2019

# Functional Organization of Molecular Memories in the CRISPR-Cas Immune System

Jon McGinn


# Functional Organization of Molecular Memories in the CRISPR-Cas Immune System

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Jon McGinn

June 2019

# FUNCTIONAL ORGANIZATION OF MOLECULAR MEMORIES IN THE CRISPR-CAS IMMUNE SYSTEM

Jon McGinn, Ph.D.

The Rockefeller University 2019

CRISPR-Cas systems endow bacteria and archaea with adaptive immunity against foreign genetic threats, like phages and plasmids. These immune systems are comprised of CRISPR-associated (Cas) protein effectors and DNA-based storage of immunological memories in the CRISPR array. The CRISPR array is a series of direct repeats intercalated by variable spacer sequences (~30bp) of foreign origin. Upon infection, spacers are excised from the foreign genome and integrated into the array. The array is then transcribed and parsed into individual CRISPR RNAs, each containing a single spacer sequence, which are used by Cas nucleases to identify foreign nucleic acids for destruction. Thus, spacer sequences represent molecular memories that serve to define the specificity of the CRISPR immune response.

New spacers are added invariably to the 5’ end of the array; therefore, the first spacer matches the most recent foreign invader. How this order is established and whether this highly polarized order of spacer insertion influences CRISPR-Cas immunity has not been explored. In my thesis work, I showed that conserved nucleotides within the leader, a sequence located immediately upstream of the CRISPR array, specify the site of new spacer integration with high fidelity. Mutation of this sequence results in erroneous incorporation of new

spacers into the middle of the array. To interrogate the importance of polarized spacer addition, I compared the immune responses generated by CRISPR systems containing wild type and mutant leader sequences. I showed that spacers added through polarized acquisition give rise to more robust immunity than spacers added to the middle of the array. This demonstrated that the CRISPR-Cas system specifies the site of spacer integration to optimize the immune response against the latest and most immediate threat to the host.

Because addition of new spacers pushes existing spacers further downstream, each spacer added to the CRISPR array weakens the immunity provided by already existing spacers within the array. How CRISPR systems address this conundrum had not been explored. In this thesis work, I showed that CRISPR systems exhibit significant natural variation in the rates of spacer acquisition and thereby can modulate the lifespan of existing spacers in the array. Fast-adapting systems can respond quickly to new invaders, but existing spacers rapidly lose their potency. In contrast, slow-adapting systems preserve potency of existing spacers at the cost of reduced rates of spacer acquisition. I showed that bacteria have overcome these tradeoffs by harboring multiple CRISPR systems that acquire new spacers at different rates. I also found that leader-repeat junctions serve as a means for spacer acquisition complexes to discriminate between related CRISPR arrays. I propose a model whereby bacteria can harbor two related CRISPR systems as a means to form both short- and long-term immunological memories against foreign invaders.

Bacteria were once thought to possess only primitive forms of innate immunity, but this notion was turned on its head by the discovery of the CRISPR-Cas immune system. My thesis work has revealed a deeper complexity of bacterial immunity and evolution by demonstrating that CRISPR systems functionally organize molecular memories of past invaders as a means to confer optimal immunity to the host.

*For Mom, Dad, Alyssa-Rae, and Olive.*

**ACKNOWLEDGMENTS**

I am deeply grateful to the countless people who helped me to reach this point in my career. None of this would have been possible without the great mentors, friends, and family in my life.

First, I would like to thank my advisor, Luciano Marraffini, for his constant support and mentorship. Luciano is an exemplary scientist, mentor, and person. His ability to think clearly and creatively about challenging scientific problems, as well as his critical eye, have been instrumental in shaping my approach to science. His genuine curiosity and passion for science are refreshing and infectious. I am particularly appreciative of the freedom Luciano granted me from my first day in the lab, as this shaped my graduate career and fostered my development as an independent scientist. But above all, what distinguishes Luciano as such an exceptional advisor is his extraordinarily caring, kind, and generous nature.

I am grateful to my thesis committee, which has included Daniel Mucida, Gaby Maimon, and David Allis. I would like to give a special thanks to Daniel and Gaby for their role in recruiting me to the graduate program at Rockefeller University. During the interview weekend, Daniel and Gaby spent hours chatting with me at Faculty Club, well past the end of the event. This engaging and insightful conversation demonstrated how committed and caring the faculty at Rockefeller can be and ultimately convinced me to join the program. It has been a privilege to have them on my thesis committee and I am extremely grateful for all of their help and encouragement.

I would also like to thank Kim Seed for graciously agreeing to serve as my external thesis examiner. Her elegant and novel work on CRISPR biology and host-pathogen interactions has been an inspiration throughout my graduate career.

Rockefeller is a remarkable place to be a graduate student, and I would like to thank everyone who works to make that so. I am deeply indebted to the Dean's Office for their assistance over the years. I would like to especially thank Emily Harms for all of her support and guidance. I am grateful to the Genomics Resource Center for experimental assistance. I thank Jeanne Garbarino for her help and her inspiring commitment to science education and outreach. I am thankful to Jesse Ausubel and Mandë Holford for establishing and organizing the exceptional science diplomacy course. I also would like to thank all of the people in the mailroom, student housing (especially Marnel Herbert), food services, cleaning, and security for making Rockefeller such a warm and welcoming place to work and live.

The Marraffini Lab is an exceptional place to be a graduate student, and I am very thankful to have had such great people as lab mates and friends over the years. In particular, I would like to thank several people who directly impacted my thesis work. I am grateful to Poulami Samai, whose mentorship and friendship has been invaluable to my development as a scientist (and a foodie). I also would like to thank Robert Heler, Joshua Modell, Andrew Varble, Wenyan Jiang, Gregory Goldberg, and Alex Meeske for invaluable advice and countless productive scientific discussions. I am also grateful to Sunandini Sridhar, an honorary member of the lab, for her perpetual optimism and her friendship.

I would like to thank my high school science teacher, Ann Grady, for inspiring my passion for science and biology. I would also like to thank Ben Czech and Jon Preall for their encouragement in my decision pursue graduate school. In particular, I learned a great deal from Ben about how to be a diligent, rigorous, and thoughtful experimentalist. I also thank Raina Thomas for her contributions to my thesis work and for being an inspiring student to work with.

On a personal note, I express my deepest gratitude to my partner and best friend, Alyssa-Rae, for her constant support and unconditional love. She makes me a better writer, scientist, thinker, and human every day. Her unwavering confidence in me has sustained me through graduate school. I wouldn't be the person or scientist I am today without her. I'm thankful to Olive for her uncanny ability to make me smile. I also would like to thank all of my family for always being there for me. I am grateful to my late grandparents, Yia Yia Lee and Papou, for their encouragement and for the example they set for me. Finally, I thank my parents for their love, support, generosity, and for making me the person I am today.

# TABLE OF CONTENTS

## LIST OF FIGURES

# CHAPTER 1: INTRODUCTION TO THE CRISPR-CAS IMMUNE RESPONSE

## 1.1 INTRODUCTION

Prokaryotic organisms are frequently exposed to both beneficial and parasitic foreign nucleic acids[1]. On one hand, this allows organisms to access and incorporate diverse genetic material, like antibiotic resistance genes encoded on plasmids or virulence factors encoded by temperate phages. On the other, this renders cells vulnerable to parasitic elements that compromise the fitness of the population, like plasmid-encoded transposons or virulent phages. To balance these costs and benefits, bacteria and archaea have evolved a number of pathways to curate the nucleic acids entering the cell. These pathways have profound implications for the evolution of prokaryotic populations[1].

Clustered regularly interspaced short palindromic repeats (CRISPR) and their associated genes (*cas*) encode one such mechanism by which cells restrict incoming nucleic acids[2,3]. CRISPR-Cas systems have the unique ability to heritably alter the host genome by incorporating small fragments of foreign nucleic acids, known as spacers, in between the repeats of the CRISPR locus (Figure 1.1.1)[2]. This process is known as spacer acquisition [4-6]. Spacers are transcribed and parsed into individual CRISPR RNAs (crRNAs), which guide effector Cas nucleases to cleave cognate nucleic acids. Thus, spacer sequences define the specificity of the CRISPR-Cas immune response, bestowing immunity to both the host and its progeny[2,3].

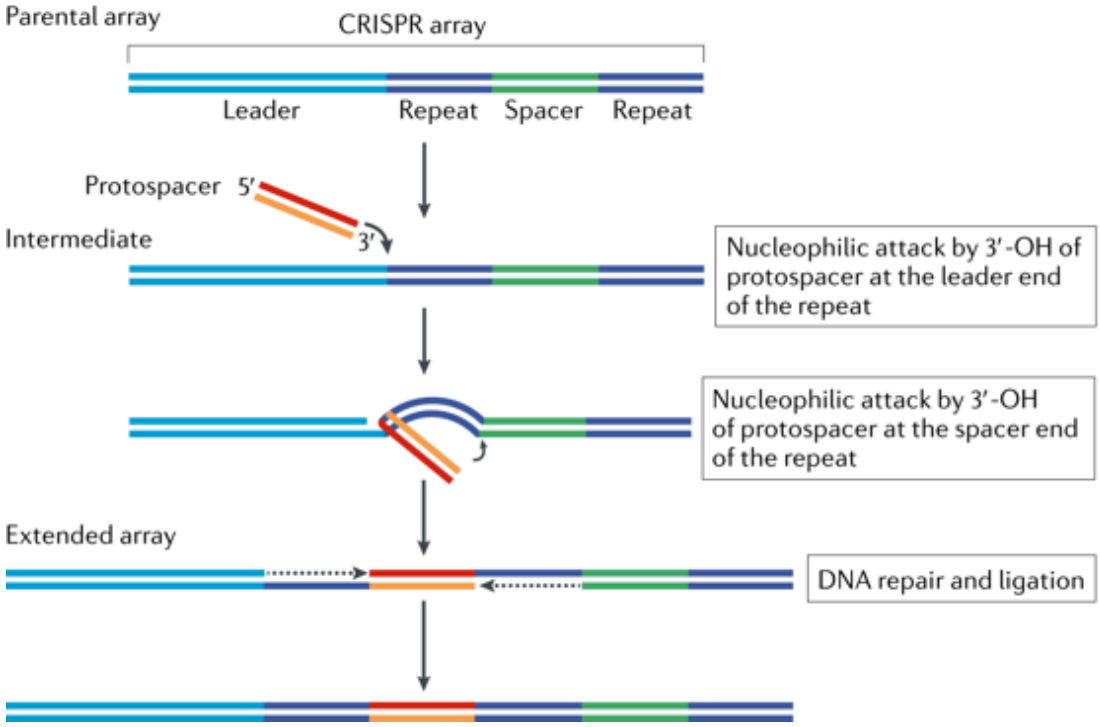



**Figure 1.1.1. General schematic of the spacer integration reaction.**
The CRISPR array comprises a series of direct, semi-palindromic repeats intercalated with variable spacer sequences of foreign origin. The array is preceded by an AT-rich leader sequence. Integration of new spacers begins with a concerted cleavage–ligation reaction that occurs preferentially at the leader end of the first repeat, whereby the terminal 3′-OH of the protospacer carries out a nucleophilic attack. Next, the repeat DNA is bent, and a second cleavage–ligation reaction takes place at the spacer side of the repeat. The product of this reaction is an intermediate in which the 3′ ends of a double-stranded (dsDNA) protospacer are ligated to single-stranded DNA (ssDNA) repeat sequences. The ssDNA repeats are presumably filled by DNA polymerase and ligated to complete the spacer integration process.

Found in approximately 45% of bacteria and 85% of archaea, CRISPR systems have been categorized by *cas* gene content into 2 classes, 6 types, and

over 20 subtypes[7]. Each of the six types utilizes functionally distinct effector complexes that mediate the destruction of foreign nucleic acids. Whereas Types I, II, and V target DNA, Type VI targets RNA, and Type III targets both RNA and DNA (Type IV systems have not been experimentally characterized yet)[7]. In contrast, the core machinery that mediates spacer acquisition, encoded primarily by *cas1* and *cas2*, is relatively well conserved across the different types. Spacer acquisition can be conceptually divided into two phases: protospacer capture and spacer integration. During the first phase, protospacers (spacer sequences in the invading genome) must be selected and extracted from foreign genomes. In the second phase, spacers must undergo processing and incorporation into the CRISPR locus. Recent studies have revealed several aspects of the molecular mechanism of spacer acquisition and how these correlate with the specific targeting mechanism of each different CRISPR Type. Here, we review the current models of spacer acquisition and discuss the future of the field, both in terms of basic science research and technological applications.

## 1.2 INTEGRATION OF NEW SPACERS

### *The Cas1-Cas2 integrase*

The Cas1-Cas2 complex is the core machinery that mediates spacer acquisition. Cas1 is the most highly conserved Cas protein and can be found in all six CRISPR Types[7]. It is believed that Cas1 evolved as the core enzyme of a class of transposons called casposons and was eventually co-opted to form the basis of the CRISPR-Cas immune system[8]. In the context of CRISPR immunity, Cas1 interacts with Cas2 to form a complex that acts as the spacer integrase[9-13].

This heterohexameric complex [$(Cas1_2\text{-}Cas2)_2$] contains two separate DNA-binding regions, one that binds the incoming protospacer and one that binds the CRISPR array. Once loaded with the incoming spacer, the Cas1-Cas2 complex catalyzes two cleavage-ligation reactions, first at the leader-end of the first repeat of the CRISPR array and subsequently at the spacer-end of the repeat[12-14] (Figure 1.1.1). In this reaction, the terminal 3'-OH of each strand of the protospacer DNA performs a nucleophilic attack on each end of the repeat DNA. The product of this reaction is an intermediate in which the 3' ends of a dsDNA protospacer is ligated to ssDNA repeat sequences. These ssDNA "gaps" are presumably filled by a DNA polymerase and then ligated, resulting in a simultaneous spacer insertion and repeat duplication[12-18].

*Site-specific integration of new spacers*

The integration of new spacers by Cas1-Cas2 is polarized, meaning that new spacers are predominantly added to the leader-end of the CRISPR array. For this reason CRISPR loci have been compared to molecular fossil records of past infections, with the newest memories at the leader-end and the most ancestral spacer sequences at the trailer-end [18]. By ordering spacers chronologically, CRISPR systems optimize their immune response against the most recent invaders since leader-end spacers provide more robust immunity relative to more downstream positions[19]. This has been proposed to be a bet-hedging strategy in which resources are prioritized to defend the host against the most recent invader, which is also most likely to be the biggest threat to the cell[19,20]. One potential mechanism underlying this phenomenon is the differential

expression of crRNAs across the CRISPR array, which has been observed in many CRISPR Types[21-25]. In one study of the *Streptococcus pyogenes* Type II-A system, a two-fold difference was observed when comparing the abundance of crRNAs originating from the same spacer sequence in the first or fifth position in the array[19]. Given that a single Cas9 ribonucleoprotein can take hours to find its target (according to single-molecule experiments)[26] and that it is reported to be a single-turnover enzyme[27], it is possible that modest differences in crRNA abundance can give rise to larger differences in the level of immunity afforded to the host [19].

Several mechanisms have the potential to "re-activate" ancestral memories that are positioned further downstream in the CRISPR array. Spacer deletions have been frequently observed in laboratory and natural contexts [28,29]. Indeed, in one study of the Type II-A system from *S. pyogenes*, deletion of four spacers from the 5' end of the array was heavily selected for during phage infection, which enabled a spacer in position 5 to be shifted to position 1 to enable maximum levels of immunity [19]. Alternatively, another study revealed that internal promoters contained within spacer sequences can enable high expression of downstream crRNAs [30]. In addition, we speculate that is possible that lower expression of downstream crRNAs that are unable to provide full immunity could enable a primed immune response (see Primed Spacer Acquisition section).

How this polarized addition of new spacers is achieved differs by CRISPR Type (Figure 1.1.2). In Type I CRISPR systems, an α-helix of Cas1 makes

sequence-specific contacts with the minor groove of the leader[12,31], but this is not sufficient to enforce leader-end spacer addition[15]. Rather, factors encoded by the host genome are required for site-specific integration. In Types I-E and I-F, a protein called Integration Host Factor (IHF) is required for polarized spacer integration in vitro and is required for spacer acquisition in vivo. These Type I leaders contain a conserved IHF binding site, and binding of this site by IHF induces a topological change of the CRISPR array DNA. This creates the ideal target substrate for the Cas1-Cas2 integrase specifically at the first repeat[12,32]. Additionally, the Cas1-Cas2 integrase makes contacts with IHF as well as upstream sequences in the leader as a result of DNA bending induced by IHF[12] (Figure 1.1.2, top). This mechanism leads to a conundrum for Type I CRISPR systems found in Gram-positive bacteria, which lack IHF homologs[32]. It is possible that related DNA-bending proteins (i.e. HU or H-NS) could fulfill this role. Indeed, an archaeal Type I-A system, whose host lacks IHF, exhibits leader specificity for spacer integration in a manner dependent on as of yet unidentified host factor(s) [33].

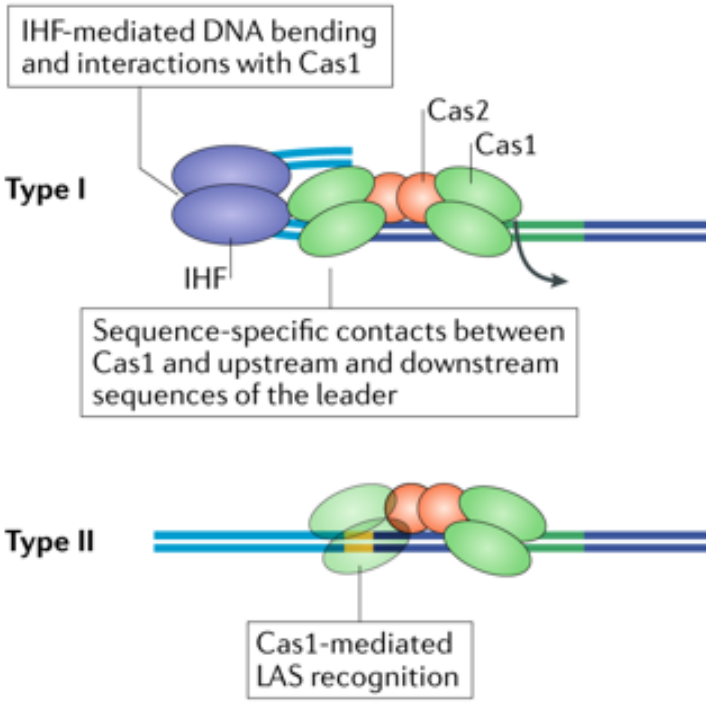



**Figure 1.1.2. Two mechanisms for preferential spacer acquisition at the leader end of the CRISPR array.**

In the type I system, integration host factor (IHF) binds to a conserved binding site in the leader and induces DNA bending, which enables the Cas1–Cas2 complex to perform the first cleavage–ligation reaction. Cas1–Cas2 makes specific contacts with upstream and downstream sequences in the leader, as well as with IHF. In the type II system, Cas1-mediated recognition of the leader anchoring sequence (LAS) in the leader is sufficient for polarized spacer integration.

