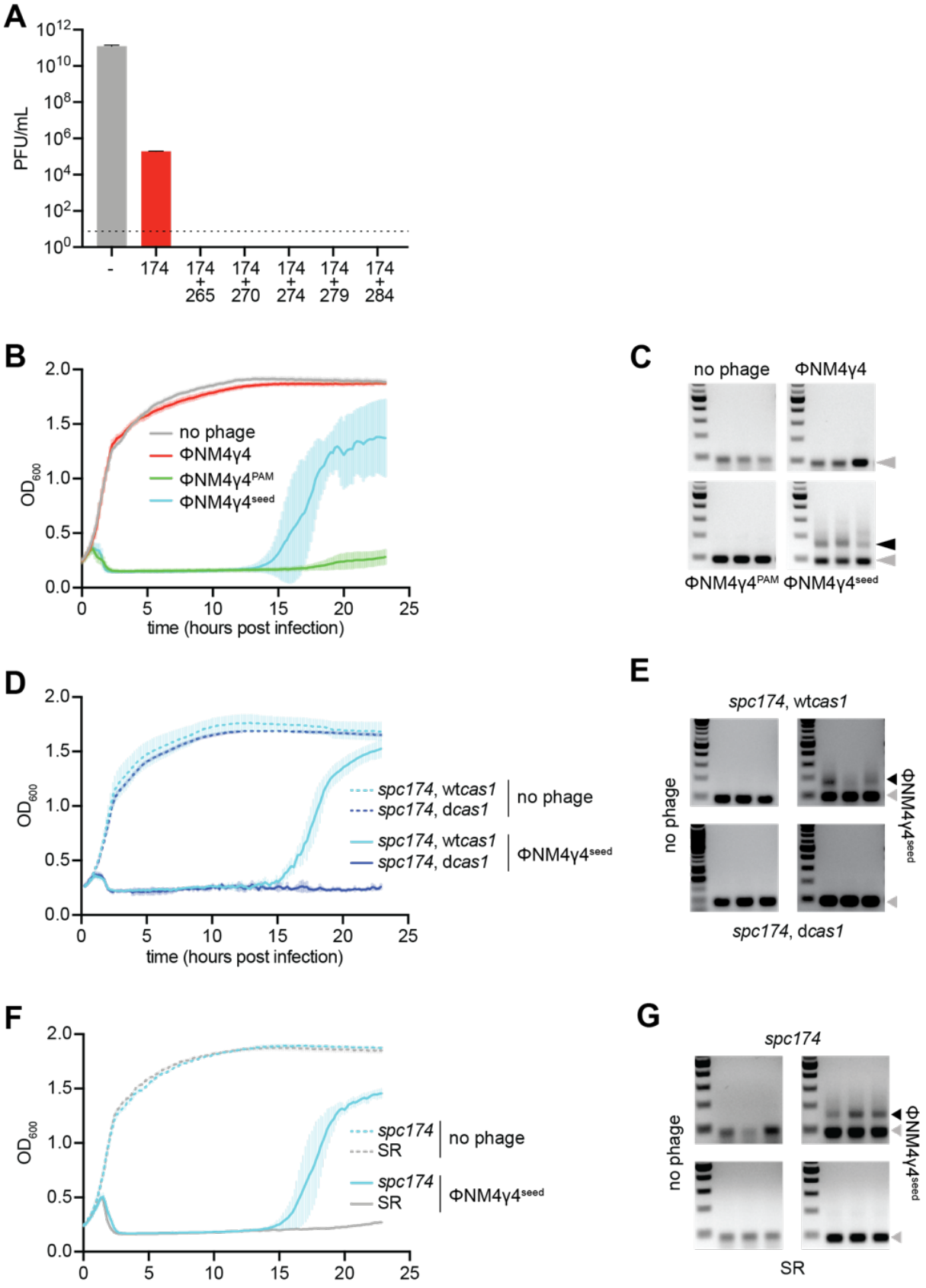
(**A**) Distance between the targets in the ΦNM4γ4 genome specified by the first and second spacers acquired after 24-hour infection of naïve staphylococci carrying the type II-A CRISPR-Cas system of *S. pyogenes* with ΦNM4γ4. The number of different spacers within 1-kb bins of the ΦNM4γ4 genome are shown; the position of first spacer acquired in each array is set as 0 kb. (**B**) Same as (**A)** but analyzed according to whether spacer targets the same or the opposite DNA strand as the first acquired spacer. (**C**) Location within the ΦNM4γ4 genome of the targets specified by the four first spacers analyzed in (*tgtA, tgtB, tgtC,* and *tgtD*) (**D-E**) Distribution of distances between the targets specified by the second spacers integrated after the acquisition of the spacers A, B, C and D, respectively; the target sequence specified by the first spacer acquired, which is given a 0 kb position, is shown.


A
% new anti-ΦNM4γ4 spacers
distance between 1st and 2nd acquired spacer (kb)
B
same strand
opposite strand
% new anti-ΦNM4γ4 spacers
distance between 1st and 2nd acquired spacer (kb)
C
C (gp06)
D (gp43)
A (gp01)
B (gp11)
ΦNM4γ4
D
tgtA
(ACACACTAAGGCATACACACATAAGTTTAT-TGG)
% new anti-ΦNM4γ4 spacers
distance between 1st and 2nd acquired spacer (kb)
E
tgtB
(CAATTTAGAAATATTAATCAGAGTGCCTGT-TGG)
% new anti-ΦNM4γ4 spacers
distance between 1st and 2nd acquired spacer (kb)
F
tgtC
(AATAGAGATACTTTATCTAACATGATACAC-GGG)
% new anti-ΦNM4γ4 spacers
distance between 1st and 2nd acquired spacer (kb)
G
tgtD
(TTTACGATAATACTTATTATTATGTATTTC-GAG)
% new anti-ΦNM4γ4 spacers
distance between 1st and 2nd acquired spacer (kb)

3.3 Primed spacer acquisition limits the propagation of existing escaper viruses

To determine if primed spacer acquisition can limit the propagation of phage escapers, we selected colonies that acquired a second spacer after priming with *spc174* (**Figure 3.2C**) and infected them with wild-type ΦNM4γ4 phage. Whereas the phage population contained a significant number of mutants that were able to propagate in staphylococci equipped with pCRISPR(*spc174*), we were unable to detect plaques on cultures harboring an additional spacer acquired through priming (**Figure 3.2A**). We then tested our first prediction; i.e. that Cas9 partial cleavage would promote the acquisition of new spacers to neutralize seed escaper phages. We infected exponentially-growing staphylococci harboring pCRISPR(*spc174*) with ΦNM4γ4, ΦNM4γ4$^{174\text{-seed}}$ or ΦNM4γ4$^{174\text{-PAM}}$ phages (or a mock infection as a control) at a MOI of 10 and we measured the optical density of the culture to monitor cell survival (**Figure 3.2B**). Due to *spc174* targeting, cultures infected with wild-type ΦNM4γ4 continue growing, similarly to the control where no phage was added. In contrast, cultures infected with both escaper phages succumbed to infection and stopped growing, and the cells that received ΦNM4γ4$^{174\text{-PAM}}$ were not able to recover. Cells infected with ΦNM4γ4$^{174\text{-seed}}$, however, regained growth at 13 hours post-infection. PCR analysis of the three replicas showed that in all cases a second new spacer was acquired (**Figure 3.2C**). Spacer acquisition was absolutely required for the recovery of the culture, as cells expressing the catalytically dead version of Cas1 (dCas1) were incapable of recovery (**Figure 3.2D-E**). In addition, the spacer acquisition that enabled the survival was primed by *spc174*, since cells harboring pCRISPR(SR) failed to regrow (**Figure 3.2F-G**). Altogether, the results presented in **Figure 3.2** demonstrate that primed spacer acquisition occurs during the type II-A CRISPR-Cas immune response, enabling the containment of seed escaper phages.

Figure 3.2: Cleavage-dependent spacer acquisition mediates a robust immune response against escaper viruses
(**A**) Quantification of phage escapers as PFU/ml after the plating of different dilutions of phage ΦNM4γ4 stock onto plates seeded with different staphylococcal strains that harbor pCRISPR either lacking a targeting spacer (-) or programmed with *spc174* or *spc174* and additional spacer acquired in the experiment of Fig. S3C. Mean ± StDev values of three independent experiments are shown. (**B**) Cell survival measured as OD_{600} after infection of cultures carrying pCRISPR(*spc174*) with different phages at MOI 10 or none as a control. The average curves of three different replicates are shown, with +/- StDev values shown in lighter colors. (**C**) Agarose gel electrophoresis of PCR products after amplification of the CRISPR array of cells obtained after the experiment in (**B**) to detect the integration of new spacers. Grey and black arrows: non-expanded and expanded, respectively, CRISPR arrays. (**D**) Cell survival measured as OD_{600} after infection of cultures carrying either the wild-type or the nuclease deficient *cas1* gene (wt*cas1* or d*cas1*, respectively) on pCRISPR(*spc174*) with ΦNM4γ4 at MOI 10 or no phage as a control. The average curves of three different replicates are shown, with +/- StDev values shown in lighter colors. (**E**) Agarose gel electrophoresis of PCR products after amplification of the CRISPR array of cells obtained after the experiment in (**D**) to detect the integration of new spacers. Grey and black arrows: non-expanded and expanded, respectively, CRISPR arrays. (**F**) Same as (**D**) but after infection of staphylococci carrying either pCRISPR(SR) or pCRISPR(*spc174*). The average curves of three different replicates are shown, with +/- StDev values shown in lighter colors. (**G**) Same as (**E**) but for the cells obtained after the experiment in (**F**).

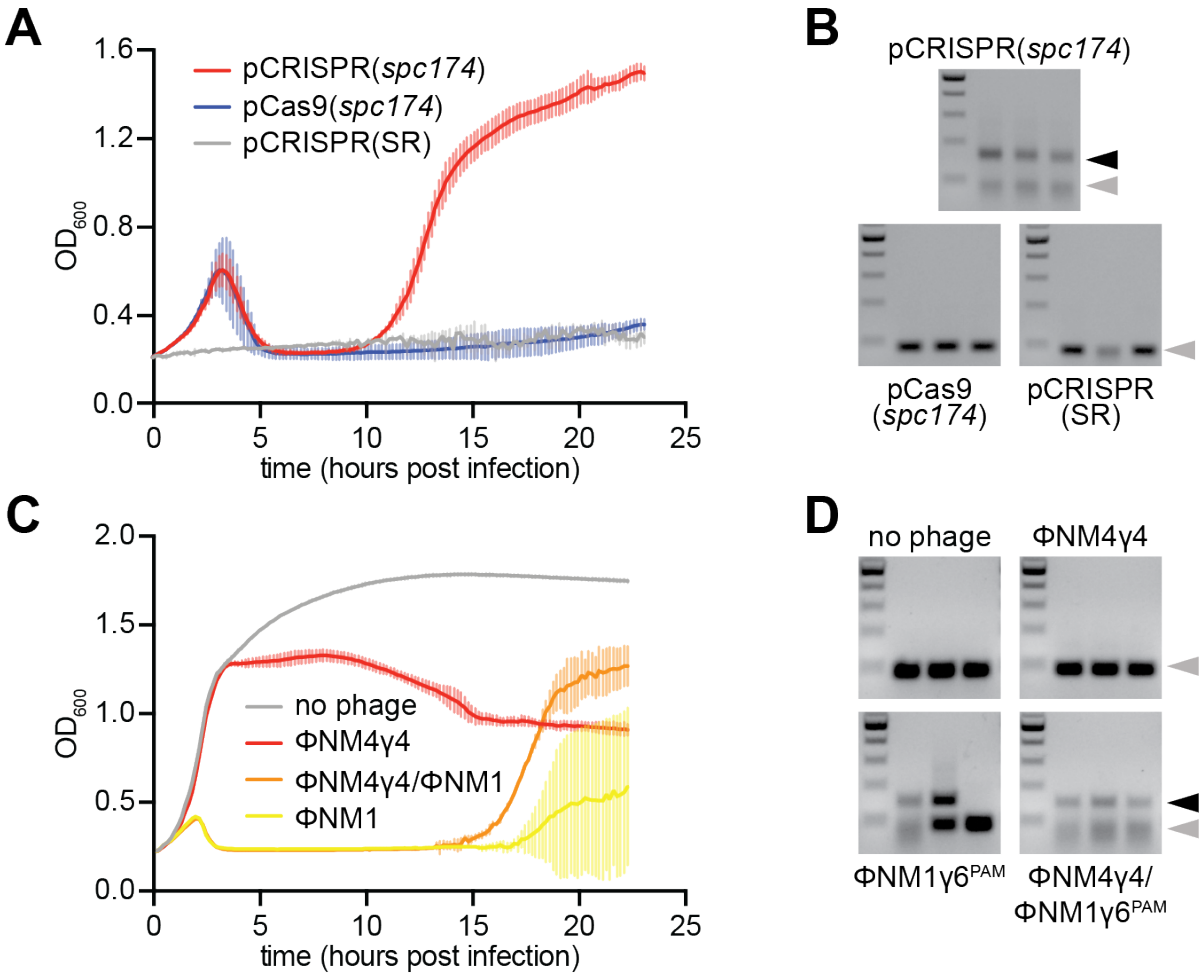

A
PFU/mL
10^{12}
10^{10}
10^{8}
10^{6}
10^{4}
10^{2}
10^{0}
-
174
174 + 265
174 + 270
174 + 274
174 + 279
174 + 284
B
OD_{600}
no phage
ΦNM4γ4
ΦNM4γ4^{PAM}
ΦNM4γ4^{seed}
time (hours post infection)
C
no phage
ΦNM4γ4
ΦNM4γ4^{PAM}
ΦNM4γ4^{seed}
D
OD_{600}
spc174, wtcas1
spc174, dcas1
no phage
spc174, wtcas1
spc174, dcas1
ΦNM4γ4^{seed}
time (hours post infection)
E
spc174, wtcas1
no phage
ΦNM4γ4^{seed}
spc174, dcas1
F
OD_{600}
spc174
SR
no phage
spc174
SR
ΦNM4γ4^{seed}
G
spc174
no phage
ΦNM4γ4^{seed}
SR

3.4. Primed spacer acquisition anticipates infection by emergent viruses

Finally, we tested the second prediction; that Cas9 cleavage within fully immune hosts prepares the cell for future escaper or related phages. First, we infected cells carrying pCRISPR(*spc174*) with ΦNM4γ4 at a high MOI (100), which increases the concentration of escapers present in the experiment. After the initial decimation of the culture, presumably by escaper phages, resistant cells that acquired new spacers were able to recover (Figs. 6D-E). Cells harboring pCRISPR(SR), lacking the priming spacer, or pCas9(*spc174*), lacking the Cas1-Cas2-Csn2 acquisition machinery, both failed to re-grow (**Figure 3.3A**) or acquire new spacers (**Figure 3.3B**), demonstrating that the survival of staphylococci carrying pCRISPR(*spc174*) depended on the expansion of the spacer repertoire mediated by *spc174*-mediated cleavage of wild-type viruses.

Second, we infected cells carrying pCRISPR(*spc174*) with ΦNM4γ4 for 30 minutes and then added the related phage ΦNM1γ6^{PAM}, which shares 73.5 % of sequence identity with ΦNM4γ4 and contains a PAM mutation in *tgt174* that prevents Cas9 cleavage. As controls, we also infected staphylococci with either ΦNM4γ4 or ΦNM1γ6^{PAM} alone. Analysis of data from **Figures 2.1C** and **Figure 2.4D** indicated that, from the 346 different spacer sequences acquired during the first 30 minutes of infection with ΦNM4γ4 in all three replicates, 343 have perfect matches on the ΦNM1γ6^{PAM} genome, and the remaining 3 have imperfect matches that could still enable some level of Cas9 targeting (**Figure 3.3E** and **Table 3.1**). Therefore, it is expected that staphylococci carrying pCRISPR(*spc174*) will use the targeting spacer not only to destroy ΦNM4γ4, but also to prepare the host population with spacers against ΦNM1γ6^{PAM} and thus ensure survival after infection with this phage. As expected, the culture resisted infection with ΦNM4γ4 but not with ΦNM1γ6^{PAM} (**Figure 3.3C**). The latter eventually recovered after about 17 hours, due to the naïve acquisition of spacers (**Figure 3.3D**). In contrast, the culture pre-infected with ΦNM4γ4 recovered from infection with ΦNM1γ6^{PAM} significantly faster, at about 13 hours (**Figure 3.3C**) through the acquisition of new spacers (**Figure 3.3D**). Next generation sequencing of the new spacers showed that, in addition to spacers acquired from the ΦNM1γ6^{PAM} genome, all 346 spacer sequences that were acquired from the ΦNM4γ4 genome after 30 minutes of infection were present at least in one replicate population that survived infection by both phages for 24 hours, including the three spacers that partially match the ΦNM1γ6^{PAM} genome (**Table 3.1**). The presence of these sequences, resulting from spacer acquisition from ΦNM4γ4 DNA following cleavage of this phage by Cas9 loaded with the *spc174* crRNA guide, and which are not required to provide immunity against this ΦNM4γ4, demonstrates that these spacers indeed can be used for the targeting of ΦNM1γ6^{PAM}. Priming within already-immune cells therefore allows a pre-emptive strike against future infection by escaper phages and/or related viruses.

Figure 3.3. Cleavage-dependent spacer acquisition provides pre-emptive immunity against viruses
(**A**) Cell survival measured as OD_{600} after infection with ΦNM4γ4 at MOI 100 of staphylococci carrying different plasmids. The average curves of three different replicates are shown, with +/- StDev values shown in lighter colors. (**C**) Agarose gel electrophoresis of PCR products after amplification of the CRISPR array of cells obtained after the experiment in (**A**) to detect the integration of new spacers. (**D**) Cell survival measured as OD_{600} after infection of cultures carrying pCRISPR(*spc174*) with different phages at MOI 1 or none as a control. "ΦNM4γ4/ΦNM1" indicates infection with ΦNM1γ6^{PAM} 30 minutes of addition of ΦNM4γ4 at MOI 10. "ΦNM1" is an abbreviation for ΦNM1γ6^{PAM}. The average curves of three different replicates are shown, with +/- StDev values shown in lighter colors. (**E**) Same as (**C**) but for the cells obtained after the experiment in (**D**).



E

	PAM	
ΦNM1γ6^{PAM}	AAGAAACATTCGAACACGCT-GGG	(31044-31066)
spacer	AAGTTAAGAAGTAACGCGCT	
	*** * *** ****	
ΦNM1γ6^{PAM}	ATCAACTTTAGTAGTCCTTC-AGG	(40476-40498)
spacer	ATCAACTTTAGTAATCCTTC	
	********** ******	
ΦNM1γ6^{PAM}	TCAACTAATTCAAGAATTAC-AGG	(42192-42214)
spacer	GCCAATGACTCAACAATTAC	
	* * * * **** ******	

Table 3.1. Spacers acquired from ΦNM4γ4 present in ΦNM1γ6^{PAM}

Spacer in NM4	NM4 PAM	NM4 position	Replicate of Figure 2-1C	Normalized Reads (30 min)	Spacer Present in NM1? (Y/N)	Position in NM1 or NM4	Reads at 24 hour of co-infection (rep #1)	Reads at 24 hour of co-infection (rep #2)	Reads at 24 hour of co-infection (rep #3)
TGAACGACACAAATGATTTA	GGG	107	1	38.98152805	Y	40912	0	2	0
TGAACGACACAAATGATTTA	GGG	107	2	30.77489057	Y	40912	0	2	0
TGAACGACACAAATGATTTA	GGG	107	3	8.206637484	Y	40912	0	2	0
CACAAATGATTTAGGGTAGG	TGG	114	1	119.0665571	Y	40919	0	0	4
CACAAATGATTTAGGGTAGG	TGG	114	2	77.65210248	Y	40919	0	0	4
CACAAATGATTTAGGGTAGG	TGG	114	3	72.47529565	Y	40919	0	0	4
GAAATCCATACCAACCATCT	GGG	190	1	165.442976	Y	40995	0	5	0
GAAATCCATACCAACCATCT	GGG	190	2	170.1699181	Y	40995	0	5	0
GAAATCCATACCAACCATCT	GGG	190	3	86.66060645	Y	40995	0	5	0
AACAATTCAACCCAGATGGT	TGG	196	1	600.3216556	Y	41001	0	2	0
AACAATTCAACCCAGATGGT	TGG	196	2	560.9304708	Y	41001	0	2	0
AACAATTCAACCCAGATGGT	TGG	196	3	379.731021	Y	41001	0	2	0
CGCTAACATAAAGAACATAT	TGG	228	1	18.90776868	Y	41033	0	0	4
CGCTAACATAAAGAACATAT	TGG	228	2	39.39118475	Y	41033	0	0	4
CGCTAACATAAAGAACATAT	TGG	228	3	14.18082651	Y	41033	0	0	4
ATGTTCTTTATGTTAGCGAC	AGG	249	1	200.6640952	Y	41054	0	0	2
ATGTTCTTTATGTTAGCGAC	AGG	249	2	278.5105988	Y	41054	0	0	2
ATGTTCTTTATGTTAGCGAC	AGG	249	3	106.0697381	Y	41054	0	0	2
ACGACAATATCCAACTTTTG	CGG	356	1	165.6578186	Y	41161	2	2	5
ACGACAATATCCAACTTTTG	CGG	356	2	103.5361366	Y	41161	2	2	5
ACGACAATATCCAACTTTTG	CGG	356	3	62.12168198	Y	41161	2	2	5
GTTTTCCCGTCAAAGTATGG	TGG	393	1	274.3707621	Y	41198	2	2	16
GTTTTCCCGTCAAAGTATGG	TGG	393	2	232.9563074	Y	41198	2	2	16
GTTTTCCCGTCAAAGTATGG	TGG	393	3	139.7737845	Y	41198	2	2	16
TCAAAGTATGGTGGCGGAGC	TGG	402	1	304.8729606	Y	41207	0	4	0
TCAAAGTATGGTGGCGGAGC	TGG	402	2	530.038545	Y	41207	0	4	0
TCAAAGTATGGTGGCGGAGC	TGG	402	3	217.722094	Y	41207	0	4	0
TTAAATACTTTCACATCATT	TGG	453	1	75.63107472	Y	41258	0	0	2
TTAAATACTTTCACATCATT	TGG	453	2	124.4761438	Y	41258	0	0	2
TTAAATACTTTCACATCATT	TGG	453	3	59.87460082	Y	41258	0	0	2
GGTCAAAACTGGAACGGTAA	AGG	474	1	28.72323119	Y	41279	0	0	3
GGTCAAAACTGGAACGGTAA	AGG	474	2	131.3061997	Y	41279	0	0	3
GGTCAAAACTGGAACGGTAA	AGG	474	3	12.30995623	Y	41279	0	0	3
ATGGCGTTGCGCAACCTGGT	TGG	505	1	846.1226484	Y	41310	0	4	0
ATGGCGTTGCGCAACCTGGT	TGG	505	2	1251.064028	Y	41310	0	4	0

ATGGCGTTGCGCAACCTGGT	TGG	505	3	438.0299744	Y	41310	0	4	0
TGGCGTTGCGCAACCTGGTT	GGG	506	1	497.9045752	Y	41311	4	0	0
TGGCGTTGCGCAACCTGGTT	GGG	506	2	765.7646315	Y	41311	4	0	0
TGGCGTTGCGCAACCTGGTT	GGG	506	3	313.5538306	Y	41311	4	0	0
ACATGTCTTGTCACAGTTTC	AGG	515	1	362.2498657	Y	41320	0	3	0
ACATGTCTTGTCACAGTTTC	AGG	515	2	350.1541939	Y	41320	0	3	0
ACATGTCTTGTCACAGTTTC	AGG	515	3	176.7828968	Y	41320	0	3	0
TTTAACCTAATAAAATACAT	TGG	554	1	15.7564739	Y	41359	0	0	2
TTTAACCTAATAAAATACAT	TGG	554	2	25.21035824	Y	41359	0	0	2
TTTAACCTAATAAAATACAT	TGG	554	3	6.30258956	Y	41359	0	0	2
ACAATCCAATGTATTTTATT	AGG	565	1	12.60517912	Y	41370	3	0	0
ACAATCCAATGTATTTTATT	AGG	565	2	17.33212129	Y	41370	3	0	0
ACAATCCAATGTATTTTATT	AGG	565	3	9.45388434	Y	41370	3	0	0
AGCGTTGGCAATAAAGCTAA	AGG	609	1	61.54978113	Y	41414	0	3	0
AGCGTTGGCAATAAAGCTAA	AGG	609	2	86.16969358	Y	41414	0	3	0
AGCGTTGGCAATAAAGCTAA	AGG	609	3	6.154978113	Y	41414	0	3	0
GTTCCGTTTCCTACTGCTCC	AGG	701	1	3236.367593	Y	41506	0	0	3
GTTCCGTTTCCTACTGCTCC	AGG	701	2	4197.198145	Y	41506	0	0	3
GTTCCGTTTCCTACTGCTCC	AGG	701	3	2094.412109	Y	41506	0	0	3
CTTAAATACTTAGCGATATT	AGG	758	1	81.93366428	Y	41563	0	0	2
CTTAAATACTTAGCGATATT	AGG	758	2	64.60154299	Y	41563	0	0	2
CTTAAATACTTAGCGATATT	AGG	758	3	34.66424258	Y	41563	0	0	2
AAAAAGATTATGGCTTATAT	TGG	889	1	77.20672211	Y	41694	0	0	3
AAAAAGATTATGGCTTATAT	TGG	889	2	26.78600563	Y	41694	0	0	3
AAAAAGATTATGGCTTATAT	TGG	889	3	23.63471085	Y	41694	0	0	3
CAGGAGAAAGCGCAAGTGGT	GGG	961	1	2366.62238	Y	41766	0	0	2
CAGGAGAAAGCGCAAGTGGT	GGG	961	2	3502.664148	Y	41766	0	0	2
CAGGAGAAAGCGCAAGTGGT	GGG	961	3	1797.813672	Y	41766	0	0	2
TATTACTAATAAAAATGATA	TGG	1154	1	2.051659371	Y	41959	0	0	2
TATTACTAATAAAAATGATA	TGG	1154	3	4.103318742	Y	41959	0	0	2
CCAGCTACCAAACCACCTAT	AGG	1223	1	527.8418757	Y	42028	0	2	0
CCAGCTACCAAACCACCTAT	AGG	1223	2	589.2921239	Y	42028	0	2	0
CCAGCTACCAAACCACCTAT	AGG	1223	3	305.6755937	Y	42028	0	2	0
GCGATTCATGGTAAGCCTAT	AGG	1224	1	516.8123439	Y	42029	0	6	0
GCGATTCATGGTAAGCCTAT	AGG	1224	2	779.9454581	Y	42029	0	6	0
GCGATTCATGGTAAGCCTAT	AGG	1224	3	354.5206628	Y	42029	0	6	0
TGGTAAGCCTATAGGTGGTT	TGG	1232	1	1635.521991	Y	42037	5	0	0
TGGTAAGCCTATAGGTGGTT	TGG	1232	2	2075.127613	Y	42037	5	0	0
TGGTAAGCCTATAGGTGGTT	TGG	1232	3	1022.595156	Y	42037	5	0	0