Type II CRISPR systems, which are commonly found in Gram-positive hosts[34], also exhibit strictly polarized spacer integration[2,14,19,35]. In contrast to the Type I machinery, the Type II Cas1-Cas2 complex can induce the necessary

DNA topology for the spacer integration reaction without any additional host factors[13,14]. Similar to the Type I system, an α-helix of the Type II Cas1 makes sequence-specific contacts with the minor groove of the leader DNA (for Type II, this is termed the Leader Anchoring Sequence, "LAS")[13,14,19]. In contrast to Type I, these contacts are sufficient to enforce leader-end spacer integration without the need of any additional host factors (Figure 1.1.2, bottom)[13,14,19]. These additional contacts between the LAS and Cas1 improve the kinetics of the cleavage-ligation reaction at the leader-repeat junction, enforcing polarized spacer addition[13,14]. Because the second cleavage-ligation reaction occurs at the spacer-repeat junction, the target substrate is variable and this requires some flexibility in the LAS-interacting domain of Cas1 for catalyzing the reaction[13]. Probably as a result of this flexibility, in the absence of a proper LAS the Type II CRISPR systems can undergo "ectopic spacer integration," or integration of new spacers in the middle of the array[19].

## 1.3 PROTOSPACER CAPTURE

*Identification of foreign nucleic acids*

CRISPR systems can acquire self-targeting spacers from the host chromosome[18,36,37], which result in autoimmunity and cell death[38-40]. To avoid this, CRISPR systems employ a variety of mechanisms to bias spacer acquisition to foreign genetic elements.

CRISPR systems utilize the DNA repair machinery of the host, RecBCD in Gram negative organisms[17,36] and its homologue AddAB in Gram positives [41], for the generation of spacer substrates. RecBCD, which binds free ends of dsDNA

to perform end resection during homologous recombination, stimulates spacer acquisition from double strand breaks. This activity is limited by *chi* sites, which are eight nucleotide sequence motifs that slow RecBCD activity. Because *chi* sites are enriched in the host chromosome relative to phage or plasmid genomes, this can serve as a mechanism to constrain spacer acquisition from the host genome and differentiate self versus non-self DNA sources[36,41]. Further, the free dsDNA end that is presented to the cell during infection by dsDNA phages is exploited by the CRISPR system to preferentially acquire spacers from the phage DNA (Figure 1.2.1), since the bacterial chromosome is circular and lacks free DNA ends (with the exception of accidental dsDNA breaks, most common at the terminus). This also biases the pool of acquired spacers to the injected end of the phage genome, which results in the immediate recognition and cleavage of the invading DNA at the very beginning of the infection resulting in more effective immunity[41]. Although RecBCD is important for efficient spacer acquisition, its degradation products are reported to be ssDNA fragments[42,43]. Given that the in vitro spacer integration studies showed that dsDNA protospacer substrates are markedly favored over ssDNA ones[10-15,44], it remains unresolved if and how RecBCD degradation products could be used for spacer integration [45,46]. Alternatively, it is possible to that the Cas1-Cas2 machinery physically associates with RecBCD [47] to either directly uptake degradation products from RecBCD [36] or to sample intact dsDNA upstream of RecBCD. Moreover, given that spacer acquisition can occur in the absence of RecBCD [36] and AddAB [41], it

is clear that alternative pathways exist for spacer generation, which will be an interesting area of future study.

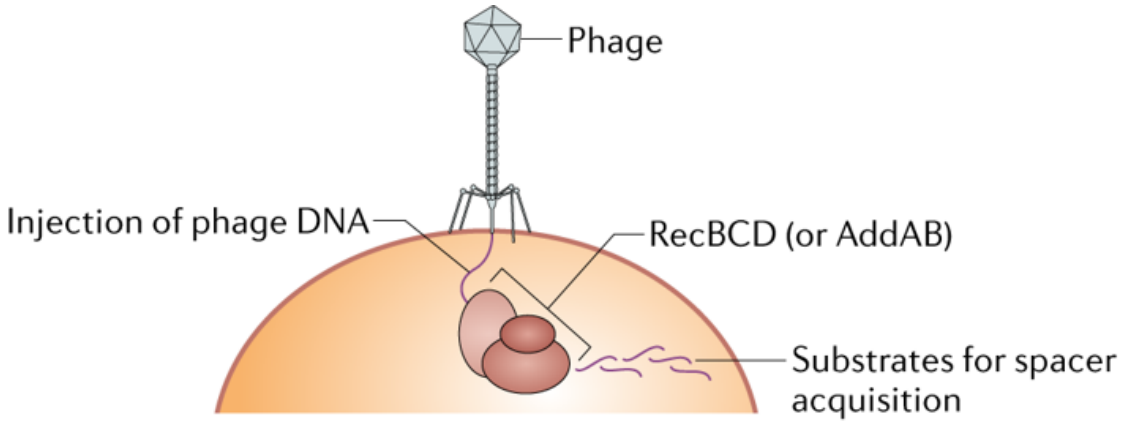



**Figure 1.2.1. Spacers are acquired during viral DNA injection.**
RecBCD in Gram-negative organisms (or AddAB in Gram-positive organisms) generates substrates for spacer acquisition following the injection of viral DNA, possibly by producing more invader DNA molecules that contain free ends.

There is also evidence that evolution has tuned CRISPR systems to avoid deleterious levels of autoimmunity by limiting the rate of spacer acquisition. In laboratory settings, successful acquisition of new spacers against phage is an extremely rare event, estimated to occur in only 1 in $10^7$ cells[19,48,49]. Spacer acquisition from the host genome is equally rare and does not pose significant fitness costs to the host. However, mutants with increased rates of spacer acquisition have been shown to lead to higher levels of toxicity[50], suggesting that the rate of spacer acquisition has been evolutionary tuned to balance protection benefits with autoimmune costs. To mitigate growth rate costs associated with autoimmunity, it is also possible for spacer acquisition to be temporally regulated.

Indeed, quorum sensing has been implicated as a regulator of CRISPR activity in at least two species of bacteria[51,52].

*Selection of functional targets*

Only a subset of sequences in foreign genomes can serve as functional spacers because of Type-specific targeting requirements. In Types I and II, a protospacer-adjacent motif (PAM) located at one of the flanks of the target is required to license target cleavage and to prevent the cleavage of the spacer sequence in the CRISPR array (repeats lack properly positioned PAMs)[53,54]. While Type III flanking sequence requirements are more flexible [55-57], transcription across the target is needed for targeting of the DNA[58-61]. To ensure functional immune responses, CRISPR systems must select protospacers that are flanked by the correct PAM or are actively transcribed.

In Types I and II, the spacer acquisition machinery preferentially samples protospacers with functional PAMs. However, the two types employ different mechanisms to accomplish this (Figure 1.2.2). In Type I-E, the Cas1-Cas2 complex has direct, sequence-specific interactions with the PAM that biases acquisition to PAM-adjacent protospacers[11] (Figure 1.2.2, top). In contrast, the Cas1-Cas2 complex from the Type II CRISPR system does not exhibit any PAM selectivity[13,48]. Instead, Cas9's PAM-interacting domain enforces PAM-specific spacer acquisition through direct interactions with the Cas1-Cas2 complex (as well as the Type II accessory protein Csn2)[48] (Figure 1.2.2, bottom). Spacer acquisition has not been observed experimentally for many of the transcription-dependent Type III CRISPR systems. A small fraction of Type III systems has

been shown to harbor reverse transcriptase-Cas1 (RT-Cas1) fusion proteins. Indeed, one such system has been observed to preferentially acquire spacers from transcribed regions of genomes[62,63]. While the exact mechanism remains unclear, it has been demonstrated that RT-Cas1 fusion proteins can acquire new spacers directly from RNA transcripts. These RNA-derived spacers provide a mechanism to ensure that their targets are transcribed and can be recognized by the Type III RNA-guided nucleases.

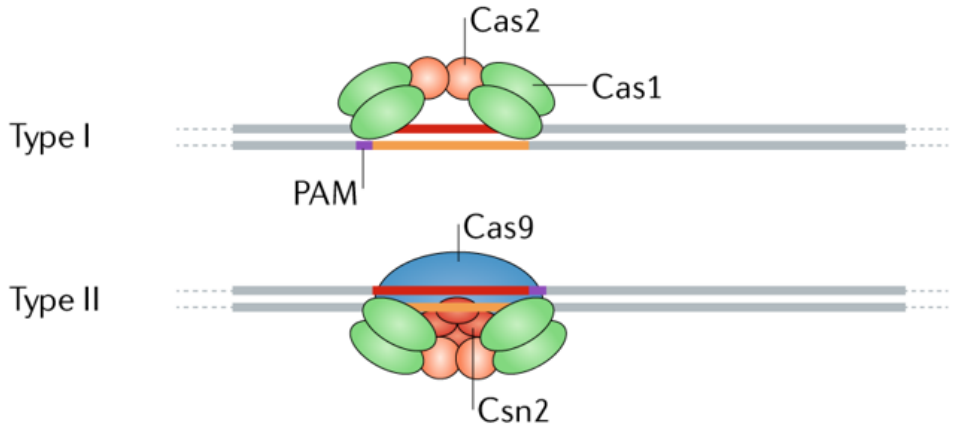



**Figure 1.2.2. Two mechanisms for selection of functional targets.**
In the type I-E system, Cas1–Cas2 has inherent substrate preference for protospacers with a canonical protospacer-adjacent motif (PAM). In type II, the PAM-interacting domain of Cas9 (loaded with trans-activating CRISPR RNA (tracrRNA), not shown) guides the Cas1–Cas2 complex (as well as the accessory protein Csn2) in selecting protospacers.

## 1.4 PRIMED SPACER ACQUISITION

Pre-existing spacers can enhance the rate of spacer acquisition in a sequence-dependent manner through a process known as primed spacer acquisition, or priming (enhanced relative to “naïve acquisition,” in which no full

or partially matching spacers are present in the CRISPR array). Priming can arise from either perfectly matching spacers or spacers against mutated phages (or related phages) with point mutations in the spacer region or PAM. Thus, once a host acquires a single spacer against a phage, it becomes more likely to subsequently acquire additional spacers from the vicinity of the priming target region in the phage genome[30,64]. This feed-forward cycle is driven by close associations and interactions between the spacer acquisition machinery and the interference machinery. Indeed, the importance of these associations is underscored by the existence of fusion proteins in which interference genes have been fused to spacer acquisition genes, like the Cas2-Cas3 fusion protein present in Type I-F systems [44,65].

As of yet, priming has only been observed for Type I CRISPR-Cas systems, with the Types I-E and I-F systems as the best studied examples. During Type I-E targeting, the crRNA-guided Cascade complex binds to a foreign target in a PAM-dependent manner, upon which it recruits the nuclease Cas3 for target destruction[66]. In addition to eliminating the foreign genome, the nuclease and helicase activities of Cas3 also drive the production of spacer substrates (Figure 1.2.3) [45]. Further, in the absence of a proper PAM, Cascade can still bind the target and recruit Cas3 in a manner dependent on Cas1-Cas2 [67]. These distinct outcomes arise from alternative conformations adopted by Cascade, in particular the Cse1 subunit, upon binding of either a perfect target or a mutated target [67-70]. In this context, the nuclease domain of Cas3 is inactive and it is believed that its helicase activity is used to translocate the Cas1-Cas2 complex

along the nearby DNA and drive primed spacer acquisition by the integrase complex (Figure 1.2.4) [65,67].

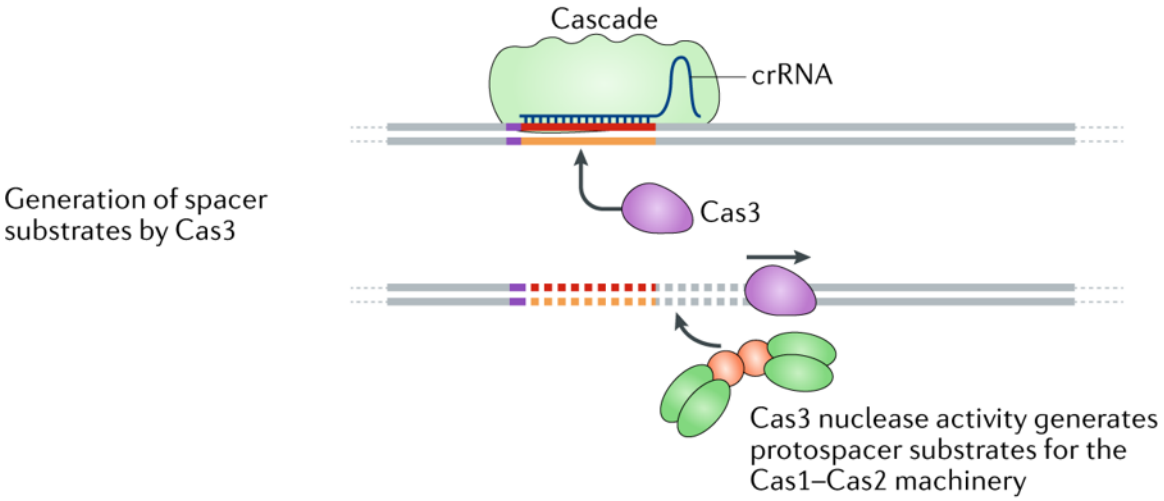



**Figure 1.2.3. Primed spacer acquisition against a perfect target.**
The CRISPR RNA (crRNA)-guided CRISPR-associated complex for antiviral defence (Cascade) binds to a foreign target in a PAM-dependent manner, and it subsequently recruits the nuclease Cas3, which results in the generation of suitable substrates for spacer acquisition.

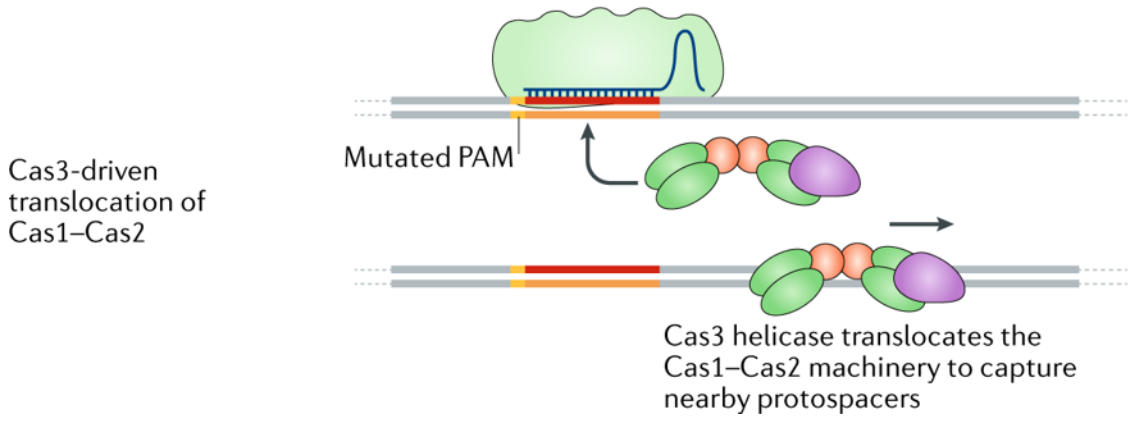



**Figure 1.2.4. Primed spacer acquisition against a mutated target.**
Imperfect target recognition by Cascade results in an altered conformation of the Cse1 subunit. This leads to the recruitment of a nuclease-inactive Cas3 in a Cas1–Cas2-dependent manner, which mediates primed spacer acquisition.

How these two outcomes relate to one another, and how they are decided between, remains unresolved. One possibility is that there are two distinct pathways, namely primed spacer acquisition and interference-driven spacer acquisition, where there is a threshold for the amount of mutations tolerated to either license an interference or priming response. In this case, both responses can result in elevated rates of spacer acquisition, though a significantly higher boost in the case of interference-driven spacer acquisition [65,71,72]. Alternatively, it is possible that the CRISPR machinery does not make such strict distinctions and rather displays a continuum of activities, ranging from interference to priming. In this case, it has been suggested that a single mechanism could explain this range of activities [73].

Primed spacer acquisition allows organisms to defend themselves against rapidly evolving phage populations, like phages that evade CRISPR-Cas immunity through the introduction of target mutations or related phages with conserved but not identical targets[64,74,75]. Priming can also serve as a mechanism to bias spacer acquisition activity to foreign DNA substrates that harbor the priming target; i.e., using the molecular memories stored in the CRISPR array to differentiate self vs. foreign DNA[74]. However, the increased rate of spacer acquisition resulting from priming poses a dilemma for the host cell: the more spacers a CRISPR system acquires, the more likely it is that a spacer with a partial match to the host chromosome is incorporated, which would elicit primed spacer acquisition from the host genome, resulting in autoimunity[72]. How CRISPR systems balance these costs and benefits remains to be addressed.

Whether priming occurs in other CRISPR Types also remains to be determined. Similarly to the interaction between Cascade and Cas3 during priming, Cas9 has been shown to interact with the Cas1-Cas2 integrase[48], an observation that opens up the possibility of priming in the Type II CRISPR-Cas immune response. For Type III systems, on the other hand, the presence of mismatches between the crRNA and the target sequence does not abrogate immunity[57], and therefore it seems unlikely that primed spacer acquisition can occur as it happens for Type I systems.

## 1.5 THE ORIGIN OF SPACER SEQUENCES

CRISPR arrays give unique insight into the genetic material encountered and selected against during the course of prokaryotic evolution. As a molecular fossil record of past invasions, we can infer not only the genomes that a given organism encountered but also in what order. Thus, there has been significant interest in analyzing the origin of spacer sequences. Of the spacers that can be mapped to sequenced genomes, 80-90% of spacers map to phage genomes (Figure 1.3.1). The remainder of mapped spacers match genes associated with mobile genetic elements. Surprisingly, though, a vast majority of spacer sequences (>90%) originate from unknown sources, comprising the CRISPR "dark matter" (Figure 1.3.1)[76].

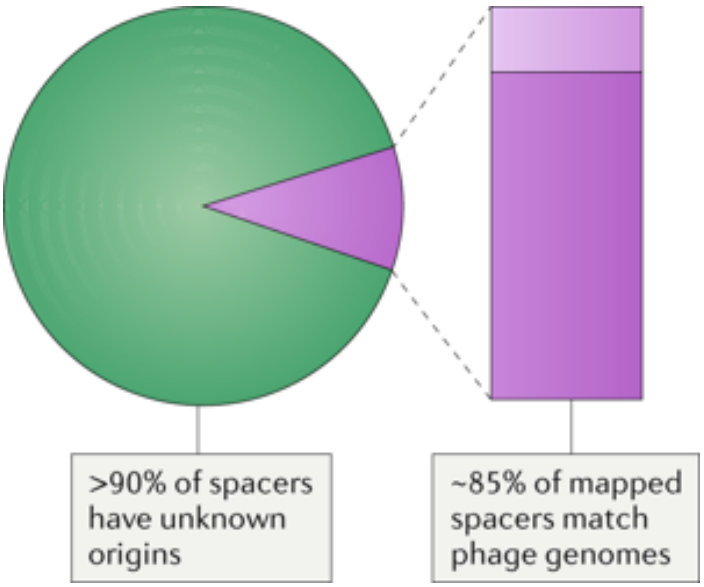



**Figure 1.3.1. Origin of spacer sequences.**
From bioinformatic analysis of CRISPR arrays in sequenced bacterial genomes, the majority of spacers are of unknown origin. But of the fraction of spacers that can be mapped, the majority of spacers map to phage genomes, though spacers matching plasmids, transposons, and other mobile genetic elements can be identified.

For these mysterious spacer sequences, perhaps the most parsimonious explanation is that sequence databases are missing a vast diversity of phage and MGE sequences. Indeed, sequence analysis has determined that many of these unmapped spacer sequences share similar properties with the mapped spacer sequences, suggesting that they would map to uncharacterized classes of phages and mobile genetic elements[77]. These dark matter spacers have also been found for a large majority of RT-Cas1-derived spacers, suggesting a diversity of RNA phages and RNA MGEs that remain to be discovered as well[78]. It is also a possibility that these spacer sequences foreshadow the discovery of unexpected sources of nucleic acids.

## 1.6 CONCLUSIONS AND OUTLOOK

More than ten years after the discovery of CRISPR's function as a form of prokaryotic adaptive immunity, we are fast approaching a comprehensive understanding of the molecular mechanisms underlying spacer acquisition. Through complementary studies utilizing genetic, biochemical and structural approaches, we now have key insights into several major steps of this process, from the selection of spacers from foreign genomes to the cleavage-ligation reactions of new spacers into the first repeat of the CRISPR array. However, gaps in our knowledge exist, particularly at the boundaries of protospacer capture and spacer integration. For instance, while it has been demonstrated that RecBCD and AddAB are involved in the generation of substrates for spacer acquisition, the mechanism of this process is largely unsolved. Further, what kind of processing protospacers undergo and how spacer size is regulated remain understudied. It is possible that relatively uncharacterized accessory proteins (like Cas4 and Csn2) that are associated with spacer acquisition modules could play a role in these processes [79]. The development of new in vivo and in vitro experimental systems will be key to gaining insight into these processes. Indeed, one such study has already suggested a role for Cas4 in PAM-dependent protospacer processing [33].

More broadly, it will be interesting to learn how spacer acquisition affects other aspects of prokaryotic life. From a cell biological perspective, it will be valuable to learn more about the interplay between CRISPR spacer acquisition and other host-encoded genes or pathways, like other phage defense

mechanisms. From an ecological standpoint, the extent to which spacer acquisition occurs in different environmental contexts remains to be explored. Finally, metagenomic studies combined with establishment of new in vivo systems will be instrumental in understanding how spacer acquisition and CRISPR immunity influence the evolution of natural prokaryotic populations in physiologically and ecologically relevant contexts.

The spacer acquisition machinery has begun to be repurposed for technological applications. Rather than being used for directed genome editing like Cas9, the Cas1-Cas2 machinery has been repurposed as a synthetic molecular recorder[80]. This could be used for storing digital information in genomes of bacterial populations[81]. Additionally, an alternative technique has been developed for recording environmental signals encountered by bacterial populations[82]. However, there are currently several factors that limit the utility and function of these techniques. Most notably, the low frequency of spacer acquisition events makes it impossible to have reliable recording within single cells, therefore the current technologies rely on deep sequencing of large populations to detect spacer acquisition. It is possible that tools like hyper-Cas9, a hyperactive spacer acquisition mutant[50], could accelerate both basic and technological research in the spacer acquisition field.

The storage of information is a fundamental aspect of all biological systems. CRISPR loci constitute a unique form of biological memory, serving to provide heritable and adaptive immunity to prokaryotes. Future research on the

biology of CRISPR and spacer acquisition is sure to illuminate our understanding of biological memory, prokaryotic evolution, and host-pathogen interactions.

## CHAPTER 2: CRISPR SYSTEMS OPTIMIZE THEIR IMMUNE RESPONSE BY SPECIFYING THE SITE OF SPACER INTEGRATION

### 2.1 INTRODUCTION

Prokaryotes are faced with the perpetual threat of invasion by foreign nucleic acids through phage infection and horizontal gene transfer. Clustered regularly interspaced short palindromic repeats (CRISPR) loci and CRISPR-associated (Cas) proteins comprise a heritable and adaptive immune system that protects bacteria and archaea from phage[2] and plasmid[3] infections. Immunological memories of these infections are stored in the CRISPR array as short spacer sequences that intercalate between repeats and specify the targets of CRISPR-Cas immunity. Upon infection, new spacer sequences matching the genome of the invading DNA are added to the 5' end of the CRISPR array[2]. Therefore the CRISPR locus constitutes a molecular fossil record of infections in which the first spacer matches the most recent foreign threat, whereas downstream spacers correspond to older infections. The CRISPR array of repeats and spacers is first transcribed as a long precursor that is processed at the repeat sequences to generate small, mature, CRISPR RNAs (crRNAs)[21,83,84]. These associate with and direct RNA-guided Cas nucleases to their targets, known as protospacers, in the genome of the invader[59,85-87]. Cleavage of the viral or plasmid target DNA prevents infection[3,88].

Based on the *cas* gene content, CRISPR-Cas systems are classified into six types (I-VI) and 19 subtypes[89,90]. The polarity of spacer incorporation, i.e. the addition of new spacers in the first position of the CRISPR array, is a feature of all CRISPR-Cas types studied so far. At the molecular level, the process of

spacer acquisition has been mostly studied using the type I-E system from *Escherichia coli* MG1655[18,74]. The Cas1-Cas2 complex from *E. coli* serves as an integrase, catalyzing a nucleophilic attack by the incoming spacer at the leader-proximal repeat[9,15,16]. During this concerted cleavage-ligation reaction, the spacer is added to the array and the repeat is duplicated simultaneously. The integration host factor (IHF) protein directs the addition of spacers into the first position of the CRISPR array[32]. IHF is a histone-like bacterial protein conserved in Gram-negative organisms that binds an AT-rich region immediately upstream of the type I-E CRISPR locus, known as the leader sequence[91], creating the proper DNA topology for Cas1-Cas2-mediated spacer integration at the first repeat. Type II-A CRISPR-Cas systems also display a stringently polarized spacer acquisition process[2,48]. However, most of these systems are present in Gram-positive bacteria, which lack IHF homologs. In these systems the leader sequence is also important for spacer acquisition[35], however how the polarity of this process is achieved is not clear. More importantly, the physiological significance of polarized spacer acquisition, a fundamental feature of CRISPR-Cas immunity, has not been explored in any CRISPR type.

Here we studied these fundamental problems of spacer acquisition in the type II-A CRISPR-Cas system of *Streptococcus pyogenes* SF370. Consistent with previous studies, we found that deletions of the array-proximal region of the leader abolish spacer integration at the leader-end of the CRISPR array. However, these deletions do not abolish all spacer acquisition activity. Instead, leader mutations result in the erroneous integration of new spacers into the

middle of the array, a phenomenon we term ectopic spacer integration. Further interrogation revealed that a short and conserved sequence at the 3' end of the leader dictates the site of spacer integration. By uncoupling the requirement for the leader during spacer acquisition from its role in positioning spacer integration, we were able to investigate the physiological significance of spacer order within the CRISPR array. We determined that wild-type, polarized, spacer integration provides a fitness advantage over ectopic spacer acquisition due to higher levels of host protection provided by spacers in the leader end of the array. This increased level of immunity is particularly critical during high titers of phage, like those that occur during CRISPR immunization. Our results demonstrate that polarized spacer acquisition ensures robust immunity against the latest invader, and thereby the most immediate threat to the host.

## 2.2 RESULTS

*Deletions within the leader sequence result in ectopic spacer integration.*

The type II-A CRISPR system of *S. pyogenes* SF370 contains four *cas* genes, a *tracrRNA* gene, and six spacers (*spc1-6*) in the CRISPR array[21] (Fig. 2.1.1-A). Immediately upstream of the first repeat there is a 102 bp, AT-rich sequence known as the leader[91]. We previously studied the roles of the different *cas* genes in spacer acquisition by cloning this CRISPR-Cas system on the pC194 staphylococcal plasmid and using *Staphylococcus aureus* RN4220, a genetically tractable Gram-positive organism without an endogenous CRISPR-Cas locus, as the host[92]. Here we used this experimental set up to investigate the function of the leader sequence. We created three strains containing different

deletions of 25, 15 or 5 bp at the 3' end of the leader sequence (Fig. 2.1.1-B). Cultures of these mutant strains as well as a wild-type control were infected with the staphylococcal lytic phage ΦNM4γ4[48] during exponential growth, at a multiplicity of infection (MOI) of 1 virus per bacterium. After 24 hours, DNA was isolated from the surviving cells in each culture and used for PCR analysis of the CRISPR locus with primers that amplify the leader-end and thus detect the acquisition of new spacers in the first position of the CRISPR array[48] (Fig. 2.1.1-A, blue arrows). All three deletions prevented any detectable insertion of new spacers in this position (Fig. 2.1.1-C). This is in agreement with the results obtained with the type II-A CRISPR1 locus of *Streptococcus thermophilus* DGCC7710[35]. All of the leader deletion strains tested gave rise to bacteriophage-resistant mutants, however, which suggested that the CRISPR-Cas immune response could still be functional. We hypothesized that spacer acquisition could occur in other positions of the CRISPR array and therefore used a different set of primers to detect the incorporation of new spacers in any position (Fig. 2.1.1-A, green arrows). Surprisingly we were able to observe *bona fide* spacer acquisition events in the middle of the CRISPR array (Fig. 2.1.1-C). Upon sequencing of many of these PCR products we determined that, in all three leader deletion mutants, the majority of spacer integration events occurred at the fifth repeat, positioning the new spacer between *spc4* and *spc5* (Fig. 2.1.1-D). We term this phenomenon, in which new spacers are added into the middle of the array, "ectopic" spacer integration. These experiments provide direct, *in vivo* evidence that the leader is involved in specifying the site of spacer integration.

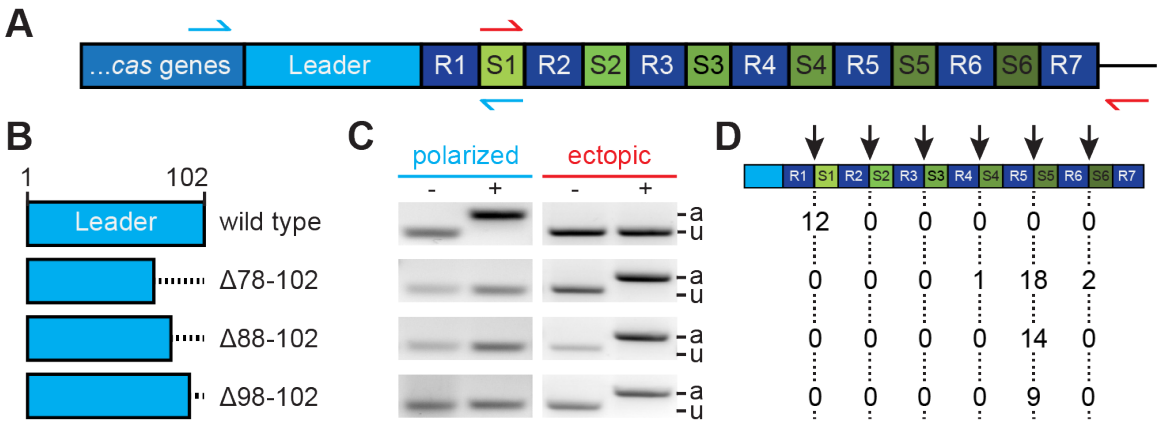



**Figure 2.1.1. Deletions within the Leader Sequence Result in Ectopic Spacer Integration**

(**A**) Type II-A CRISPR locus from *Streptococcus pyogenes* SF370. A 102 bp leader sequence separates the *cas* genes from the CRISPR array, which contains seven repeats (R1–7) flanking six spacers (*spc1-6*). Blue and red arrows indicate primers used to detect spacer integration at the leader-end (polarized) or at the middle of the array (ectopic), respectively.

(**B**) Deletions of the leader sequence analyzed in this study.

(**C**) PCR-based detection of polarized or ectopic spacer integration using the primers described in (A). DNA for PCR was extracted from colonies obtained from cultures incubated with (+) or without (–) phage ΦNM4γ4. PCR products were separated by agarose gel electrophoresis and stained with ethidium bromide. A gel image representative of many PCRs is shown. The size of the PCR product reflects the presence (a, “adapted”) or absence (u, “unadapted”) of integration of new spacers.

(**D**) Position of spacer integration events (marked by the black arrows) after Sanger sequencing of PCR products obtained in (C) for cells infected with phage carrying different leader sequence deletions. Numbers in table represent totals from three replicates.

*A leader-anchoring sequence specifies the site of integration for new spacers.*

To investigate the leader sequences that specify the site of integration of new spacers in more detail, we introduced A-C and G-T transversion mutations in the 10 bp at the 3'-end of the leader. We tested two mutants, one with mutations in the -5 to -1 region of the leader and one with mutations in the -10 to -6 region (Fig. 2.2.1-A). PCR analysis after phage infection with each set of primers used in Figure 1 showed that only the -5 to -1 leader mutant resulted in ectopic spacer integration (Fig. 2.2.1-B). We have therefore termed this region of the leader the "leader-anchoring sequence," or LAS. This sequence, especially the 3'-end GAG, is highly conserved in related type II-A CRISPR systems[93] (Fig. 2.2.2-A). To precisely determine the effect of the LAS on the position of spacer integration within the CRISPR locus, we performed next-generation sequencing of the PCR products containing the full array obtained after infection of wild-type and LAS mutant cultures. We found that while wild-type cells displayed polarized spacer integration almost exclusively (>99.999% of new spacers were integrated into the first repeat, Fig. 2.2.1-C), LAS mutant bacteria acquired new spacers at different positions in the array, with only ~ 0.65 % of the integration events occurring at the first repeat and more than 99% at the fifth repeat, positioning the new spacer between *spc4* and *spc5* (Fig. 2.2.1-C).

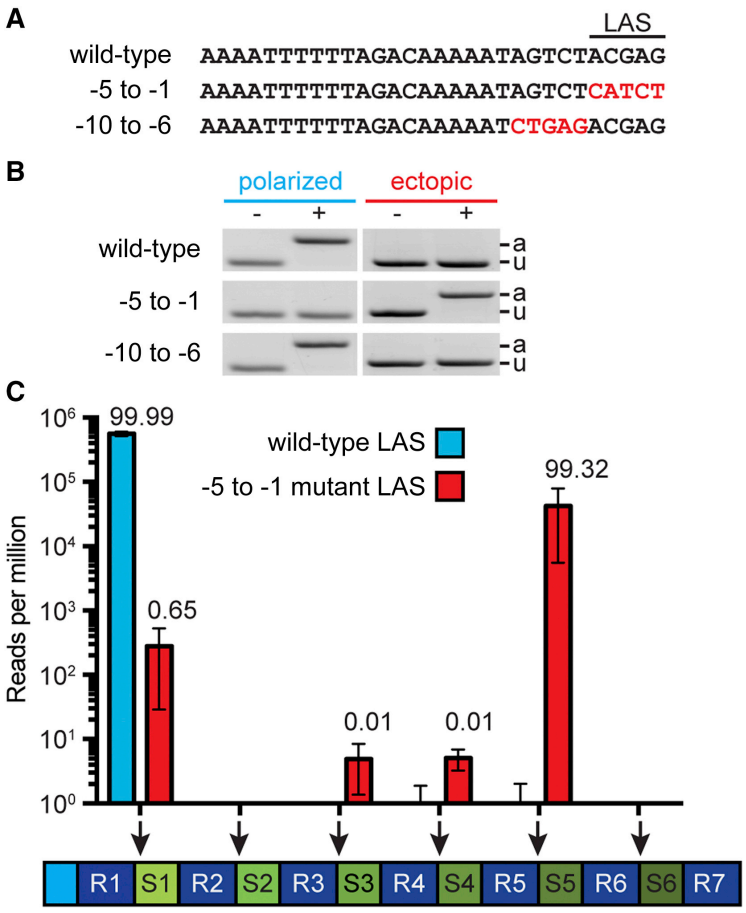



**Figure 2.2.1. A Sequence within the Leader Specifies the Site of Spacer Integration**

(**A**) Mutations were introduced at the 3′ end of the leader to define the leader-anchoring sequence (LAS). (**B**) Strains containing the leader mutations described in (A) were infected with phage ΦNM4γ4 and analyzed by PCR for polarized and ectopic spacer integration. (**C**) Analysis of the site of spacer integration using next-generation sequencing. Liquid cultures harboring a wild-type or mutant LAS were infected at an MOI of 1, and the DNA isolated from surviving cells at the end of infection was used for PCR amplification of the entire array. The expanded PCR amplicons were purified from the gel and used for MiSeq next-generation sequencing. Bars show the number of normalized reads for the integration of new spacers in each possible position of the CRISPR array (marked by the black arrows). Mean ± SEM of three replicates are reported. Percentage of spacer integration events is indicated above each bar.

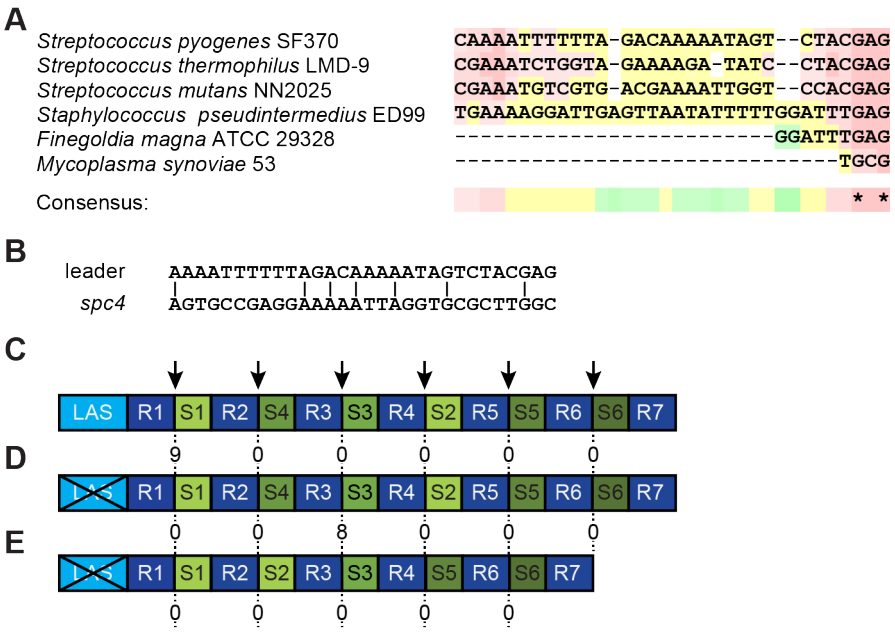



**Figure 2.2.2. The leader anchoring sequence is conserved in related Type II-A CRISPR systems.**

(**A**) Alignment of leader sequences upstream of the first repeat of type II-A CRISPR-Cas systems present in different bacteria. Red indicates strong conservation, yellow indicates average conservation, green indicates poor conservation. (**B**) Alignment of the 30 nucleotide leader sequence upstream of the first repeat of the S. pyogenes CRISPR locus and the spc4 sequence. (**C**) Spacer integration in cells harboring a CRISPR locus in which the sequences of spc2 and spc4 were swapped, downstream of a wild-type LAS. (**D**) Spacer integration in cells harboring a CRISPR locus in which the sequences of spc2 and spc4 were swapped, downstream of a mutant LAS. (**E**) Spacer integration in cells harboring a CRISPR locus in which the sequences of repeat #5 and spc4 were deleted, downstream of a mutant LAS.

We wondered whether the sequence of *spc4* could work as a “pseudo-LAS” upstream of the fifth repeat that would direct the integration of new spacers in the absence of the wild-type LAS upstream of the first repeat. Comparison of

the LAS and *spc4* sequences indicated that there is very little homology between them (Fig. 2.2.2-B). However, it is possible that different sequences could serve as anchors for spacer integration. To test this, we switched the positions of *spc2* and *spc4* and determined the location of newly acquired spacers. As expected, switching the spacers did not affect polarized acquisition in the strain harboring the wild-type LAS (Fig. 2.2.2-C). However, all of the LAS-mutant, phage-resistant colonies tested contained the new spacer integrated in the third repeat, i.e. immediately downstream from the new location of *spc4* in this strain (Fig. 2.2.2-D). In addition, the combination of the LAS mutation and the deletion of *spc4* prevented the detection of spacer integration (Fig. 2.2.2-E). These results suggests that in the absence of a proper LAS, other sequences within the type II CRISPR array (*spc4* in the case of the *S. pyogenes* CRISPR-Cas system) can anchor spacer integration. Altogether our experiments show that short sequences immediately upstream of repeats are able to specify the site of spacer integration. In particular a short conserved sequence within the CRISPR leader immediately upstream of the first repeat, here named the LAS, specifies the acquisition of new spacers in the first position of the array with high fidelity.