CCTATAGGTGGTTTGGTAGC	TGG	1239	1	882.6739023	Y	42044	5	0	0
CCTATAGGTGGTTTGGTAGC	TGG	1239	2	1021.619056	Y	42044	5	0	0
CCTATAGGTGGTTTGGTAGC	TGG	1239	3	494.6819656	Y	42044	5	0	0
AGTGTATAACCTGCTGGCAC	TGG	1283	1	15613.34139	Y	42088	7	4	1
AGTGTATAACCTGCTGGCAC	TGG	1283	2	26265.59643	Y	42088	7	4	1
AGTGTATAACCTGCTGGCAC	TGG	1283	3	11487.47669	Y	42088	7	4	1
TCAACTAATTCAAGAATTAC	AGG	1407	1	35.35657936	Y	42212	0	1	2
TCAACTAATTCAAGAATTAC	AGG	1407	2	63.57981377	Y	42212	0	1	2
TCAACTAATTCAAGAATTAC	AGG	1407	3	22.64061661	Y	42212	0	1	2
ATTACTTATATTGCTAATAG	TGG	1494	1	25.88403416	Y	42299	0	3	0
ATTACTTATATTGCTAATAG	TGG	1494	2	10.35361366	Y	42299	0	3	0
TCGTTATATAGCGACAGGAG	AGG	1523	1	471.0894217	Y	42328	0	2	2
TCGTTATATAGCGACAGGAG	AGG	1523	2	243.3099211	Y	42328	0	2	2
TCGTTATATAGCGACAGGAG	AGG	1523	3	269.1939553	Y	42328	0	2	2
AGCGACAGGAGAGGTAGACA	AGG	1532	1	404.1768961	Y	42337	2	0	3
AGCGACAGGAGAGGTAGACA	AGG	1532	2	619.60113	Y	42337	2	0	3
AGCGACAGGAGAGGTAGACA	AGG	1532	3	211.3209152	Y	42337	2	0	3
ACAGGAGAGGTAGACAAGGC	AGG	1536	1	597.9603948	Y	42341	9	0	0
ACAGGAGAGGTAGACAAGGC	AGG	1536	2	540.2733443	Y	42341	9	0	0
ACAGGAGAGGTAGACAAGGC	AGG	1536	3	157.8640254	Y	42341	9	0	0
ACAGTTATTAAATAACTATT	TGG	1646	1	29.93730041	Y	42451	0	2	2
ACAGTTATTAAATAACTATT	TGG	1646	2	23.63471085	Y	42451	0	2	2
ACAGTTATTAAATAACTATT	TGG	1646	3	14.18082651	Y	42451	0	2	2
AGAAATGTTAAAAAGTGTAT	AGG	1668	1	36.23988997	Y	42473	0	3	0
AGAAATGTTAAAAAGTGTAT	AGG	1668	2	33.08859519	Y	42473	0	3	0
AGAAATGTTAAAAAGTGTAT	AGG	1668	3	28.36165302	Y	42473	0	3	0
TAGATAACAGGCAGGTACTT	CGG	1722	1	113.4466121	Y	42527	4	0	0
TAGATAACAGGCAGGTACTT	CGG	1722	2	129.203086	Y	42527	4	0	0
TAGATAACAGGCAGGTACTT	CGG	1722	3	94.5388434	Y	42527	4	0	0
GTGACATGCTTGGGTGAACA	AGG	1817	1	326.21384	Y	42622	0	0	2
GTGACATGCTTGGGTGAACA	AGG	1817	2	894.5234858	Y	42622	0	0	2
GTGACATGCTTGGGTGAACA	AGG	1817	3	254.405762	Y	42622	0	0	2
AAAACATCCAGTGACATGCT	TGG	1827	1	91.38754862	Y	42632	0	4	0
AAAACATCCAGTGACATGCT	TGG	1827	2	97.69013818	Y	42632	0	4	0
AAAACATCCAGTGACATGCT	TGG	1827	3	72.47977994	Y	42632	0	4	0
TGCTAAAGTCATATATACTA	CGG	1886	1	6.154978113	Y	42691	0	1	2
TGCTAAAGTCATATATACTA	CGG	1886	2	20.51659371	Y	42691	0	1	2
TGCTAAAGTCATATATACTA	CGG	1886	3	20.51659371	Y	42691	0	1	2
CATTCCCGTATAACAGTTTA	CGG	1923	1	63.6014405	Y	42728	6	0	2

CATTCCCGTATAACAGTTTA	CGG	1923	2	123.0995623	Y	42728	6	0	2
CATTCCCGTATAACAGTTTA	CGG	1923	3	104.6346279	Y	42728	6	0	2
TCAGATACATGGCCTCTGCC	CGG	1980	1	1930.965466	Y	42785	0	0	2
TCAGATACATGGCCTCTGCC	CGG	1980	2	3023.917972	Y	42785	0	0	2
TCAGATACATGGCCTCTGCC	CGG	1980	3	958.9696788	Y	42785	0	0	2
GGGACCAACAGTCAGATACA	TGG	1991	1	45.13650616	Y	42796	0	6	0
GGGACCAACAGTCAGATACA	TGG	1991	2	59.49812176	Y	42796	0	6	0
GGGACCAACAGTCAGATACA	TGG	1991	3	22.56825308	Y	42796	0	6	0
TAACCGTATCTTAATCGATA	CGG	2099	1	2.051659371	Y	42904	0	1	0
TAACCGTATCTTAATCGATA	CGG	2099	2	26.67157182	Y	42904	0	1	0
TAACCGTATCTTAATCGATA	CGG	2099	3	10.25829686	Y	42904	0	1	0
ACTAAGACATCAATTTTAGT	TGG	4576	1	47.2694217	Y	4974	0	0	1
ACTAAGACATCAATTTTAGT	TGG	4576	2	47.2694217	Y	4974	0	0	1
ACTAAGACATCAATTTTAGT	TGG	4576	3	18.90776868	Y	4974	0	0	1
AACAAAACGGTATAAACATC	GGG	4619	1	55.82617794	Y	5017	3	0	0
AACAAAACGGTATAAACATC	GGG	4619	2	120.9567189	Y	5017	3	0	0
AACAAAACGGTATAAACATC	GGG	4619	3	46.21166952	Y	5017	3	0	0
GGATTCCTTATTAAACGCAA	GGG	4675	1	36.92986868	Y	5073	0	0	3
GGATTCCTTATTAAACGCAA	GGG	4675	2	55.39480302	Y	5073	0	0	3
GGATTCCTTATTAAACGCAA	GGG	4675	3	38.98152805	Y	5073	0	0	3
CGTTCCATTGAATACTGTGT	AGG	4698	1	794.1262846	Y	5096	1100731	12	8
CGTTCCATTGAATACTGTGT	AGG	4698	2	885.5138332	Y	5096	1100731	12	8
CGTTCCATTGAATACTGTGT	AGG	4698	3	392.3362001	Y	5096	1100731	12	8
CATGCCTACACAGTATTCAA	TGG	4710	1	45.13650616	Y	5108	0	5	0
CATGCCTACACAGTATTCAA	TGG	4710	2	143.616156	Y	5108	0	5	0
CATGCCTACACAGTATTCAA	TGG	4710	3	24.61991245	Y	5108	0	5	0
AGAAACATCAATCACACATT	CGG	4755	1	7.87823695	Y	5153	6	0	0
AGAAACATCAATCACACATT	CGG	4755	2	9.45388434	Y	5153	6	0	0
AGAAACATCAATCACACATT	CGG	4755	3	17.33212129	Y	5153	6	0	0
ACATCAATCACACATTCGGA	CGG	4759	1	137.4611779	Y	5157	0	5	0
ACATCAATCACACATTCGGA	CGG	4759	2	164.1327497	Y	5157	0	5	0
ACATCAATCACACATTCGGA	CGG	4759	3	67.70475924	Y	5157	0	5	0
AGTAAGACGCCAAAAGTAAC	AGG	4795	1	125.2987549	Y	5193	0	0	1
AGTAAGACGCCAAAAGTAAC	AGG	4795	2	234.7800928	Y	5193	0	0	1
AGTAAGACGCCAAAAGTAAC	AGG	4795	3	89.63203014	Y	5193	0	0	1
CAACTGCCATTGTGATGAGG	AGG	4895	3	25.88403416	Y	5293	7	0	0
TCACAACTGCCATTGTGATG	AGG	4898	1	150.1273981	Y	5296	8	11	0
TCACAACTGCCATTGTGATG	AGG	4898	2	181.1882391	Y	5296	8	11	0
TCACAACTGCCATTGTGATG	AGG	4898	3	113.8897503	Y	5296	8	11	0

AATGGCAGTTGTGACGTGGA	AGG	4923	1	129.2545404	Y	5321	9	37	626632
AATGGCAGTTGTGACGTGGA	AGG	4923	2	373.4020055	Y	5321	9	37	626632
AATGGCAGTTGTGACGTGGA	AGG	4923	3	55.39480302	Y	5321	9	37	626632
TGAAGCATAATACTGCTACT	AGG	5418	1	672.8014355	Y	5959	8	0	4
TGAAGCATAATACTGCTACT	AGG	5418	2	842.9713537	Y	5959	8	0	4
TGAAGCATAATACTGCTACT	AGG	5418	3	414.3952636	Y	5959	8	0	4
GCTGTAGTGAAGTATAGAAA	CGG	5464	1	34.87820931	Y	6005	0	2	0
GCTGTAGTGAAGTATAGAAA	CGG	5464	2	4.103318742	Y	6005	0	2	0
GCTGTAGTGAAGTATAGAAA	CGG	5464	3	8.206637484	Y	6005	0	2	0
TTCTATACTTCACTACAGCA	TGG	5484	1	369.2986868	Y	6025	0	0	3
TTCTATACTTCACTACAGCA	TGG	5484	2	373.4020055	Y	6025	0	0	3
TTCTATACTTCACTACAGCA	TGG	5484	3	178.4943653	Y	6025	0	0	3
ATAAAAAAACTGCTACTTGT	TGG	5569	1	286.767825	Y	6110	4	11	239434
ATAAAAAAACTGCTACTTGT	TGG	5569	2	300.9486515	Y	6110	4	11	239434
ATAAAAAAACTGCTACTTGT	TGG	5569	3	209.5611029	Y	6110	4	11	239434
GTCAAGATGTATTACGAAAT	AGG	5658	1	29.93730041	Y	6199	0	3	0
GTCAAGATGTATTACGAAAT	AGG	5658	2	31.5129478	Y	6199	0	3	0
GTCAAGATGTATTACGAAAT	AGG	5658	3	23.63471085	Y	6199	0	3	0
ATTCATTTTAAAAGGTCATA	TGG	5729	1	2.051659371	Y	6270	0	2	0
ATTCATTTTAAAAGGTCATA	TGG	5729	2	6.154978113	Y	6270	0	2	0
ATTCATTTTAAAAGGTCATA	TGG	5729	3	2.051659371	Y	6270	0	2	0
TTCATTTTAAAAGGTCATAT	GGG	5730	1	6.30258956	Y	6271	0	0	1
TTCATTTTAAAAGGTCATAT	GGG	5730	2	22.05906346	Y	6271	0	0	1
TTCATTTTAAAAGGTCATAT	GGG	5730	3	6.30258956	Y	6271	0	0	1
ACGACATAAGCATGTTTAAT	TGG	5771	1	99.26578557	Y	6312	0	0	8
ACGACATAAGCATGTTTAAT	TGG	5771	2	86.66060645	Y	6312	0	0	8
ACGACATAAGCATGTTTAAT	TGG	5771	3	34.66424258	Y	6312	0	0	8
AGATGAGAATGACTTAGATA	TGG	5810	1	30.77489057	Y	6351	2	0	0
AGATGAGAATGACTTAGATA	TGG	5810	2	75.91139673	Y	6351	2	0	0
AGATGAGAATGACTTAGATA	TGG	5810	3	26.67157182	Y	6351	2	0	0
CAATCCATTCATCTATTGCT	TGG	5833	1	406.5170266	Y	6374	26	744762	6
CAATCCATTCATCTATTGCT	TGG	5833	2	477.4211592	Y	6374	26	744762	6
CAATCCATTCATCTATTGCT	TGG	5833	3	263.1331141	Y	6374	26	744762	6
TGAAGAGAACACAGACGAAC	AGG	5870	1	181.4350783	Y	6411	6	0	0
TGAAGAGAACACAGACGAAC	AGG	5870	2	412.1832805	Y	6411	6	0	0
TGAAGAGAACACAGACGAAC	AGG	5870	3	186.7075507	Y	6411	6	0	0
ATATCTGAATTGTTATCAGT	TGG	5981	1	132.3543808	Y	6522	0	0	4
ATATCTGAATTGTTATCAGT	TGG	5981	2	138.6569703	Y	6522	0	0	4
ATATCTGAATTGTTATCAGT	TGG	5981	3	47.2694217	Y	6522	0	0	4

AATAGAGCTAGGGAGTTTAA	CGG	6039	1	63.6014405	Y	6580	188	0	0
AATAGAGCTAGGGAGTTTAA	CGG	6039	2	106.6862873	Y	6580	188	0	0
AATAGAGCTAGGGAGTTTAA	CGG	6039	3	32.82654994	Y	6580	188	0	0
GACAGAAACTATTGAGTACG	AGG	6116	1	176.0114323	Y	6657	8	0	6
GACAGAAACTATTGAGTACG	AGG	6116	2	253.6635348	Y	6657	8	0	6
GACAGAAACTATTGAGTACG	AGG	6116	3	144.9505913	Y	6657	8	0	6
AGAAACTATTGAGTACGAGG	AGG	6119	1	274.3707621	Y	6660	51483	2	6
AGAAACTATTGAGTACGAGG	AGG	6119	2	238.1331143	Y	6660	51483	2	6
AGAAACTATTGAGTACGAGG	AGG	6119	3	124.243364	Y	6660	51483	2	6
AGTATGCTTACTTTTTCTTG	TGG	6153	1	429.6749671	Y	6694	3	0	0
AGTATGCTTACTTTTTCTTG	TGG	6153	2	647.100854	Y	6694	3	0	0
AGTATGCTTACTTTTTCTTG	TGG	6153	3	393.4373192	Y	6694	3	0	0
TTTGCAGTATACGGAAAAAT	TGG	6247	1	3.15129478	Y	6788	0	0	2
TTTGCAGTATACGGAAAAAT	TGG	6247	2	14.18082651	Y	6788	0	0	2
GTCATTGACATTAACGAAGG	TGG	6313	1	25.88403416	Y	6854	0	7	1
GTCATTGACATTAACGAAGG	TGG	6313	2	31.06084099	Y	6854	0	7	1
CATTAACGAAGGTGGAACAA	CGG	6321	1	55.39480302	Y	6862	0	4	0
CATTAACGAAGGTGGAACAA	CGG	6321	2	149.7711341	Y	6862	0	4	0
CATTAACGAAGGTGGAACAA	CGG	6321	3	51.29148428	Y	6862	0	4	0
TTTTTTACCTCAAATTTTAC	AGG	6405	1	42.48992432	Y	6946	0	0	1
TTTTTTACCTCAAATTTTAC	AGG	6405	2	41.55948802	Y	6946	0	0	1
TTTTTTACCTCAAATTTTAC	AGG	6405	3	21.71018031	Y	6946	0	0	1
TCTCCCCAATCATTAAACGT	TGG	6516	1	23.63471085	Y	7057	0	8	4
TCTCCCCAATCATTAAACGT	TGG	6516	2	70.90413255	Y	7057	0	8	4
TCTCCCCAATCATTAAACGT	TGG	6516	3	34.66424258	Y	7057	0	8	4
AAAAACCAACGTTTAATGAT	TGG	6527	1	12.60517912	Y	7068	0	1	0
AAAAACCAACGTTTAATGAT	TGG	6527	2	11.02953173	Y	7068	0	1	0
AAAAACCAACGTTTAATGAT	TGG	6527	3	7.87823695	Y	7068	0	1	0
GTCAGTATGTACAGATTAAT	AGG	6571	1	91.38754862	Y	7112	9	0	1
GTCAGTATGTACAGATTAAT	AGG	6571	2	96.11449079	Y	7112	9	0	1
GTCAGTATGTACAGATTAAT	AGG	6571	3	66.17719038	Y	7112	9	0	1
CATGACCTGTAATAACAAAG	TGG	6602	1	15.5304205	Y	7143	0	2	0
CATGACCTGTAATAACAAAG	TGG	6602	2	10.35361366	Y	7143	0	2	0
CATGACCTGTAATAACAAAG	TGG	6602	3	10.35361366	Y	7143	0	2	0
GCGCTTCAATAGTGATAGTA	GGG	6668	1	244.1474651	Y	7209	0	0	2
GCGCTTCAATAGTGATAGTA	GGG	6668	2	352.8854118	Y	7209	0	0	2
GCGCTTCAATAGTGATAGTA	GGG	6668	3	157.9777716	Y	7209	0	0	2
TGCGCTTCAATAGTGATAGT	AGG	6669	1	61.45024821	Y	7210	0	0	8
TGCGCTTCAATAGTGATAGT	AGG	6669	2	69.32848516	Y	7210	0	0	8

TGCGCTTCAATAGTGATAGT	AGG	6669	3	29.93730041	Y	7210	0	0	8
TCTACTACGTCCGTAATGCT	AGG	6876	1	477.4211592	Y	7417	37028	6	12
TCTACTACGTCCGTAATGCT	AGG	6876	2	475.8455118	Y	7417	37028	6	12
TCTACTACGTCCGTAATGCT	AGG	6876	3	278.889588	Y	7417	37028	6	12
ATTTGCAAATCCTAGCATTA	CGG	6882	1	67.70475924	Y	7423	0	3	3
ATTTGCAAATCCTAGCATTA	CGG	6882	2	86.16969358	Y	7423	0	3	3
ATTTGCAAATCCTAGCATTA	CGG	6882	3	28.72323119	Y	7423	0	3	3
GTAGTAGAAGCAATTAGAAA	TGG	6907	1	14.3616156	Y	7448	4	2	0
GTAGTAGAAGCAATTAGAAA	TGG	6907	2	12.30995623	Y	7448	4	2	0
GTAGTAGAAGCAATTAGAAA	TGG	6907	3	2.051659371	Y	7448	4	2	0
AGTTTTATAACGGCTCAGCA	GGG	7014	1	377.5053243	Y	7555	45282	18	165780
AGTTTTATAACGGCTCAGCA	GGG	7014	2	315.9555431	Y	7555	45282	18	165780
AGTTTTATAACGGCTCAGCA	GGG	7014	3	80.01471547	Y	7555	45282	18	165780
GGCTCAGCAGGGTTTCAAGC	TGG	7025	1	96.14508423	Y	7566	0	0	6
GGCTCAGCAGGGTTTCAAGC	TGG	7025	2	161.895916	Y	7566	0	0	6
GGCTCAGCAGGGTTTCAAGC	TGG	7025	3	73.19432219	Y	7566	0	0	6
ATCGTATTTGAAAATGATGA	AGG	7109	1	4.103318742	Y	7650	0	0	1
ATCGTATTTGAAAATGATGA	AGG	7109	2	4.103318742	Y	7650	0	0	1
ATCGTATTTGAAAATGATGA	AGG	7109	3	8.206637484	Y	7650	0	0	1
TCTTGGAAATCATATTTATA	CGG	7147	1	2.051659371	Y	7688	177817	5	64548
TCTTGGAAATCATATTTATA	CGG	7147	2	16.41327497	Y	7688	177817	5	64548
TCTTGGAAATCATATTTATA	CGG	7147	3	8.206637484	Y	7688	177817	5	64548
ATTCAATCAATTGTTTTTCT	TGG	7164	1	9.45388434	Y	7705	0	5	0
ATTCAATCAATTGTTTTTCT	TGG	7164	2	11.02953173	Y	7705	0	5	0
ATTCAATCAATTGTTTTTCT	TGG	7164	3	7.87823695	Y	7705	0	5	0
TTGGTATCAAAATCTAAGCT	AGG	7219	1	1.57564739	Y	7760	0	4	0
TTGGTATCAAAATCTAAGCT	AGG	7219	2	12.60517912	Y	7760	0	4	0
AAACTTACCAATAAGATCAT	TGG	7238	1	45.69377431	Y	7779	0	0	2
AAACTTACCAATAAGATCAT	TGG	7238	2	63.0258956	Y	7779	0	0	2
AAACTTACCAATAAGATCAT	TGG	7238	3	15.7564739	Y	7779	0	0	2
TTTGATACCAATGATCTTAT	TGG	7247	1	66.17719038	Y	7788	4	0	0
TTTGATACCAATGATCTTAT	TGG	7247	2	85.08495906	Y	7788	4	0	0
TTTGATACCAATGATCTTAT	TGG	7247	3	36.23988997	Y	7788	4	0	0
TATTGGTAAGTTTTGTCACT	TGG	7264	1	72.47977994	Y	7805	18995	55709	3
TATTGGTAAGTTTTGTCACT	TGG	7264	2	61.45024821	Y	7805	18995	55709	3
TATTGGTAAGTTTTGTCACT	TGG	7264	3	34.66424258	Y	7805	18995	55709	3
TGGAAATTCAATGAAGATGA	AGG	7295	1	90.27301232	Y	7836	60971	0	0
TGGAAATTCAATGAAGATGA	AGG	7295	2	38.98152805	Y	7836	60971	0	0
TGGAAATTCAATGAAGATGA	AGG	7295	3	16.41327497	Y	7836	60971	0	0

ACATCATCGCCCTTTTTGTA	AGG	7333	1	221.5792121	Y	7874	27	0	3
ACATCATCGCCCTTTTTGTA	AGG	7333	2	644.2210425	Y	7874	27	0	3
ACATCATCGCCCTTTTTGTA	AGG	7333	3	285.1806526	Y	7874	27	0	3
TTTATTAAACCTTACAAAAA	GGG	7340	1	2.051659371	Y	7881	3	0	0
TTTATTAAACCTTACAAAAA	GGG	7340	2	6.154978113	Y	7881	3	0	0
GCTTTTTGCTTATCTGTCTT	CGG	7372	1	228.4688716	Y	7913	1	1	3
GCTTTTTGCTTATCTGTCTT	CGG	7372	2	299.3730041	Y	7913	1	1	3
GCTTTTTGCTTATCTGTCTT	CGG	7372	3	127.6274386	Y	7913	1	1	3
AGCAATCCATTTGAAAGCAG	TGG	7457	1	15.5304205	Y	7998	0	3	0
AGCAATCCATTTGAAAGCAG	TGG	7457	2	20.70722733	Y	7998	0	3	0
CTTGGTCGTCATATCCAAAT	TGG	7464	1	1.57564739	Y	8005	0	2	3
TTTGAAAGCAGTGGCCAATT	TGG	7466	1	31.5129478	Y	8007	6	0	0
TTTGAAAGCAGTGGCCAATT	TGG	7466	2	15.7564739	Y	8007	6	0	0
TTTGAAAGCAGTGGCCAATT	TGG	7466	3	3.15129478	Y	8007	6	0	0
CACCTTAAAACGCTAAATCT	TGG	7482	1	12.60517912	Y	8023	4	0	3
CACCTTAAAACGCTAAATCT	TGG	7482	2	15.7564739	Y	8023	4	0	3
CACCTTAAAACGCTAAATCT	TGG	7482	3	11.02953173	Y	8023	4	0	3
GACCAAGATTTAGCGTTTTA	AGG	7496	1	110.789606	Y	8037	0	5	0
GACCAAGATTTAGCGTTTTA	AGG	7496	2	145.6678153	Y	8037	0	5	0
GACCAAGATTTAGCGTTTTA	AGG	7496	3	59.49812176	Y	8037	0	5	0
AAGTACCGTCGTTATCTTTC	TGG	7533	1	314.7976145	Y	8074	0	5	0
AAGTACCGTCGTTATCTTTC	TGG	7533	2	548.647271	Y	8074	0	5	0
AAGTACCGTCGTTATCTTTC	TGG	7533	3	219.893112	Y	8074	0	5	0
ACTTATTCCGTCGTTGCTAC	TGG	7568	1	739.3867123	Y	8109	0	7	0
ACTTATTCCGTCGTTGCTAC	TGG	7568	2	2208.235483	Y	8109	0	7	0
ACTTATTCCGTCGTTGCTAC	TGG	7568	3	791.491145	Y	8109	0	7	0
ACCTCTACTTCTGCTTTTAG	TGG	7621	1	150.1273981	Y	8162	0	0	2
ACCTCTACTTCTGCTTTTAG	TGG	7621	2	103.5361366	Y	8162	0	0	2
ACCTCTACTTCTGCTTTTAG	TGG	7621	3	98.35932981	Y	8162	0	0	2
AAAAGCAGAAGTAGAGGTTC	CGG	7642	1	138.6350086	Y	8183	1	0	0
AAAAGCAGAAGTAGAGGTTC	CGG	7642	2	232.6090748	Y	8183	1	0	0
AAAAGCAGAAGTAGAGGTTC	CGG	7642	3	123.4378823	Y	8183	1	0	0
TAGAGATATAGAACTTCACT	GGG	7714	1	148.1108547	Y	8255	0	6130	5171
TAGAGATATAGAACTTCACT	GGG	7714	2	242.6496981	Y	8255	0	6130	5171
TAGAGATATAGAACTTCACT	GGG	7714	3	75.63107472	Y	8255	0	6130	5171
AGAGATATAGAACTTCACTG	GGG	7715	1	56.94487515	Y	8256	0	3	0
AGAGATATAGAACTTCACTG	GGG	7715	2	31.06084099	Y	8256	0	3	0
AGAGATATAGAACTTCACTG	GGG	7715	3	51.76806832	Y	8256	0	3	0
AATTTTCTAGTTGATTCTAC	TGG	7726	1	115.9943919	Y	8267	0	2	0