*The LAS provides a competitive advantage during spacer acquisition.*

Our experiments with LAS mutant cells showed that the CRISPR-Cas immune response does not absolutely require the addition of new spacers in the first position. In spite of this, all CRISPR-Cas systems studied so far display an invariable specificity for spacer integration in this position. We wondered if polarized spacer acquisition provided an advantage versus ectopic acquisition.

To test this we carried out a series of competition analyses between strains that acquire spacers in different positions within the CRISPR array. First we performed a pairwise competition assay between strains harboring wild-type or mutant LAS. Each of the two naïve strains were grown to exponential phase and mixed in a 1:1 ratio. One aliquot of the mixed culture was uninfected as a control, while another was infected with phage ΦNM4γ4 at an MOI of 1. The cultures were grown for 24 hours and then streaked onto agar plates. DNA from individual colonies (originating from cells that survived phage infection) was used for PCR and Sanger sequencing to determine the population composition after the experiment (16 colonies were analyzed per replicate; 48 colonies total), Fig. 2.3.1-A). The control showed a 1:1 ratio of wild-type to LAS mutant cells in the absence of phage infection, indicating that there is no intrinsic selective advantage for any of the strains (16 colonies were analyzed per replicate; 48 colonies total). In contrast, in the presence of phage only the strain harboring the wild-type leader was able to generate resistant colonies, suggesting a strong fitness defect for the LAS mutant cells. As expected, all the colonies contained new spacers in the first position of the CRISPR array (Fig. 2.3.2-A). This fitness defect was also observed during monoculture phage challenges, where the survival of cells is measured as the optical density of the culture after addition of phage (Fig. 2.3.1-B). In this experiment naïve cells succumb to viral infection and the optical density decreases dramatically. However, cultures can regain growth upon acquisition of new spacer sequences that direct phage destruction. While

cells harboring a wild-type leader sequence restart growing at ~16 hours, LAS mutants take ~25 hours to regrow.

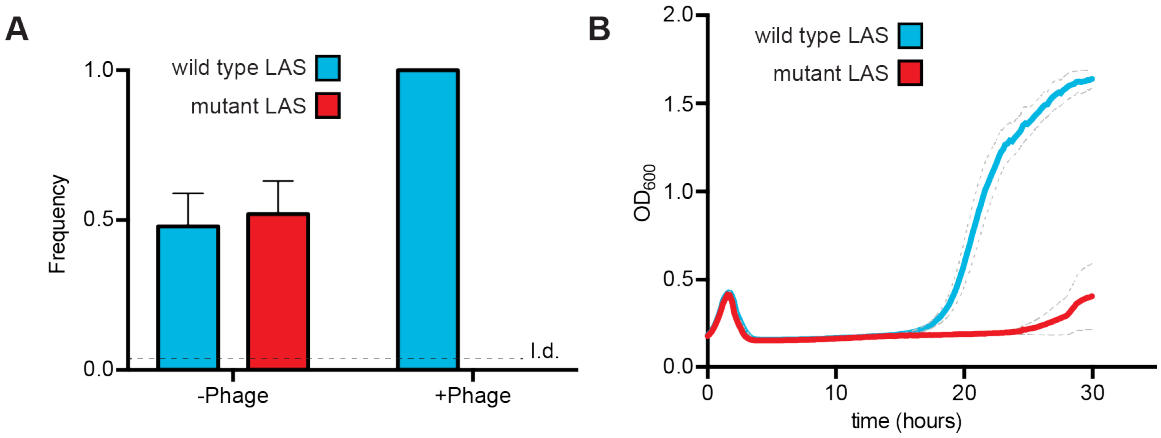



**Figure 2.3.1. The LAS Confers a Fitness Benefit during CRISPR-Cas Immunity**

(**A**) Analysis of culture composition following direct competition between strains harboring the wild-type or mutant LAS. Strains were mixed in a 1:1 ratio and infected with phage ΦNM4γ4 at an MOI of 1. Once the infection completed, the cultures were streaked onto a plate and colonies were picked to determine their LAS by Sanger sequencing (n = 16 per condition per replicate, 96 colonies tested in total). Mean + SEM of three replicates are reported. L.d, limit of detection. (**B**) Growth of cultures infected with phage ΦNM4γ4 followed by the measurement of optical density at 600 nm ($OD_{600}$). Cells containing a wild-type or mutant LAS were infected, and their $OD_{600}$ was followed over time. While most cells die after infection, a small fraction can acquire new spacers and resume growth after viral clearance through CRISPR-Cas immunity. Mean ± SEM (gray dotted line) of three replicates are reported.

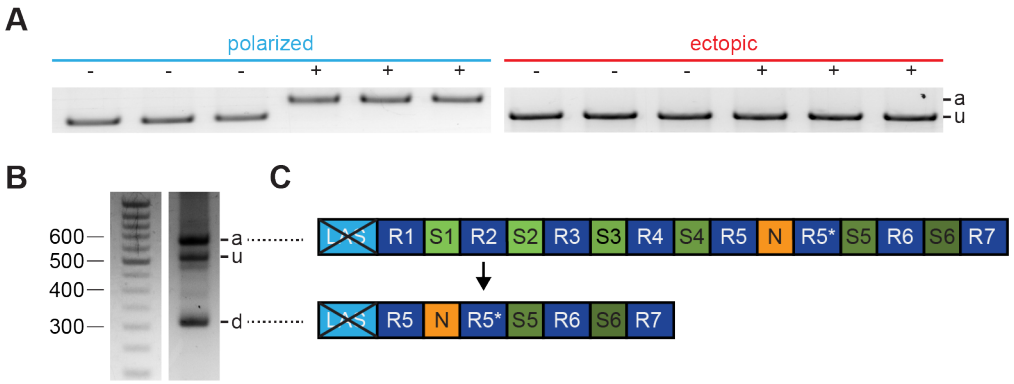



**Figure 2.3.2. Deletions of repeat-spacer units within the CRISPR locus shift ectopically integrated spacers to the leader-end of the array.**

(**A**) DNA from mixed cultures incubated with (+) or without (-) phage ΦNM4γ4 (from Fig. 3A, which contains the average of the three biological replicates for each condition reported here) was extracted 24 hours post-infection and used as template for PCRbased detection of polarized or ectopic spacer integration as in Fig. 1A. (**B**) DNA from cells harboring a mutant LAS that recovered after phage infection (from Fig. 2C) was extracted and used as template for PCR to detect expansion of the CRISPR array. In addition to PCR products corresponding to expanded, or adapted (a), and unadapted (u) CRISPR loci, a third PCR product (d), smaller than the other two, was observed. (**C**) Next-generation sequencing of the DNA rendered sequences corresponding to the smaller band in which the first four repeat-spacer units were deleted and the fifth, newly acquired spacer was repositioned as the first spacer.

Interestingly, when preparing the samples for next-generation sequencing (Fig. 2.2.1-C), we noticed an additional PCR product from the LAS mutant sample, smaller than both the naïve and adapted CRISPR arrays present in both populations (Fig. 2.3.2-B). Sequencing of these PCR products revealed the presence of CRISPR loci that had lost the first four repeat-spacer units to

relocate the new spacer in the first position (Fig. 2.3.2-C). This result shows that, in the event of ectopic adaptation, there is selective pressure to reposition new spacers to the leader-end of the array. Altogether these experiments demonstrate that, by enabling the acquisition of new spacers in the first position of the CRISPR array, the LAS confers a selective advantage during the CRISPR-Cas immune response.

*Spacers in the first position of the CRISPR array provide a more robust immune response.*

There are at least two possible explanations for the fitness advantage of cells harboring a wild-type LAS observed in Figure 3. One scenario is that *spc4* has poor LAS properties and spacer integration at position 5 (in the LAS mutant) is less efficient than at the first position (in the wild-type CRISPR locus). Another possibility is that spacers integrated into repeat 5 provide a weaker CRISPR immune response and take more time to clear the virus and regrow. To explore the first scenario, we replaced *spc4* with a 30 bp fragment of the leader containing the LAS (Fig. 2.4.1-A). In this way the LAS is present upstream of both the first and fifth repeats, and there is an optimal LAS for the acquisition of spacers in the first or fifth position of the CRISPR array. This set up allowed us to perform an “intracellular competition experiment” in which integration at repeat 1 (the wild-type position) and repeat 5 (the ectopic position) should have similar rates. We first corroborated that, in the absence of a wild-type in the LAS upstream of the first repeat, the replacement of *spc4* for the LAS resulted in the exclusive acquisition of new spacers in position 5 (Fig. 2.4.2-A). This also

demonstrates that the LAS positively directs spacer acquisition, and therefore that the particular LAS mutations that we introduced are not inhibitory for spacer incorporation. Next, we infected cells harboring a duplicated LAS and analyzed the surviving colonies (Fig. 2.4.1-A). Spacer acquisition events were detectable at both leader-repeat junctions, but spacers integrated into the first repeat were highly enriched after phage selection (in fact, ectopic spacer integration was observed only in 1 of 6 replicates). These data show that polarized spacer acquisition is favored over ectopic spacer acquisition even when both integration events are mediated by the same, wild-type, LAS. Although it is still possible that upstream leader sequences not included in the 30 bp LAS spacer have a minimal positive effect on the efficiency of spacer acquisition, this result suggests that the selective advantage of the wild-type CRIPSR-Cas system is not due to a low rate of spacer incorporation in LAS mutant cells, but rather reflects a difference in the level of immunity provided from different positions within the CRISPR array.

**Figure 2.4.1. Leader-End Spacers Provide More Robust Immunity than Spacers in the Middle of the CRISPR Array**

(**A**) Analysis of the site of spacer integration in a strain harboring two copies of the LAS, upstream of the first and fifth repeat. A culture was infected with ΦNM4γ4 at an MOI of 1. The CRISPR array from surviving cells was amplified and subjected to Sanger sequencing to determine the position of integration (marked by the black arrows) of new spacers. Mean + SEM of six replicates are reported.

(**B**) Two strains were engineered to test the levels of CRISPR-Cas immunity provided by the same spacer sequence located in the first (Position 1 strain) or fifth (Position 5 strain) position. In addition, the *cas1* gene was mutated to prevent the acquisition of new spacers.

(**C**) Growth of Position 1 and Position 5 cultures infected with phage ΦNM4γ4 at MOI 5 or 50 followed by the measurement of optical density at 600 nm ($OD_{600}$). Mean of three replicates is reported.

(**D**) Simulation of CRISPR immunization with Position 1 and Position 5 strains. Position 1 or Position 5 cells were diluted with cells lacking CRISPR-Cas in a 1:10,000 ratio and infected with phage ΦNM4γ4 at an MOI of 1. Cell growth that results from CRISPR-Cas immunity was monitored by optical density measurements at 600 nm ($OD_{600}$) for 24 hr. Cells lacking a CRISPR-Cas system were used as control. Mean ± SEM (gray dotted line) of three replicates is reported.

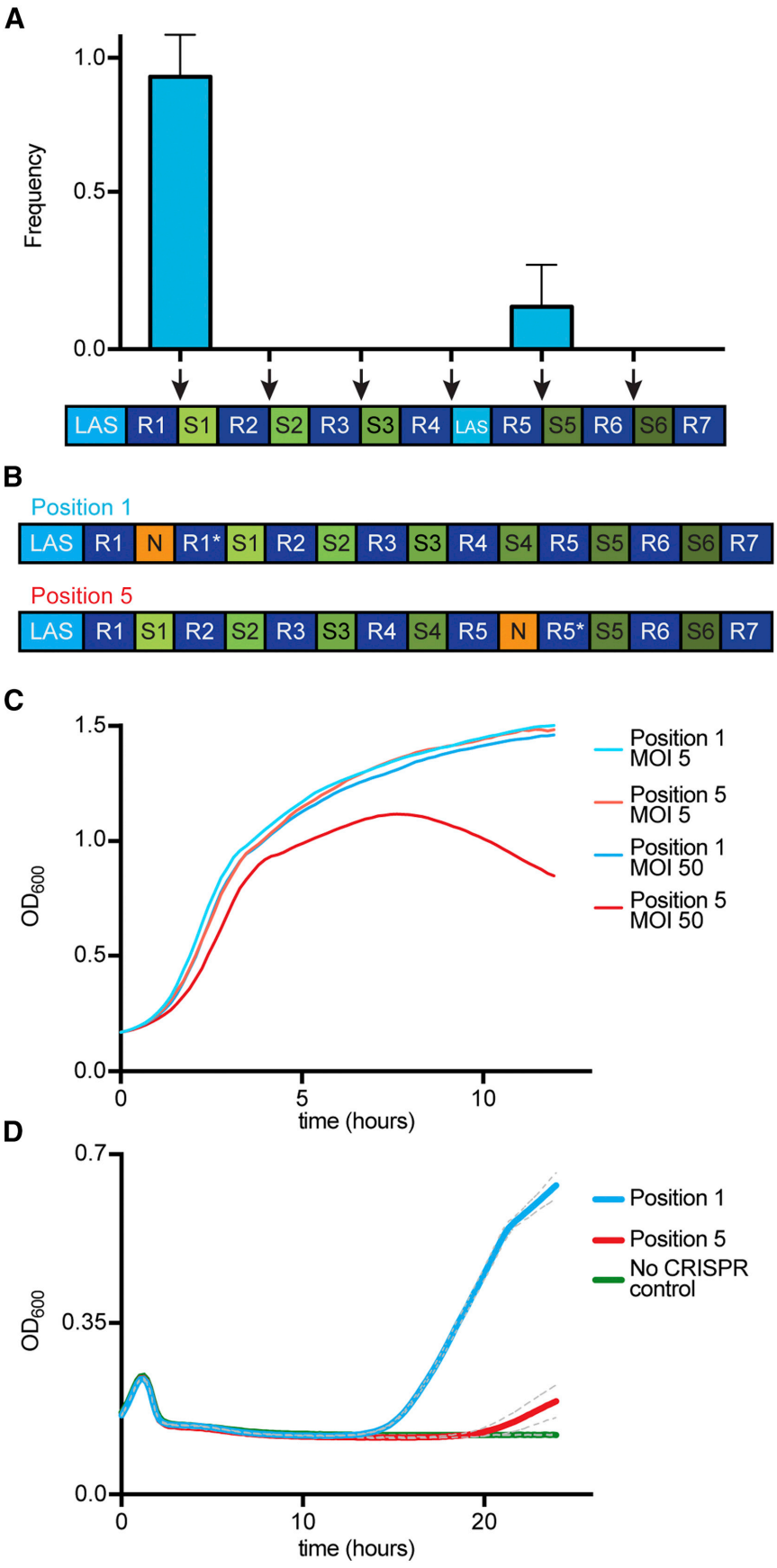
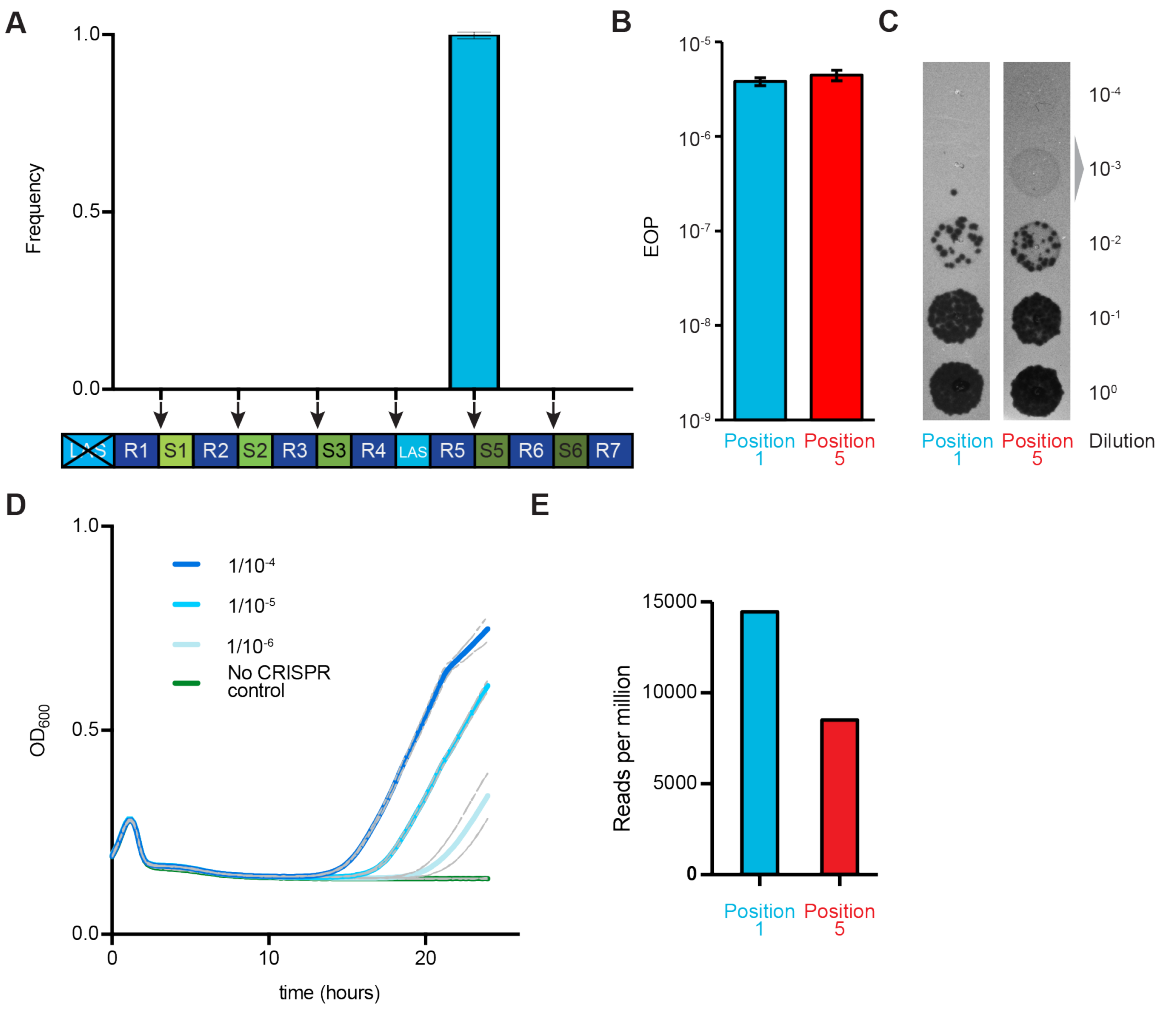

A
1.0
0.5
0.0
Frequency
LAS R1 S1 R2 S2 R3 S3 R4 LAS R5 S5 R6 S6 R7
B
Position 1
LAS R1 N R1* S1 R2 S2 R3 S3 R4 S4 R5 S5 R6 S6 R7
Position 5
LAS R1 S1 R2 S2 R3 S3 R4 S4 R5 N R5* S5 R6 S6 R7
C
1.5
1.0
0.5
0.0
OD600
0
5
10
time (hours)
Position 1 MOI 5
Position 5 MOI 5
Position 1 MOI 50
Position 5 MOI 50
D
0.7
0.35
0.0
OD600
0
10
20
time (hours)
Position 1
Position 5
No CRISPR control

**Figure 2.4.2. Spacer sequences in position 1 provide better CRISPRCas immunity than spacers in position 5.**

(A) Spacer acquisition after phage infection of cells containing a mutant LAS upstream the first repeat and a wild-type LAS instead of spc4. Surviving cells were plated and DNA from colonies was extracted for PCR. Sanger sequence of PCR products was used to determine the position of integration (marked by the black arrows) of the new spacer. Mean ± SEM of three replicates are reported. (B) Efficiency of plaquing (EOP) of the phage ΦNM4γ4 on plates containing cells with a targeting spacer in position 1 or 5. To calculate the EOP, the number of plaques formed on each of these strains is divided the number of plaques obtained using cells without CRISPR-Cas immunity that support full viral propagation. EOP = 10-9 is the lower limit of detection of the assay. Mean ± SEM of three replicates are reported. (C) Picture of the plates used for panel (B), in which 10-fold serial dilution of the phage ΦNM4γ4 were spotted on Position 1 or Position 5 cells. The inhibition of growth zone observed in the 10-3 dilution is marked. (D) Assay that simulates CRISPR-Cas immunization. Cells already immunized (with a spacer targeting the phage ΦNM4γ4) are mixed with an excess of non-immune cells at different proportions and infected with phage. The time that it takes for each mixed culture to regain growth correlates with the amount of immunized cells in the sample. Mean ± SEM (gray dotted line) of three replicates is reported. (E) RNA-seq of Position 1 and Position 5 cells showing the read count (normalized to total reads) for the crRNA derived from the same spacer sequence integrated in each of these positions.




We examined the second possibility by directly comparing the levels of CRISPR immunity provided by polarized and ectopic spacer acquisition. We engineered two CRISPR-Cas systems containing the same spacer sequence (GTGTTCTCTTCAATCCATTCATCTATTGCT) in two different positions within the array (Fig. 2.4.1-B), one mimicking polarized spacer integration ("Position 1") and the other mimicking ectopic spacer integration ("Position 5"). To prevent additional immunization events, spacer integration was abrogated in both strains by the introduction of an inactivating mutation (E220A) in the Cas1 integrase[48,94]. Both strains were grown to exponential phase and infected with phage ΦNM4γ4

at two MOIs, 5 and 50, and growth was monitored by measuring the optical density of the cultures (Fig. 2.4.1-C). Position 1 and Position 5 strains exhibited comparable levels of immunity at an MOI of 5. In contrast, at an MOI of 50, the Position 5 strain showed a severe growth defect. A similar result was obtained when phage propagation was measured on plates containing Position 1 or Position 5 cells, seeded with 10-fold serial dilutions of the phage ΦNM4γ4 stock. Whereas the number of plaques originating from phage escapers (usually harboring mutations in the target sequence that makes them refractory to CRISPR-Cas immunity [53]) were similar for both strains (Fig. 2.4.2-B), Position 5 plates showed an inhibition of growth zone (most notable at the $10^{-3}$ dilution, Fig. 2.4.2-C), suggestive of some level of phage propagation due to a poor CRISPR-Cas defense. Altogether these experiments demonstrate that while spacers in any location of the array confer some level of immunity to the host, positioning the immunity-conferring spacer at the leader-end of the array enables a more robust immunity at higher titers of phage.

CRISPR-Cas immunization is a rare event, calculated to happen only in 1 in $10^7$ cells of the infected population in our experimental set up[48]. As a consequence of this, most cells in the culture succumb to viral infection, creating very high titers of phage. We estimated that the small fraction of cells that is able to acquire a new spacer against the phage face an extremely high MOI (on the order of 10,000). Given the results obtained in Figure 2.3.1-A, we speculated that the position of the targeting spacer in the CRISPR array could be critical for the CRISPR immune response under these extreme phage stresses. To test this we

developed an assay that simulates the CRISPR immunization process in which a small proportion of CRISPR-immune cells (already harboring a phage-targeting spacer sequence in either position 1 or position 5) is mixed with a majority of non-CRISPR cells that enables exponential phage propagation. In this assay, the time at which the culture resumes growth after viral infection is delayed proportionately to the fraction of CRISPR-immune cells (Fig. 2.4.2-C). When the Position 1 CRISPR-immune strain is tested, growth resumes at ~15 hours (Fig. 2.4.1-D). In stark contrast, when the Position 5 CRISPR-immune strain is used growth resumes at ~20 hours. This pattern mirrors the growth curves produced during CRISPR immunization of wild-type and LAS mutant cultures (Fig. 2.3.1-B). Together, these results demonstrate that cells in which spacers are integrated ectopically suffer a severe growth defect due to compromised immunity.

## 2.3 DISCUSSION

A hallmark feature of CRISPR-Cas systems is the integration of short viral spacer sequences into the 5'-end of the CRISPR locus[2]. However, the physiological significance of this highly polarized process has remained unknown. Here we studied this problem in the type II-A CRISPR-Cas system of *S. pyogenes*. We found that a short sequence immediately upstream of the first CRISPR repeat, which we called the leader-anchoring sequence or LAS, is required for the exclusive insertion of spacer sequences in the first position of the CRISPR array. Mutations in the LAS result in the integration of new spacers in the middle of the array, a phenomenon we called ectopic spacer integration. The phenotype of the LAS mutant allowed us to determine the importance of ordered

spacer addition during the CRISPR immune response against phage infection. We found that polarized spacer integration bestows the host with a competitive advantage by positioning the new spacer in the first position of the array, where spacers provide more robust CRISPR-Cas immunity. Since the first spacer derives from the most recent invader, polarized spacer acquisition allows CRISPR-Cas systems to prioritize immunity against the most immediate threat to the host.

Given that Cas9 acts as a single-turnover enzyme[27], each crRNA molecule that directs the cleavage of one invading phage cannot be re-used to cleave a second viral genome. Therefore, the abundance of a targeting crRNA could be critical during CRISPR immunization, when cells are challenged by an extremely high number of phages. This is because upon infection of a naïve bacterial population, the viral titers rise to extraordinary levels due to the initially unconstrained transmission of the virus. In this exceptional condition in which a few newly-immunized cells are infected by thousands of phages at the same time, the abundance of the crRNA guide produced from the new spacer could be decisive for the success of the CRISPR-Cas immune response. Higher abundance of leader-end crRNAs has been observed in many CRISPR-Cas systems[21-25]. Importantly for our study, the *S. pyogenes* type II-A CRISPR-Cas system produces higher levels of *spc1* crRNA than the other crRNAs derived from downstream spacers[21]. Further supporting this scenario, we found that the levels of the ΦNM4γ4-targeting crRNA produced from Position 1 is ~2-fold higher than when it originates from Position 5 during exponential growth (Fig. 2.4.2-E).

We believe that it is conceivable that the same spacer sequence integrated in the first or a more downstream position of the CRISPR array could produce different levels of mature crRNA due to asymmetric transcription and/or differential processing of the crRNA precursor. The molecular mechanisms that lead to the uneven distribution of crRNAs, and how small differences in crRNA abundance affect the CRISPR immune response, will require further investigation, in type II and other CRISPR types.

While the LAS is critical for the integration of new spacers into the first repeat, we have found that the *spc4* sequence can also specify the addition of spacers into the repeat that follows it. In addition, the mutations inserted in the LAS (-1 to -5) do not completely abrogate the integration of spacers in the first position of the CRISPR array (Fig. 2.2.1-C). These results suggest that other sequences or sequence motifs can perform the LAS function. Moreover, it is possible that upstream sequences within the leader could contribute to spacer integration, though such sequences were not detectable in a related Type II-A CRISPR system (Wei et al.). We propose that the Cas1-Cas2 integrase complex samples the nucleotides immediately upstream from the repeat and that the wild-type LAS provides the optimal sequence for anchoring the complex, biasing its activity toward the leader-end of the array. In the type I CRISPR-Cas system, binding of IHF to the leader creates the required DNA topology for spacer integration at the 5'-end of the array. Given the absence of IHF homologs in Gram-positive bacteria (including the host used in our studies, *S. aureus*) it is possible that there are other factors that perform a similar function for type II

CRISPR-Cas systems[95]. Alternatively, the type II Cas1-Cas2 complex could be sufficient to catalyze polarized spacer integration without a requirement for additional host factors[31]. Additional work employing biochemical and structural techniques will address these questions.

The prioritization of the CRISPR-Cas immune response against the most recent invader has been proposed as a bet-hedging strategy[20]. This is analogous to the mammalian adaptive immune response, where effector T-cell populations and antibody titers against a virus are highest immediately post-immunization and gradually decrease over time[96]. By devoting more resources to defending the host against the most recent infection, immune systems provide robust protection against the infectious agents that are most likely to be present at high titers. This is an efficient way to conserve resources while still providing robust immunity against the most threatening attackers. For CRISPR-Cas systems, the benefit of deprioritizing immunity against past invaders would be two-fold. First, during infection of a population that has been immunized in the past, i.e. harboring the invader-matching spacer in the middle of the array, transmission of the re-infecting virus will be immediately contained and viral titers will remain low. In this situation the CRISPR-Cas immune response does not need to be at peak levels to efficiently protect the population, as shown in Figure 4C in which a spacer in position 5 provides full immunity in conditions of low MOI. Second, it has been reported that phage mutations in target sequences are highly abundant in viral populations and are selected for their ability to enable the escape from CRISPR-Cas immunity[53,97]. Therefore there is a high probability that spacers in the middle

of the array would not be able to provide immunity against re-infecting, mutated phages. In this scenario, maintaining full expression of these spacers would also be wasteful. Interestingly, in a scenario of re-infection by high titers of a non-mutated virus, an old spacer sequence can regain full potency through its repositioning as the first spacer sequence (similar to our results of Figure 2.3.2). Recombination and deletion within the CRISPR array have been extensively described in natural populations[97,98] and could represent a functional, rather than accidental, feature of CRISPR-Cas loci. Our results ascribe a physiological role to the establishment and preservation of the timeline of infection that is a hallmark of CRISPR-Cas immune systems and further our understanding of the selective pressures that guide the evolution of CRISPR systems.

# CHAPTER 3: CRISPR SYSTEMS MODULATE LIFESPAN OF IMMUNOLOGICAL MEMORIES BY TUNING RATES OF SPACER ACQUISITION

## 3.1 INTRODUCTION

Bacteria are under constant evolutionary pressure from phage threats. CRISPR-Cas systems confer adaptive and heritable immunity to bacteria by incorporating short segments of viral DNA, termed spacers, into their genome. Spacers are intercalated between semi-palindromic, direct repeats in host loci known as CRISPR arrays. Spacers, once transcribed into CRISPR RNAs (crRNAs), serve as immunological memories that guide effector nucleases to seek and destroy foreign cognate nucleic acids, thereby neutralizing viral threats during the interference stage of CRISPR immunity. CRISPR-associated (*cas*) genes encode the protein components of the immune system and are the basis by which CRISPR systems are categorized. To date, there are six types and over twenty subtypes of CRISPR systems[99,100].

*cas1* and *cas2* are the most conserved *cas* genes and can be found in all six types[100]. Together, they form the Cas1-Cas2 integrase, a heterohexameric complex that mediates the spacer integration process. New spacers are added unidirectionally to the CRISPR array and as a result, the array represents a chronological record of past infections. The most recently acquired spacers are found at the 5'-end of the array and the most ancestral ones are downstream at the 3'-end of the array. The leader is a stretch of DNA that lies immediately upstream of the CRISPR array and plays an important role in enforcing polarized spacer integration. In Type II CRISPR systems, the Cas1-Cas2 complex makes

sequence-specific contacts with the last five nucleotides of the leader, known as the leader-anchoring sequence (LAS)[5,19,100]. This is thought to stabilize the integrase complex and enable faster reaction kinetics for spacer integration at the leader-end of the array[5,13,14,19,100]. Spacers at the leader-end of the CRISPR array give rise to a more robust immune response relative to downstream spacers. Thus, the chronological storage of spacers allows CRISPR systems to prioritize resources against the most recent invader[19]. However, as a consequence, addition of new spacers shifts existing spacers further downstream within the array, thereby reducing the potency of older spacers. Furthermore, it has been shown that an increased number of spacers within a CRISPR array weakens the interference provided by individual spacers, likely by a dilution effect[101]. How CRISPR systems balance the benefits of acquiring new spacers with the cost of weakening existing spacers has not been explored.

We hypothesized that natural variation in spacer acquisition rates could exist and be a means for CRISPR systems to modulate the lifespan of existing spacers. Indeed, our lab previously demonstrated that the rate of spacer acquisition could be boosted by artificially introducing mutations in Cas9, the hallmark protein of Type II-A CRISPR systems[50]. Here, we compare the immune responses generated by a variety of Type II-A CRISPR systems and show significant variation in immunization rates across these systems. CRISPR systems that acquire new spacers at a high frequency can rapidly respond to new threats, but existing spacers quickly lose their potency. In contrast, CRISPR systems that exhibit low-frequency spacer acquisition prioritize the immunity

conferred by existing spacers at the cost of a slower uptake of new spacers. We found that bacteria have overcome these tradeoffs by harboring multiple CRISPR systems that acquire spacers at different rates. Furthermore, we have demonstrated that these related CRISPR systems can acquire new spacers independently of each other as a result of specific interactions between the acquisition machinery and the leader-repeat junction. We propose a model in which bacteria can utilize multiple CRISPR systems as a means to form both short- and long-term immunological memories to defend against a diversity of threats.

## 3.2 RESULTS

*Type II-A CRISPR systems exhibit natural variation in their rates of spacer acquisition*

Natural variation is a key component of evolution and allows for selective pressures to refine the functionality of biological phenomena[102]. Given that previous work from our lab demonstrated that artificially introduced mutations in Cas9 could alter the rate of spacer acquisition[50], we sought to determine if CRISPR systems naturally exhibit variation in their rates of spacer acquisition. We reasoned that analysis of closely related CRISPR systems would help to minimize confounding variables that could be introduced when comparing more evolutionary distant systems. We started by analyzing the spacer content of a variety of Type II-A CRISPR systems (N=80)[103]. Interestingly, we observed a wide range in the number of spacers in CRISPR arrays (min: 1, max: 69, mean: 21.5; Figure 3.1.1-A). This range in the number of spacers suggested that the

frequency of spacer addition could vary significantly across different Type II-A CRISPR systems.

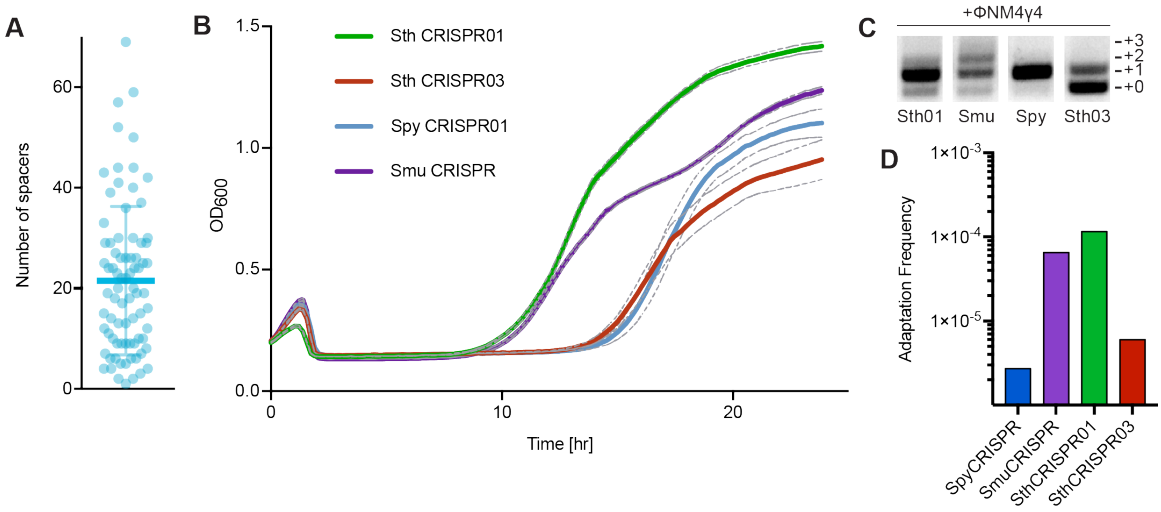



**Figure 3.1.1. Type II-A CRISPR systems exhibit natural variation in their rates of spacer acquisition**.
(**A**) Analysis of the number of spacers in Type II-A CRISPR arrays from previously sequenced genomes. (**B**) PCR analysis of CRISPR arrays at 24 hours post infection (MOI 1 pfu/cfu). (**C**) Growth curves of cell cultures harboring different CRISPR systems and infected with phage at MOI 1 pfu/cfu. (**D**) Analysis of bacteriophage-insensitive mutants at 24 hours post infection in a soft-agar phage infection assay.

To test this experimentally, we chose four Type II-A CRISPR systems originating from streptococcal hosts: *Streptococcus pyogenes* SF370 (SpyCRISPR), *Streptococcus mutans* NN2025 (SmuCRISPR), and *Streptococcus thermophilus* LMD-9 CRISPR01 (SthCRISPR01) and CRISPR03 (SthCRISPR03). We individually cloned and tested each CRISPR system in *Staphylococcus aureus* RN4220 as done previously[19,41,48,50]. This experimental system allowed us to compare the immune responses of CRISPR systems originating from different organisms in a more controlled manner (i.e. infection with the same phage, growth in identical media, same host genetic background).

In addition, we truncated the CRISPR arrays so that each only contained a single spacer in order to minimize spacer-specific effects. We monitored the growth of each CRISPR-containing strain after infection with the lytic virus ΦNM4γ4 [48] at a multiplicity of infection (MOI) of 1 pfu/cfu (Figure 3.1.1-B). We observed significant variability in the time elapsed before the cells recovered from the phage infection suggesting variability in their immune responses. Using PCR-based analysis of the CRISPR arrays 24 hours post infection, we were able to detect spacer acquisition in all four systems with qualitative differences consistent with the variability we observed in the growth curves (Figure 3.1.1-C). SmuCRISPR and SthCRISPR01 exhibited multiple bands corresponding to up to 2 or 3 newly added spacers, while SpyCRISPR and SthCRISPR03 only had a prominent band corresponding to the addition of a single new spacer. We then tested these CRISPR systems using a soft agar spacer acquisition assay, in which each colony originates from a single BIM (bacteriophage-insensitive mutant). Because all liquid cultures reached peak $OD_{600}$ values by approximately 22 hours post infection, we quantified phage-resistant colonies at 24 hours post infection. Again, we observed variability (>1 log) in the rates of spacer acquisition (Figure 3.1.1-D) and these results were consistent with our liquid phage infection assays. To test the generalizability of these results, we performed the same assay for each Type II-A CRISPR system using another lytic staphylococcal phage, ΦNM1γ6[58]. We observed a similar pattern in both liquid and soft agar phage infections (Figure 3.1.2-A-C). Notably, while no correlation was observed in the soft agar infection with ΦNM4γ4 ($R^2$ = 0.09604; Figure 3.1.2-D), the colony

count for the soft agar experiment with ΦNM1γ6 had a strong correlation with the number of spacers naturally found in the array of each system ($R^2$=0.9826; Figure 3.1.2-E). Based on these results, we conclude that Type II-A CRISPR systems exhibit natural variation in their rates of spacer acquisition and that this results in differentially expanded CRISPR arrays in nature.