AATTTTCTAGTTGATTCTAC	TGG	7726	2	72.26388589	Y	8267	0	2	0
AATTTTCTAGTTGATTCTAC	TGG	7726	3	86.84072124	Y	8267	0	2	0
ACTGTTCTATGAAAGTTGCA	AGG	7818	1	162.0810903	Y	8359	0	7	0
ACTGTTCTATGAAAGTTGCA	AGG	7818	2	223.6308714	Y	8359	0	7	0
ACTGTTCTATGAAAGTTGCA	AGG	7818	3	102.5829686	Y	8359	0	7	0
CTGTTCTATGAAAGTTGCAA	GGG	7819	1	205.1659371	Y	8360	0	0	1
CTGTTCTATGAAAGTTGCAA	GGG	7819	2	240.0441464	Y	8360	0	0	1
CTGTTCTATGAAAGTTGCAA	GGG	7819	3	141.5644966	Y	8360	0	0	1
TTACTCGTTTCTACACTCAT	AGG	7870	1	97.69013818	Y	8411	0	0	1
TTACTCGTTTCTACACTCAT	AGG	7870	2	77.20672211	Y	8411	0	0	1
TTACTCGTTTCTACACTCAT	AGG	7870	3	47.2694217	Y	8411	0	0	1
AACCGCAACTGTGTAATATG	CGG	7952	1	10.35361366	Y	8493	6	0	0
AACCGCAACTGTGTAATATG	CGG	7952	2	10.35361366	Y	8493	6	0	0
CGGAAAGCCTCACGCAGACC	TGG	7972	1	832.1201967	Y	8513	25592	6	4
CGGAAAGCCTCACGCAGACC	TGG	7972	2	1130.480103	Y	8513	25592	6	4
CGGAAAGCCTCACGCAGACC	TGG	7972	3	499.0240017	Y	8513	25592	6	4
CTGGCACATTATGAAGCAGT	CGG	7991	1	189.0776868	Y	8532	8	0	4
CTGGCACATTATGAAGCAGT	CGG	7991	2	237.9227559	Y	8532	8	0	4
CTGGCACATTATGAAGCAGT	CGG	7991	3	72.47977994	Y	8532	8	0	4
CTAATACATGTTTGTCATAG	TGG	8025	1	170.8346255	Y	8566	0	3	0
CTAATACATGTTTGTCATAG	TGG	8025	2	186.365046	Y	8566	0	3	0
CTAATACATGTTTGTCATAG	TGG	8025	3	93.18252298	Y	8566	0	3	0
CACAACGAGCAACATGCGAT	TGG	8081	1	182.7750972	Y	8622	0	2	0
CACAACGAGCAACATGCGAT	TGG	8081	2	181.1994499	Y	8622	0	2	0
CACAACGAGCAACATGCGAT	TGG	8081	3	102.4170804	Y	8622	0	2	0
AATACCACTTGCATGACTCG	TGG	8124	1	253.6635348	Y	8665	0	0	1
AATACCACTTGCATGACTCG	TGG	8124	2	310.6084099	Y	8665	0	0	1
AATACCACTTGCATGACTCG	TGG	8124	3	93.18252298	Y	8665	0	0	1
CTTCCGCCAAGATGACGATT	AGG	8213	1	17.33212129	Y	8754	2181	0	0
CTTCCGCCAAGATGACGATT	AGG	8213	2	22.05906346	Y	8754	2181	0	0
CTTCCGCCAAGATGACGATT	AGG	8213	3	9.45388434	Y	8754	2181	0	0
AGCACTCCTAATCGTCATCT	TGG	8223	1	167.0186233	Y	8764	0	1	2
AGCACTCCTAATCGTCATCT	TGG	8223	2	189.0776868	Y	8764	0	1	2
AGCACTCCTAATCGTCATCT	TGG	8223	3	107.1440225	Y	8764	0	1	2
ACTCCTAATCGTCATCTTGG	CGG	8226	1	165.6578186	Y	8767	1	489	0
ACTCCTAATCGTCATCTTGG	CGG	8226	2	238.1331143	Y	8767	1	489	0
ACTCCTAATCGTCATCTTGG	CGG	8226	3	72.47529565	Y	8767	1	489	0
AAAAACCTGTATTAACTAAA	CGG	8286	2	6.154978113	Y	8827	0	1	0
AAAAACCTGTATTAACTAAA	CGG	8286	3	2.051659371	Y	8827	0	1	0

AATCTCCGTTTAGTTAATAC	AGG	8297	1	8.373926691	Y	8838	0	5	0
AATCTCCGTTTAGTTAATAC	AGG	8297	2	35.97687023	Y	8838	0	5	0
AATCTCCGTTTAGTTAATAC	AGG	8297	3	5.582617794	Y	8838	0	5	0
TTTTACAAAAGCTTTACCAT	AGG	8321	1	83.50931167	Y	8862	15	24	215581
TCAATTAGTTTGTCCGCCTA	TGG	8321	1	1025.829686	Y	8862	0	7	7
TTTTACAAAAGCTTTACCAT	AGG	8321	2	81.93366428	Y	8862	15	24	215581
TCAATTAGTTTGTCCGCCTA	TGG	8321	2	1489.504703	Y	8862	0	7	7
TTTTACAAAAGCTTTACCAT	AGG	8321	3	26.78600563	Y	8862	15	24	215581
TCAATTAGTTTGTCCGCCTA	TGG	8321	3	841.1803421	Y	8862	0	7	7
TTTTTTGATGTCTATTACCC	AGG	8363	1	82.80883061	Y	8904	0	5	0
TTTTTTGATGTCTATTACCC	AGG	8363	2	113.5132285	Y	8904	0	5	0
TTTTTTGATGTCTATTACCC	AGG	8363	3	33.8058522	Y	8904	0	5	0
AAAGTTACATTACAGCCCCT	GGG	8364	1	419.1222057	Y	8905	3	0	2
AAAGTTACATTACAGCCCCT	GGG	8364	2	564.0817656	Y	8905	3	0	2
AAAGTTACATTACAGCCCCT	GGG	8364	3	173.3212129	Y	8905	3	0	2
GAAACAATAAGTAAACTTTC	TGG	8405	1	18.60872598	Y	8946	0	1	5
GAAACAATAAGTAAACTTTC	TGG	8405	2	44.66094235	Y	8946	0	1	5
GAAACAATAAGTAAACTTTC	TGG	8405	3	23.57105291	Y	8946	0	1	5
AAAGTTTACTTATTGTTTCT	AGG	8425	1	45.69377431	Y	8966	0	6	11
AAAGTTTACTTATTGTTTCT	AGG	8425	2	55.14765865	Y	8966	0	6	11
AAAGTTTACTTATTGTTTCT	AGG	8425	3	33.08859519	Y	8966	0	6	11
GTTAAAAGAATGTTAAAGTC	AGG	8438	1	20.46959858	Y	8979	0	6	0
GTTAAAAGAATGTTAAAGTC	AGG	8438	2	24.81163464	Y	8979	0	6	0
GTTAAAAGAATGTTAAAGTC	AGG	8438	3	25.74207094	Y	8979	0	6	0
GAAAAACAAAGATTTGTTTC	GGG	8471	1	37.21745196	Y	9012	0	0	2
GAAAAACAAAGATTTGTTTC	GGG	8471	2	30.084107	Y	9012	0	0	2
GAAAAACAAAGATTTGTTTC	GGG	8471	3	19.84930771	Y	9012	0	0	2
ATTAATAACCTAGTAAGGTT	AGG	8594	1	14.18082651	Y	9135	0	0	1
ATTAATAACCTAGTAAGGTT	AGG	8594	2	29.93730041	Y	9135	0	0	1
ATTAATAACCTAGTAAGGTT	AGG	8594	3	11.02953173	Y	9135	0	0	1
TTAAAACCCCTAACCTTACT	AGG	8602	1	85.08495906	Y	9143	3	15	158780
TTAAAACCCCTAACCTTACT	AGG	8602	2	34.66424258	Y	9143	3	15	158780
TTAAAACCCCTAACCTTACT	AGG	8602	3	25.21035824	Y	9143	3	15	158780
TCGGTTGATTCTATATCTAA	CGG	8640	1	53.34314365	Y	9181	0	4	2
TCGGTTGATTCTATATCTAA	CGG	8640	2	149.7711341	Y	9181	0	4	2
TCGGTTGATTCTATATCTAA	CGG	8640	3	51.29148428	Y	9181	0	4	2
GAATGACGGTATGAATATAT	CGG	8667	1	9.45388434	Y	9208	0	0	6
GAATGACGGTATGAATATAT	CGG	8667	2	29.93730041	Y	9208	0	0	6
GAATGACGGTATGAATATAT	CGG	8667	3	17.33212129	Y	9208	0	0	6

GTTTTAAACCCAAAGAATGA	CGG	8681	1	32.82654994	Y	9222	0	0	64
GTTTTAAACCCAAAGAATGA	CGG	8681	2	125.1512216	Y	9222	0	0	64
GTTTTAAACCCAAAGAATGA	CGG	8681	3	28.72323119	Y	9222	0	0	64
ATATTCATACCGTCATTCTT	TGG	8688	1	160.7160338	Y	9229	0	1	0
ATATTCATACCGTCATTCTT	TGG	8688	2	159.1403864	Y	9229	0	1	0
ATATTCATACCGTCATTCTT	TGG	8688	3	53.57201126	Y	9229	0	1	0
TTAAAAGAATAGCATCATTT	GGG	8767	1	3.15129478	Y	9308	0	0	2
TTAAAAGAATAGCATCATTT	GGG	8767	2	11.02953173	Y	9308	0	0	2
TTAAAAGAATAGCATCATTT	GGG	8767	3	1.57564739	Y	9308	0	0	2
TTATTTCAACAAATGAATGG	TGG	8803	1	10.35361366	Y	9344	1	0	5049
TTATTTCAACAAATGAATGG	TGG	8803	2	5.176806832	Y	9344	1	0	5049
TTATTTCAACAAATGAATGG	TGG	8803	3	10.35361366	Y	9344	1	0	5049
TGGGAAATTCCACCTTTATG	TGG	9030	1	10.35361366	Y	9571	0	0	3
TGGGAAATTCCACCTTTATG	TGG	9030	2	82.82890931	Y	9571	0	0	3
TGGGAAATTCCACCTTTATG	TGG	9030	3	20.70722733	Y	9571	0	0	3
AAATAACTTTATCCACATAA	AGG	9034	1	34.87820931	Y	9575	8	0	1
AAATAACTTTATCCACATAA	AGG	9034	2	28.72323119	Y	9575	8	0	1
AAATAACTTTATCCACATAA	AGG	9034	3	43.08484679	Y	9575	8	0	1
AAACTTGGGAACTTGTCTTT	GGG	9049	1	12.60517912	Y	9590	0	2	0
AAACTTGGGAACTTGTCTTT	GGG	9049	2	9.45388434	Y	9590	0	2	0
AAACTTGGGAACTTGTCTTT	GGG	9049	3	15.7564739	Y	9590	0	2	0
TTATTGAAGGCGCATCAAAA	GGG	9091	2	12.30995623	Y	9632	4	0	0
ATTATTGAAGGCGCATCAAA	AGG	9092	1	2.051659371	Y	9633	0	1	0
ATTATTGAAGGCGCATCAAA	AGG	9092	2	4.103318742	Y	9633	0	1	0
TCAATACCTTTACCTAAAAT	AGG	9135	1	3.15129478	Y	9676	0	10	0
TCAATACCTTTACCTAAAAT	AGG	9135	2	1.57564739	Y	9676	0	10	0
TCAATACCTTTACCTAAAAT	AGG	9135	3	3.15129478	Y	9676	0	10	0
CCATTTCGTTATCTCCTTTC	TGG	9193	1	342.400558	Y	9734	8	32	6895
CCATTTCGTTATCTCCTTTC	TGG	9193	2	357.9078297	Y	9734	8	32	6895
CCATTTCGTTATCTCCTTTC	TGG	9193	3	130.5712273	Y	9734	8	32	6895
ATTTAAGTGCTAAAGCTAAA	GGG	9301	1	26.67157182	Y	9842	0	0	4
ATTTAAGTGCTAAAGCTAAA	GGG	9301	2	26.67157182	Y	9842	0	0	4
ATTTAAGTGCTAAAGCTAAA	GGG	9301	3	12.30995623	Y	9842	0	0	4
GTGTATATTTGCCAATTGTC	AGG	9332	1	586.7951592	Y	9873	0	2	0
GTGTATATTTGCCAATTGTC	AGG	9332	2	744.6591846	Y	9873	0	2	0
GTGTATATTTGCCAATTGTC	AGG	9332	3	238.1916925	Y	9873	0	2	0
TCCTAAGTCGTCCTGACAAT	TGG	9337	1	53.57201126	Y	9878	5	0	0
TCCTAAGTCGTCCTGACAAT	TGG	9337	2	92.96319601	Y	9878	5	0	0
TCCTAAGTCGTCCTGACAAT	TGG	9337	3	37.81553736	Y	9878	5	0	0

ACCAACCGAAATGCCATTTT	CGG	9518	1	7.87823695	Y	10059	0	4	0
ACCAACCGAAATGCCATTTT	CGG	9518	2	45.69377431	Y	10059	0	4	0
ACCAACCGAAATGCCATTTT	CGG	9518	3	1.57564739	Y	10059	0	4	0
AAAACGGATTATCGGCAAAC	GGG	9541	1	41.55948802	Y	10082	0	0	1
AAAACGGATTATCGGCAAAC	GGG	9541	2	49.00297841	Y	10082	0	0	1
AAAACGGATTATCGGCAAAC	GGG	9541	3	10.23479929	Y	10082	0	0	1
GGAAAAACGGAAAACCGAAA	AGG	9576	2	6.154978113	Y	10117	0	0	4
GAAAAACGGAAAACCGAAAA	GGG	9577	1	2.051659371	Y	10118	0	0	2
GAAAAACGGAAAACCGAAAA	GGG	9577	2	2.051659371	Y	10118	0	0	2
TAATGATGGAAGTATATTGT	CGG	9656	1	7.87823695	Y	10197	0	12	0
TAATGATGGAAGTATATTGT	CGG	9656	2	25.21035824	Y	10197	0	12	0
AGGGAATGGAATACACAGTC	GGG	9667	1	115.3741011	Y	10208	0	2	0
AGGGAATGGAATACACAGTC	GGG	9667	2	191.980023	Y	10208	0	2	0
AGGGAATGGAATACACAGTC	GGG	9667	3	92.42333903	Y	10208	0	2	0
GTATTCGATAATTTCTTTAT	AGG	9687	3	1.57564739	Y	10228	0	3	0
ATGGAATCAAGATTTTAGGT	TGG	9794	1	14.18082651	Y	10335	18	784259	5
ATGGAATCAAGATTTTAGGT	TGG	9794	2	15.7564739	Y	10335	18	784259	5
ATGGAATCAAGATTTTAGGT	TGG	9794	3	17.33212129	Y	10335	18	784259	5
CAAAAAGTGTTTCTGGTCTA	AGG	9865	1	36.92986868	Y	10406	0	1	0
CAAAAAGTGTTTCTGGTCTA	AGG	9865	2	86.16969358	Y	10406	0	1	0
CAAAAAGTGTTTCTGGTCTA	AGG	9865	3	51.29148428	Y	10406	0	1	0
ACTTTTTGGCAGTAAATTTG	AGG	9896	2	20.70722733	Y	10437	0	7	0
ACTTTTTGGCAGTAAATTTG	AGG	9896	3	10.35361366	Y	10437	0	7	0
TCCAATTGATCCGTGCCAGT	TGG	9926	1	918.6024284	Y	10467	26467	9	5
TCCAATTGATCCGTGCCAGT	TGG	9926	2	1435.414772	Y	10467	26467	9	5
TCCAATTGATCCGTGCCAGT	TGG	9926	3	502.6315174	Y	10467	26467	9	5
ACCAACTGGCACGGATCAAT	TGG	9941	1	283.6165302	Y	10482	3	12	0
ACCAACTGGCACGGATCAAT	TGG	9941	2	267.8600563	Y	10482	3	12	0
ACCAACTGGCACGGATCAAT	TGG	9941	3	215.8636924	Y	10482	3	12	0
GAAGTACGACGAAAGTTATT	GGG	9971	1	28.36165302	Y	10512	0	0	1
GAAGTACGACGAAAGTTATT	GGG	9971	2	23.63471085	Y	10512	0	0	1
GAAGTACGACGAAAGTTATT	GGG	9971	3	9.45388434	Y	10512	0	0	1
TTTATCTTTTCGCTGAATAG	TGG	9998	1	36.23764782	Y	10539	0	8	0
TTTATCTTTTCGCTGAATAG	TGG	9998	2	25.88403416	Y	10539	0	8	0
TTTATCTTTTCGCTGAATAG	TGG	9998	3	41.41445466	Y	10539	0	8	0
AATCCTTTTTCGACATGAGT	AGG	10043	1	163.8673286	Y	10584	7	1	0
AATCCTTTTTCGACATGAGT	AGG	10043	2	162.2916812	Y	10584	7	1	0
AATCCTTTTTCGACATGAGT	AGG	10043	3	86.66060645	Y	10584	7	1	0
GGATTGAAATGTGAGAGATG	TGG	10077	1	72.47529565	Y	10618	0	4	0

GGATTGAAATGTGAGAGATG	TGG	10077	2	62.12168198	Y	10618	0	4	0
GGATTGAAATGTGAGAGATG	TGG	10077	3	25.88403416	Y	10618	0	4	0
CTTTATACTCGTAACCATTC	GGG	10126	1	119.7161371	Y	10667	2	1	2
CTTTATACTCGTAACCATTC	GGG	10126	2	357.9078297	Y	10667	2	1	2
CTTTATACTCGTAACCATTC	GGG	10126	3	52.10443274	Y	10667	2	1	2
TCTTTATACTCGTAACCATT	CGG	10127	1	154.4134442	Y	10668	7	0	0
TCTTTATACTCGTAACCATT	CGG	10127	2	151.2621494	Y	10668	7	0	0
TCTTTATACTCGTAACCATT	CGG	10127	3	78.7823695	Y	10668	7	0	0
CCTACTAAAAAACACCCGAA	TGG	10128	1	65.65309987	Y	10669	2	0	0
CCTACTAAAAAACACCCGAA	TGG	10128	2	71.80807799	Y	10669	2	0	0
CCTACTAAAAAACACCCGAA	TGG	10128	3	2.051659371	Y	10669	2	0	0
ACAGCAATAGAGTACGTACA	AGG	10320	1	184.6493434	Y	10861	0	9	0
ACAGCAATAGAGTACGTACA	AGG	10320	2	441.1067648	Y	10861	0	9	0
ACAGCAATAGAGTACGTACA	AGG	10320	3	100.5313092	Y	10861	0	9	0
TTGCAAGGTTCATACGGAAC	TGG	10374	1	49.62326928	Y	10915	0	0	2
TTGCAAGGTTCATACGGAAC	TGG	10374	2	61.0986503	Y	10915	0	0	2
TTGCAAGGTTCATACGGAAC	TGG	10374	3	57.99719597	Y	10915	0	0	2
CTTTTGCGATAGCGTATGCT	AGG	10390	1	223.7419294	Y	10931	0	0	6
CTTTTGCGATAGCGTATGCT	AGG	10390	2	316.7051254	Y	10931	0	0	6
CTTTTGCGATAGCGTATGCT	AGG	10390	3	168.5942707	Y	10931	0	0	6
TTGATACGATCCATCAACAT	TGG	10454	1	195.3802764	Y	10995	11	1	7
TTGATACGATCCATCAACAT	TGG	10454	2	253.6792298	Y	10995	11	1	7
TTGATACGATCCATCAACAT	TGG	10454	3	116.5979069	Y	10995	11	1	7
TATGCACATACCAATGTTGA	TGG	10460	1	172.3393872	Y	11001	0	0	4
TATGCACATACCAATGTTGA	TGG	10460	2	260.5607401	Y	11001	0	0	4
TATGCACATACCAATGTTGA	TGG	10460	3	149.7711341	Y	11001	0	0	4
TTTACTTGTACTAGATGATA	TGG	10559	1	73.85973736	Y	11100	4	0	0
TTTACTTGTACTAGATGATA	TGG	10559	2	96.42799044	Y	11100	4	0	0
TTTACTTGTACTAGATGATA	TGG	10559	3	69.75641861	Y	11100	4	0	0
TTACTTGTACTAGATGATAT	GGG	10560	1	70.90413255	Y	11101	1	0	7
TTACTTGTACTAGATGATAT	GGG	10560	2	80.35801689	Y	11101	1	0	7
TTACTTGTACTAGATGATAT	GGG	10560	3	56.72330604	Y	11101	1	0	7
AGCATTGTTGATAACAGAGT	AGG	10620	1	55.14765865	Y	11161	7	0	0
AGCATTGTTGATAACAGAGT	AGG	10620	2	69.32848516	Y	11161	7	0	0
AGCATTGTTGATAACAGAGT	AGG	10620	3	75.63107472	Y	11161	7	0	0
AACTAAATCAAAATATGAAC	TGG	10681	1	11.16523559	Y	11222	0	0	1
AACTAAATCAAAATATGAAC	TGG	10681	2	36.59716109	Y	11222	0	0	1
AACTAAATCAAAATATGAAC	TGG	10681	3	12.71596275	Y	11222	0	0	1
GCGAGAAAAGTAAGAGTAAT	CGG	10734	1	77.20672211	Y	11275	9	0	0

GCGAGAAAAGTAAGAGTAAT	CGG	10734	2	91.38754862	Y	11275	9	0	0
GCGAGAAAAGTAAGAGTAAT	CGG	10734	3	42.54247953	Y	11275	9	0	0
TTTAGTTTTTAAAAATTCTT	TGG	10772	1	4.72694217	Y	11313	0	2	0
TTTAGTTTTTAAAAATTCTT	TGG	10772	2	4.72694217	Y	11313	0	2	0
TTTAGTTTTTAAAAATTCTT	TGG	10772	3	1.57564739	Y	11313	0	2	0
GAAACTCATAGATGAGGCAC	AGG	10837	1	295.5685976	Y	11378	0	11	0
GAAACTCATAGATGAGGCAC	AGG	10837	2	494.3718202	Y	11378	0	11	0
GAAACTCATAGATGAGGCAC	AGG	10837	3	228.2670387	Y	11378	0	11	0
CCTTAATATTCGACGATAGC	GGG	10903	1	102.6581383	Y	11444	0	0	2
CCTTAATATTCGACGATAGC	GGG	10903	2	131.1915182	Y	11444	0	0	2
CCTTAATATTCGACGATAGC	GGG	10903	3	76.60592195	Y	11444	0	0	2
TATCTTCCGTGCCATCTTCT	CGG	10962	1	623.9563664	Y	11503	7	15	166
TATCTTCCGTGCCATCTTCT	CGG	10962	2	1011.565624	Y	11503	7	15	166
TATCTTCCGTGCCATCTTCT	CGG	10962	3	504.2071648	Y	11503	7	15	166
TATTACTTATACCGAGAAGA	TGG	10967	1	8.206637484	Y	11508	0	10	0
TATTACTTATACCGAGAAGA	TGG	10967	2	24.61991245	Y	11508	0	10	0
TATTACTTATACCGAGAAGA	TGG	10967	3	8.206637484	Y	11508	0	10	0
AGATGGCACGGAAGATATTA	AGG	10984	1	10.25829686	Y	11525	0	3	0
AGATGGCACGGAAGATATTA	AGG	10984	2	10.25829686	Y	11525	0	3	0
AGATGGCACGGAAGATATTA	AGG	10984	3	12.30995623	Y	11525	0	3	0
AATGAAGTTTATTCGCTCAC	AGG	11030	1	693.4851882	Y	11571	0	0	5
AATGAAGTTTATTCGCTCAC	AGG	11030	2	1286.793402	Y	11571	0	0	5
AATGAAGTTTATTCGCTCAC	AGG	11030	3	390.7832456	Y	11571	0	0	5
TTTCTTTTCGTCGCTGAAAT	GGG	11039	1	3.15129478	Y	11580	0	4	0
TTTCTTTTCGTCGCTGAAAT	GGG	11039	3	9.45388434	Y	11580	0	4	0
CTTTTATATGAGCAAGAGCT	AGG	11123	1	14.18082651	Y	11664	0	3	0
CTTTTATATGAGCAAGAGCT	AGG	11123	2	25.21035824	Y	11664	0	3	0
CTTTTATATGAGCAAGAGCT	AGG	11123	3	48.84506909	Y	11664	0	3	0
CGTGGACGAGGCGAGCCCAT	AGG	11206	1	182.7750972	Y	11747	0	0	3
CGTGGACGAGGCGAGCCCAT	AGG	11206	2	277.3139406	Y	11747	0	0	3
CGTGGACGAGGCGAGCCCAT	AGG	11206	3	64.60154299	Y	11747	0	0	3
GAAACGTTCGAGGCACCTAT	GGG	11207	1	167.0186233	Y	11748	0	5	0
GAAACGTTCGAGGCACCTAT	GGG	11207	2	130.7787334	Y	11748	0	5	0
GAAACGTTCGAGGCACCTAT	GGG	11207	3	88.23625384	Y	11748	0	5	0
GTATTACGAAAGCGTGGACG	AGG	11218	1	170.8346255	Y	11759	0	5	0
GTATTACGAAAGCGTGGACG	AGG	11218	2	77.65210248	Y	11759	0	5	0
GTATTACGAAAGCGTGGACG	AGG	11218	3	82.82890931	Y	11759	0	5	0
ATCATAAAGCGTATATACAA	GGG	11298	1	16.41327497	Y	11839	0	2	1
ATCATAAAGCGTATATACAA	GGG	11298	2	38.98152805	Y	11839	0	2	1

ATCATAAAGCGTATATACAA	GGG	11298	3	2.051659371	Y	11839	0	2	1
TTCGACCATGATTTAAGTAA	TGG	11362	1	34.87820931	Y	11903	0	0	2
TTCGACCATGATTTAAGTAA	TGG	11362	2	71.80807799	Y	11903	0	0	2
TTCGACCATGATTTAAGTAA	TGG	11362	3	34.87820931	Y	11903	0	0	2
ACTTTCCATTACTTAAATCA	TGG	11373	1	184.6493434	Y	11914	6	177225	1
ACTTTCCATTACTTAAATCA	TGG	11373	2	155.9261122	Y	11914	6	177225	1
ACTTTCCATTACTTAAATCA	TGG	11373	3	38.98152805	Y	11914	6	177225	1
TTAATTAAGTTATCGATATC	CGG	11428	1	54.27545078	Y	11969	0	6	1
TTAATTAAGTTATCGATATC	CGG	11428	2	78.15664912	Y	11969	0	6	1
TTAATTAAGTTATCGATATC	CGG	11428	3	41.86963346	Y	11969	0	6	1
ACAGTATTAGACGCATGTAA	TGG	11468	1	102.5829686	Y	12009	0	3	0
ACAGTATTAGACGCATGTAA	TGG	11468	2	160.0294309	Y	12009	0	3	0
ACAGTATTAGACGCATGTAA	TGG	11468	3	92.3246717	Y	12009	0	3	0
AGCTAGTTATTTCTGTAATT	TGG	11496	1	20.48341607	Y	12037	0	3	0
AGCTAGTTATTTCTGTAATT	TGG	11496	2	29.93730041	Y	12037	0	3	0
AGCTAGTTATTTCTGTAATT	TGG	11496	3	3.15129478	Y	12037	0	3	0
TACTGATATTCAACTTCTGT	AGG	11619	1	154.4134442	Y	12160	0	1	0
TACTGATATTCAACTTCTGT	AGG	11619	2	196.9559238	Y	12160	0	1	0
TACTGATATTCAACTTCTGT	AGG	11619	3	72.47977994	Y	12160	0	1	0
GGATATCGAAAAAGAAGCGC	TGG	11670	1	570.3574513	Y	12211	0	2	0
GGATATCGAAAAAGAAGCGC	TGG	11670	2	928.265281	Y	12211	0	2	0
GGATATCGAAAAAGAAGCGC	TGG	11670	3	375.8962648	Y	12211	0	2	0
TCATACTCTAGTAATTCGTC	TGG	11697	1	648.5141004	Y	12238	2	8	10
TCATACTCTAGTAATTCGTC	TGG	11697	2	867.4767761	Y	12238	2	8	10
TCATACTCTAGTAATTCGTC	TGG	11697	3	417.4557528	Y	12238	2	8	10
AAAAAGAACGGGTTAACTCC	TGG	11839	1	42.17977889	Y	12380	0	10	0
AAAAAGAACGGGTTAACTCC	TGG	11839	2	32.8754159	Y	12380	0	10	0
AAAAAGAACGGGTTAACTCC	TGG	11839	3	26.67250724	Y	12380	0	10	0
ACTCTTTTTGCAACCATTCC	AGG	11841	1	461.1862589	Y	12382	17	0	1
ACTCTTTTTGCAACCATTCC	AGG	11841	2	789.0099816	Y	12382	17	0	1
ACTCTTTTTGCAACCATTCC	AGG	11841	3	220.8235483	Y	12382	17	0	1
GAACGGGTTAACTCCTGGAA	TGG	11844	1	141.5644966	Y	12385	0	7	0
GAACGGGTTAACTCCTGGAA	TGG	11844	2	184.6493434	Y	12385	0	7	0
GAACGGGTTAACTCCTGGAA	TGG	11844	3	22.56825308	Y	12385	0	7	0
ATGGTTGCAAAAAGAGTACG	TGG	11863	1	20.70722733	Y	12404	4	0	0
ATGGTTGCAAAAAGAGTACG	TGG	11863	2	20.70722733	Y	12404	4	0	0
ATGGTTGCAAAAAGAGTACG	TGG	11863	3	15.5304205	Y	12404	4	0	0
AAAAAGAGTACGTGGTGGCT	GGG	11871	1	269.4357037	Y	12412	0	0	4
AAAAAGAGTACGTGGTGGCT	GGG	11871	2	438.0299744	Y	12412	0	0	4