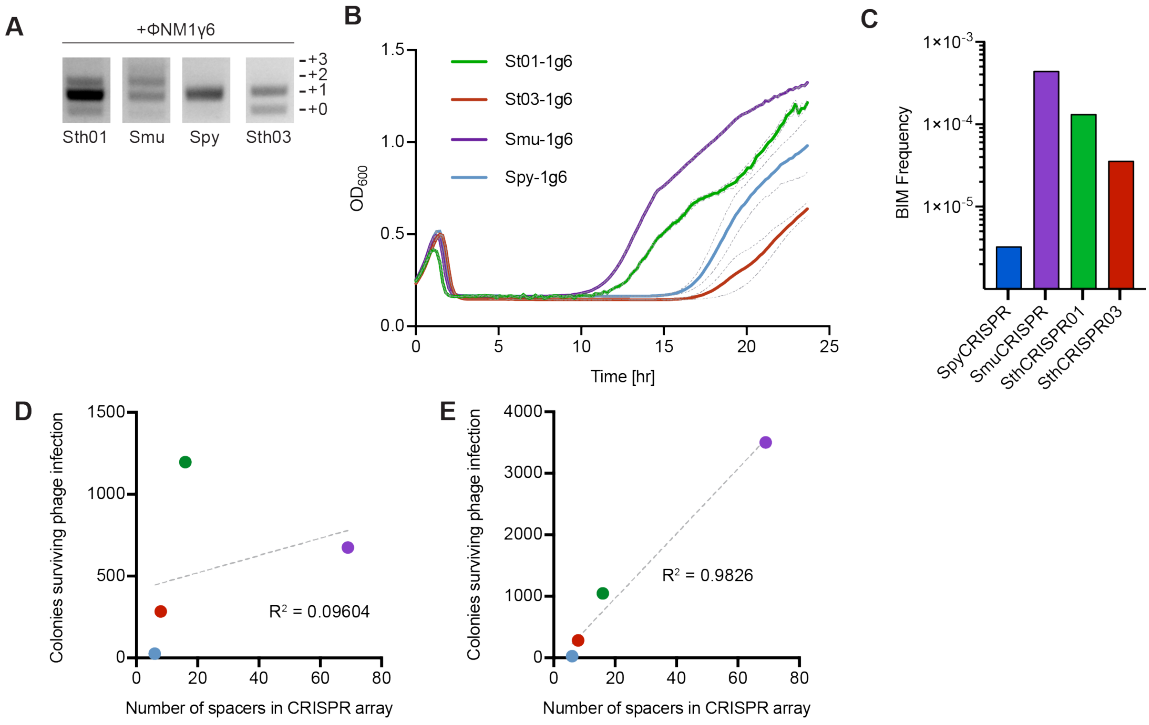



**Figure 3.1.2. Natural variation in the rates of spacer acquisition against phage ΦNM1γ6.**

(**A**) PCR analysis of CRISPR arrays at 24 hours post infection (MOI 1 pfu/cfu). (**B**) Growth curves of cell cultures harboring different CRISPR systems and infected with phage at MOI 1 pfu/cfu. (**C**) Analysis of bacteriophage-insensitive mutants at 24 hours post infection in a soft-agar phage infection assay. (**D**&**E**) Comparison of the number of spacers found in WT array with soft-agar BIM formation rate for (D) ΦNM4γ4 and (E) ΦNM1γ6.

*S. thermophilus LMD-9 utilizes two Type II-A CRISPR systems to form short- and long-term immunological memories*

Because adding new spacers and shifting existing spacers further downstream within an array weaken their level of immunity conferred[19,101], we reasoned that CRISPR systems could regulate the potency and lifespan of existing spacers by evolutionarily tuning their rate of spacer acquisition. High frequency spacer acquisition is beneficial for fast responses to new challenges, whereas low frequency spacer acquisition preserves potency of existing spacers and retains immunity against recurring threats. Therefore, both high- and low-frequency systems have costs and benefits, but harboring both varieties would allow bacteria to overcome these tradeoffs. We hypothesized that this could be the case for *Streptococcus thermophilus* LMD-9, which harbors two Type II-A CRISPR systems that exhibit high- and low-frequency spacer acquisition[2,35].

We developed an experimental system to study the interplay between the two Type II-A CRISPR systems of *S. thermophilus* LMD-9 by introducing them into *S. aureus* RN4220. Upon infection of this double CRISPR-containing strain with ΦNM4γ4, we observed regrowth of the population approximately 11 hours post infection (Figure 3.2.1-A) and detected addition of spacers to both arrays, with more spacer addition detectable in the CRISPR01 locus (Figure 3.2.1-B). This was consistent with the results from a soft agar infection assay when we challenged the strain with ΦNM4γ4 and analyzed both CRISPR arrays of each colony at 24 hours post infection. We found that 92.49% of the surviving colonies had acquired at least one spacer in one of the CRISPR arrays and 87.45% of

new spacers were added to CRISPR01 (Figure 3.2.1-C). These results are consistent with what has been observed in its native host, suggesting that our experimental system is a faithful recapitulation.

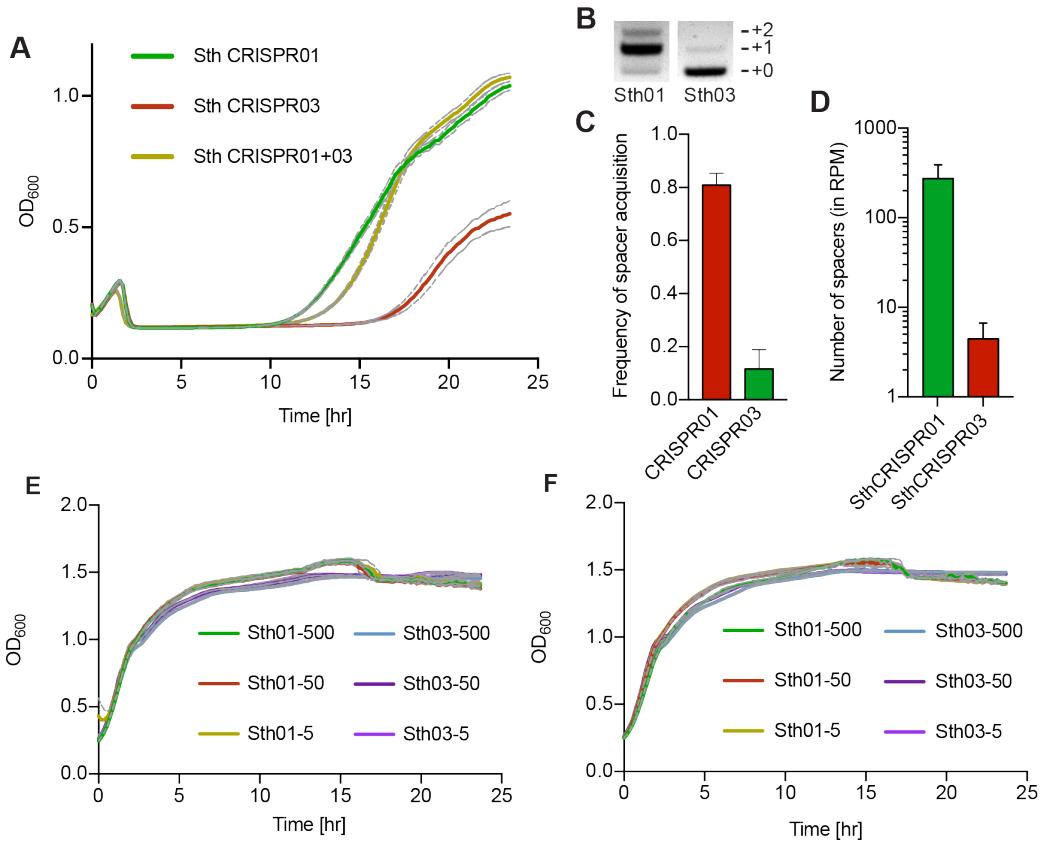



**Figure 3.2.1. The two Type II-A CRISPR systems of *S. thermophilus* acquire spacers at different rates when expressed in the same cell.**

(**A**) Growth curves of cell cultures harboring different configurations of CRISPR systems from *S. thermophilus* LMD-9 and infected with phage at MOI 1 pfu/cfu. (**B**) PCR analysis of CRISPR arrays at 24 hours post infection (MOI 1 pfu/cfu). (**C**) Analysis of spacer acquisition in cells harboring the two Type II-A CRISPR systems of *S. thermophilus* at 24 hours post infection in a soft-agar phage infection assay. (**D**) Deep sequencing analysis of spacer acquisition in the two CRISPR arrays at an early time point. (**E**&**F**) Growth curves of cells harboring two sets of overlapping spacers in the two CRISPR systems of *S. thermophilus* infected by phage at different MOIs.

To further confirm that the differences in spacer addition we observe is due to differences in spacer acquisition activity rather than interference-related effects, we amplified and isolated CRISPR loci at 30 minutes post infection using a high-sensitivity detection method[48]. This allowed us to measure spacer acquisition independent of interference and survival of the phage infection (lytic cycle of ΦNM4γ4 is ~45 minutes). Indeed, we observed greater rates of array expansion in CRISPR01 relative to CRISPR03 (more abundant species corresponding to double and triple spacer addition in CRISPR01; Figure 3.2.2-A). Upon high-throughput sequencing and analysis of phage-targeting spacers, we observed that the CRISPR01 locus acquired more spacers than CRISPR03 by greater than one order of magnitude (Figure 3.2.1-D). We next tested for differences in interference by comparing the immunity provided by two different sets of spacers for CRISPR01 and CRISPR03. Because each Cas9 has different protospacer adjacent motif (PAM) sequence requirements[34], it is impossible to test identical spacer sequences against the same phage, but we were able to design two sets of spacers that share 29/30 bp of homology with overlapping PAMs. Indeed, both sets of spacers provided similar protection at various MOIs (Figure 3.2.1-E&F). We also tested the rate of escaper phage generation against three different spacers for each CRISPR system and found no significant difference between the two systems ($p = 0.3745$; Figure 3.2.2-B). Therefore, we conclude that CRISPR01 and CRISPR03 acquire spacers at different rates as a result of differences in spacer acquisition activity, rather than differences in interference.

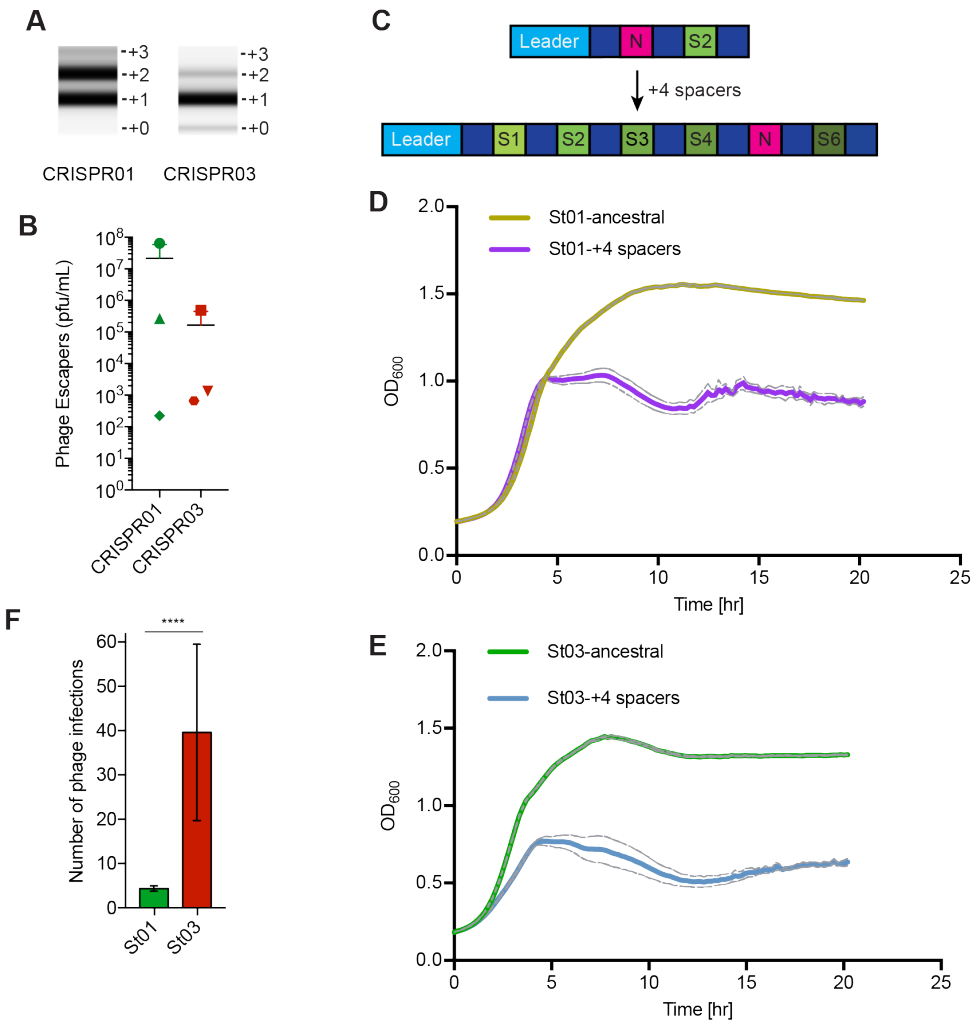



**Figure 3.2.2. Short- and long-term storage of immunological memories in the Type II-A CRISPR systems of *S. thermophilus*.**

(**A**) PCR analysis of spacer acquisition in the two CRISPR arrays from an early time point phage infection visualized via tapestation. (**B**) Analysis of phage escape rate against three different spacers in each CRISPR system. (**C**) Schematic of CRISPR arrays used to test immunity provided by the same spacer upon addition of four additional spacers with random sequences. (**D**&**E**) Growth curves of cell cultures harboring a phage-targeting spacer in either an ancestral or +4 spacers position within the (D) CRISPR01 or (E) CRISPR03 system upon infection with phage at MOI 100 or 1 pfu/cfu, respectively. (**F**) Probabilistic model predicting the number of phage infections needed to shift the current spacer 1 to position 5 within the CRISPR array for the two CRISPR systems.

We hypothesized the different rates of spacer acquisition observed for the two Type II-A CRISPR systems allows *S. thermophilus* to form both short- and long-term immunological memories. To support this model, we set out to determine whether spacers in the two Type II-A CRISPR systems of *Streptococcus thermophilus* lose potency upon addition of new spacers. For each CRISPR system, we tested the same phage-targeting spacer in two different arrays, one with the spacer as the first spacer in the array (ancestral) and one with four additional non-targeting spacers upstream of the phage-targeting spacer (+4 spacers) (Figure 3.2.2-C). As predicted, the addition of the four new spacers resulted in a weaker immune response for both SthCRISPR01 and SthCRISPR03 (Figure 3.2.2-D&E). Based on our calculated difference in spacer acquisition rates for the two CRISPR systems (Figure 3.2.1-C), we generated a probabilistic model to predict how many phage infections it would take to reach this point of diminished immunity (Figure 3.2.2-F). Based on this model, we predict that while spacers in SthCRISPR01 would become significantly weaker after approximately 4 new phage challenges, the spacers stored in SthCRISPR03 could maintain their potency for approximately 40 rounds of new phage challenges. Taken together, our data support a model in which *Streptococcus thermophilus* can form both long- and short-term immunological memories by harboring two CRISPR systems that acquire new spacers at different rates.

*Leader-repeat junction specificity underlies orthogonal spacer storage in S. thermophilus*

For the two Type II-A CRISPR systems to acquire spacers at different rates within the same cell, the two CRISPR systems have to acquire spacers independently of one another. If crosstalk were possible, then the more active spacer acquisition machinery could exhibit a dominant phenotype where both systems rapidly acquire new spacers. Given their ability to acquire spacers at different rates even when we expressed them in the same cell, we speculated that the two related CRISPR systems have evolved mechanisms that allow them to acquire new spacers independently of one another and avoid crosstalk during spacer integration.

While it has been previously demonstrated that the two systems are able to process crRNAs and interfere independently, the possibility of crosstalk during spacer acquisition has not yet been addressed. To test our hypothesis, we analyzed the spacer sequences inserted into each CRISPR locus in our early time point next-generation sequencing experiment. We were able to use PAM sequences (CRISPR01: NNAGAAW; CRISPR03: NGGNG[34]) to assign spacers inserted into each array to the spacer integration machinery of each system. We found that spacer acquisition was highly accurate for both CRISPR01 and CRISPR03 (Figure 3.3.1-A&B). For phage-targeting spacers, both systems were greater than 90% accurate (Figure 3.3.1-A; means of 275.103 vs. 0.395 spacers for CRISPR01 and 4.460 vs. 0.345 spacers for CRISPR03, mean values in RPM for n = 3). Because of the high sensitivity of the PCR-based assay we used, we

were also able to detect self-targeting spacer acquisition against the host chromosome. For spacers matching the host chromosome, we also observed high specificity for both systems (CRISPR01=96.686% and CRISPR03=99.306%; Figure 3.3.1-B). This experiment demonstrates that while crosstalk is possible between the two systems during spacer acquisition, mechanisms have evolved to minimize crosstalk and thereby increase the efficiency of the system.

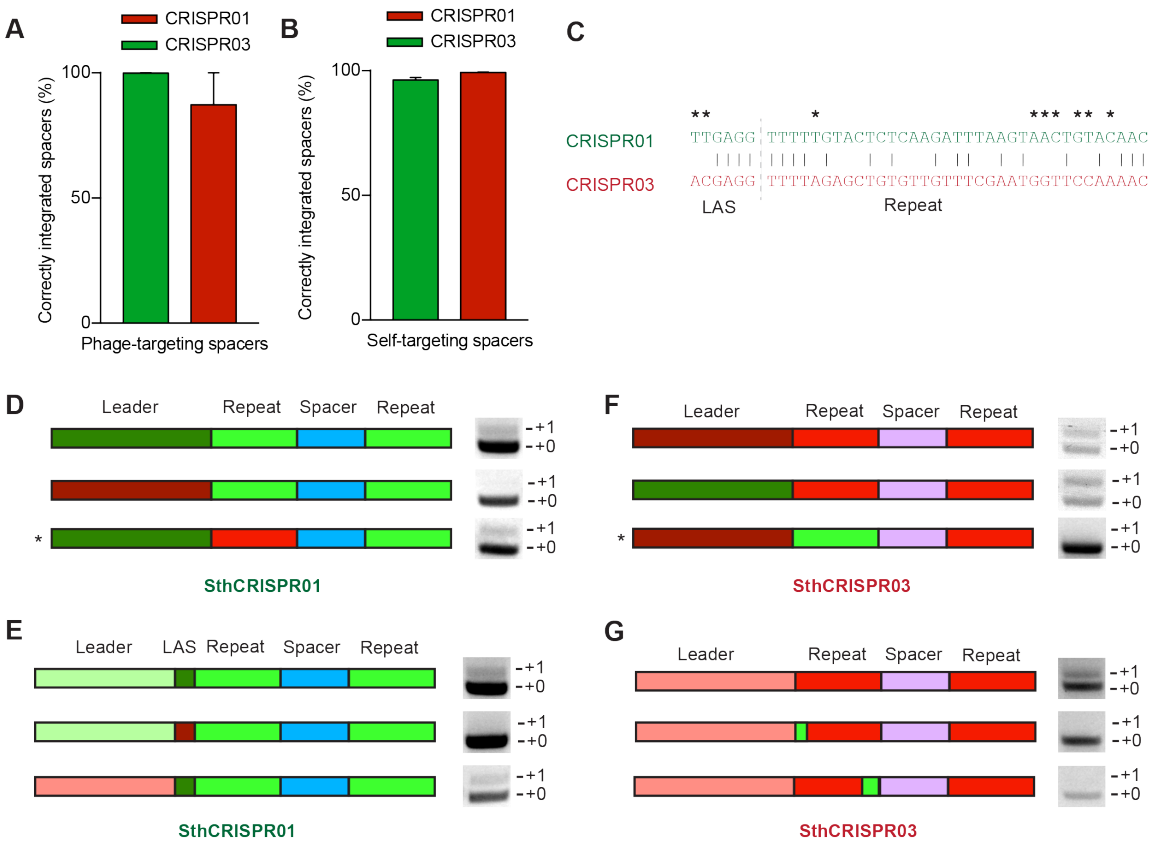



**Figure 3.3.1. Sequences at the leader-repeat junction prevent crosstalk between Type II-A CRISPR systems during spacer acquisition.**
(**A**&**B**) Deep sequencing analysis of spacer integration accuracy in the two CRISPR systems at an early time point against (A) the phage genome and (B) the host chromosome. (**C**) Comparison of the leader-repeat junctions in the two Type II-A CRISPR systems of *S. thermophilus* LMD-9. (**D**-**G**) PCR analysis of spacer acquisition in arrays with mutations in the leader-repeat junction of the (D&E) CRISPR01 and (F&G) CRISPR03 loci (* denotes mutant tracrRNA background).