AAAAAGAGTACGTGGTGGCT	GGG	11871	3	173.3212129	Y	12412	0	0	4
GAAGCCTTAAACGCACCTTA	TGG	11902	1	240.0441464	Y	12443	0	2	0
GAAGCCTTAAACGCACCTTA	TGG	11902	2	878.1102108	Y	12443	0	2	0
GAAGCCTTAAACGCACCTTA	TGG	11902	3	320.0588619	Y	12443	0	2	0
GCAAAGAACGTGAAATAGCT	AGG	11981	1	33.08859519	Y	12522	0	4	0
GCAAAGAACGTGAAATAGCT	AGG	11981	2	33.08859519	Y	12522	0	4	0
GCAAAGAACGTGAAATAGCT	AGG	11981	3	11.02953173	Y	12522	0	4	0
AGCTAGGCAACGACGTAAAG	AGG	11997	2	15.5304205	Y	12538	0	0	8
AAGAGGTTGAACTACGTAAG	AGG	12014	1	181.1882391	Y	12555	0	3	0
AAGAGGTTGAACTACGTAAG	AGG	12014	2	165.6578186	Y	12555	0	3	0
AAGAGGTTGAACTACGTAAG	AGG	12014	3	36.23764782	Y	12555	0	3	0
AACATTCACGTGATCCGTAC	TGG	12065	1	731.0127856	Y	12606	2	3	0
AACATTCACGTGATCCGTAC	TGG	12065	2	915.2391728	Y	12606	2	3	0
AACATTCACGTGATCCGTAC	TGG	12065	3	467.699313	Y	12606	2	3	0
CACTCCATTTCTTGAACATT	TGG	12089	1	20.48341607	Y	12630	0	1	0
CACTCCATTTCTTGAACATT	TGG	12089	2	25.21035824	Y	12630	0	1	0
CACTCCATTTCTTGAACATT	TGG	12089	3	14.18082651	Y	12630	0	1	0
TCGCCGTATGTGTAATGTGC	TGG	12179	1	1674.165047	Y	12720	0	1	12
TCGCCGTATGTGTAATGTGC	TGG	12179	2	2046.029422	Y	12720	0	1	12
TCGCCGTATGTGTAATGTGC	TGG	12179	3	1027.201674	Y	12720	0	1	12
CCACAATTAGCATTTGCAAT	AGG	12264	1	63.0258956	Y	12805	10	0	1
CCACAATTAGCATTTGCAAT	AGG	12264	2	58.29895343	Y	12805	10	0	1
CCACAATTAGCATTTGCAAT	AGG	12264	3	42.54247953	Y	12805	10	0	1
TCCTCATGACCATTCTTTAA	CGG	12299	1	88.22135295	Y	12840	0	0	5
TCCTCATGACCATTCTTTAA	CGG	12299	2	182.597684	Y	12840	0	0	5
TCCTCATGACCATTCTTTAA	CGG	12299	3	65.65309987	Y	12840	0	0	5
TTTACGCGCGTTATCTGTCA	TGG	12377	1	328.2654994	Y	12918	0	1	0
TTTACGCGCGTTATCTGTCA	TGG	12377	2	309.800565	Y	12918	0	1	0
TTTACGCGCGTTATCTGTCA	TGG	12377	3	131.3061997	Y	12918	0	1	0
GTATCAGGATAACGAGCGAG	TGG	12454	1	243.3099211	Y	12995	12	0	0
GTATCAGGATAACGAGCGAG	TGG	12454	2	129.4201708	Y	12995	12	0	0
GTATCAGGATAACGAGCGAG	TGG	12454	3	93.18252298	Y	12995	12	0	0
AATAAGCGCCATTCACTACA	TGG	12468	1	147.7194747	Y	13009	0	4	1
AATAAGCGCCATTCACTACA	TGG	12468	2	326.21384	Y	13009	0	4	1
AATAAGCGCCATTCACTACA	TGG	12468	3	59.49812176	Y	13009	0	4	1
GCACATATCCATGTAGTGAA	TGG	12476	1	500.6048865	Y	13017	0	0	16
GCACATATCCATGTAGTGAA	TGG	12476	2	900.6784639	Y	13017	0	0	16
GCACATATCCATGTAGTGAA	TGG	12476	3	451.3650616	Y	13017	0	0	16
ATGGCGCTTATTACTTTCAC	GGG	12495	1	123.1277369	Y	13036	0	9	0

ATGGCGCTTATTACTTTCAC	GGG	12495	2	308.9048513	Y	13036	0	9	0
ATGGCGCTTATTACTTTCAC	GGG	12495	3	25.74207094	Y	13036	0	9	0
TTTTTCACACCTTGCCAATC	TGG	12511	1	127.469773	Y	13052	0	4	0
TTTTTCACACCTTGCCAATC	TGG	12511	2	299.6004883	Y	13052	0	4	0
TTTTTCACACCTTGCCAATC	TGG	12511	3	118.4755554	Y	13052	0	4	0
ACGGGCATATCGTACCAGAT	TGG	12513	1	124.4761438	Y	13054	1	1	0
ACGGGCATATCGTACCAGAT	TGG	12513	2	334.0372467	Y	13054	1	1	0
ACGGGCATATCGTACCAGAT	TGG	12513	3	132.3543808	Y	13054	1	1	0
CATATCGTACCAGATTGGCA	AGG	12518	1	668.8409549	Y	13059	8	4	1
CATATCGTACCAGATTGGCA	AGG	12518	2	1009.416411	Y	13059	8	4	1
CATATCGTACCAGATTGGCA	AGG	12518	3	262.6123995	Y	13059	8	4	1
ATTAGACGAGTTAATTAAAT	GGG	12681	1	4.72694217	Y	13222	4	0	3
ATTAGACGAGTTAATTAAAT	GGG	12681	2	9.45388434	Y	13222	4	0	3
ATTAGACGAGTTAATTAAAT	GGG	12681	3	4.72694217	Y	13222	4	0	3
TTTCAATATCAACTATGAAG	GGG	12812	1	10.35361366	Y	13353	0	9	0
TTTCAATATCAACTATGAAG	GGG	12812	2	31.06084099	Y	13353	0	9	0
CTTATCAAATACCGTGTCTT	TGG	12844	1	187.5020394	Y	13385	0	12	0
CTTATCAAATACCGTGTCTT	TGG	12844	2	248.9522876	Y	13385	0	12	0
CTTATCAAATACCGTGTCTT	TGG	12844	3	85.08495906	Y	13385	0	12	0
AAAAGAAGTAACCAAAGACA	CGG	12849	1	30.77489057	Y	13390	3	5	0
AAAAGAAGTAACCAAAGACA	CGG	12849	2	90.27301232	Y	13390	3	5	0
AAAAGAAGTAACCAAAGACA	CGG	12849	3	61.54978113	Y	13390	3	5	0
TTTGAAATGTACGAGATGGA	AGG	12886	1	434.9517867	Y	13427	0	6	0
TTTGAAATGTACGAGATGGA	AGG	12886	2	869.9035733	Y	13427	0	6	0
TTTGAAATGTACGAGATGGA	AGG	12886	3	473.9333147	Y	13427	0	6	0
AGTATAAAAGAATGTTTATA	TGG	12943	1	4.103318742	Y	13484	4	0	3
AGTATAAAAGAATGTTTATA	TGG	12943	2	36.92986868	Y	13484	4	0	3
AGTATAAAAGAATGTTTATA	TGG	12943	3	4.103318742	Y	13484	4	0	3
GTTTAAGATATAGAATGCTT	TGG	12964	1	69.32848516	Y	13505	4	0	988
GTTTAAGATATAGAATGCTT	TGG	12964	2	130.7787334	Y	13505	4	0	988
GTTTAAGATATAGAATGCTT	TGG	12964	3	48.84506909	Y	13505	4	0	988
ATATGACGATGACGTTAATC	TGG	13007	1	50.24356015	Y	13548	0	2	0
ATATGACGATGACGTTAATC	TGG	13007	2	35.35657936	Y	13548	0	2	0
ATATGACGATGACGTTAATC	TGG	13007	3	33.8058522	Y	13548	0	2	0
AATCTGGAAAGATGGGGAGT	TGG	13023	1	99.26578557	Y	13564	0	0	2
AATCTGGAAAGATGGGGAGT	TGG	13023	2	88.23625384	Y	13564	0	0	2
AATCTGGAAAGATGGGGAGT	TGG	13023	3	25.21035824	Y	13564	0	0	2
GATACAAAGACTTACTTTAT	AGG	13044	1	17.33212129	Y	13585	0	9	0
GATACAAAGACTTACTTTAT	AGG	13044	2	37.81553736	Y	13585	0	9	0

GATACAAAGACTTACTTTAT	AGG	13044	3	18.90776868	Y	13585	0	9	0
TTTTGTGCAAAGTACGAATG	AGG	13127	2	5.176806832	Y	13668	1	0	0
CAGTTTTAATACCGTATTCG	TGG	13140	1	93.18252298	Y	13681	15816	2	12974
CAGTTTTAATACCGTATTCG	TGG	13140	2	82.82890931	Y	13681	15816	2	12974
CAGTTTTAATACCGTATTCG	TGG	13140	3	41.41445466	Y	13681	15816	2	12974
CGACCTCTATGCTTGCAGTT	TGG	13183	1	115.0222595	Y	13724	0	8	2
CGACCTCTATGCTTGCAGTT	TGG	13183	2	55.14765865	Y	13724	0	8	2
CGACCTCTATGCTTGCAGTT	TGG	13183	3	36.23988997	Y	13724	0	8	2
GTTCCAAACTGCAAGCATAG	AGG	13196	1	15.5304205	Y	13737	0	1	0
GTTCCAAACTGCAAGCATAG	AGG	13196	2	51.76806832	Y	13737	0	1	0
GTTCCAAACTGCAAGCATAG	AGG	13196	3	5.176806832	Y	13737	0	1	0
TCGTCCCACACTCGATATTT	CGG	13231	1	26.78600563	Y	13772	1	0	1
TCGTCCCACACTCGATATTT	CGG	13231	2	37.81553736	Y	13772	1	0	1
TCGTCCCACACTCGATATTT	CGG	13231	3	33.08859519	Y	13772	1	0	1
CGAGTGTGGGACGAATATAC	AGG	13256	1	180.8147874	Y	13797	0	9	0
CGAGTGTGGGACGAATATAC	AGG	13256	2	231.9887839	Y	13797	0	9	0
CGAGTGTGGGACGAATATAC	AGG	13256	3	110.1016287	Y	13797	0	9	0
GGAAGAATACACGATGTTGT	AGG	13277	1	672.8014355	Y	13818	40	24899	2
GGAAGAATACACGATGTTGT	AGG	13277	2	761.0376894	Y	13818	40	24899	2
GGAAGAATACACGATGTTGT	AGG	13277	3	340.3398362	Y	13818	40	24899	2
AGTTCACTATGAAAACTACG	CGG	13321	1	108.7129435	Y	13862	14	0	5
AGTTCACTATGAAAACTACG	CGG	13321	2	82.82890931	Y	13862	14	0	5
AGTTCACTATGAAAACTACG	CGG	13321	3	82.82890931	Y	13862	14	0	5
GTAGAACTTATGCAAAGTAC	AGG	13382	1	215.2409305	Y	13923	10	3	0
GTAGAACTTATGCAAAGTAC	AGG	13382	2	375.2759739	Y	13923	10	3	0
GTAGAACTTATGCAAAGTAC	AGG	13382	3	277.2700171	Y	13923	10	3	0
CCTATTACAGATTCATCGTC	TGG	13462	1	571.9081785	Y	14003	11	98	4
CCTATTACAGATTCATCGTC	TGG	13462	2	829.9491787	Y	14003	11	98	4
CCTATTACAGATTCATCGTC	TGG	13462	3	350.7744847	Y	14003	11	98	4
TTAAAAGATTTTATGTTTGA	GGG	13550	1	24.61991245	Y	14091	1	0	0
TTAAAAGATTTTATGTTTGA	GGG	13550	2	24.61991245	Y	14091	1	0	0
TTAAAAGATTTTATGTTTGA	GGG	13550	3	6.154978113	Y	14091	1	0	0
AGATATGACTGTGAGGTTAA	AGG	13625	1	123.0995623	Y	14166	0	6	0
AGATATGACTGTGAGGTTAA	AGG	13625	2	367.2470274	Y	14166	0	6	0
AGATATGACTGTGAGGTTAA	AGG	13625	3	75.91139673	Y	14166	0	6	0
ATTTGAATCATCACATTTAT	TGG	13654	1	15.7564739	Y	14195	0	0	3
ATTTGAATCATCACATTTAT	TGG	13654	2	18.90776868	Y	14195	0	0	3
ATTTGAATCATCACATTTAT	TGG	13654	3	14.18082651	Y	14195	0	0	3
TGAATCATCACATTTATTGG	AGG	13657	1	51.76806832	Y	14198	0	0	10

TGAATCATCACATTTATTGG	AGG	13657	2	36.23764782	Y	14198	0	0	10
TGAATCATCACATTTATTGG	AGG	13657	3	5.176806832	Y	14198	0	0	10
TAGCGTCCATTACACCTAGT	TGG	13691	1	297.7973567	Y	14232	0	2	0
TAGCGTCCATTACACCTAGT	TGG	13691	2	431.7273849	Y	14232	0	2	0
TAGCGTCCATTACACCTAGT	TGG	13691	3	234.7714611	Y	14232	0	2	0
AAGAGAACGCAACAAAGAGC	TGG	13806	1	101.1074112	Y	14347	5	0	0
AAGAGAACGCAACAAAGAGC	TGG	13806	2	231.368493	Y	14347	5	0	0
AAGAGAACGCAACAAAGAGC	TGG	13806	3	76.29577652	Y	14347	5	0	0
AAAGATTTAATAAACGAATT	TGG	13873	1	4.72694217	Y	14414	0	0	6
AAAGATTTAATAAACGAATT	TGG	13873	2	1.57564739	Y	14414	0	0	6
AAAGATTTAATAAACGAATT	TGG	13873	3	9.45388434	Y	14414	0	0	6
CGGATTGTTCTATTTGTTCA	CGG	13956	1	498.5532272	Y	14497	0	12	0
CGGATTGTTCTATTTGTTCA	CGG	13956	2	718.0807799	Y	14497	0	12	0
CGGATTGTTCTATTTGTTCA	CGG	13956	3	432.9001273	Y	14497	0	12	0
ATACGCACTAGCACTTATAA	CGG	13976	1	84.11803421	Y	14517	0	0	2
ATACGCACTAGCACTTATAA	CGG	13976	2	147.7194747	Y	14517	0	0	2
ATACGCACTAGCACTTATAA	CGG	13976	3	22.56825308	Y	14517	0	0	2
GCGTTTGATGAAATACTTGA	GGG	14075	1	129.2545404	Y	14616	0	4	0
GCGTTTGATGAAATACTTGA	GGG	14075	2	203.1142777	Y	14616	0	4	0
GCGTTTGATGAAATACTTGA	GGG	14075	3	102.5829686	Y	14616	0	4	0
ATTCAACATTCAGTTAAAGA	AGG	14111	1	22.56825308	Y	14652	0	13479	0
ATTCAACATTCAGTTAAAGA	AGG	14111	2	20.51659371	Y	14652	0	13479	0
ATTCAACATTCAGTTAAAGA	AGG	14111	3	12.30995623	Y	14652	0	13479	0
AGTTGTCTATAAATATGAGG	AGG	14173	1	77.65210248	Y	14714	0	8	1
AGTTGTCTATAAATATGAGG	AGG	14173	2	36.23764782	Y	14714	0	8	1
AGTTGTCTATAAATATGAGG	AGG	14173	3	5.176806832	Y	14714	0	8	1
CTATCTCGTAAGTTCAGCGT	TGG	14261	1	816.185348	Y	14802	0	18	0
CTATCTCGTAAGTTCAGCGT	TGG	14261	2	800.4288741	Y	14802	0	18	0
CTATCTCGTAAGTTCAGCGT	TGG	14261	3	346.6424258	Y	14802	0	18	0
CTTTGAATGGTTTAATACAT	TGG	14324	1	53.57201126	Y	14865	0	7756	0
CTTTGAATGGTTTAATACAT	TGG	14324	2	29.93730041	Y	14865	0	7756	0
CTTTGAATGGTTTAATACAT	TGG	14324	3	14.18082651	Y	14865	0	7756	0
TGTACGTCTAACGGCTTACC	TGG	14357	1	291.2265616	Y	14898	1	0	0
TGTACGTCTAACGGCTTACC	TGG	14357	2	564.4646881	Y	14898	1	0	0
TGTACGTCTAACGGCTTACC	TGG	14357	3	136.774136	Y	14898	1	0	0
TCATCAAGTTGTACGTCTAA	CGG	14366	1	164.1327497	Y	14907	0	3	0
TCATCAAGTTGTACGTCTAA	CGG	14366	2	330.3171587	Y	14907	0	3	0
TCATCAAGTTGTACGTCTAA	CGG	14366	3	131.3061997	Y	14907	0	3	0
TGATGAATTAGCTGACATGT	TGG	14399	1	215.8636924	Y	14940	42	10	0

TGATGAATTAGCTGACATGT	TGG	14399	2	315.129478	Y	14940	42	10	0
TGATGAATTAGCTGACATGT	TGG	14399	3	92.96319601	Y	14940	42	10	0
TTGAGTATTGCGAATCAAGT	AGG	14430	1	75.63107472	Y	14971	0	2	0
TTGAGTATTGCGAATCAAGT	AGG	14430	2	127.6274386	Y	14971	0	2	0
TTGAGTATTGCGAATCAAGT	AGG	14430	3	64.60154299	Y	14971	0	2	0
ATAACAATTGCTTGTTGGTC	GGG	14579	1	159.724898	Y	15120	0	2	0
ATAACAATTGCTTGTTGGTC	GGG	14579	2	352.3252119	Y	15120	0	2	0
ATAACAATTGCTTGTTGGTC	GGG	14579	3	152.9016985	Y	15120	0	2	0
TCACTATAACAATTGCTTGT	TGG	14584	1	584.5651817	Y	15125	14	0	0
TCACTATAACAATTGCTTGT	TGG	14584	2	606.6242452	Y	15125	14	0	0
TCACTATAACAATTGCTTGT	TGG	14584	3	321.4320676	Y	15125	14	0	0
TTTTGTATGCGTCAATGAGT	TGG	14638	1	294.6460619	Y	15179	0	3	3
TTTTGTATGCGTCAATGAGT	TGG	14638	2	311.9781832	Y	15179	0	3	3
TTTTGTATGCGTCAATGAGT	TGG	14638	3	135.5056755	Y	15179	0	3	3
CAAGATGGAACAGCAGACGC	AGG	14703	1	250.2873644	Y	15244	0	0	3
CAAGATGGAACAGCAGACGC	AGG	14703	2	538.4124717	Y	15244	0	0	3
CAAGATGGAACAGCAGACGC	AGG	14703	3	104.2088655	Y	15244	0	0	3
AGATCGAGTCAAGGAGGTTT	TGG	14751	1	29.93730041	Y	15292	0	0	2
AGATCGAGTCAAGGAGGTTT	TGG	14751	2	53.57201126	Y	15292	0	0	2
AGATCGAGTCAAGGAGGTTT	TGG	14751	3	17.33212129	Y	15292	0	0	2
GACGGCTAATGATGATGTAG	AGG	16066	1	31.06084099	Y	15771	0	0	6
GACGGCTAATGATGATGTAG	AGG	16066	2	134.5969776	Y	15771	0	0	6
GACGGCTAATGATGATGTAG	AGG	16066	3	119.0665571	Y	15771	0	0	6
CGATCAGTCTGATTTGATGA	GGG	16114	1	18.46493434	Y	15819	8	0	0
CGATCAGTCTGATTTGATGA	GGG	16114	2	88.22135295	Y	15819	8	0	0
AAGTTAAGAAGTAACGCGCT	AGG	21588	1	59.87460082	N	21588	8	0	0
AAGTTAAGAAGTAACGCGCT	AGG	21588	2	129.203086	N	21588	8	0	0
AAGTTAAGAAGTAACGCGCT	AGG	21588	3	113.4466121	N	21588	8	0	0
GGAGACCTAATCATATTACA	AGG	31489	1	36.92986868	Y	31867	0	2	0
GGAGACCTAATCATATTACA	AGG	31489	2	49.2398249	Y	31867	0	2	0
GGAGACCTAATCATATTACA	AGG	31489	3	4.103318742	Y	31867	0	2	0
TCCGTTTTATCAGTGCCTAT	CGG	31536	1	565.657413	Y	31914	0	5	0
TCCGTTTTATCAGTGCCTAT	CGG	31536	2	962.7205553	Y	31914	0	5	0
TCCGTTTTATCAGTGCCTAT	CGG	31536	3	397.0631423	Y	31914	0	5	0
AACTTCAATTGCAGAACAAA	AGG	31579	1	14.3616156	Y	31957	0	3	0
AACTTCAATTGCAGAACAAA	AGG	31579	2	18.46493434	Y	31957	0	3	0
AACTTCAATTGCAGAACAAA	AGG	31579	3	22.56825308	Y	31957	0	3	0
ATATAAATAATTTTCGTTCT	AGG	32433	1	25.21035824	Y	32825	0	0	4
ATATAAATAATTTTCGTTCT	AGG	32433	2	6.30258956	Y	32825	0	0	4

ATATAAATAATTTTCGTTCT	AGG	32433	3	20.48341607	Y	32825	0	0	4
TATATGACGCAAGCTCGTCC	AGG	32467	1	1179.172936	Y	32859	0	7	6
TATATGACGCAAGCTCGTCC	AGG	32467	2	1712.312936	Y	32859	0	7	6
TATATGACGCAAGCTCGTCC	AGG	32467	3	800.1752171	Y	32859	0	7	6
CTATCAATAAATTGTCCGTT	GGG	32502	1	42.54247953	Y	32894	0	0	4
CTATCAATAAATTGTCCGTT	GGG	32502	2	66.17719038	Y	32894	0	0	4
CTATCAATAAATTGTCCGTT	GGG	32502	3	50.42071648	Y	32894	0	0	4
CTTGTTAAAAACGGCGGTCA	CGG	32545	1	1460.781472	Y	32937	0	1	8
CTTGTTAAAAACGGCGGTCA	CGG	32545	2	1920.353171	Y	32937	0	1	8
CTTGTTAAAAACGGCGGTCA	CGG	32545	3	699.6158455	Y	32937	0	1	8
TTATTTCTCCAGTTCTATAT	TGG	32642	1	111.8709647	Y	33034	0	12	0
TTATTTCTCCAGTTCTATAT	TGG	32642	2	77.20672211	Y	33034	0	12	0
TTATTTCTCCAGTTCTATAT	TGG	32642	3	48.84506909	Y	33034	0	12	0
TATCTGTCGTTAAATATATT	CGG	32694	1	25.21035824	Y	33086	0	10	0
TATCTGTCGTTAAATATATT	CGG	32694	2	31.5129478	Y	33086	0	10	0
TATCTGTCGTTAAATATATT	CGG	32694	3	12.60517912	Y	33086	0	10	0
GTATCTGAAGTGTATTCCAT	AGG	32877	1	174.8968603	Y	33269	0	6	0
GTATCTGAAGTGTATTCCAT	AGG	32877	2	116.5979069	Y	33269	0	6	0
GTATCTGAAGTGTATTCCAT	AGG	32877	3	80.35801689	Y	33269	0	6	0
TCAGATACACAACCTATGCA	AGG	32908	1	1760.32374	Y	33300	0	4	0
TCAGATACACAACCTATGCA	AGG	32908	2	2168.603955	Y	33300	0	4	0
TCAGATACACAACCTATGCA	AGG	32908	3	1520.279594	Y	33300	0	4	0
GCCAACCCTAACTACTTACA	AGG	32983	1	94.37633107	Y	33375	0	3	0
GCCAACCCTAACTACTTACA	AGG	32983	2	162.0810903	Y	33375	0	3	0
GCCAACCCTAACTACTTACA	AGG	32983	3	49.2398249	Y	33375	0	3	0
TTTAAACGACGTATCGATAT	TGG	33034	1	144.9595599	Y	33426	0	0	2
TTTAAACGACGTATCGATAT	TGG	33034	2	124.4761438	Y	33426	0	0	2
TTTAAACGACGTATCGATAT	TGG	33034	3	61.45024821	Y	33426	0	0	2
GGTGTGAATAATAACTTTAA	AGG	33058	2	2.051659371	Y	33450	0	3	0
AGAATTGGTTAACACCTCTT	TGG	33188	1	185.926392	Y	33580	0	4	4
AGAATTGGTTAACACCTCTT	TGG	33188	2	518.3879913	Y	33580	0	4	4
AGAATTGGTTAACACCTCTT	TGG	33188	3	85.08495906	Y	33580	0	4	4
AGATATGCTGGGTTCTGTAT	TGG	33270	1	764.1889842	Y	33662	0	0	3
AGATATGCTGGGTTCTGTAT	TGG	33270	2	540.4470548	Y	33662	0	0	3
AGATATGCTGGGTTCTGTAT	TGG	33270	3	248.9522876	Y	33662	0	0	3
TAATATCACTTAGATATGCT	GGG	33281	1	48.84506909	Y	33673	0	1	0
TAATATCACTTAGATATGCT	GGG	33281	2	116.5979069	Y	33673	0	1	0
TAATATCACTTAGATATGCT	GGG	33281	3	33.08859519	Y	33673	0	1	0
ATTTCTACCTGTGCTGTTTC	TGG	33445	1	316.9686325	Y	33837	0	6	0

ATTTCTACCTGTGCTGTTTC	TGG	33445	2	347.0527395	Y	33837	0	6	0
ATTTCTACCTGTGCTGTTTC	TGG	33445	3	326.5831409	Y	33837	0	6	0
GTACTTACCAGAAACAGCAC	AGG	33454	1	123.7480278	Y	33846	13	0	2
GTACTTACCAGAAACAGCAC	AGG	33454	2	204.3858403	Y	33846	13	0	2
GTACTTACCAGAAACAGCAC	AGG	33454	3	148.8698078	Y	33846	13	0	2
AGGGATATGTTCCCAATAAC	CGG	33544	1	59.85806857	Y	33936	0	4	0
AGGGATATGTTCCCAATAAC	CGG	33544	2	144.5277718	Y	33936	0	4	0
AGGGATATGTTCCCAATAAC	CGG	33544	3	52.72472361	Y	33936	0	4	0
GTCCGCAAAACGCCGGTTAT	TGG	33548	1	100.841433	Y	33940	0	0	2
GTCCGCAAAACGCCGGTTAT	TGG	33548	2	155.9890916	Y	33940	0	0	2
GTCCGCAAAACGCCGGTTAT	TGG	33548	3	75.63107472	Y	33940	0	0	2
TCCGCAAAACGCCGGTTATT	GGG	33549	1	214.288045	Y	33941	0	0	2
TCCGCAAAACGCCGGTTATT	GGG	33549	2	178.0481551	Y	33941	0	0	2
TCCGCAAAACGCCGGTTATT	GGG	33549	3	66.17719038	Y	33941	0	0	2
TTATCAGATTTAAAAATCGT	TGG	33598	1	20.48341607	Y	33990	0	3	1
TTATCAGATTTAAAAATCGT	TGG	33598	2	15.7564739	Y	33990	0	3	1
TTATCAGATTTAAAAATCGT	TGG	33598	3	7.87823695	Y	33990	0	3	1
GCAATACCTTTAAAGTCTTT	AGG	33654	1	28.36165302	Y	34046	0	0	5
GCAATACCTTTAAAGTCTTT	AGG	33654	2	74.05542733	Y	34046	0	0	5
GCAATACCTTTAAAGTCTTT	AGG	33654	3	17.33212129	Y	34046	0	0	5
ACTTTAAAGGTATTGCAGGT	TGG	33677	1	119.7492016	Y	34069	0	4	0
ACTTTAAAGGTATTGCAGGT	TGG	33677	2	115.0222595	Y	34069	0	4	0
ACTTTAAAGGTATTGCAGGT	TGG	33677	3	51.99636387	Y	34069	0	4	0
AATACTTGTGTTGTGTTACC	CGG	33708	1	810.7201619	Y	34100	0	0	6
AATACTTGTGTTGTGTTACC	CGG	33708	2	1213.288934	Y	34100	0	0	6
AATACTTGTGTTGTGTTACC	CGG	33708	3	829.6390333	Y	34100	0	0	6
AACTTTGGTACTGGTGGCGT	TGG	33790	1	1002.11174	Y	34182	3	0	0
AACTTTGGTACTGGTGGCGT	TGG	33790	2	720.0708572	Y	34182	3	0	0
AACTTTGGTACTGGTGGCGT	TGG	33790	3	348.2180732	Y	34182	3	0	0
GAGTTTATTCGAGGGAAAGG	TGG	33819	1	72.47529565	Y	34211	0	0	4
GAGTTTATTCGAGGGAAAGG	TGG	33819	2	88.00571614	Y	34211	0	0	4
GAGTTTATTCGAGGGAAAGG	TGG	33819	3	51.76806832	Y	34211	0	0	4
CGATAGTAGACGCAATTAAT	GGG	34096	1	39.39118475	Y	34488	2	0	0
CGATAGTAGACGCAATTAAT	GGG	34096	2	61.45024821	Y	34488	2	0	0
CGATAGTAGACGCAATTAAT	GGG	34096	3	61.45024821	Y	34488	2	0	0
ATTTTTAAAACATTCAGGCA	AGG	34145	1	217.4758933	Y	34537	16	0	0
ATTTTTAAAACATTCAGGCA	AGG	34145	2	26.67157182	Y	34537	16	0	0
ATTTTTAAAACATTCAGGCA	AGG	34145	3	121.0479029	Y	34537	16	0	0
CAAGATACTATCGAAGCTGT	CGG	34302	1	244.2253455	Y	34694	0	1	1

CAAGATACTATCGAAGCTGT	CGG	34302	2	300.9486515	Y	34694	0	1	1
CAAGATACTATCGAAGCTGT	CGG	34302	3	154.4134442	Y	34694	0	1	1
GAAGTTGAACAACAAATCAA	TGG	34542	1	59.49812176	Y	34934	3	3	0
GAAGTTGAACAACAAATCAA	TGG	34542	2	92.3246717	Y	34934	3	3	0
GAAGTTGAACAACAAATCAA	TGG	34542	3	34.87820931	Y	34934	3	3	0
TCTAAACTTACAGATGATTA	CGG	34608	1	77.9630561	Y	35000	0	0	1
TCTAAACTTACAGATGATTA	CGG	34608	2	203.1142777	Y	35000	0	0	1
TCTAAACTTACAGATGATTA	CGG	34608	3	88.22135295	Y	35000	0	0	1
TAAATCTTTTTAGAGTTATA	AGG	35231	1	28.72323119	Y	35623	0	4	3
TAAATCTTTTTAGAGTTATA	AGG	35231	2	75.91139673	Y	35623	0	4	3
TAAATCTTTTTAGAGTTATA	AGG	35231	3	36.92986868	Y	35623	0	4	3
TACACACGATCAATCACAAA	CGG	35268	1	24.61991245	Y	35660	0	0	3
TACACACGATCAATCACAAA	CGG	35268	2	108.7379467	Y	35660	0	0	3
TACACACGATCAATCACAAA	CGG	35268	3	26.67157182	Y	35660	0	0	3
CACAAACTATTCTATTTTGT	TGG	35396	1	64.60154299	Y	35788	0	3	0
CACAAACTATTCTATTTTGT	TGG	35396	2	53.57201126	Y	35788	0	3	0
CACAAACTATTCTATTTTGT	TGG	35396	3	17.33212129	Y	35788	0	3	0
TTGGTAAGTGGAACTTATCC	AGG	35415	1	271.6873993	Y	35807	0	7	0
TTGGTAAGTGGAACTTATCC	AGG	35415	2	428.6209884	Y	35807	0	7	0
TTGGTAAGTGGAACTTATCC	AGG	35415	3	202.2148223	Y	35807	0	7	0
AATCCCTCAATAACGCCACC	TGG	35417	1	98.31610226	Y	35809	0	13	0
AATCCCTCAATAACGCCACC	TGG	35417	2	243.7743103	Y	35809	0	13	0
AATCCCTCAATAACGCCACC	TGG	35417	3	130.2610819	Y	35809	0	13	0
TTACTCAATTGAATCGCGTT	AGG	35456	1	138.6569703	Y	35848	12	0	0
TTACTCAATTGAATCGCGTT	AGG	35456	2	151.2621494	Y	35848	12	0	0
TTACTCAATTGAATCGCGTT	AGG	35456	3	78.7823695	Y	35848	12	0	0
GTTGACTCAGACGGCAACGG	TGG	35505	1	88.00571614	Y	35897	0	0	3
GTTGACTCAGACGGCAACGG	TGG	35505	2	62.12168198	Y	35897	0	0	3
GTTGACTCAGACGGCAACGG	TGG	35505	3	15.5304205	Y	35897	0	0	3
AACGATGTGTACTTTGATTT	AGG	35583	1	140.2326177	Y	35975	0	0	1
AACGATGTGTACTTTGATTT	AGG	35583	2	143.3839125	Y	35975	0	0	1
AACGATGTGTACTTTGATTT	AGG	35583	3	67.75283777	Y	35975	0	0	1
TTTAGTTATAGTAACTTTGT	TGG	35616	1	66.17719038	Y	36008	0	3	0
TTTAGTTATAGTAACTTTGT	TGG	35616	2	80.35801689	Y	36008	0	3	0
TTTAGTTATAGTAACTTTGT	TGG	35616	3	33.08859519	Y	36008	0	3	0
GTTACTATAACTAAAATTAT	GGG	35640	1	1.57564739	Y	36032	0	0	1
GTTACTATAACTAAAATTAT	GGG	35640	2	4.72694217	Y	36032	0	0	1
GTTACTATAACTAAAATTAT	GGG	35640	3	1.57564739	Y	36032	0	0	1
CTTTGTTGGTTTGATGCATT	CGG	35842	1	209.5611029	Y	36234	0	0	2

CTTTGTTGGTTTGATGCATT	CGG	35842	2	318.2807728	Y	36234	0	0	2
CTTTGTTGGTTTGATGCATT	CGG	35842	3	155.9890916	Y	36234	0	0	2
AACTAACAAAACAAATACTG	AGG	36016	2	5.176806832	Y	36408	3	0	0
AACTAACAAAACAAATACTG	AGG	36016	3	5.176806832	Y	36408	3	0	0
CATTAAAACTTTTTATGTGT	GGG	36082	1	25.21035824	Y	36474	0	0	1
CATTAAAACTTTTTATGTGT	GGG	36082	2	23.63471085	Y	36474	0	0	1
CATTAAAACTTTTTATGTGT	GGG	36082	3	1.57564739	Y	36474	0	0	1
ATTAAGTGGTACGTAGACAT	GGG	36125	1	132.3543808	Y	36517	9	0	0
ATTAAGTGGTACGTAGACAT	GGG	36125	2	88.23625384	Y	36517	9	0	0
ATTAAGTGGTACGTAGACAT	GGG	36125	3	20.48341607	Y	36517	9	0	0
ATGTGGATTCTCGGTTTGAT	AGG	36489	1	29.93730041	Y	36881	0	0	5
ATGTGGATTCTCGGTTTGAT	AGG	36489	2	39.39118475	Y	36881	0	0	5
ATGTGGATTCTCGGTTTGAT	AGG	36489	3	59.87460082	Y	36881	0	0	5
TCCTAAAATCCCTTTAAGCA	TGG	36563	1	432.9001273	Y	36955	0	4	0
TCCTAAAATCCCTTTAAGCA	TGG	36563	2	572.4129645	Y	36955	0	4	0
TCCTAAAATCCCTTTAAGCA	TGG	36563	3	240.0441464	Y	36955	0	4	0
GATTTTAGGATATAGCTTCT	GGG	36592	1	185.926392	Y	36984	0	5	0
GATTTTAGGATATAGCTTCT	GGG	36592	2	160.7160338	Y	36984	0	5	0
GATTTTAGGATATAGCTTCT	GGG	36592	3	81.93366428	Y	36984	0	5	0
ATAGCTTCTGGGCGTGCTTC	TGG	36603	1	144.2176263	Y	36995	0	0	1
ATAGCTTCTGGGCGTGCTTC	TGG	36603	2	520.1138911	Y	36995	0	0	1
ATAGCTTCTGGGCGTGCTTC	TGG	36603	3	303.3222335	Y	36995	0	0	1
TTCTGGGCGTGCTTCTGGTT	TGG	36608	1	304.0999463	Y	37000	0	0	5
TTCTGGGCGTGCTTCTGGTT	TGG	36608	2	467.9672748	Y	37000	0	0	5
TTCTGGGCGTGCTTCTGGTT	TGG	36608	3	283.6165302	Y	37000	0	0	5
GCCAATGACTCAACAATTAC	TGG	37201	1	59.85806857	N	37201	0	9	0
GCCAATGACTCAACAATTAC	TGG	37201	2	77.53635825	N	37201	0	9	0
GCCAATGACTCAACAATTAC	TGG	37201	3	19.84930771	N	37201	0	9	0
CCGCCCATGATTGCTTTTGC	TGG	38337	1	714.8852231	Y	38756	1	0	0
CCGCCCATGATTGCTTTTGC	TGG	38337	2	1022.549493	Y	38756	1	0	0
CCGCCCATGATTGCTTTTGC	TGG	38337	3	644.1720643	Y	38756	1	0	0
TCTTAGGATTCCATCTAATT	CGG	38411	1	7.87823695	Y	38830	5	0	0
TCTTAGGATTCCATCTAATT	CGG	38411	2	18.90776868	Y	38830	5	0	0
TCTTAGGATTCCATCTAATT	CGG	38411	3	4.72694217	Y	38830	5	0	0
CGATTGTACTTGCTTGATGT	TGG	38453	1	699.5874412	Y	38893	0	0	12
CGATTGTACTTGCTTGATGT	TGG	38453	2	715.3439151	Y	38893	0	0	12
CGATTGTACTTGCTTGATGT	TGG	38453	3	499.4802226	Y	38893	0	0	12
TAAAAGGTATCTACTTCACA	AGG	38516	1	75.91139673	Y	38956	8	0	0
TAAAAGGTATCTACTTCACA	AGG	38516	2	43.08484679	Y	38956	8	0	0

TAAAAGGTATCTACTTCACA	AGG	38516	3	61.54978113	Y	38956	8	0	0
AAAAGGTATCTACTTCACAA	GGG	38517	1	36.92986868	Y	38957	9	0	0
AAAAGGTATCTACTTCACAA	GGG	38517	2	192.8559809	Y	38957	9	0	0
AAAAGGTATCTACTTCACAA	GGG	38517	3	61.54978113	Y	38957	9	0	0
TCCATAAACAAATGTAATCT	AGG	38712	1	12.60517912	Y	39152	0	4	0
TCCATAAACAAATGTAATCT	AGG	38712	2	18.90776868	Y	39152	0	4	0
TCCATAAACAAATGTAATCT	AGG	38712	3	1.57564739	Y	39152	0	4	0
ACCTAGATTACATTTGTTTA	TGG	38727	1	182.597684	Y	39167	5	0	3
ACCTAGATTACATTTGTTTA	TGG	38727	2	262.6123995	Y	39167	5	0	3
ACCTAGATTACATTTGTTTA	TGG	38727	3	65.65309987	Y	39167	5	0	3
TACAACATCGTCGATAATAA	GGG	38761	1	14.3616156	Y	39201	0	0	4
TACAACATCGTCGATAATAA	GGG	38761	2	34.87820931	Y	39201	0	0	4
TACAACATCGTCGATAATAA	GGG	38761	3	18.46493434	Y	39201	0	0	4
TCTTTCAATATCGTTGATAG	TGG	38911	1	139.7737845	Y	39351	0	0	2
TCTTTCAATATCGTTGATAG	TGG	38911	2	103.5361366	Y	39351	0	0	2
TCTTTCAATATCGTTGATAG	TGG	38911	3	93.18252298	Y	39351	0	0	2
ACTACAGTACCGTTTTTACC	GGG	39159	1	256.1801277	Y	39599	0	7	1
ACTACAGTACCGTTTTTACC	GGG	39159	2	258.0410003	Y	39599	0	7	1
ACTACAGTACCGTTTTTACC	GGG	39159	3	177.4031877	Y	39599	0	7	1
ATTTGCCACATTTTAGTGTC	AGG	39192	1	184.5365326	Y	39632	0	4	0
ATTTGCCACATTTTAGTGTC	AGG	39192	2	174.9220242	Y	39632	0	4	0
TCAATCCTGACACTAAAATG	TGG	39203	2	10.35361366	Y	39643	0	0	1
TCAATCCTGACACTAAAATG	TGG	39203	3	15.5304205	Y	39643	0	0	1
ACTAAAATGTGGCAAATTGA	TGG	39214	1	151.8227935	Y	39654	0	2	2
ACTAAAATGTGGCAAATTGA	TGG	39214	2	320.0588619	Y	39654	0	2	2
ACTAAAATGTGGCAAATTGA	TGG	39214	3	75.91139673	Y	39654	0	2	2
TCGTCATGATTATGATTTTT	TGG	39534	1	63.0258956	Y	39974	0	5	1
TCGTCATGATTATGATTTTT	TGG	39534	2	45.69377431	Y	39974	0	5	1
TCGTCATGATTATGATTTTT	TGG	39534	3	33.08859519	Y	39974	0	5	1
AAAATCATAATCATGACGAG	CGG	39554	1	62.12168198	Y	39994	0	0	1
AAAATCATAATCATGACGAG	CGG	39554	2	41.41445466	Y	39994	0	0	1
AAAATCATAATCATGACGAG	CGG	39554	3	31.06084099	Y	39994	0	0	1
CCTGTCCAAATTTTAACCGT	CGG	39633	1	77.20672211	Y	40073	3	0	0
CCTGTCCAAATTTTAACCGT	CGG	39633	2	36.23988997	Y	40073	3	0	0
CCTGTCCAAATTTTAACCGT	CGG	39633	3	31.5129478	Y	40073	3	0	0
CTCAACCGACGGTTAAAATT	TGG	39644	1	3.15129478	Y	40084	10	0	0
GGTTTTATTACAAAAGATGA	AGG	39821	1	47.18816553	Y	40261	0	0	5
GGTTTTATTACAAAAGATGA	AGG	39821	2	86.16969358	Y	40261	0	0	5
GGTTTTATTACAAAAGATGA	AGG	39821	3	18.46493434	Y	40261	0	0	5

CTATCCGGATATTTATTTTT	AGG	39848	1	29.93730041	Y	40288	0	6	301
CTATCCGGATATTTATTTTT	AGG	39848	2	22.05906346	Y	40288	0	6	301
CTATCCGGATATTTATTTTT	AGG	39848	3	25.21035824	Y	40288	0	6	301
GTTGATTCTTCTATGCTATC	CGG	39863	1	519.1834548	Y	40303	0	0	8
GTTGATTCTTCTATGCTATC	CGG	39863	2	829.6390333	Y	40303	0	0	8
GTTGATTCTTCTATGCTATC	CGG	39863	3	345.5020124	Y	40303	0	0	8
TGAACCTAATAAAAGTTATC	AGG	39926	1	49.00297841	Y	40366	53	0	3
TGAACCTAATAAAAGTTATC	AGG	39926	2	70.71315872	Y	40366	53	0	3
TGAACCTAATAAAAGTTATC	AGG	39926	3	16.74785338	Y	40366	53	0	3
ACAATAGAAAATGTACGTAG	CGG	39951	1	103.5361366	Y	40391	0	4	0
ACAATAGAAAATGTACGTAG	CGG	39951	2	93.18252298	Y	40391	0	4	0
ACAATAGAAAATGTACGTAG	CGG	39951	3	103.5361366	Y	40391	0	4	0
TCAAGTGATTTAGGAATATC	AGG	40005	1	70.71315872	Y	40445	11	0	0
TCAAGTGATTTAGGAATATC	AGG	40005	2	84.97984864	Y	40445	11	0	0
TCAAGTGATTTAGGAATATC	AGG	40005	3	62.64937747	Y	40445	11	0	0
GTGACATACAACATCCCTGA	AGG	40055	1	379.5569836	Y	40495	0	2	0
GTGACATACAACATCCCTGA	AGG	40055	2	379.5569836	Y	40495	0	2	0
GTGACATACAACATCCCTGA	AGG	40055	3	162.0810903	Y	40495	0	2	0
ATCAACTTTAGTAATCCTTC	AGG	40056	1	17.98843511	N	40056	0	0	3
ATCAACTTTAGTAATCCTTC	AGG	40056	2	42.48992432	N	40056	0	0	3
ATCAACTTTAGTAATCCTTC	AGG	40056	3	13.95654449	N	40056	0	0	3
ACTATAAGTGATGTTTATTC	AGG	40092	1	61.71894117	Y	40532	0	0	2
ACTATAAGTGATGTTTATTC	AGG	40092	2	77.84650368	Y	40532	0	0	2
ACTATAAGTGATGTTTATTC	AGG	40092	3	49.93341471	Y	40532	0	0	2
GGAAAAAAGGAGCAAACAAA	TGG	40176	1	32.82654994	Y	40616	0	1	1
GGAAAAAAGGAGCAAACAAA	TGG	40176	2	45.13650616	Y	40616	0	1	1
GGAAAAAAGGAGCAAACAAA	TGG	40176	3	16.41327497	Y	40616	0	1	1
GAAACACTTTCTTCATCTAC	TGG	40266	1	1017.587166	Y	40706	4	0	0
GAAACACTTTCTTCATCTAC	TGG	40266	2	1563.132982	Y	40706	4	0	0
GAAACACTTTCTTCATCTAC	TGG	40266	3	678.9083528	Y	40706	4	0	0
AACATCTCAAGAAGGGAAAT	GGG	40356	1	7.87823695	Y	40796	3	0	0
AACATCTCAAGAAGGGAAAT	GGG	40356	2	7.87823695	Y	40796	3	0	0
AACATCTCAAGAAGGGAAAT	GGG	40356	3	1.57564739	Y	40796	3	0	0

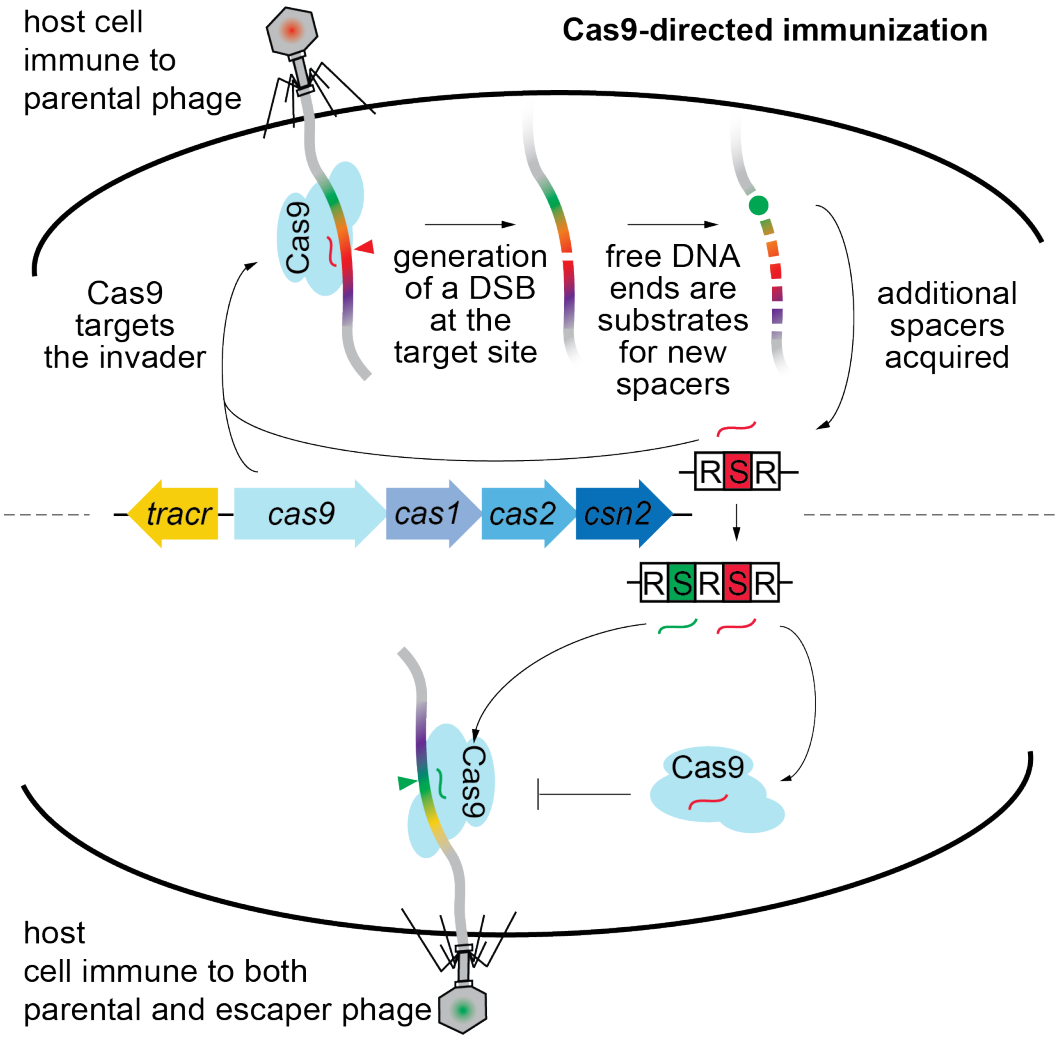


Figure 3.4. A model for type II-A CRISPR-Cas primed immunity against bacteriophages

A schematic for II-A primed spacer acquisition in pre-immune cells. Following phage infection, Cas9 utilizes a pre-existing spacer (red) to produce a dsDNA DNA break (DSB) that is used a substrate to generate new spacers for acquisition. The newly acquired spacer (green) can then be used provide immunity against the original, parental virus as well as related escaper phages.

Primed spacer acquisition has clear advantages for the bacterial population: it increases the diversity of spacer sequences to prevent the raise of escaper phages (van Houte et al., 2016) and to generate new immunity against related viruses. In type I systems, the presence of a target mutation would trigger additional spacer acquisition from other regions of the escaper or related phage to neutralize these

threats. In type II systems, however, given the strong priming effect of fully targeting spacers, spacer diversity is generated not only after, but also before the cell is infected with an escaper or a related phage. This is a unique feature of type II primed spacer acquisition that allows "preventive" immunity against future threats that could overcome the defense provided by the initial spacer acquired. We believe that cleavage-mediated spacer acquisition is an integral part of the type II-A CRISPR-Cas immune response that immediately follows the acquisition of the first spacers by naïve cells (an argument supported by the data in **Figure 2.11A**) and leads to a more rapid increase in the diversity and abundance of spacers in the population that limits the rise of escapers. Phage escapers, however, are only one of the many threats that bacteria face in their natural habitats. How cleavage-mediated spacer acquisition impacts the dynamics of the virus-host co-evolution in ecological contexts remains largely unknown. In **Figure 2.13C** we explored conditions in which the host is exposed to two phages and showed that this mechanism can immunize the host against both infections if the viral genomes are related. In more complex environments, the number of co-existing phages, their genetic divergence and their relative frequencies will most likely affect the benefits of type II priming. Finally, since the strength of the immunity provided by type II spacers decreases as their positions within the CRISPR array upon new integration events (McGinn and Marraffini, 2016), the accelerated acquisition mediated by Cas9 cleavage will also "downgrade" the previously acquired spacers. How this potential cost of priming affects the overall type II CRISPR-Cas immune response is not known.

CHAPTER 4: PERSEPCTIVES

The ever-present competition between bacteria and bacteriophages has given rise to an "evolutionary arms race", known as Red Queen interactions (McLaughlin and Malik, 2017), that has led to the development of diverse anti-phage mechanisms (Rostol and Marraffini, 2019a). A key driver in the diversification of these systems is the inherent ability of phages to eventually escape these defenses, thus forcing further bacterial innovation (Fernandez et al., 2018). By acquiring spacers, CRISPR-Cas immune systems are therefore unique among prokaryotic defense strategies in that they are able to adapt to the viruses that challenge them without fundamentally altering their molecular machinery (Barrangou et al., 2007). Once added to the CRISPR array, spacers function as molecular memories that can be repeatedly utilized upon subsequent infections to guide CRISPR-associated (Cas) nucleases to execute sequence-specific nucleic acid cleavage (Barrangou et al., 2007; Marraffini and Sontheimer, 2008). Yet, despite the potency of this immune response, viruses rapidly and readily escape CRISPR by mutating their genomes, thereby precluding spacer-mediated recognition (Deveau et al., 2010; Pyenson et al., 2017; Semenova et al., 2011). This problem of CRISPR-escape by viruses is particularly acute for type II CRISPR systems because the Cas9 immunity can be abrogated with single nucleotide polymorphisms (SNPs) in the target's PAM or seed sequence (Deveau et al., 2008; Pyenson et al., 2017).

In my thesis work, I have explored how type II-A CRISPR-Cas9 systems combat rapidly evolving bacteriophages that readily escape initial Cas9 targeting. By studying the CRISPR-Cas loci from *Streptococcus pyogenes* and *Streptococcus thermophilus*, I have identified a previously unidentified phase of the CRISPR-Cas9 immune response known as primed spacer acquisition or "priming". This primed immune response equips these systems with a means of combating phage escapers by using their preexisting spacers to direct additional spacer acquisition at an enhanced rate. I demonstrated that the mechanism that mediates primed acquisition requires CRISPR RNA (crRNA) mediated recognition of the viral genome and subsequent Cas9-generated dsDNA breaks (DSBs). This mechanism allows CRISPR immune cells to acquire additional spacers at the same time as they destroy the infecting virus. A consequence of Cas9-mediated priming is that the spacers acquired through this mechanism are located within the immediate vicinity of the target site cleaved by Cas9. I have shown these newly acquired spacers are capable of combating existing escaper phages as well as preemptively immunizing against escapers and related viruses that would otherwise prove lethal to them in the absence of priming.