We next sought to probe the mechanisms that enable orthogonal spacer acquisition in these two co-existing CRISPR systems. Type II-A CRISPR systems have been shown to not require host factors for site-specific spacer

integration but rather rely on interactions between the Cas1-Cas2 integrase, the leader anchoring sequence (LAS), and the repeat sequence[13,14,19]. We wondered whether differences in the leader-repeat junction of the two CRISPR systems were important for reducing crosstalk during spacer acquisition (Figure 3.3.1-C). To test determinants of specificity for the spacer acquisition machinery of each system, we introduced a variety of plasmids containing mutations in either the leader or repeat sequences. To test the importance of the leader, we replaced the entire leader sequence with the leader of the other system. For the repeat, we swapped bases that differed in the first six and last twelve nucleotides of the repeat sequence, as these regions were identified as important for spacer integration. We also made compensatory mutations in the tracrRNA to preserve base pairing interactions[104], which allowed for proper interference from these arrays (Figure 3.3.2). Replacement of the SthCRISPR01 leader with the sequence from the SthCRISPR03 leader significantly diminished spacer acquisition activity (Figure 3.3.1-D). We were then able to narrow down the relevant region to the LAS, as just swapping the two nucleotides that differ between the LAS's was sufficient (Figure 3.3.1-E) and reverting those two nucleotides in the SthCRISPR03 leader was also sufficient to restore spacer acquisition activity in the leader swap construct (Figure 3.3.1-E). Mutations in the repeat sequence, however, had little or no effect on spacer acquisition activity (Figure 3.3.1-D). These results demonstrate the importance of the leader-repeat junction in site-specific spacer integration for SthCRISPR01 and suggest a primary role for the LAS in aiding the integrase to discriminate between CRISPR

loci. In contrast, we found that the leader sequences were interchangeable for SthCRISPR03 (Figure 3.3.1-F). Rather, altering the repeat sequence diminished spacer integration activity for SthCRISPR03 (Figure 3.3.1-F). We tested the importance of both the 5' and 3' ends of the repeat individually, and both appeared to be sufficient for reducing spacer integration activity (Figure 3.3.1-G). Thus, we conclude that SthCRISPR01 and SthCRISPR03 have evolved distinct mechanisms by which they distinguish between CRISPR arrays during spacer integration, allowing the two coexisting CRISPR systems to efficiently defend the host.

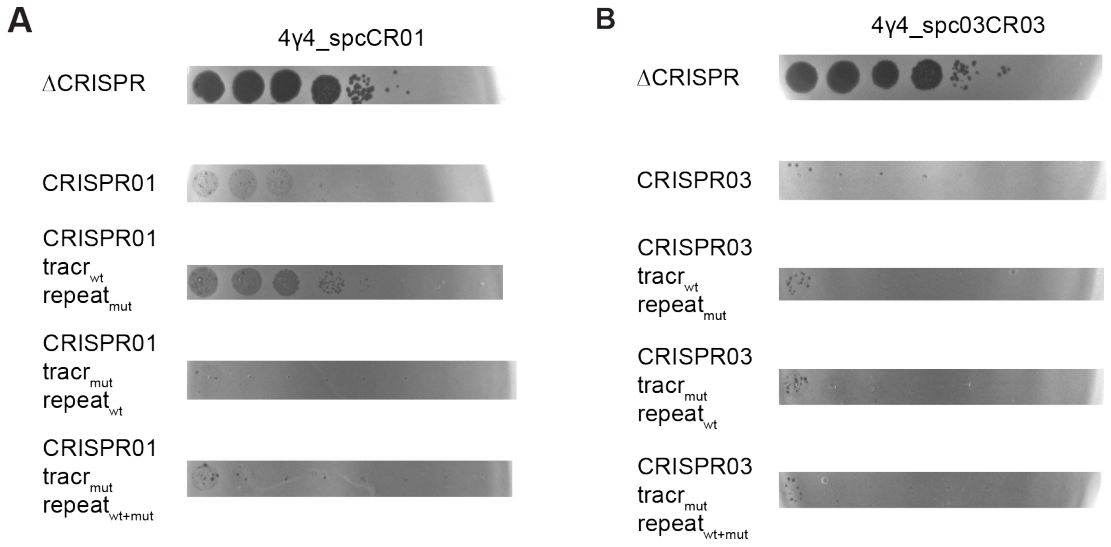



**Figure 3.3.2. Compensatory mutations in the tracrRNA enable interference with mutated repeat sequences.**

(**A**&**B**) Efficiency of plaquing analysis of interference from the native spacer 1 in (A) CRISPR01 or (B) CRISPR03 in various configurations of repeat and tracrRNA mutations.

## 3.3 DISCUSSION

By stably incorporating segments of foreign DNA into arrays, spacer acquisition enables a specific and adaptive immune response to foreign threats for both the prokaryotic host and its progeny[100]. However, addition of new spacers weakens the immunity provided by existing spacers by pushing them further downstream within the array as well as by a dilution effect from the increased number of total spacers[19,101]. We propose that for bacteria that face a diverse or rapidly mutating population of phages, evolutionary pressures will select for CRISPR systems that acquire new spacers at higher rates. In contrast, for bacteria that are faced with recurring infections by similar or related phages, it would be more beneficial to preserve the potency of existing spacers by acquiring new spacers less frequently.

Infection of a population of CRISPR-containing cells yields a heterogeneous population with a rich diversity of spacer sequences[48]. Selective pressures imposed by the phage population can then shape the distribution of spacers in the bacterial population[20]. It is common for bacteria to harbor multiple CRISPR loci (Figure 3.4.1-A). In that scenario, these phage-enforced selective pressures can also shape the distribution of spacers across the different CRISPR arrays. When the coexisting CRISPR systems belong to different types, it is reasonable to speculate that selective advantages conferred by distinct targeting mechanisms of the different effector complexes would determine which system adapts to a particular threat[62]. However, it is less clear what advantages would be conferred by having multiple CRISPR systems of the same type and even

subtype. Here, we have demonstrated that two Type II-A CRISPR systems of *Streptococcus thermophilus* can independently add new spacers at different rates to form both short- and long-term immunological memories. We postulate that selective pressures could determine which CRISPR system is used by the bacterial population to adapt to new phage threats. For spacers that target a broad range of phages or a recurring threat, spacers stored in the low-frequency system will have an advantage and will thus be selected for on a population level. In contrast, for spacers that respond to a new threat that is not persistent, the high-frequency system will be able to rapidly neutralize the threat but potency of this spacer will not need to be preserved. Each CRISPR system could also benefit the cell differently depending on the ecological niche currently occupied by the cell. In niches with high dispersal of phages, a fast-responding CRISPR system would be more beneficial due to the diversity of phage likely to be encountered. In contrast, in ecological niches with low phage dispersal, a low-frequency system would be more beneficial because phage infections are more likely to be recurring and maintaining the potency of acquired spacers would be more beneficial for the population. Our results reveal that intrinsic differences in spacer acquisition endows each system with a unique set of benefits and enable them to work together to form a robust anti-phage defense system.

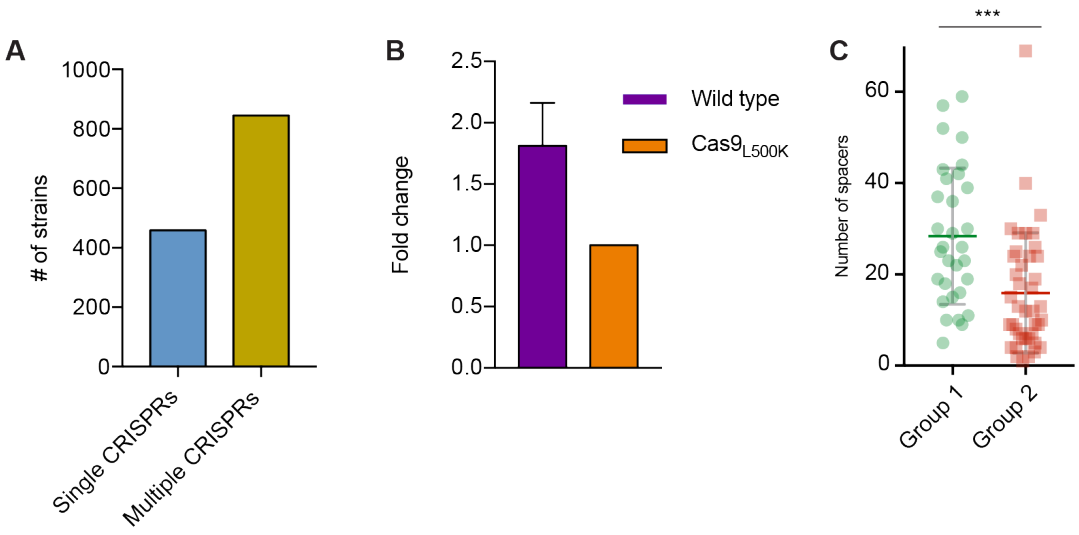



**Figure 3.4.1. Natural variation in rates of spacer acquisition may be genetically encoded.**

(**A**) Analysis of the number of CRISPR loci harbored in various prokaryotic genomes. (**B**) Comparison of BIM formation in the CRISPR system of *Streptococcus mutans* NN2025 harboring either WT or mutated Cas9. (**C**) Comparison of the number of spacers in native CRISPR arrays from Group 1 and Group 2 Type II-A CRISPR systems.

We have also demonstrated that the two Type II-A CRISPR systems of *S. thermophilus* have evolved genetic isolation from one another. We have shown that specific interactions between the Cas1-Cas2 integrases and the leader-repeat junctions enable orthogonal storage of spacers. This allows for the two Type II-A systems to maintain different rates of spacer uptake and give their existing spacers different longevities. CRISPR systems have been categorized by their effector *cas* genes that deal predominantly with the interference phase of CRISPR immunity. However, our results reveal that functional differences in spacer acquisition can exist among similarly classified CRISPR systems.

Previously, our lab demonstrated that by introducing mutations in the cas9 gene of the Type II-A CRISPR system from *Streptococcus pyogenes* it was possible to increase the rate of spacer acquisition[50]. Here, we report that natural variation in the rates of spacer acquisition also exists among Type II-A CRISPR systems. While it remains technically challenging to experimentally address the differences in rates of spacer acquisition we observe, we have some hints at mechanisms that underlie this variation. For instance, we noticed a leucine residue in the 500$^{th}$ amino acid of the Cas9 from the CRISPR system from *Streptococcus mutans* that exhibits a high frequency of spacer acquisition. This was similar to a mutation ($Cas9_{K500I}$) that we previously reported as offering a small increase in the rate of spacer acquisition[50]. We introduced the reciprocal mutation into the *Streptococcus mutans* system ($Cas9_{L500K}$) and noticed a modest decrease in the rate of spacer acquisition (Figure 3.4.1-B), suggesting that this is a naturally occurring hyper-Cas9 mutation. For the *Streptococcus thermophilus* CRISPR systems, we noticed that they belong to two separate clades of Type II-A CRISPR systems (SthCRISPR01 and SthCRISPR03 belong to Group 1 and Group 2, respectively). When we divide our spacer content data (Figure 3.1.1-A) by Type II-A Groups, we find that the mean number of spacers for Group 1 Type II-A systems is double that of Group 2 ($p = 0.0004$; Figure 3.4.1-C). This suggests that these two clades of CRISPR systems have diverged for rates of spacer acquisition as well as interactions between Cas1 and the leader-repeat junction. Notably, SthCRISPR03 and SpyCRISPR share the most sequence homology among this collection of Type II-A CRISPR systems and

exhibit comparable rates of spacer acquisition. Together, these observations suggest a genetic basis for the differences we observed in spacer acquisition rates. As our understanding of the molecular details of spacer acquisition increases, it will be interesting for future studies to more directly address the molecular basis for the variation we observe in rates of spacer acquisition. Similarly, comparative analysis of spacer acquisition in related but functionally distinct CRISPR systems could yield a better understanding of the molecular basis of CRISPR immunity.

## CHAPTER 4: PERSPECTIVES

CRISPR-Cas systems comprise a genetic immune system to safeguard bacteria and archaea from invading threats by storing fragments of foreign DNA in CRISPR arrays within the host genome. Once transcribed, these foreign-derived sequences, called spacers, provide sequence-specificity for CRISPR-associated (Cas) nucleases to destroy viruses, plasmids, and other potentially lethal threats. By stably storing spacers in the genome, CRISPR immunity is both adaptive and heritable[99,100].

In my thesis work, I have explored how CRISPR systems store and organize their molecular memories of past invaders and the functional consequences of this process. By studying the Type II-A CRISPR locus from *Streptococcus pyogenes*, I identified a crucial regulatory region of the CRISPR locus that determines the site of spacer integration within the CRISPR array, which I termed the leader-anchoring sequence (LAS). The LAS enforces polarized addition of spacers to the array such that the chronological order of spacer acquisition is preserved. Mutation of the LAS leads to erroneous integration of new spacers into the middle of the array. Further, I demonstrated that spacers positioned near the LAS provide more robust immunity than spacers positioned further downstream. Therefore, CRISPR systems can utilize the chronological organization of spacers within the array to prioritize their immune resources against the most recent invader[19].

Two complementary studies corroborated my findings through biochemical and structural approaches[13,14]. Wright et al. (2016) demonstrated the importance

of the LAS during spacer integration in an in vitro spacer integration assay[14]. This study revealed that improved reaction kinetics on the leader-end of the array, presumably from additional contacts made between the spacer integration machinery and the LAS, resulted in spacer integration preferentially occurring at the leader-end of the array. This was subsequently confirmed by Xiao et al. (2017) in their structure of the Cas1-Cas2 complex from a Type II-A system in complex with the integration target substrate[13]. This structure revealed that an α-helix from Cas1 inserts itself into the minor groove of the leader DNA to make direct, sequence-specific contacts with the LAS. Taken together, we now have a mechanistic understanding of how spacer acquisition is polarized in Type II CRISPR immune systems.

Less clear, however, is the precise mechanism by which position within a CRISPR array affects the level of immunity that it provides. In Chapter 2, I demonstrated that altering the position of the same spacer sequence within the array modulated the abundance of the corresponding crRNA by two-fold[19]. While this difference is modest, it is possible that other aspects of the CRISPR immune response could amplify this difference (i.e. that Cas9 is a single-turnover enzyme[27]). Further experimentation will be needed to fully tease apart this mechanism. For example, it would be informative to look at differences in crRNA loading into Cas9 that arise from positioning of the spacer within the CRISPR array. In addition, little is known about crRNA transcription, processing, and loading into Cas9 for Type II-A CRISPR systems, so it remains possible that other aspects of these processes are affected by spacer position within the array.

In particular, it could be valuable to determine the kinetics of crRNA biogenesis after the spacer has been newly integrated into different positions within the array. Of course, in vitro reconstitution would provide crucial insights into this process. The uneven distribution of crRNA expression has been observed in many CRISPR systems and appears to be a general trend for CRISPR immune systems[19], so it could be informative to look at the generality (or variation) in mechanisms that underlie this feature of CRISPR immunity. Notably, a strain harboring a natural mutation in its LAS has recently been identified and it was confirmed to result in ectopic spacer integration[105]. Why (and if) a mutation in the LAS would be selected for is unclear, but it could imaginably serve as a way to bypass strict chronological storage of spacers and give rise to a more complex organizational system.

While spacer acquisition provides the immediate and significant benefit of immunity, it can also impart a cost on the system by weakening the immunity provided by existing spacers by shifting them further downstream within the array. Also, it has been shown that increasing the absolute number of spacers in an array has a detrimental effect on the immunity provided by a single spacer, presumably by a dilution effect[101]. To explore how bacteria deal with this conundrum, I examined differences in spacer acquisition rates across a variety of related CRISPR systems. I found striking differences in the rates of spacer acquisition across Type II-A CRISPR systems, including two systems that coexist in *Streptococcus thermophilus*. By harboring CRISPR systems that acquire new spacers at different rates, the bacterial population can store spacers for short-

and long-term utility. Thus, immunological memories are organized not only within CRISPR arrays but also across CRISPR arrays to ensure optimal immunity.

On the flip side of spacer acquisition, spacers can also be lost from the array[28,29], though this process is poorly understood. The rates of spacer loss from CRISPR arrays, as well as the mechanistic basis of spacer loss, have not been properly studied. However, my work has suggested that spacer loss can be functional and selected for. When phage-targeting spacers are downstream in the array, deletion of upstream spacers can boost the immunity provided by the spacer by repositioning it within the array[19]. Future studies investigating deletion and recombination within and across arrays will provide valuable insights into CRISPR evolution and biological memory.

Many bacteria harbor multiple CRISPR systems, both of the same type and different types. However, many questions remain regarding the functional consequences and benefits of these configurations. Further, what advantages and disadvantages correspond to the different types of CRISPR systems has not been studied. Exploration of the benefits and costs associated with different CRISPR systems can shed light on why bacteria harbor multiple types of CRISPR systems as well as reasons underlying the diversification of CRISPR systems.

That around half of sequenced bacterial genomes do not contain CRISPR systems hints at there being deleterious consequences associated with having a CRISPR system, but this has not been fully addressed[100]. The fact that some

CRISPR systems, i.e. Type III loci, can frequently be found without spacer acquisition-related genes suggests that the adaptive nature of CRISPR systems can itself be detrimental to the cell[62,89]. This could possibly be due to costs associated with autoimmunity resulting from acquisition of spacers against the host genome[50,106]. Another potential cost of having a CRISPR system could be limiting beneficial horizontal gene transfer, as we know that a number of sequenced spacers map to plasmids and other mobile genetic elements[3,76].

CRISPR-Cas systems, and spacer acquisition in particular, represent a unique form of bacterial evolution, in that they provide a quasi-Lamarckian mechanism to adapt against genetic threats. What impacts CRISPR immunity and spacer acquisition have on the evolution of bacterial genomes remains unaddressed. Perhaps most interestingly is that a spacer acquisition event is a clear example of a hard selective sweep in a bacterial population. While the initial population recovering from a new phage infection is heterogeneous (with many spacer sequences in the population[48]), eventually purifying selection results in homogenized populations (uniformity can be observed in downstream spacers within the array[20]). Are beneficial traits unrelated to CRISPR immunity lost because of this loss of genetic diversity? Conversely, what traits or genes can use this as a means for genetic hitchhiking? If CRISPR immunity does have a large impact on shaping genome evolution or limiting genetic diversity in bacterial populations, it is tempting to speculate that this could be a reason for bacteria to not harbor CRISPR systems (or at least spacer acquisition competent ones) in their genomes.