During our study a complementary bioinformatic study was published corroborating our *in vivo* findings (Nicholson et al., 2018). Nicholson et al (2018) analyzed the genomes of 35,240 individual archaea and bacterial strains and examined strains with one CRISPR locus containing spacers matching targets in archaeal viruses, bacteriophages and prophages. From this group of CRISPR loci they identified 1,488 systems with arrays containing two or more spacers against the same viral genome. Similar to our own analysis of type II-A CRISPR-Cas system of *S. pyogenes* with two anti-viral spacers after 24 hour infections with ΦNM4γ4 (**Figure 3.1**), Nicholson et al (2018) then compared the position of each spacer target relative to the first targeting spacer acquired in the array (ie the

most leader-distal anti-viral spacer). In agreement with our findings in **Figure 3.1A-B**, this analysis revealed that more recently acquired anti-viral spacers tended to be located in close proximity to the original targeting spacer, suggesting that Cas9-priming is occurring against these viruses. Together with our work on the *S. pyogenes* and *S. thermophilus* loci, this suggests that Cas9-priming is a widespread phenomenon among type II-A CRISPR-Cas systems. In addition, the presence of multiple spacers targeting the same virus in these strains suggests that, in agreement with our own data (**Figure 3.2A**) as well as the work of others (Pyenson et al., 2017; van Houte et al., 2016), having multiple targeting spacers is fundamental to mounting a successful anti-viral CRISPR immune response. Intriguingly, Nicholson et al (2018) also identified a similar number of type II-C CRISPR arrays that appeared to have also undergone anti-viral priming. Altogether, these results, in combination with the work presented in this thesis, provides evidence for a model in which priming is a hallmark feature of all type II CRISPR-Cas immune responses.

While the genetic basis for Cas9-directed primed spacer acquisition is now established, the precise mechanistic basis of how Cas9-generated dsDNA breaks (DSBs) lead to the recruitment of the type II spacer acquisition machinery is less clear. As discussed in Chapter 1.5 and Chapter 2.7, Cas9 has been shown to physically associate with Csn2 and the Cas1-Cas2 complex (Heler et al., 2015; Wilkinson et al., 2019). In addition, it is well established from work in our own lab as well as others that DSBs serve as "hotspots" where spacer acquisition is highly active (Levy et al., 2015; Modell et al., 2017). Given this data as well as results presented here, we propose a mechanistic model of spacer acquisition in which the creation of free DNA ends generated either during the injection of a phage genome (Modell et al., 2017), the replication of a plasmid (Levy et al., 2015), or in our case the cleavage of a viral genome by Cas9 acts as a universal signal for the recruitment of the acquisition machinery in all type I and II CRISPRs. In the case of type II priming (**Figure 4.1**), after Cas9 produces a DSB it remains firmly bound to its target (Sternberg et al., 2014) allowing sufficient time for Cas1-Cas2-Csn2 complex (Wilkinson et al., 2019) recruitment to the viral genome. As has been previously reported, Cas9 in association with Cas1-Cas2-Csn2 proceeds to select functional viral targets (pre-spacers) (Heler et al., 2015) in the proximity of the Cas9 cut site. Cas9 and Csn2 then dissociate following Cas1-Cas2 binding of the pre-spacer forming the integration complex (Wright et al., 2017; Xiao et al., 2017b) that coordinates spacer placement in the CRISPR array following LAS recognition (McGinn and Marraffini, 2016; Wei et al., 2015a).

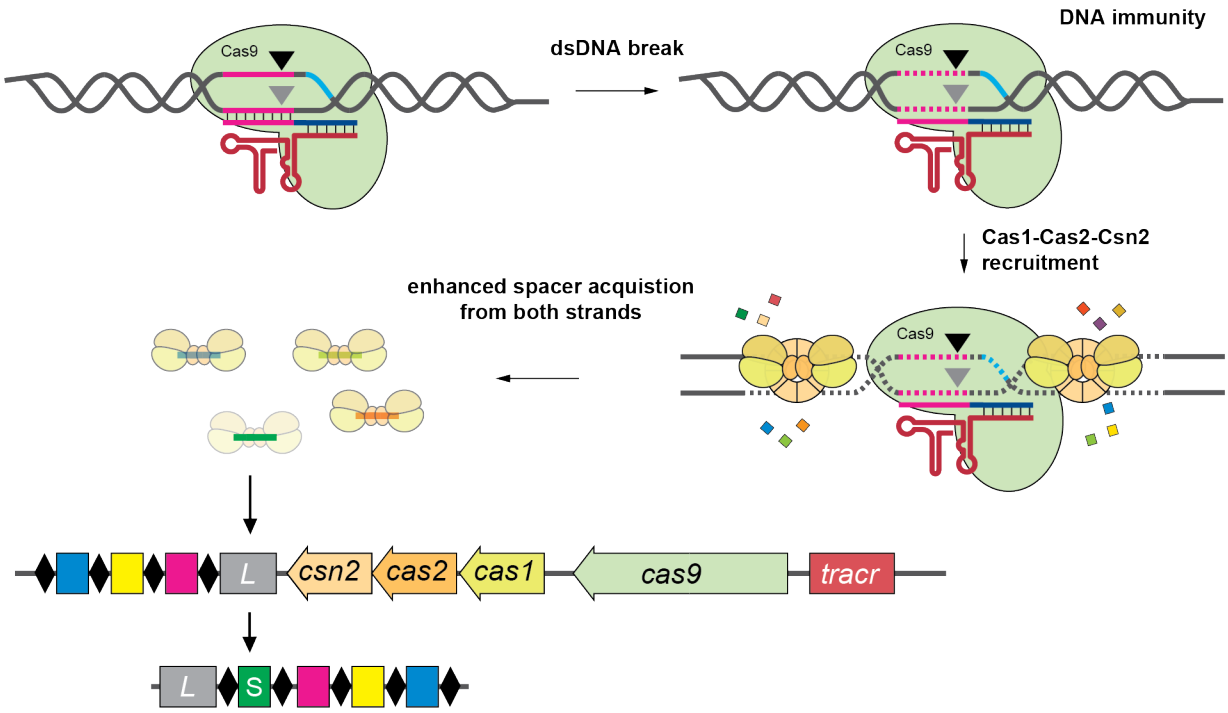


Figure 4.1. A model for the mechanism of Cas9-primed spacer acquisition. In the type II-A system, Cas9 binds PAM-containing targets and introduces dsDNA breaks. The spacer acquisition complex (Cas1-Cas2-Csn2) is recruited to the free DNA ends generated by Cas9 and pre-spacers (colored squares) are captured from both DNA strands. The abundance of pre-spacers allows for enhanced rate of integration and spacer acquisition.

This model (summarized above in **Figure 4.1**) possess many questions for future experimentation both *in vivo* and *in vitro*. In particular, after DNA cleavage, does Cas9 directly recruit Cas1-Cas2 and Csn2 or is the complex innately attracted to the DSB itself? Experiments in which type II-A spacer acquisition against bacterial genome is monitored after cutting with the restriction enzyme I-sceI (Modell et al., 2017) suggest that DSB are sufficient to generate hotspots for Cas1-Cas2 recruitment, but whether this is true for Cas9 priming has yet to be tested. Alternatively, Cas9 primed spacer acquisition may be more efficient than naive or "I-sceI-assisted" spacer acquisition because Cas9 is able to physically associate with (Heler et al., 2015) and therefore recruit Cas1-Cas2-Csn2. It is has been reported that cleavage of phage 2972 with LlaDCHI restriction endonuclease system can increase the rate of spacer acquisition in *Streptococcus thermophilus* DGCC7710 (Hynes et al., 2014) but the mechanistic basis for this phenomenon is unknown. Whether cleavage of viral genomes by restriction modification systems (R-Ms) can phenocopy Cas9's role in priming requires further study. Of particular interest will be investigating if phages cut by R-Ms can generate the same pattern of spacer acquisition seen during Cas9 targeting with new spacers clustered within 1 kb of the cut site (**Figure 2.1D-H**, **Figure 2.10F,** and **Figure 3.1A**). Cas9 participates in spacer acquisition at two stages: (1) directing primed spacer acquisition and (2) selecting pre-spacers for integration (Heler et al., 2015; Wei et al., 2015b) but how these two processes are

coordinated in the cell is unknown. Does the same Cas9 responsible for DNA cleavage also participate in spacer selection *in cis* or are multiple Cas9s required *in trans*? Single-molecule studies utilizing DNA curtains, like those performed on the type I-E systems (Dillard et al., 2018; Redding et al., 2015), would be highly informative. In addition, the question of how pre-spacers are produced within the 1 kb region around the Cas9 cut site is fundamental. During naive spacer acquisition our lab has shown that the functional analog of RecBCD, AddAB processes free DNA ends to generate substrates for the type II-A spacer acquisition machinery (Modell et al., 2017). The hotspots of spacer acquisition generated from Cas9-directed priming in both *S. pyogenes* and *S. thermophilus* however showed no evidence of AddAB activity (**Figure 2.1D-H** and **Figure 2.10F**). This implies that there is an AddAB-independent means of generating pre-spacers. In type I systems it is believed that Cas3 in concert with Cas1-Cas2 and Cascade can generate pre-spacers (Künne et al., 2016; Shiriaeva et al., 2019) (reviewed in **Chapter 1.6**). Whether the type II-A spacer acquisition machinery has a role in producing pre-spacers has yet to be investigated. Of particular interest is Csn2, whose role in spacer acquisition is currently unknown. Although required for spacer acquisition *in vivo* (Heler et al.; Wei et al., 2015b), Csn2 is not need for spacer integration *in vitro* (Wright and Doudna, 2016). In addition, it has been shown that Csn2 associates with Cas1-Cas2 and forms a central channel capable of threading dsDNA ends (Arslan et al., 2013; Bernheim et al., 2017; Wilkinson et al., 2019) suggesting that Csn2 may be important in the potential translocation of Cas1-Cas2 in a manner analogous to Cas3's hypothesized role in the type I primed acquisition complex (PAC) (Dillard et al., 2018; Redding et al., 2015) (**Figure 1.7**). Altogether the work presented in this thesis provides a platform for additional mechanistic studies that will clarify the molecular events that coordinate type II priming.

Prokaryotes commonly carry multiple CRISPR loci both of the same and different type (Makarova et al., 2020; Nicholson et al., 2018). The potential interplay between the systems and its functional consequences have only begun to be explored. In the *Marinomonas mediterranea*, the type I-F and III-B systems can exchange and successful deploy each other's crRNAs during CRISPR targeting (Silas et al., 2017). If CRISPR loci can synergize during spacer acquisition is unknown. One possibility, in line with work presented here, is that Cas nucleases from one CRISPR system can generate dsDNA breaks that could be recognized by the Cas1-Cas2 machinery derived from another locus. In this scenario, CRISPR targeting by one system enhances the acquisition of spacers by another, thereby imparting the host with two targeting spacers and enhanced immunity. Cas1-Cas2 complexes from different loci may also compete for the same substrates. Future work will be needed to determine the costs and benefits of carrying multiple CRISPRs. The *S. thermophilus* DGCC710 strain presented in this work, that naturally contains two type II-A CRISPR loci as well as a III-A and I-E locus (Carte et al., 2014), provides a potential model for answering these question.

The acquisition of spacers provides CRISPR-containing hosts with a multitude of advantages especially during the context of infection with genetic parasites but, unlike CRISPR targeting (reviewed in **Chapter 1.4.3**), appears to have no dedicated mechanism for distinguishing between self from non-self nucleic acids (Levy et al., 2015). This is particularly striking given that spurious

acquisition of spacers derived from the host chromosome often leads to the generation of lethal DSBs that cannot be repaired in many prokaryotes (Bernheim et al., 2017; Jiang et al., 2013). While cells that undergo self-acquisition are likely rapidly removed from the population, these auto-immune events likely impose a heavy fitness cost to hosts carrying active CRISPR systems. The data presented here provide a mechanism by which type II-A systems can prioritize the acquisition non-self spacers. By utilizing base-pairing between the crRNA and its targets, Cas9 preferentially localizes the spacer acquisition machinery to foreign DNA elements. Future studies will be needed to further investigate if self-acquisition is limited by active priming.

CRISPR-Cas9 systems have been widely adopted for a variety of genome editing and biotechnologies (Doudna and Charpentier, 2014). In addition, spacer acquisition has recently been repurposed as a technology that can be used as a memory storage system (Sheth et al., 2017) and more recently as a means of generating comprehensive crRNA libraries that can be used to perturb bacterial genomes of interest (Jiang et al., 2020). The observation that Cas9 cleavage can specify 1 kb regions of DNA to be acquired may therefore be leveraged as a means of controlling which DNA sequences are preferentially utilized by these technologies. Thus, the continued study of the fundamental biology of CRISPR-Cas9 will undoubtably lead to further biotechnological advances.

The type II primed spacer acquisition response presented here provides prokaryotes with potent solutions to the problem posed by the Red Queen hypothesis (**Figure 1.1)**. Firstly, by continually updating their immune repertoire, priming allows the host to maintain its fitness in the face of rapidly evolving viruses that threaten it. This aspect of priming is particularly useful when confronting bacteriophages that have escaped full CRISPR targeting but can still be recognized by Cas9 (ie seed sequence escapers shown in **Figure 2.3B**). While such escapers can be partially cleaved by Cas9, this level of DNA cleavage is not sufficient on its own to protect the population from phage infection (**Figure 2.4A-B** and **Figure 3.2A-E**). Primed spacer acquisition however allows the host to rapidly counteract this class of escapers by rapidly immunizing the hosts with secondary targeting spacers, and in doing so allowing the population to recover from the infection (**Figure 3.2B-C**). This therefore perfectly fits in the existing paradigm of Red Queen interactions where viral escape compels the host to update is anti-viral defense strategy (**Figure 1.1**). Priming against viral escapers requires base-paring between the crRNA and the target (**Figure 2.4** and **Figure 3.2G-F**). As a result, spacers that are no longer useful for anti-viral targeting still retain vital functionality. This lies in sharp contrast to many Red Queen interactions where the emergence of virulent escapers renders the existing host defenses completely obsolete, thereby forcing *de novo* adaptation and replacement of the exiting defense mechanism (**Figure 1.1**). Additionally, the strong priming effect observed with fully targeting spacers means that the host can acquire secondary immunity prior to infection with an escaper or related phage. As a result, rather than responding to the evolution of viral populations, the immune system anticipates viruses' evolution. This novel form of "anticipatory immunity" inverts the traditional reading of the Red Queen hypothesis, allowing the host rather than the pathogen to drive the evolutionary arms race. A limitation of our current study is that our experiments are carried out on relatively short evolutionary timescales. More long-term passaging

experiments will be needed to determine how priming influences the coevolution of bacteria and phages. Furthermore, this study exclusively examines Cas9-priming in the context of lytic phage infection but in the environment, bacteria are thought to commonly encounter lysogenic viruses that frequently integrate into their genomes as prophages (Canchaya et al., 2004) (**Figure 1.3**). Such prophages can provide their host with protection against related phages through superinfection exclusion in a CRISPR-independent manner (Zinder, 1958). If priming affects the rates of phage lysogenization and conversely if presence of lysogens affect the stability of type II CRISPRs systems, as is the case for type I systems (Rollie et al., 2020), requires further exploration.

The microbial world has provided us with endless innovation as a consequence of the interplay between prokaryotes and their viruses. The discovery of type II priming and the novel anticipatory immunity it provides represents only one of an ever-growing list of sophisticated innovations developed by these organisms. Continued exploration of the microbial world will undoubtably lie at the vanguard of science and further our own advancement.

CHAPTER 5: MATERIALS AND METHODS

5.1 Bacterial strains and growth conditions

Cultivation of *S. aureus* RN4220 (Kreiswirth et al., 1983) was carried out in brain-heart infusion (BHI) broth medium or tryptic soy broth (TSB) at 37°C with shaking. Wherever applicable, media were supplemented with chloramphenicol at 10 μg/ml and or erythromycin at 10 μg/ml to ensure pC194-derived and pE194-derived plasmid maintenance respectively. Cultivation of *S. thermophilus* DG7710 (Horvath et al., 2008) was carried out in M17 broth supplemented with 10% lactose (LM17) at 42 °C without shaking. Wherever applicable, media were supplemented with 200 μg/ml spectinomycin. Strains and plasmids used in this study are listed in **Table 4.1** and **Table 4.2**.

5.2 Plasmid Construction

All cloning was performed with electrocompetent *S. aureus* RN4220 cells as described previously (Goldberg et al., 2014). The sequences and oligonucleotides used in this study are listed in Supplementary Data File 2. To construct pPN86, pE194 was amplified with oPN283 and oPN284, pGG32 (Goldberg et al., 2014) was amplified with oPN285 and oPN286, and a two-piece Gibson assembly was preformed. To construct pPN249, pC194 was amplified with oPN548 and oPN549, ϕNM4γ4 genomic DNA was amplified with oPN550 and oPN551, and a two-piece Gibson assembly was preformed. pPN250 was then constructed by amplifying pPN249 with AV186 and oPN565 as well as AV187 and oPN564 and a two-piece Gibson assembly was preformed. pPN23 was constructed by amplifying pJM62 (Modell et al., 2017) with H237 and H238 (Heler et al., 2015) and preforming a one-piece Gibson assembly. BsaI cloning was used to make pPN183 by inserting spacer174 (annealed primers RH486-RH487) into the BsaI site of pPN23. pPN77 was constructed by amplifying pJM62 with JM110 and JM115 (McGinn and Marraffini, 2016) and preforming a one piece Gibson assembly. pPN91 then was constructed in a two-piece Gibson by amplifying pPN77 with JM90 and oPN280 as well as JM91 and oPN279 and a two-piece Gibson assembly was preformed. BsaI cloning was used to make pPN174 and pPN256 by inserting spacer174 (annealed primers RH486-RH487) and spacer256 (annealed oPN91-oPN92) into the BsaI site of pPN91. The E220A mutation was introduced into pPN174 to generate pPN285 by amplifying pPN174 with PS285 and H293 as well as PS284 and H294 and then preforming a two-piece Gibson assembly as described previously (Heler et al., 2015). pPN136 was constructed by amplying pGG32 (Heler et al., 2015) with JM91 and oGG140 and JM90 and JM115 and preforming a two piece Gibson assembly (McGinn and Marraffini, 2016). pPN290 was constructed by replacing the wild-type Cas9 with the nuclease-dead form of Cas9 (D10A, H840A mutations) by amplying pPN91 with H294 and H295 (Heler et al., 2015), amplyfing pDB182 (Bikard et al., 2014) with H293 and H296 (Heler et al., 2015), and preforming a two-piece Gibson assembly. pPN294 was then made by inserting spacer174 (annealed primers RH486-RH487) into the BsaI site of pPN290. pPN92 was constructed by amplifying pJM62 (Modell et al., 2017) with JM90 and oPN280, amplifying pPN77 with JM91 and oPN279, and preforming a two piece Gibson assembly (McGinn and Marraffini, 2016). BsaI

cloning was used to make pPN182 by inserting spacer174 (annealed primers RH486-RH487) into the BsaI site of pPN92.

5.3 Strain Construction

To create the *S. thermophilus* strain JAV25 we employed the method described previously (Varble et al., 2019). Briefly, deletions were made by transforming PCR amplicons with 2-kb homology arms flanking an spectinomycin resistance cassette into wild-type strains. The sequences and oligonucleotides used in this study are listed in Supplementary Data File 2. To delete CRISPR1, DG7710 genomic DNA was amplified with AV664 and AV665 as well as AV666-AV667 (Varble et al., 2019)and pLZ12spec55 was amplified with AV672 and AV673 (Varble et al., 2019) followed by a thee-piece Gibson assembly to create the final PCR amplicon for transformation.

5.4 Isolation of JAV25 CRISPR BIMs

BIMs derived from JAV25 (CRISPR3-naïve in the text) were isolated using a previously described method (Hynes et al., 2017) with minor modification. Overnight cultures of JAV25 were infected with Φ2972 at a MOI of 0.1 and mixed with with LM17 media soft agar supplemented with 10 mM $CaCl_2$ and then plated LM17 agar plates supplements on 10 mM $CaCl_2$. Plates were incubated at 42 °C overnight after drying at room temperature from 30 minutes. The resulting colonies were then screened for spacer acquisition by TopTaq PCR amplification (Qiagen) with the primers oPN737 with oPN738 and three of these colonies containing a single, unique spacer targeting Φ2972 were saved for subsequent study (CRISPR3α, CRISPR3β, and CRISPR3γ).

5.5 Plaque formation assay in *S. aureus*

Serial dilutions of phage stock were prepared in triplicate, spotted on fresh top agar lawns of RN4220 in BHI agar supplemented with appropriate antibiotic and 5 mM $CaCl_2$. Plates were incubated at 37 °C overnight after drying at room temperature from 30 minutes.

5.6 Plaque formation assay in *S. thermophilus*

Plaque formation of was measured as described previously (Hynes et al., 2017) with minor variations. Briefly, overnight cultures were launched from single colonies, infected with serials dilutions of phage stock Φ2972 were prepared in triplicate, and mixed with with LM17 media soft agar supplemented with 10 mM $CaCl_2$ and then plated LM17 agar plates supplements on 10 mM $CaCl_2$. Plates were incubated at 42 °C overnight after drying at room temperature from 30 minutes.

5.7 Quantifying CRISPR phage escapers

Thirty escaper plaques were isolated from soft agar lawns of RN4220 harboring a type II-A CRISPR plasmid targeting ϕNM4γ4 in BHI soft agar supplemented with the appropriate antibiotic, 5 mM $CaCl_2$, and resuspended in 20 μL BHI. 2 μL of the phage mixture was added to 30 μL of Colony Lysis Buffer (Pyenson et al., 2017) and boiled at 98 °C degrees for 10 minutes. 1 μL of the resulting phage lysate was then used as template for PCR amplification. For escapers of CRISPR174, oPN246 and oPN247 (all oligonucleotides used in this study are listed in Supplementary Data File 2) were used to amplify a portion of ϕNM4γ4 amidase gene (*gp68*) (Bae et al., 2006). PCR products were purified with the QIAquick PCR Purification Kit (Qiagen) and submitted for Sanger sequencing.

5.8 Isolation of spontaneous CRISPR escaper phage

ϕNM4γ4^{PAM}, ϕNM4γ4^{Seed}, and ϕNM1γ6^{PAM} were isolated from spontaneous escaper plaques following infection of ϕNM4γ4 on soft agar lawns of *S. aureus* RN4220 cells carrying pPN174 that encodes a type II-A CRISPR system targeting the amidase gene (*gp68*) of ϕNM4γ4 (CRISPR174) (**Figure 2.4**). PCR and Sanger sequencing of the resulting PCR amplicons confirmed escape mutation in the PAM and seed of the target of CRISPR174 (**Figure 2.3**).

5.9 Phage construction

We used a variation of a previously described method (Lemay et al., 2017). ϕNM4γ4^{256rev} was created by propagating ϕNM4γ4 liquid culture of cells harboring pPN250 which contains the protospacer and PAM of CRISPR256 flipped onto the opposite strand as well as the surrounding upstream and downstream homology regions for recombination with the ϕNM4γ4 genome. Recombinant ϕNM4γ4^{256rev} plaques were isolated on a soft agar lawn of RN4220 pGG14, which encodes a type III-A CRISPR-Cas system targeting the sequence present in ϕNM4γ4 and not in ϕNM4γ4^{256rev}. Subsequent PCR and Sanger sequencing of the resulting PCR amplicons confirmed the flipped target.

5.10 Spacer Acquisition in *S. aureus* Liquid Culture

Overnight cultures were launched from single colonies and diluted to an optical density at 600 nm (OD_{600}) of ~0.1 in BHI, 5 mM $CaCl_2$, and the appropriate antibiotics. After 1 hour and ten minutes, optical density (OD_{600}) was measured for each culture, and each sample was normalized to an equal cell density. Cultures were then infected with ϕNM4γ4, ϕNM4γ4^{PAM}, ϕNM4γ4^{Seed}, ϕNM4γ4^{256rev}, or mock infected at a multiplicity of infection (MOI) of 100 for thirty minutes prior pelleting infected cells, removing the supernatant, and flash freezing the pellets in liquid nitrogen. Frozen pellets were kept at -80°C until ready for DNA extraction. For 24 hour spacer acquisition assay (**Figure 3.1**), three cultures of pWJ40 prepared as above and each normalized to an OD_{600} of ~0.5 prior to infection with ϕNM4γ4 at a MOI of 1 for 24 hours. Infected cultures

were pelleted, the supernatant was removed, and immediately prepared for DNA extraction.

5.11 Spacer Acquisition in *S. thermophilus* Liquid Culture

Overnight cultures were launched from single colonies and diluted to an optical density at 600 nm (OD_{600}) of ~0.1 in LM17 with 10 mM $CaCl_2$. After 1 hour and 30 minutes of outgrowth, the optical density (OD_{600}) was measured for each culture, and each sample was normalized to an equal cell density. Cultures were then infected with Φ2972 or mock infected at a multiplicity of infection (MOI) of 10 for thirty minutes prior pelleting infected cells, removing the supernatant, and flash freezing the pellets in liquid nitrogen. Frozen pellets were kept at -80°C until ready for DNA extraction.

5.12 PCR amplification of expanded CRISPR loci for high-throughput sequencing in *S. aureus*

CRISPR plasmids were isolated from RN4220 cells with modified QIAprep Spin Miniprep Kit protocol (Qiagen) as previously described (Modell et al., 2017). We used 200 ng (log phase) of plasmid as input for encrichment PCR of the CRISPR locus with Phusion DNA Polymerase (Thermo) with the following primer mix: oPN287 and cocktail containing an equal mixture of oPN288, oPN289, and oPN290. For 24 hour infection assay, H188 and JM257, JM248 and JM258, and JM249 and JM259 were used to amplify the CRISPR for the three cultures respectively. To differentiate between samples during multiplexed high-throughput sequencing, variants of oPN287 containing randomized 5 nucleotides (NNNNN) followed by 3-6 nucleotide barcode at 5' end. Amplicons corresponding to the size of expanded CRISPR arrays were gel purified allowing for the removal of unexpanded CRISPR arrays. Purified amplicons were then prepared for sequencing with the TrueSeq Nano DNA Library Prep protocol (Illumina), using a final concentration of 1.36x Sample Purification Beads (Ilumina) following end repair for further size selection, followed by high-throughput sequencing with the MiSeq platform.

5.13 PCR amplification of expanded CRISPR loci for high-throughput sequencing in *S. thermophilus*

Genomic DNA was isolated from DG7710 cells by first treating with modified Wizard Genomic DNA Purification protocol for gram-positive bacteria (Promega): bacterial cell pellets were resuspended in 0.5 M EDTA, pH 8.0 supplemented with 200 μg ml^{-1} lysozyme (Sigma) and incubated at 37 °C for 25 min prior to pelleting and removing the supernatant. The standard Wizard protocol for gram-positives was then followed as described by the manufacturer (Promega). We used 200 ng (log phase) of plasmid as input for PCR of the CRISPR locus with Phusion DNA Polymerase (Thermo) with the following primer mix: oPN737 with oPN738, oPN757, oPN758, or oPN759 depending on whether JAV25, JAV25-BIM01, JAV25-BIM02, or JAV25-BIM03 was used. To

differentiate between samples during multiplexed high-throughput sequencing, variants of oPN737 containing randomized 5 nucleotides (NNNNN) followed by 3-6 nucleotide barcode at 5′ end. Amplicons corresponding to the size of expanded CRISPR arrays were gel purified allowing for the removal of unexpanded CRISPR arrays. Purified amplicons were then prepared for sequencing with the TrueSeq Nano DNA Library Prep protocol (Illumina), using a final concentration of 1.36x Sample Purification Beads (Ilumina) following end repair for further size selection, followed by high-throughput sequencing with the MiSeq platform.

5.14 Spacer acquisition in soft agar

Overnight cultures were launched from single colonies and diluted to an optical density at 600 nm (OD_{600}) of ~0.1 in BHI and the appropriate antibiotics. After 1 hour and ten minutes, optical density (OD_{600}) was measured for each culture, and each sample was normalized to an equal cell density. Cultures were then concentrated to a final density of OD_{600} ~125, infected at with ϕNM4γ4, ϕNM4γ4^{PAM} or ϕNM4γ4^{Seed} at MOI 2.5, mixed with soft BHI soft agar containing 5 mM $CaCl_2$ and then plated BHI agar with the appropriate antibiotics. In order to quantify initial number of cells infected, prior to infection an aliquot of each culture was serially diluted and plated on BHI plates. Following overnight incubation at 37 °C, soft agar plates were photographed and the number of surviving bacteriophage-immune mutant colonies (BIMs) were quantified with ImageJ32. To assay for spacer acquisition, individual colonies were picked and lysed with lysis buffer containing 50 ng/μl lysostaphin (Ambi) as described before (Heler et al., 2015). Following centrifugation (21,000*g*), the supernatant of each sample was used a template for TopTaq PCR amplification (Qiagen) with the primers oPN479 and oPN292. The resultant PCR amplicons were then analyzed on 2% agarose gels and the percent of BIMs containing expanded arrays (CRISPR BIMs) was determined (**Figure 2.5**). The spacer acquisition rate for each infection was then calculated based on the percent of CRISPR BIMs divided by the initial number of cells infected (**Figure 2.4C** and **Figure 2.5**).

5.15 *In vitro* CRISPR-Cas9 cleavage assay

174CRISPR RNA and tracrRNA (IDT cat# 1072532) were purchased from IDT and annealed to form crRNA174:tracrRNA RNA duplexes according Alt-R CRISPR-Cas9: *In vitro* cleavage of target DNA with ribonucleoprotein complex protocol (IDT). All RNA oligonucleotides used in this study are listed in Supplementary Data File 2.Target DNA substrates were generated by PCR of purified ϕNM4γ4 genomic DNA with Phusion DNA Polymerase (Thermo) using oPN144 and oPN562 and then purified by gel extraction. Cleavage assays were preformed similarly as described previously (Jinek et al., 2012). Briefly, crRNA:tracrRNA and Cas9 (NEB #MO386) were allowed form RNP complexes at room temperature for 10 minutes and then diluted to a final concentration of 6.25, 12.5, 25, 50 and 100 nM following to the addition of target DNA. All reactions were incubated at 37°C for 5 minutes before the digestion with proteinase K (NEB #P8107) to stop the reactions and liberate target DNA and

then stored at -80°C until ready for further analyzed. Samples were visualized on 2% agarose gel with SYBR Gold Nucleic Acid Gel Stain (Invitrogen #S11494) and the abundance of cleavage products quantified by automated electrophoresis and imaging using a Tapestation 4200 (Agilent).

5.16 *In vivo* CRISPR-Cas9 cleavage of viral DNA

To observe CRISPR-Cas9 cleavage of anti-viral targets, overnight culture of RN4220 cells carrying the pC194 (CRISPR-), pPN136 (SR), or pPN174 (CRISPR174) were diluted to an OD_{600} of ~0.1 with the appropriate antibiotic and 5 mM $CaCl_2$. After 1 hour and ten minutes of growth, optical density (OD_{600}) was measured for each culture, and each sample was normalized to equal cell density. Cultures were then infected at MOI 5 with ϕNM4γ4, ϕNM4γ4^{PAM} or ϕNM4γ4^{Seed} for 20 minutes prior to centrifugation and flash-freezing of cell pellets. All samples were stored at -80°C until ready for genomic DNA (gDNA) purification using the DNeasy Blood and Soft Tissue protocol for Gram-Positive organisms (Qiagen). In order assay for viral gDNA cleavage we preformed modified Anchor PCR by utilizing components of the 5′ RACE System for Rapid Amplification of cDNA Ends Version 2.0 kit (Invitrogen Life Sciences). In brief, 200 ng of purified genomic DNA was incubated with dCTP at final concentration of 200 μM for 2.5 minutes at 94°C degrees to denature dsDNA. The reactions were chilled on ice for 1 minute prior to the addition of recombinant terminal deoxynucleotidyl transferase (TdT) (Invitrogen Life Sciences). Incubation at 37°C for 10 minutes allows for the addition dC-homopolymeric tail to the 3′ ends of dsDNA ends generated by Cas9-mediated cleavage and creates abridged anchor primer (AAP) binding site (Invitrogen Life Sciences). TdT is then heat inactivated at 65°C for 10 minutes. dC-tailed DNA was then amplified with *Taq* DNA polymerase (Invitrogen Life Sciences) using the AAP and oPN656 upstream of CRISPR174 protospacer. Finally amplicons were visualized on 2% agarose gel and amplicons were purified by gel extraction for Zero Blunt TOPO PCR cloning (Invitrogen Life Sciences).

5.17 Bacterial infection growth curves

Viral infections were preformed in a microplate reader as previously described (Goldberg et al., 2014; McGinn and Marraffini, 2016) with minor alterations. Overnight cultures were launched in triplicate from single colonies and diluted to an OD_{600} of ~0.1 with the appropriate antibiotic and 5 mM $CaCl_2$. After 1 hour and ten minutes of growth, optical density (OD_{600}) was measured for each culture, and each sample was normalized to equal cell density (OD_{600} ~ 0.4) and loaded into a 96-well plate (Cellstar, 655180). Individual cultures were then infected with ϕNM4γ4, ϕNM4γ4^{PAM} or ϕNM4γ4^{Seed} at an MOI 1. For infections at high MOIs (Fig. 5D), cells were prepared as above but infected at an MOI 100 in bulk culture (5 ml) and the infection were allowed to proceed in the shaking incubator for 30 minutes before an aliquiot of each culture was then loaded into a 96-well plate. Optical density measurements were then taken every 10 minutes for 24 hours in a microplate reader (TECAN Infinite 200 PRO) to generate growth curves. At the end of 24 hours, an aliquot of each culture was lysed in colony

lysis buffer contains 50 ng/μl lysostaphin (Ambi) as described before (Heler et al., 2015) and the supernatant was used as template in PCR to assay for spacer acquisition using oPN479 and oPN292 or oPN635 and oPN292. The resulting amplicons were then analyzed on a 2% agarose gel.

5.18 Bacterial co-infection growth curve assays

As before, overnight cultures were launched, diluted to an OD_{600} of ~0.025, and then normalized to equal cell density (OD_{600} ~ 0.1) prior to infection with ϕNM4γ4 at a MOI 10. The infections were allowed to proceed in the shaking incubator for 30 minutes before the optical density (OD_{600}) was re-measured. Cultures were then loaded into a 96-well plate and infected with ϕNM1γ6^{PAM} at MOI 1. Optical density measurements were then taken every 10 minutes for 24 hours in a microplate reader (TECAN Infinite 200 PRO) to generate growth curves. At the end of 24 hours, an aliquot of each culture was lysed in colony lysis buffer contains 50 ng/μl lysostaphin as described before (Heler et al., 2015) and the supernatant was used as template in PCR to assay for spacer acquisition using oPN635 and oPN292. The resulting amplicons were then analyzed on a 2% agarose gel.

5.19 High Throughput Sequencing Data Analysis

Spacers were extracted from raw MiSeq FASTQ files and aligned to reference phage genomes using Python. Spacers that aligned perfectly to each reference genome were assigned to genomic position at the 5′ end of the alignment and reads were aggregated based on each unique spacer sequence. In order to avoid bias introduced into the data set due to duplication of the CRISPR array, reads containing expanded CRISPR arrays comprising of the original priming spacer (*spc174, spc256, spc300, spc303, spc305, spcα, spcβ, or spcγ*) were discarded. Following aggregation, the flanking protospacer adjacent motif (PAM) and strand for each spacer was determined based on the reference genome. To account sequencing bias introduced as result of enrichment PCR, all spacer read counts were normalized as described previously (Modell et al., 2017). When conventional spacers primers were used to amplify CRISPR loci this normalization step was skipped. Each unique spacer was then sorted into 1-kb bins and each bin was divided by the number of 5′-NGG-3′ sequences CRISPR loci derived from *S. pyogenes* SF370 or 5′-NGGNG-3′ for CRISPR loci derived from *S. thermophilus* DGCC7710 CRISPR3 within each bin in order to account for observation that spacer acquisition occurs primarily from sequences immediately upstream of PAMs (Heler et al., 2015). Reads per million (RPM) were calculated as RPM_{tot} or RPM_{phage} as desired previously (Modell et al., 2017).

When assessing expanded CRISPR arrays with two anti-viral spacers (Figs. 5A and S5A-E), both spacers were extracted pooled from three liquid cultures, aligned to ϕNM4γ4 genome, and the genomic position, PAM, and strand were determined for each spacer as before. Each anti-viral spacer pair was then grouped into 1-kb bins based on the distance between their respective positions in the ϕNM4γ4 genome. The proportion of anti-viral spacer pairs in

each bin was then calculated by dividing by the total number of pairs. To examine groups of doubly expanded arrays that all arise from primed acquisition against the same viral target within ϕNM4γ4 (**Figure 3.1**), CRISPR loci were sorted into groups based on the first acquired spacer in the array and the distance between each anti-viral spacer was calculated.

5.20 Statistical analysis

The basis for the statistical analysis used in this thesis are listed in the section bellow

5.20.1 Growth Curve Analysis

For growth curves in **Figure 3.2B-G** and **Figure 3.3A-D**, error bars represent the standard error of four biological replicates

5.20.2 *In vitro* CRISPR-Cas9 Cleavage Assay Analysis

For the cleavage assay in **Figure 2.8**, error bars represent the standard error of 3 biological replicates.

5.20.3 Efficiency of Plaquing Assay Analysis

For EOP assays in **Figure 2.2**, **Figure 2.4**, **Figure 2.10** and **Figure 3.2**, error bars represent the standard error of 3 biological replicates.

Table 5.1. Strains used in this study

Strain	Description	Source s
RN4220	*S. aureus* strain	Kreiswirth et al., 1983
DGCC7710	*S. thermophilus* strain	Barrangou et al., 2007
JAV25	CRISPR1 knockout DGC7710 (CRISPR3-naïve)	This study
JAV25-BIM01	JAV25 BIM with spacer alpha	This study
JAV25-BIM02	JAV25 BIM with spacer beta	This study
JAV25-BIM03	JAV25 BIM with spacer gamma	This study
JAV27	CRISPR1 and CRISPR3 knockout DGC7710	Varble et al., 2019

Table 5.2. Plasmids used in this study

Plasmid	Description	Source
pGG14	Type III-A CRISPR-Cas system targeting the sequence present in ϕNM4γ4 and not in ϕNM4γ4256rev	G. Goldberg PhD thesis
pGG32	Type II-A CRISPR-Cas system with a CRISPR array comprising of a single repeat	Heler et al 2015
pPN23	Cas9 only CRISPR-Cas system with BsaI spacer for oligo cloning	This study
pPN91	Type II-A CRISPR-Cas system with BsaI spacer for oligo cloning	This study
pPN92	Δleader Type II-A CRISPR-Cas system with BsaI spacer for oligo cloning	This study
pPN86	pLR	This study
pPN136	Δleader SR	This study
pPN174	Δleader CRISPR174	This study
pPN182	WT-CRISPR174	This study
pPN183	Cas9-CRISPR	This study
pPN249	656 bp from ϕNM4γ4 for construction of pPN250	This study
pPN250	656 bp from ϕNM4γ4 for construction of ϕNM4γ256rev	This study
pPN256	Δleader CRISPR256	This study
pPN285	Δleader dcas1 CRISPR174	This study
pPN290	dcas9 Type II-A CRISPR-Cas system with BsaI spacer for oligo cloning	This study
pPN294	Δleader dcas9 CRISPR174	This study
pWJ40	Type II-A CRISPR-Cas system with a CRISPR array from SF370	Heler et al 2015

Table 5.3. Primers used in this study

Oligo	Sequence	Function
oPN246	TTTCTATTTCCATtcgtgctataattatactaattttataaggagg	Check for escapers of CRISPR174 targeting
oPN247	attatagcacgaATGGAAATAGAAATTAAATTTAACGAAACGTTC	Check for escapers of CRISPR174 targeting
oPN287	AGTGCGATTACAAAATTTTTTAGAC	Forward primer for enchrichment PCR
oPN288	AAAACAGCATAGCTCTAAAACT	Primer #1 for reverse primer cocktal for enchrichment PCR
oPN289	AAAACAGCATAGCTCTAAAACG	Primer #2 for reverse primer cocktal for enchrichment PCR
oPN290	AAAACAGCATAGCTCTAAAACA	Primer #3 for reverse primer cocktal for enchrichment PCR
oGG38	AAGATAAAGAATTTGCTCAAGACG	For confirming 256 flip target
oGG40	ACCATTAAAACTCGTCATTCTTTC	For confirming 256 flip target
oPN292	CTTTCTCAAGTTATCATCGGCAATG	Assay CRISPR174 spacer acquistion
oPN479	AAAATTTTTTAGACAAAAATAGTCTACGAG	Assay CRISPR174 spacer acquistion
oPN635	GTAATGATTTACTAAATGTTAATGTATCA	Assay CRISPR174 spacer acquistion
oPN144	TATCTCATATTTAATAGTTATTTAATAACTGTAAGATTCCCTATAATTAATG	For creation of the in vitro cleavage assay substrate from NM4g4 genomic DNA
oPN562	attatagcacgaATACGGTGGCTCAAGTCAATC	For creation of the in vitro cleavage assay substrate from NM4g4 genomic DNA
oPN547	CATTATTACCTTTAACATTAGCTACTGTG	Anchor PCR reverse primer
oPN283	GATGGAACAAACttattgaataaaagatatgagagatttatctaattc	Gibson assembly of pPN86 fragment #1
oPN284	AGCCACTTCATCCtcgtgctataattatactaattttataaggag	Gibson assembly of pPN86 fragment #1
oPN285	attatagcacgaGGATGAAGTGGCTAGTTTAC	Gibson assembly of pPN86 fragment #2
oPN286	ttttattcaataaGTTTGTTCCATCACACTACTC	Gibson assembly of pPN86 fragment #2
oPN548	CATTAGTTCAACGTTTTAAAGATACAGAATTTCACAAAATGTTT	Gibson assembly of pPN249 fragment #1
oPN549	AACATAAATATATTTGATTCAATCGTATCTGTTTTAATATACGT	Gibson assembly of pPN249 fragment #1
oPN550	CGATTGAATCAAATATATTTATGTTACAGTAATATTGACTTTTAAAAAAGG	Gibson assembly of pPN249 fragment #2
oPN551	GTATCTTTAAAACGTTGAACTAATGGGTGCTTTAG	Gibson assembly of pPN249 fragment #2
AV186	AATCGATAACCACATAACAGTCATAAAAC	Gibson assembly of pPN250 fragment #1
oPN565	ACACACATAACTAAAGCTAAGAGTAATCAAAGGATGTTTACGTCCTGTTG	Gibson assembly of pPN250 fragment #1
AV187	GTTTTATGACTGTTATGTGGTTATCGATT	Gibson assembly of pPN250 fragment #2
oPN564	CCTTTGATTACTCTTAGCTTTAGTTATGTGTGTTTTACAGTCGTTGAGGCAG	Gibson assembly of pPN250 fragment #2
H237	ggcgtactgatgaagattatttcttaataactaaaaatatgg	Gibson assembly of pPN23
H238	tttagttattaagaaataatcttcatcagtacgccaaccagcc	Gibson assembly of pPN23
RH486	aaacTGATACATTAACATTTAGTAAATCATTACGg	BsaI cloning to construct pPN174, pPN182, pPN183, pPN294
RH487	aaaacCGTAATGATTTACTAAATGTTAATGTATCA	BsaI cloning to construct pPN174, pPN182, pPN183, pPN294
JM110	Gttttagagctatgctgttttgaatggtcccaaaactgagaccagtctcggaagctcaaa	Gibson assembly of pPN77
JM115	attcaaaacagcatagctctaaaaCttataccatatttttagttattaagaaataatctt	Gibson assembly of pPN77 and pPN136 fragment #2

Table 5.3. Primers used in this study (continued).

JM90	GAGAAGATTGAAAAAATCTTGACTTTTCG	Gibson assembly of pPN91, pPN92 fragment #1, and pPN136 fragment #2
oPN280	**GTCTCTTGAATGAGACAAATGCACCATGTTAAAAATACC**	**Gibson assembly of pPN91 and pPN92 fragment #1**
JM91	**CGAAAAGTCAAGATTTTTTCAATCTTCTC**	**Gibson assembly of pPN91 fragment #2 and pPN136 fragment #1**
PS285	**CTAAAGGCCTAAATGGTgCCATAATATCGCTAGC**	**Gibson assembly of pPN285 fragment #1**
H293	**GCAAAAATGGATAAGAAATACTCAATAGGC**	**Gibson assembly of pPN285 fragment #1 and gibson assembly of pPN290 fragment #2**
PS284	**GCTAGCGATATTATGGcACCATTTAGGCCTTTAG**	**Gibson assembly of pPN285 fragment #2**
H294	**TATTGAGTATTTCTTATCCATTTTTGCCTCC**	**Gibson assembly of pPN285 fragment #2 and gibson assembly of pPN290 fragment #1**
H295	**AACACGCATTGATTTGAGTCAGC**	**Gibson assembly of pPN290 fragment #1**
H296	**TCCTAGCTGACTCAAATCAATGCG**	**Gibson assembly of pPN290 fragment #2**
oGG140	**aaaGGTCTCGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACAACATTGCCG**	**Gibson assembly of pPN136 fragment #1**
H188	**NNNNNAGTCAAAATTTTTTAGACAAAAATAGTC**	**Applification of CRISPR arrays in 24 hour infection assay replicate #1**
JM248	**NNNNNTAGCAAAATTTTTTAGACAAAAATAGTC**	**Applification of CRISPR arrays in 24 hour infection assay replicate #1**
JM249	**NNNNNCAGCAAAATTTTTTAGACAAAAATAGTC**	**Applification of CRISPR arrays in 24 hour infection assay replicate #2**
JM257	**NNNNNAGTCTTTCTCAAGTTATCATCGGC**	**Applification of CRISPR arrays in 24 hour infection assay replicate #2**
JM258	**NNNNNTAGCTTTCTCAAGTTATCATCGGC**	**Applification of CRISPR arrays in 24 hour infection assay replicate #3**
JM259	**NNNNNCAGCTTTCTCAAGTTATCATCGGC**	**Applification of CRISPR arrays in 24 hour infection assay replicate #3**
oPN737	**AGAAAAGATATCCTACGAGG**	**Amplifcation of CRISPR3 array of *S. thermophilus* DGC7710**
oPN738	**CTCTTTAGCGTTTAGAATTTG**	**Amplifcation of CRISPR3 array of *S. thermophilus* DGC7710**
oPN757	**TCGTCCAAGTTGCCCTCATG**	**Amplifcation of CRISPR3 array of *S. thermophilus* JAV25-BIM01**
oPN758	**ACGTCAATGACAAAGAATGT**	**Amplifcation of CRISPR3 array of *S. thermophilus* JAV25-BIM02**
oPN759	**TCTTTTTGTTGGTAATAAACG**	**Amplifcation of CRISPR3 array of *S. thermophilus* JAV25-BIM03**
AV664	**cttgggcagaaaaccttgtagatg**	**Gibson assembly for JAV25 construction fragment #1**
AV665	**AGTCACGTTACGTTATgaacttggctttttaaaatacacg**	**Gibson assembly for JAV25 construction fragment #1**
AV666	**ACCCTTGGACTTTCGTCactactaacttgttggcaagg**	**Gibson assembly for JAV25 construction fragment #2**
AV667	**caagggcgatgaccttcaagg**	**Gibson assembly for JAV25 construction fragment #2**
AV672	**ATAACGTAACGTGACTGGCAAGA**	**Gibson assembly for JAV25 construction fragment #3**
AV673	**GACGAAAGTCCAAGGGTTTATTG**	**Gibson assembly for JAV25 construction fragment #3**

Table 5.4. Spacers used in this study

Spacer name/CRISPR#	Sequence (5'-3')	position in phage genome	strand	PAM (5'-3')
174	TGATACATTAACATTTAGTAAATCATTACG	39099-39128	bottom	AGG
256	ACACACATAACTAAAGCTAAGAGTAATCAA	12480-12509	top	AGG
300	TACTAAATCACCTTACAACACTTCAACTAG	19911-19940	top	TGG
303	AGGAATTGAGACACCTCAATATATACTTGC	15220-15249	top	TGG
305	TAATACAGGTTTTTACAAAAGCTTTACCAT	5962-5991	top	AGG
a	CATGAGGGCAACTTGGACGATTGATAAGGT	24038-24067	top	TGGCG
b	TACATTCTTTGTCATTGACGTGGACATGCA	30072-30171	top	TGGCG
γ	GTTTATTACCAACAAAAAGAAAAAAATATA	29393-29422	top	TGGAG

Table 5.5 Spacers used in this study

Spacer name/CRISPR #	Sequence (5'-3')	position in phage genome	strand	PAM (5'-3')
174	TGATACATTAACATTTAGTAAATCATTACG	39099-39128	bottom	AGG
256	ACACACATAACTAAAGCTAAGAGTAATCAA	12480-12509	top	AGG
300	TACTAAATCACCTTACAACACTTCAACTAG	19911-19940	top	TGG
303	AGGAATTGAGACACCTCAATATATACTTGC	15220-15249	top	TGG
305	TAATACAGGTTTTTACAAAAGCTTTACCAT	5962-5991	top	AGG
a	CATGAGGGCAACTTGGACGATTGATAAGGT	24038-24067	top	TGGCG
b	TACATTCTTTGTCATTGACGTGGACATGCA	30072-30171	top	TGGCG
γ	GTTTATTACCAACAAAAAGAAAAAAATATA	29393-29422	top	TGGAG

REFERENCES

Al-Shayeb, B., Sachdeva, R., Chen, L.X., Ward, F., Munk, P., Devoto, A., Castelle, C.J., Olm, M.R., Bouma-Gregson, K., Amano, Y., *et al.* (2020). Clades of huge phages from across Earth's ecosystems. Nature.

Almendros, C., Nobrega, F.L., McKenzie, R.E., and Brouns, S.J.J. (2019). Cas4-Cas1 fusions drive efficient PAM selection and control CRISPR adaptation. Nucleic Acids Res *47*, 5223-5230.

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

Andersson, A.F., and Banfield, J.F. (2008). Virus population dynamics and acquired virus resistance in natural microbial communities. Science *320*, 1047-1050.

Arslan, Z., Hermanns, V., Wurm, R., Wagner, R., and Pul, U. (2014). Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. Nucleic Acids Res *42*, 7884-7893.

Arslan, Z., Wurm, R., Brener, O., Ellinger, P., Nagel-Steger, L., Oesterhelt, F., Schmitt, L., Willbold, D., Wagner, R., Gohlke, H., *et al.* (2013). Double-strand DNA end-binding and sliding of the toroidal CRISPR-associated protein Csn2. Nucleic Acids Res *41*, 6347-6359.

Athukoralage, J.S., Rouillon, C., Graham, S., Gruschow, S., and White, M.F. (2018). Ring nucleases deactivate type III CRISPR ribonucleases by degrading cyclic oligoadenylate. Nature.

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

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

Bernheim, A., Calvo-Villamanan, A., Basier, C., Cui, L., Rocha, E.P.C., Touchon, M., and Bikard, D. (2017). Inhibition of NHEJ repair by type II-A CRISPR-Cas systems in bacteria. Nat Commun *8*, 2094.

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

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

Blosser, T.R., Loeff, L., Westra, E.R., Vlot, M., Kunne, T., Sobota, M., Dekker, C., Brouns, S.J., and Joo, C. (2015). Two Distinct DNA Binding Modes Guide Dual Roles of a CRISPR-Cas Protein Complex. Mol Cell *58*, 60-70.

Bobay, L.M., Touchon, M., and Rocha, E.P. (2013). Manipulating or superseding host recombination functions: a dilemma that shapes phage evolvability. PLoS Genet *9*, e1003825.

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

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

Brum, J.R., Schenck, R.O., and Sullivan, M.B. (2013). Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses. ISME J *7*, 1738-1751.

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

Burroughs, A.M., Zhang, D., Schaffer, D.E., Iyer, L.M., and Aravind, L. (2015). Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. Nucleic Acids Res *43*, 10633-10654.

Canchaya, C., Fournous, G., and Brussow, H. (2004). The impact of prophages on bacterial chromosomes. Mol Microbiol *53*, 9-18.

Carte, J., Christopher, R.T., Smith, J.T., Olson, S., Barrangou, R., Moineau, S., Glover, C.V., 3rd, Graveley, B.R., Terns, R.M., and Terns, M.P. (2014). The three major types of CRISPR-Cas systems function independently in CRISPR RNA biogenesis in Streptococcus thermophilus. Mol Microbiol *93*, 98-112.

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

Chen, J.S., Ma, E., Harrington, L.B., Da Costa, M., Tian, X., Palefsky, J.M., and Doudna, J.A. (2018). CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. Science *360*, 436-439.

Clarke, R., Heler, R., MacDougall, M.S., Yeo, N.C., Chavez, A., Regan, M., Hanakahi, L., Church, G.M., Marraffini, L.A., and Merrill, B.J. (2018). Enhanced Bacterial Immunity and Mammalian Genome Editing via RNA-Polymerase-Mediated Dislodging of Cas9 from Double-Strand DNA Breaks. Mol Cell *71*, 42-55 e48.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., *et al.* (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science *339*, 819-823.

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

Daugherty, M.D., and Malik, H.S. (2012). Rules of engagement: molecular insights from host-virus arms races. Annu Rev Genet *46*, 677-700.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602-607.

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

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

Deveau, H., Garneau, J.E., and Moineau, S. (2010). CRISPR/Cas system and its role in phage-bacteria interactions. Annu Rev Microbiol *64*, 475-493.

Dillard, K.E., Brown, M.W., Johnson, N.V., Xiao, Y., Dolan, A., Hernandez, E., Dahlhauser, S.D., Kim, Y., Myler, L.R., Anslyn, E.V., *et al.* (2018). Assembly and Translocation of a CRISPR-Cas Primed Acquisition Complex. Cell *175*, 934-946 e915.

Dion, M.B., Oechslin, F., and Moineau, S. (2020). Phage diversity, genomics and phylogeny. Nat Rev Microbiol *18*, 125-138.

Dong, D., Ren, K., Qiu, X., Zheng, J., Guo, M., Guan, X., Liu, H., Li, N., Zhang, B., Yang, D., *et al.* (2016). The crystal structure of Cpf1 in complex with CRISPR RNA. Nature *532*, 522-526.

Doudna, J.A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science *346*, 1258096.

Drabavicius, G., Sinkunas, T., Silanskas, A., Gasiunas, G., Venclovas, C., and Siksnys, V. (2018). DnaQ exonuclease-like domain of Cas2 promotes spacer integration in a type I-E CRISPR-Cas system. EMBO Rep *19*.

East-Seletsky, A., O'Connell, M.R., Burstein, D., Knott, G.J., and Doudna, J.A. (2017). RNA Targeting by Functionally Orthogonal Type VI-A CRISPR-Cas Enzymes. Mol Cell *66*, 373-383 e373.

East-Seletsky, A., O'Connell, M.R., Knight, S.C., Burstein, D., Cate, J.H., Tjian, R., and Doudna, J.A. (2016). Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. Nature *538*, 270-273.

Fagerlund, R.D., Wilkinson, M.E., Klykov, O., Barendregt, A., Pearce, F.G., Kieper, S.N., Maxwell, H.W.R., Capolupo, A., Heck, A.J.R., Krause, K.L., *et al.* (2017). Spacer capture and integration by a type I-F Cas1-Cas2-3 CRISPR adaptation complex. Proc Natl Acad Sci U S A *114*, E5122-E5128.

Fernandez, L., Rodriguez, A., and Garcia, P. (2018). Phage or foe: an insight into the impact of viral predation on microbial communities. ISME J *12*, 1171-1179.

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

Fonfara, I., Le Rhun, A., Chylinski, K., Makarova, K.S., Lecrivain, A.L., Bzdrenga, J., Koonin, E.V., and Charpentier, E. (2014). Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. Nucleic Acids Res *42*, 2577-2590.

Fonfara, I., Richter, H., Bratovic, M., Le Rhun, A., and Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. Nature *532*, 517-521.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol *31*, 822-826.

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

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

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

Goren, M.G., Doron, S., Globus, R., Amitai, G., Sorek, R., and Qimron, U. (2016). Repeat Size Determination by Two Molecular Rulers in the Type I-E CRISPR Array. Cell Rep *16*, 2811-2818.

Haft, D.H., Selengut, J., Mongodin, E.F., and Nelson, K.E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput Biol *1*, e60.

Hale, C., Kleppe, K., Terns, R.M., and Terns, M.P. (2008). Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. RNA *14*, 2572-2579.

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

Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K., and Doudna, J.A. (2010). Sequence- and structure-specific RNA processing by a CRISPR endonuclease. Science *329*, 1355-1358.

Hayes, R.P., Xiao, Y., Ding, F., van Erp, P.B., Rajashankar, K., Bailey, S., Wiedenheft, B., and Ke, A. (2016). Structural basis for promiscuous PAM recognition in type I-E Cascade from E. coli. Nature *530*, 499-503.

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

Heler, R., Wright, A.V., Vucelja, M., Doudna, J.A., and Marraffini, L.A. (2019). Spacer Acquisition Rates Determine the Immunological Diversity of the Type II CRISPR-Cas Immune Response. Cell Host Microbe *25*, 242-249 e243.

Hershey, A.D., and Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. J Gen Physiol *36*, 39-56.

Herskowitz, I., and Hagen, D. (1980). The lysis-lysogeny decision of phage lambda: explicit programming and responsiveness. Annu Rev Genet *14*, 399-445.

Hochstrasser, M.L., Taylor, D.W., Bhat, P., Guegler, C.K., Sternberg, S.H., Nogales, E., and Doudna, J.A. (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. Proc Natl Acad Sci U S A *111*, 6618-6623.

Horvath, P., Romero, D.A., Coute-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., and Barrangou, R. (2008). Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. J Bacteriol *190*, 1401-1412.

Hsu, B.B., Gibson, T.E., Yeliseyev, V., Liu, Q., Lyon, L., Bry, L., Silver, P.A., and Gerber, G.K. (2019). Dynamic Modulation of the Gut Microbiota and Metabolome by Bacteriophages in a Mouse Model. Cell Host Microbe *25*, 803-814 e805.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., *et al.* (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol *31*, 827-832.

Huo, Y., Nam, K.H., Ding, F., Lee, H., Wu, L., Xiao, Y., Farchione, M.D., Jr., Zhou, S., Rajashankar, K., Kurinov, I., *et al.* (2014). Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. Nat Struct Mol Biol *21*, 771-777.

Hynes, A.P., Lemay, M.L., Trudel, L., Deveau, H., Frenette, M., Tremblay, D.M., and Moineau, S. (2017). Detecting natural adaptation of the Streptococcus thermophilus CRISPR-Cas systems in research and classroom settings. Nat Protoc *12*, 547-565.

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

Ivancic-Bace, I., Cass, S.D., Wearne, S.J., and Bolt, E.L. (2015). Different genome stability proteins underpin primed and naive adaptation in E. coli CRISPR-Cas immunity. Nucleic Acids Res *43*, 10821-10830.

Jackson, R.N., Golden, S.M., van Erp, P.B., Carter, J., Westra, E.R., Brouns, S.J., van der Oost, J., Terwilliger, T.C., Read, R.J., and Wiedenheft, B. (2014). Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. Science *345*, 1473-1479.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol *3*, 318-356.

Jansen, R., Embden, J.D., Gaastra, W., and Schouls, L.M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol *43*, 1565-1575.

Jeon, Y., Choi, Y.H., Jang, Y., Yu, J., Goo, J., Lee, G., Jeong, Y.K., Lee, S.H., Kim, I.S., Kim, J.S., *et al.* (2018). Direct observation of DNA target searching and cleavage by CRISPR-Cas12a. Nat Commun *9*, 2777.

Jia, N., Jones, R., Yang, G., Ouerfelli, O., and Patel, D.J. (2019a). CRISPR-Cas III-A Csm6 CARF Domain Is a Ring Nuclease Triggering Stepwise cA4 Cleavage with ApA>p Formation Terminating RNase Activity. Mol Cell.

Jia, N., Mo, C.Y., Wang, C., Eng, E.T., Marraffini, L.A., and Patel, D.J. (2019b). Type III-A CRISPR-Cas Csm Complexes: Assembly, Periodic RNA Cleavage, DNase Activity Regulation, and Autoimmunity. Mol Cell *73*, 264-277 e265.

Jiang, F., Taylor, D.W., Chen, J.S., Kornfeld, J.E., Zhou, K., Thompson, A.J., Nogales, E., and Doudna, J.A. (2016a). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. Science *351*, 867-871.

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

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

Jiang, W., Oikonomou, P., and Tavazoie, S. (2020). Comprehensive Genome-wide Perturbations via CRISPR Adaptation Reveal Complex Genetics of Antibiotic Sensitivity. Cell *180*, 1002-1017 e1031.

Jiang, W., Samai, P., and Marraffini, L.A. (2016b). Degradation of phage transcripts by CRISPR-associated RNases enables type III CRISPR-Cas immunity. Cell *164*, 710-721.

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

Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., *et al.* (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science *343*, 1247997.

Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R., *et al.* (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol *18*, 529-536.

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

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

Kim, S., Loeff, L., Colombo, S., Jergic, S., Brouns, S.J.J., and Joo, C. (2020). Selective loading and processing of prespacers for precise CRISPR adaptation. Nature.

Kreiswirth, B.N., Lofdahl, S., Betley, M.J., O'Reilly, M., Schlievert, P.M., Bergdoll, M.S., and Novick, R.P. (1983). The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage. Nature *305*, 709-712.

Krivoy, A., Rutkauskas, M., Kuznedelov, K., Musharova, O., Rouillon, C., Severinov, K., and Seidel, R. (2018). Primed CRISPR adaptation in Escherichia coli cells does not depend on conformational changes in the Cascade effector complex detected in Vitro. Nucleic Acids Res *46*, 4087-4098.

Krupovic, M., Beguin, P., and Koonin, E.V. (2017). Casposons: mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery. Curr Opin Microbiol *38*, 36-43.

Künne, T., Kieper, S.N., Bannenberg, J.W., Vogel, A.I., Miellet, W.R., Klein, M., Depken, M., Suarez-Diez, M., and Brouns, S.J. (2016). Cas3-derived target DNA degradation fragments fuel primed CRISPR adaptation. Mol Cell.

Kupczok, A., Neve, H., Huang, K.D., Hoeppner, M.P., Heller, K.J., Franz, C., and Dagan, T. (2018). Rates of Mutation and Recombination in Siphoviridae Phage Genome Evolution over Three Decades. Mol Biol Evol *35*, 1147-1159.

Lau, R.K., Ye, Q., Birkholz, E.A., Berg, K.R., Patel, L., Mathews, I.T., Watrous, J.D., Ego, K., Whiteley, A.T., Lowey, B., *et al.* (2020). Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity. Mol Cell.

Lee, H., Dhingra, Y., and Sashital, D.G. (2019). The Cas4-Cas1-Cas2 complex mediates precise prespacer processing during CRISPR adaptation. Elife *8*.

Lee, H., Zhou, Y., Taylor, D.W., and Sashital, D.G. (2018). Cas4-Dependent Prespacer Processing Ensures High-Fidelity Programming of CRISPR Arrays. Mol Cell *70*, 48-59 e45.

Lemay, M.L., Tremblay, D.M., and Moineau, S. (2017). Genome Engineering of Virulent Lactococcal Phages Using CRISPR-Cas9. ACS Synth Biol *6*, 1351-1358.

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

Li, M., Wang, R., Zhao, D., and Xiang, H. (2014). Adaptation of the *Haloarcula hispanica* CRISPR-Cas system to a purified virus strictly requires a priming process. Nucleic Acids Res *42*, 2483-2492.

Lillestøl, R.K., Shah, S.A., Brugger, K., Redder, P., Phan, H., Christiansen, J., and Garrett, R.A. (2009). CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. Mol Microbiol *72*, 259-272.

Liu, L., Li, X., Ma, J., Li, Z., You, L., Wang, J., Wang, M., Zhang, X., and Wang, Y. (2017a). The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a. Cell *170*, 714-726 e710.

Liu, L., Li, X., Wang, J., Wang, M., Chen, P., Yin, M., Li, J., Sheng, G., and Wang, Y. (2017b). Two Distant Catalytic Sites Are Responsible for C2c2 RNase Activities. Cell *168*, 121-134 e112.

Liu, T.Y., Liu, J.J., Aditham, A.J., Nogales, E., and Doudna, J.A. (2019). Target preference of Type III-A CRISPR-Cas complexes at the transcription bubble. Nat Commun *10*, 3001.

Loeff, L., Brouns, S.J.J., and Joo, C. (2018). Repetitive DNA Reeling by the Cascade-Cas3 Complex in Nucleotide Unwinding Steps. Mol Cell *70*, 385-394 e383.

Luria, S.E., and Delbruck, M. (1943). Mutations of Bacteria from Virus Sensitivity to Virus Resistance. Genetics *28*, 491-511.

Makarova, K.S., Anantharaman, V., Grishin, N.V., Koonin, E.V., and Aravind, L. (2014). CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems. Front Genet *5*, 102.

Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., *et al.* (2019). Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol.

Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., *et al.* (2020). Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol *18*, 67-83.

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

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

McDonald, N.D., Regmi, A., Morreale, D.P., Borowski, J.D., and Boyd, E.F. (2019). CRISPR-Cas systems are present predominantly on mobile genetic elements in Vibrio species. BMC Genomics *20*, 105.

McGinn, J., and Marraffini, L.A. (2016). CRISPR-Cas systems optimize their immune response by specifying the site of spacer integration. Mol Cell *64*, 616-623.

McLaughlin, R.N., Jr., and Malik, H.S. (2017). Genetic conflicts: the usual suspects and beyond. J Exp Biol *220*, 6-17.

McMahon, S.A., Zhu, W., Graham, S., Rambo, R., White, M.F., and Gloster, T.M. (2020). Structure and mechanism of a Type III CRISPR defence DNA nuclease activated by cyclic oligoadenylate. Nat Commun *11*, 500.

Meeske, A.J., and Marraffini, L.A. (2018). RNA Guide Complementarity Prevents Self-Targeting in Type VI CRISPR Systems. Mol Cell *71*, 791-801 e793.

Meeske, A.J., Nakandakari-Higa, S., and Marraffini, L.A. (2019). Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. Nature *570*, 241-245.

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

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology *155*, 733-740.

Mulepati, S., and Bailey, S. (2013). In vitro reconstitution of an Escherichia coli RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. J Biol Chem *288*, 22184-22192.

Mulepati, S., Heroux, A., and Bailey, S. (2014). Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. Science *345*, 1479-1484.

Muller, H.P., and Varmus, H.E. (1994). DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. EMBO J *13*, 4704-4714.

Musharova, O., Sitnik, V., Vlot, M., Savitskaya, E., Datsenko, K.A., Krivoy, A., Fedorov, I., Semenova, E., Brouns, S.J.J., and Severinov, K. (2019). Systematic analysis of Type I-E Escherichia coli CRISPR-Cas PAM sequences ability to promote interference and primed adaptation. Mol Microbiol.

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

Nam, K.H., Haitjema, C., Liu, X., Ding, F., Wang, H., DeLisa, M.P., and Ke, A. (2012). Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system. Structure *20*, 1574-1584.

Nicholson, T.J., Jackson, S.A., Croft, B.I., Staals, R.H.J., Fineran, P.C., and Brown, C.M. (2018). Bioinformatic evidence of widespread priming in type I and II CRISPR-Cas systems. RNA Biol, 1-11.

Niewoehner, O., Garcia-Doval, C., Rostol, J.T., Berk, C., Schwede, F., Bigler, L., Hall, J.,

Marraffini, L.A., and Jinek, M. (2017). Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. Nature *548*, 543-548.

Nobrega, F.L., Costa, A.R., Kluskens, L.D., and Azeredo, J. (2015). Revisiting phage therapy: new applications for old resources. Trends Microbiol *23*, 185-191.

Novick, R.P. (1998). Contrasting lifestyles of rolling-circle phages and plasmids. Trends Biochem Sci *23*, 434-438.

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

Nunez, J.K., Harrington, L.B., Kranzusch, P.J., Engelman, A.N., and Doudna, J.A. (2015a). Foreign DNA capture during CRISPR-Cas adaptive immunity. Nature *527*, 535-538.

Nunez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W., and Doudna, J.A. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. Nat Struct Mol Biol *21*, 528-534.

Nunez, J.K., Lee, A.S., Engelman, A., and Doudna, J.A. (2015b). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. Nature *519*, 193-198.

Nussenzweig, P.M., McGinn, J., and Marraffini, L.A. (2019). Cas9 Cleavage of Viral Genomes Primes the Acquisition of New Immunological Memories. Cell Host Microbe *26*, 515-526 e516.

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

Oppenheim, A.B., Kobiler, O., Stavans, J., Court, D.L., and Adhya, S. (2005). Switches in bacteriophage lambda development. Annu Rev Genet *39*, 409-429.

Paez-Espino, D., Morovic, W., Sun, C.L., Thomas, B.C., Ueda, K., Stahl, B., Barrangou, R., and Banfield, J.F. (2013). Strong bias in the bacterial CRISPR elements that confer immunity to phage. Nature communications *4*, 1430.

Peng, W., Feng, M., Feng, X., Liang, Y.X., and She, Q. (2015). An archaeal CRISPR type III-B system exhibiting distinctive RNA targeting features and mediating dual RNA and DNA interference. Nucleic Acids Res *43*, 406-417.

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

Pyenson, N.C., Gayvert, K., Varble, A., Elemento, O., and Marraffini, L.A. (2017). Broad Targeting Specificity during Bacterial Type III CRISPR-Cas Immunity Constrains Viral Escape. Cell Host Microbe *22*, 343-353 e343.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173-1183.

Ramachandran, A., Summerville, L., Learn, B.A., DeBell, L., and Bailey, S. (2019). Processing and integration of functionally oriented prespacers in the E. coli CRISPR system depends on bacterial host exonucleases. J Biol Chem.

Redding, S., Sternberg, S.H., Marshall, M., Gibb, B., Bhat, P., Guegler, C.K., Wiedenheft, B., Doudna, J.A., and Greene, E.C. (2015). Surveillance and Processing of Foreign DNA by the *Escherichia coli* CRISPR-Cas System. Cell *163*, 854-865.

Richter, C., Dy, R.L., McKenzie, R.E., Watson, B.N., Taylor, C., Chang, J.T., McNeil,

M.B., Staals, R.H., and Fineran, P.C. (2014). 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.

Rollie, C., Chevallereau, A., Watson, B.N.J., Chyou, T.Y., Fradet, O., McLeod, I., Fineran, P.C., Brown, C.M., Gandon, S., and Westra, E.R. (2020). Targeting of temperate phages drives loss of type I CRISPR-Cas systems. Nature *578*, 149-153.

Rollins, M.F., Chowdhury, S., Carter, J., Golden, S.M., Wilkinson, R.A., Bondy-Denomy, J., Lander, G.C., and Wiedenheft, B. (2017). Cas1 and the Csy complex are opposing regulators of Cas2/3 nuclease activity. Proc Natl Acad Sci U S A *114*, E5113-E5121.

Rollins, M.F., Schuman, J.T., Paulus, K., Bukhari, H.S., and Wiedenheft, B. (2015). Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from Pseudomonas aeruginosa. Nucleic Acids Res *43*, 2216-2222.

Rostol, J.T., and Marraffini, L. (2019a). (Ph)ighting Phages: How Bacteria Resist Their Parasites. Cell Host Microbe *25*, 184-194.

Rostol, J.T., and Marraffini, L.A. (2019b). Non-specific degradation of transcripts promotes plasmid clearance during type III-A CRISPR-Cas immunity. Nat Microbiol *4*, 656-662.

Rouillon, C., Athukoralage, J.S., Graham, S., Gruschow, S., and White, M.F. (2018). Control of cyclic oligoadenylate synthesis in a type III CRISPR system. Elife *7*.

Rouillon, C., Zhou, M., Zhang, J., Politis, A., Beilsten-Edmands, V., Cannone, G., Graham, S., Robinson, C.V., Spagnolo, L., and White, M.F. (2013). Structure of the CRISPR interference complex CSM reveals key similarities with cascade. Mol Cell *52*, 124-134.

Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M.S., Siksnys, V., and Seidel, R. (2015). Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection. Cell Rep.

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

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

Sashital, D.G., Jinek, M., and Doudna, J.A. (2011). An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. Nat Struct Mol Biol *18*, 680-687.

Sashital, D.G., Wiedenheft, B., and Doudna, J.A. (2012). Mechanism of foreign DNA selection in a bacterial adaptive immune system. Mol Cell *46*, 606-615.

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

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

Semenova, E., Savitskaya, E., Musharova, O., Strotskaya, A., Vorontsova, D., Datsenko, K.A., Logacheva, M.D., and Severinov, K. (2016). Highly efficient primed spacer acquisition from targets destroyed by the Escherichia coli type I-E CRISPR-Cas interfering complex. Proc Natl Acad Sci U S A *113*, 7626-7631.

Shah, S.A., Alkhnbashi, O.S., Behler, J., Han, W., She, Q., Hess, W.R., Garrett, R.A., and Backofen, R. (2018). Comprehensive search for accessory proteins encoded with archaeal and bacterial type III CRISPR-cas gene cassettes reveals 39 new cas gene families. RNA Biol, 1-13.

Shao, Y., Richter, H., Sun, S., Sharma, K., Urlaub, H., Randau, L., and Li, H. (2016). A Non-Stem-Loop CRISPR RNA Is Processed by Dual Binding Cas6. Structure *24*, 547-554.

Sheth, R.U., Yim, S.S., Wu, F.L., and Wang, H.H. (2017). Multiplex recording of cellular events over time on CRISPR biological tape. Science.

Shiimori, M., Garrett, S.C., Chambers, D.P., Glover, C.V.C., 3rd, Graveley, B.R., and
Terns, M.P. (2017). Role of free DNA ends and protospacer adjacent motifs for CRISPR DNA uptake in Pyrococcus furiosus. Nucleic Acids Res *45*, 11281-11294.

Shiimori, M., Garrett, S.C., Graveley, B.R., and Terns, M.P. (2018). Cas4 Nucleases Define the PAM, Length, and Orientation of DNA Fragments Integrated at CRISPR Loci. Mol Cell *70*, 814-824 e816.

Shiriaeva, A.A., Savitskaya, E., Datsenko, K.A., Vvedenskaya, I.O., Fedorova, I., Morozova, N., Metlitskaya, A., Sabantsev, A., Nickels, B.E., Severinov, K., *et al.* (2019). Detection of spacer precursors formed in vivo during primed CRISPR adaptation. Nat Commun *10*, 4603.

Shmakov, S.A., Makarova, K.S., Wolf, Y.I., Severinov, K.V., and Koonin, E.V. (2018). Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. Proc Natl Acad Sci U S A *115*, E5307-E5316.

Shmakov, S.A., Sitnik, V., Makarova, K.S., Wolf, Y.I., Severinov, K.V., and Koonin, E.V. (2017). The CRISPR spacer space is dominated by sequences from species-specific mobilomes. MBio *8*.

Silas, S., Lucas-Elio, P., Jackson, S.A., Aroca-Crevillen, A., Hansen, L.L., Fineran, P.C., Fire, A.Z., and Sanchez-Amat, A. (2017). Type III CRISPR-Cas systems can provide redundancy to counteract viral escape from type I systems. Elife *6*.

Silas, S., Mohr, G., Sidote, D.J., Markham, L.M., Sanchez-Amat, A., Bhaya, D., Lambowitz, A.M., and Fire, A.Z. (2016). Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. Science *351*, aad4234.

Singh, D., Mallon, J., Poddar, A., Wang, Y., Tippana, R., Yang, O., Bailey, S., and Ha, T. (2018). Real-time observation of DNA target interrogation and product release by the RNA-guided endonuclease CRISPR Cpf1 (Cas12a). Proc Natl Acad Sci U S A *115*, 5444-5449.

Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. EMBO J *30*, 1335-1342.

Sinkunas, T., Gasiunas, G., Waghmare, S.P., Dickman, M.J., Barrangou, R., Horvath, P., and Siksnys, V. (2013). In vitro reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. EMBO J *32*, 385-394.

Sokolowski, R.D., Graham, S., and White, M.F. (2014). Cas6 specificity and CRISPR RNA loading in a complex CRISPR-Cas system. Nucleic Acids Res *42*, 6532-6541.

Staals, R.H., Jackson, S.A., Biswas, A., Brouns, S.J., Brown, C.M., and Fineran, P.C. (2016). Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. Nat Commun *7*, 12853.

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

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

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

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

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

Swarts, D.C., and Jinek, M. (2019). Mechanistic Insights into the cis- and trans-Acting DNase Activities of Cas12a. Mol Cell *73*, 589-600 e584.

Swarts, D.C., Mosterd, C., van Passel, M.W., and Brouns, S.J. (2012). CRISPR interference directs strand specific spacer acquisition. PLoS One *7*, e35888.

Swarts, D.C., van der Oost, J., and Jinek, M. (2017). Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. Mol Cell *66*, 221-233 e224.

Szczelkun, M.D., Tikhomirova, M.S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., and Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. Proc Natl Acad Sci USA *111*, 9798-9803.

Tamulaitis, G., Kazlauskiene, M., Manakova, E., Venclovas, C., Nwokeoji, A.O., Dickman, M.J., Horvath, P., and Siksnys, V. (2014). Programmable RNA

Shredding by the Type III-A CRISPR-Cas System of *Streptococcus thermophilus*. Mol Cell *56*, 506-517.

Toro, N., Mestre, M.R., Martinez-Abarca, F., and Gonzalez-Delgado, A. (2019). Recruitment of Reverse Transcriptase-Cas1 Fusion Proteins by Type VI-A CRISPR-Cas Systems. Front Microbiol *10*, 2160.

van Houte, S., Ekroth, A.K., Broniewski, J.M., Chabas, H., Ashby, B., Bondy-Denomy, J., Gandon, S., Boots, M., Paterson, S., Buckling, A.*, et al.* (2016). The diversity-generating benefits of a prokaryotic adaptive immune system. Nature *532*, 385-388.

Varble, A., Meaden, S., Barrangou, R., Westra, E.R., and Marraffini, L.A. (2019). Recombination between phages and CRISPR-cas loci facilitates horizontal gene transfer in staphylococci. Nat Microbiol *4*, 956-963.

Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M., and Wang, Y. (2015). Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. Cell *163*, 840-853.

Wang, L., Mo, C.Y., Wasserman, M.R., Rostol, J.T., Marraffini, L.A., and Liu, S. (2019). Dynamics of Cas10 Govern Discrimination between Self and Non-self in Type III CRISPR-Cas Immunity. Mol Cell *73*, 278-290 e274.

Wang, R., Li, M., Gong, L., Hu, S., and Xiang, H. (2016). DNA motifs determining the accuracy of repeat duplication during CRISPR adaptation in Haloarcula hispanica. Nucleic Acids Res *44*, 4266-4277.

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

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

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

Westra, E.R., van Houte, S., Oyesiku-Blakemore, S., Makin, B., Broniewski, J.M., Best, A., Bondy-Denomy, J., Davidson, A., Boots, M., and Buckling, A. (2015). Parasite Exposure Drives Selective Evolution of Constitutive versus Inducible Defense. Curr Biol *25*, 1043-1049.

Wilkinson, M., Drabavicius, G., Silanskas, A., Gasiunas, G., Siksnys, V., and Wigley, D.B. (2019). Structure of the DNA-Bound Spacer Capture Complex of a Type II CRISPR-Cas System. Mol Cell *75*, 90-101 e105.

Wright, A.V., and Doudna, J.A. (2016). Protecting genome integrity during CRISPR immune adaptation. Nat Struct Mol Biol *23*, 876-883.

Wright, A.V., Liu, J.J., Knott, G.J., Doxzen, K.W., Nogales, E., and Doudna, J.A. (2017). Structures of the CRISPR genome integration complex. Science *357*, 1113-1118.

Xiao, Y., Luo, M., Hayes, R.P., Kim, J., Ng, S., Ding, F., Liao, M., and Ke, A. (2017a). Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. Cell *170*, 48-60 e11.

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

Xue, C., Seetharam, A.S., Musharova, O., Severinov, K., Brouns, S.J., Severin, A.J., and Sashital, D.G. (2015). CRISPR interference and priming varies with individual spacer sequences. Nucleic Acids Res *43*, 10831-10847.

Xue, C., Whitis, N.R., and Sashital, D.G. (2016). Conformational control of cascade interference and priming activities in CRISPR immunity. Mol Cell *64*, 826-834.

Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I.M., Li, Y., Fedorova, I., Nakane, T., Makarova, K.S., Koonin, E.V., *et al.* (2016). Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. Cell *165*, 949-962.

Yan, W.X., Hunnewell, P., Alfonse, L.E., Carte, J.M., Keston-Smith, E., Sothiselvam, S., Garrity, A.J., Chong, S., Makarova, K.S., Koonin, E.V., *et al.* (2018). Functionally diverse type V CRISPR-Cas systems. Science.

Yoganand, K.N., Sivathanu, R., Nimkar, S., and Anand, B. (2017). Asymmetric positioning of Cas1-2 complex and Integration Host Factor induced DNA bending guide the unidirectional homing of protospacer in CRISPR-Cas type I-E system. Nucleic Acids Res *45*, 367-381.

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

You, L., Ma, J., Wang, J., Artamonova, D., Wang, M., Liu, L., Xiang, H., Severinov, K.,

Zhang, X., and Wang, Y. (2019). Structure Studies of the CRISPR-Csm Complex Reveal Mechanism of Co-transcriptional Interference. Cell *176*, 239-253 e216.

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

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

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

Zhang, Y., Heidrich, N., Ampattu, B.J., Gunderson, C.W., Seifert, H.S., Schoen, C., Vogel, J., and Sontheimer, E.J. (2013). Processing-Independent CRISPR RNAs Limit Natural Transformation in *Neisseria meningitidis*. Mol Cell *50*, 488-503.
Zinder, N.D. (1958). Lysogenization and superinfection immunity in Salmonella. Virology *5*, 291-326.