More ecological and metagenomic approaches will expand the depth and breadth of our understanding of CRISPR biology. A great majority of CRISPR studies are conducted under specific sets of idealized conditions that likely do not faithfully recapitulate natural environmental contexts (this is not exclusive to CRISPR and is a general limitation of many microbiological studies). Studies of bacteria-phage interactions using log-phase cultures grown in rich media are not likely to accurately reflect the majority of such interactions in nature. For instance, the vast majority of bacteria on the planet are in a non-replicative state[107]. This has significant implications for how we view CRISPR biology. For instance, it has been experimentally demonstrated that lytic phage programs cannot be initiated in dormant cells, but the lytic cycle can resume when the cell exits dormancy[108]. How does this impact spacer acquisition? Is our assessment of spacer acquisition as a rare event (~$10^{-7}$) a gross underestimate? If spacer acquisition could still proceed in a non-replicative cell in which phage replication is stalled, then perhaps the cell can take advantage of this extra time to acquire new spacers. In this case, spacer acquisition could actually be quite common within a population. Relatedly, in populations that are exponentially growing, do persister cells represent a non-genetic phenotypic subpopulation that acquires spacers at higher rates? Generally, how cell metabolism and growth state affect CRISPR immunity remains largely unexplored but is a promising area for future study.

Another limitation of current experimental studies of spacer acquisition is that the majority ignores lysogeny as an option for cells to gain phage resistance.

Prophages are rampant in sequenced bacterial genomes and can offer protection against related phages through superinfection exclusion and against other phages by a variety of other phage exclusion mechanisms[109]. It will be informative to study spacer acquisition against lysogeny-competent temperate phages. For example, it would be interesting to explore the balance between lysogeny-mediated and CRISPR-mediated phage resistance and how environmental conditions affect the balance between these two mechanisms.

Recently, spacer acquisition has been the basis of a variety of new DNA-based information and memory storage systems[110]. In addition to this, CRISPR-Cas systems have been widely repurposed for a variety of technological applications, ranging from genome editing[111] to point-of-care diagnostics[112]. Altogether, these CRISPR-enabled technological advances have already revolutionized the world of biology and medicine, ushering in a new era of genome manipulation. But still, the unique and surprising biology of CRISPR immune systems in their native microbial world has a wealth of mysteries waiting to be unraveled.

## CHAPTER 5: MATERIALS AND METHODS

### Bacterial strains and growth conditions

Cultivation of *S. aureus* RN4220[92] was carried out in brain-heart infusion (BHI) or heart infusion (HI) media (BD) at 37°C. When applicable, media was supplemented with chloramphenicol (10 μg/mL), erythromycin (10 μg/mL) or spectinomycin (250 μg/mL) to ensure maintenance of the pC194-derived[113], pE194[114], or pLZ12-derived[115] plasmids, respectively.

### Simulation of CRISPR immunization

Overnight cultures were diluted 1:100 into fresh BHI supplemented with appropriate antibiotics and 5 mM $CaCl_2$. These cultures were then grown to an $OD_{600}$ of 0.2-0.4, upon which they were normalized to $OD_{600}$ = 0.2. Position 1 and Position 5 strains were diluted 1:10,000 into the sensitive strain RN4220 in triplicate. The mixed cultures were infected with phage ΦNM4γ4 at an MOI of 1. Growth of the cultures was detected measuring $OD_{600}$ using plate reader.

### Plasmid construction

The plasmid containing the wild-type *Streptococcus pyogenes* CRISPR (pWJ40) was described elsewhere[48]. The sequences of all the oligonucleotides used in this study are shown in Table S1. DNA sequences obtained by gene synthesis (Genewiz, Inc.) are shown in the accompanying Supplemental sequences file. Leader deletion 78-102 (pJM13) was constructed by two-piece Gibson assembly[116] from pWJ40 using the primer pairs JM118-JM91 and JM90-JM112, respectively. Leader deletions 88-102 (pJM23), 98-102 (pJM25) were constructed by one-piece Gibson assembly from pWJ40 using the primer pairs

JM118-JM132 and JM118-JM134, respectively. Leader mutation -5 to -1 (pJM28) and -10 to -6 (pJM29) were constructed using one-piece Gibson assembly from pWJ40 using the primer pairs JM139-JM140 and JM141-JM142, respectively. pJM37 (swap of *spc2* and *spc4*) was constructed by inserting the array with BsaI-compatible overhangs in place of flanking repeats into pDB184 [58]. The swapped array sequence for pJM37 was obtained by gene synthesis. pJM70 (pJM37 with LAS mutation) was constructed by two-piece Gibson assembly from pJM37 using the primer pairs JM139-JM91 and JM140-JM90. pJM62 (pWJ40 with BsaI-cloning site for array cloning) was constructed by two-piece Gibson assembly from pWJ40 using the primer pairs JM210-JM173 and JM211-JM172. pJM73 (insertion of LAS in place of *spc4*) was constructed by inserting the array with BsaI-compatible overhangs in place of flanking repeats into pJM62. The array sequence for pJM73 was obtained by gene synthesis. pJM111 (LAS mutant Δ*spc4*) was constructed by around-the-horn PCR[117] using primers JM318 and JM319 to remove spacer 4 and one repeat from the pJM28 CRISPR array template. pJM74 (pJM73 with LAS mutation) was constructed by two-piece Gibson assembly from pJM73 using the primer pairs JM139-JM91 and JM140-JM90. pJM86 [pJM62 with dCas1 (E220A)] was constructed by two-piece Gibson assembly from pJM62 using the primer pairs JM172-PS285 and JM173-PS284. pRH253 (pWJ40 with spacer targeting phage ΦNM4γ4) was isolated during an on-plate spacer acquisition assay[48]. pJM87 (Position 5, dCas1, pC194-derived) was constructed by inserting the array with BsaI-compatible overhangs in place of flanking repeats into pJM86 The array sequence for pJM87 was obtained by

gene synthesis. pJM90 (Position 1, dCas1, pC194-derived) was constructed by two-piece Gibson assembly from pRH253 using the primer pairs JM172-PS285 and JM173-PS284. pJM141 (Position 5, dCas1, pLZ12-derived) and pJM142 (Position 1, dCas1, pLZ12-derived) were constructed by three-piece Gibson assembly from pJM87 and pJM90, respectively, and pLZ12 using the primer pairs JM350-JM173, JM351-JM172, and JM352-JM353.

The plasmid containing the *Streptococcus thermophilus* LMD-9 CRISPR03 was described elsewhere[48]. The plasmid containing *Streptococcus thermophilus* LMD-9 CRISPR01 was cloned using Gibson assembly using pC194 and *Streptococcus thermophilus* LMD-9 genomic DNA as templates with primers JM409, JM411, JW155, JW156. The plasmid containing the *Streptococcus mutans* NN2025 CRISPR locus was cloned using Gibson assembly using pC194 and *Streptococcus mutans* NN2025 genomic DNA as templates with primers JM386, JM389, JM511, JM512. Derivatives of these plasmids were constructed using Gibson assembly (details available upon request). *Streptococcus thermophilus* CRISPR arrays with different spacers were constructed using Gibson assembly and synthesized DNA constructs purchased from Genewiz.

**Spacer acquisition assay in liquid cultures**

In-liquid spacer acquisition assays were performed using a slightly modified method described in [48]. Overnight cultures were diluted 1:100 in fresh BHI with appropriate antibiotics and 5 mM $CaCl_2$ and grown to $OD_{600}$ = 0.5. Cultures were then infected with phage ΦNM4γ4 at a MOI of 1 phage per bacterium, unless otherwise noted, and incubated with shaking at 37°C for 24

hours. For Sanger sequencing, cultures were streaked onto a fresh BHI-agar plate containing the appropriate antibiotics. Plates were incubated overnight at 37°C. Single colonies were picked and incubated in lysis buffer (250 mM KCl, 5 mM $MgCl_2$ 50 mM Tris-HCl at pH 9.0, 0.5% Triton X-100) for 20 minutes at 37°C and then 98°C for 10 minutes. 1 µl of supernatant was then used as template for PCR amplification using TopTaq Master Mix Kit (Qiagen).

For *Streptococcus pyogenes* CRISPR spacer acquisition, primers JM116/H050 were used for detecting wild-type spacer integration, JM171/JM119 were used for detecting ectopic spacer integration, and JM116/JM119 were used for amplification of the entire array.

For *Streptococcus thermophilus* CRISPR spacer acquisition, primers JM439/JM440 and H169/H267 were used for CRISPR01 and CRISPR03 respectively.

**Adaptation competition assay**

Overnight cultures were diluted 1:100 into fresh BHI with appropriate antibiotics and 5 mM $CaCl_2$ and grown to $OD_{600}$ = 0.4-0.6. The cultures were then normalized by OD600 and mixed in a 1:1 ratio. This mixed culture was then infected with phage ΦNM4γ4, in triplicate, at an MOI of 1 for 24 hours, upon which the culture was streaked onto a BHI agar plate containing appropriate antibiotics. Colonies were subjected to PCR using the primers JM116/JM119 for identification of the strain and analysis of spacer integration via Sanger sequencing of PCR products.

**Bacterial growth curves during phage infection**

Phage infections and plate reader growth curves were performed as previously described[58] with slight modifications. In brief, overnight cultures were diluted 1:100 into 2 ml of fresh BHI supplemented with appropriate antibiotics and 5 mM $CaCl_2$. These cultures were then grown to an $OD_{600}$ of 0.4-0.6, upon which they were infected with phage ΦNM4γ4. To produce plate reader growth curves, 150 µl of infected cultures, normalized for $OD_{600}$, was transferred to a 96-well plate in triplicate. $OD_{600}$ measurements were taken every 10 minutes.

**Phage titer assay**

Phage titer assays were performed as previously described[48].

**Detection of spacer integration events using next-generation sequencing**

Overnight cultures were diluted 1:100 in fresh BHI containing the appropriate antibiotics. At $OD_{600}$ = 0.5, the cultures were infected with phage ΦNM4γ4 at an MOI of 1. After 24 hours, plasmids were isolated from the cultures using the QIAprep Spin Miniprep Kit protocol as previously described[48]. 100 ng of plasmid DNA served as template for PCR amplification using Phusion High-Fidelity DNA Polymerase (Thermo Scientific) and 6 barcoding primer sets (H188/JM257, JM248/JM258, JM249/JM259, JM250/JM260, JM251/JM261, and JM252/JM262). After gel extraction and purification of the DNA bands corresponding to expanded CRISPR arrays, the DNA was subjected to

sequencing by Illumina MiSeq (600 cycles). Reads were filtered by quality score, sorted by barcode, and subsequently converted to .txt format using a custom Python script that utilized Biopython. Other custom Python scripts were employed to create a list of reads, normalized to total number of reads per barcode, that met the following criteria: reads must contain 7 spacers and 8 repeats, the pre-existing spacers must be in the correct order, and the new spacer must have at least 15 bp matching the phage genome.

For early time point deep sequencing, cultures were harvested at 30 minutes post infection with ΦNM4γ4 at an MOI of 1 and then miniprepped as described above using the modified lysostaphin protocol. PCR with Phusion polymerase was used to amplify the CRISPR arrays using a special primer set designed to amplify low-frequency spacer integration events as done previously[48]. The primers used here were JM472-479. Size selection was done using a modified Illumina library preparation protocol as described previously[41,48]. Custom Python scripts and BWA[118] were used to analyze the deep sequencing data and assign PAMs to unique spacer reads based on genome alignment.

**Preparation of electrocompetent *S. aureus* cells**

Electrocompetent *S. aureus* cells were prepared and transformed with DNA using a previously described method[58], with the exception of using BHI medium instead of TSB medium.

**Mathematical modeling of dual CRISPR system spacer acquisition**

A probabilistic model was generated using a custom Python script which assumed a 9:1 spacer acquisition ratio for CRISPR01:CRISPR03. The average values over one hundred iterations was used for the model presented here.

**Leader sequence alignments**

Related type II-A CRISPR systems were selected based on previous work that classified these systems[93]. Multiple sequence alignment was performed and visualized using T-Coffee[119].

**Bioinformatic analysis of Type II-A CRISPR systems**

The databases entitled CRISPRdb (http://crispr.i2bc.paris-saclay.fr/crispr/)[120] and CRISPI (http://crispi.genouest.org/)[121] were used to collect data on CRISPR array length and other characteristics of CRISPR systems. The list of CRISPR systems was compiled based on a previous analysis of Type II-A CRISPR systems[103].

**Total RNA extraction**

Total RNA was extracted as described in Jiang et al. 2016[122] with minor modifications. Briefly, 6 ml of *S. aureus* culture was pelleted and was resuspended in 1 ml of TRIzol (Thermo Fisher Scientific). The cell suspension was transferred to 2 ml-tubes pre-filled with 0.25 $cm^3$ of 0.1 mm glass beads on ice. Cells were disrupted twice at 4°C, using MiniBeadbeater-1 (BioSpec

Products) at an intensity setting of 42 for 30 seconds. 200 µl of chloroform was added to the mixture and the remainder of the RNA extraction was carried out according to the Trizol manufacturer's instructions.

**Small RNA library preparation and next-generation sequencing of RNA**

Small RNA libraries were prepared with the TruSeq Small RNA Library Preparation Kit (Illumina) using 2 µg of total RNA extracted from cultures at $OD_{600}$ = 0.5. Reverse transcribed and PCR-amplified RNA species between 30 and 55 nt were purified by gel extraction and ethanol precipitation. The purified DNA was subjected to next-generation sequencing by Illumina MiSeq (150 cycles). Reads were sorted by index and converted to .txt format using a custom Python script that utilized Biopython. Other custom Python scripts were used analyze crRNA abundance.

## Oligonucleotides used in this study

| Name | Sequence |
|---|---|
| JM1 | CTATGCAGAAGACCTTCAAGAAAG |
| JM2 | CATCTTAGCTTGGTAATCTTCCTTG |
| JM3 | TCTCACTTACTAAGTCACAAGGAAGATT |
| JM4 | AATCTTTTTTCAACAATTGATAAAAATC |
| JM5 | CACGATGGATGGTTGGCTATT |
| JM6 | ATAGAAATTGAGAATAAACTACATTTGATG |
| JM7 | TTGATCCCGAAACAGTTCATAGT |
| JM8 | GCTAATCGTCCTAACGTACCTAGAC |
| JM9 | TCATAGGGCATCTGGAACGC |
| JM10 | TCAAATCCACTAGTTCTCCTGTTTCTA |
| JM11 | AGGAAATCATTGGACTTATGAAAGAC |
| JM12 | ATGTAATCAGCGAAGCGTTCC |
| JM13 | TAAACAATCTCGCATGATATCAAAA |
| JM14 | AAATAAAGAATAGCGTGTCAATGCT |
| JM15 | CCACAGTTTATGGCATCAATATAAC |
| JM16 | ATCTATTGGTTGGTCGGGATTT |
| JM17 | GATTGGACATTACAACAATTAGTAGATTAC |
| JM18 | AATCTGCATAACCATTTTCTTTTAAT |
| JM19 | TTAACAATTAAACCATGTGAGACATTT |
| JM20 | CTTGTTCACTATCTAAATCAATGATTGTT |
| JM21 | TGAAAATGCAGCATGTCATTTAG |
| JM22 | TAATCCGCTTGCAATCTTTTC |
| JM23 | GGATCTGCTTACGCCTTTAATATT |
| JM24 | TGAAATCGACATGTACATTTGAATC |
| JM25 | GTTTGACGTATCATACATACCTTGC |
| JM26 | CTAAATGCGAACAAACCACGA |
| JM27 | TAAACAAATACTTAAAGATGACGGTGA |
| JM28 | GTACATTCCAAATTGCGACATACTT |
| JM29 | AATCAGATGAGTATTTAATAACAGAATTATCA |
| JM30 | GCAGAATAAATATTTGTAATGATAGCG |
| JM31 | AATTATTAGGTATTGAGAAATATAAACCACA |
| JM32 | TTTTGATAGTTCTCTAACGTTTCATGG |
| JM33 | GTCAAAATATTAAACCAGTTTATAAAATGG |
| JM34 | TCATTTTCATCTTCATACGTACCATAA |
| JM35 | ATACATGTGCAGCTGAATTTGAAT |
| JM36 | CTAGTACTAGAATTTTCTCTTTATCAGTTACTATAG |
| JM37 | aaacATTGTTTATCTACCATGTAATAGTAAAAAAg |
| JM38 | aaaacTTTTTTACTATTACATGGTAGATAAACAAT |

| | |
|---|---|
| JM39 | aaacCCCATTATTATATACGCATTTTGTAAAAAAg |
| JM40 | aaaacTTTTTTACAAAATGCGTATATAATAATGGG |
| JM41 | aaacCATTTTCATCAAATAAAGTATTGTAAAAAAg |
| JM42 | aaaacTTTTTTACAATACTTTATTTGATGAAAATG |
| JM43 | CAGATCAAGAAATCAAACTCCAA |
| JM44 | CAGCATCATCTACAGAAACTC |
| JM45 | CGTTATAAGGGGCTAGGTGAAA |
| JM46 | CAGCATCATCTACAGAAACTCTAGC |
| JM47 | CCGATTGCTCTAGTAAAAGTCCTG |
| JM48 | CGTAATGGTAAAATCGCCTGC |
| JM49 | AGTTGATATTCAAGAAAAAATGGGA |
| JM50 | AGATACTTTGTATCCGCCACCG |
| JM51 | TGTTGATAATAGTATTGACGAGGCATTAGC |
| JM52 | GTGAAGACCGCCAGAT |
| JM53 | AGAAAAGATGGGACGCCCT |
| JM54 | GTGAAGACCGCCAGATACTTTG |
| JM55 | GATTGAAAATGCAGCTTACTTTGTAAAAAAgttttagagctatgctgttttgaatg |
| JM56 | TTTTTTACAAAGTAAGCTGCATTTTCAATCgttttgggaccattcaaaacagc |
| JM57 | AGGGCTTAACAAAACTAGATTTGTAAAAAAgttttagagctatgctgttttgaatg |
| JM58 | TTTTTTACAAATCTAGTTTTGTTAAGCCCTgttttgggaccattcaaaacagc |
| JM59 | GCCCTATGATATAAAGGTATTGAGTTCCATgttttagagctatgctgttttgaatg |
| JM60 | ATGGAACTCAATACCTTTATATCATAGGGCgttttgggaccattcaaaacagc |
| JM61 | tcaaggtaagtatgtgattaaagcaac |
| JM62 | attgggctcgactatatcttgaa |
| JM63 | caatttgcgcttaactcttttg |
| JM64 | gaataagttcatcgaattcattaccac |
| JM65 | CACTAATTTCCCTTCCATCGCTCTTTGG |
| JM66 | CGATTGCTTCTATTATGCGTTGTTGTTCGG |
| JM67 | CGTTCTAGGGCTTAACTTCTGTGTTCG |
| JM68 | CAGCAAACTTCCCATAATGTTTCCCATCATCACCAC |
| JM69 | CCAGACCATTGATTACATTTCTGAGCAATCACAGGG |
| JM70 | ATGACACCAGAACAACTTAAAGCA |
| JM71 | TATACTTTTCACCCTTTTCAGTATCAC |
| JM72 | TACTCTTAATAAATGCAGTAATACAGGGGCTTTTCAAGACTGAAGTCTAGCTGAGACAAATAGTGCGATTACAAAATTTTTTAGACAAAAATAGTCTACGAG |
| JM73 | CTCGTAGACTATTTTTGTCTAAAAAATTTTGTAATCGCACTATTTGTCTCAGCTAGACTTCAGTCTTGAAAAGCCCCTGTATTACTGCATTTATTAAGAGTA |
| JM74 | GAAAT**TAATACGACTCACTATA**GGGGTTGGAACCATTCAAAACAGCATAGCA |
| JM75 | AAAAAAAGCACCGACTCGGTGC |
| JM76 | GAUUUCUUCUUGCGCUUUUUGUUUUAGAGCUAUGCUGUUUUG |
| JM77 | TAATGAATTCCCCAATACCC**CTAAAGAAGAACGCGAAAAA**AACCAGCGCA |
| JM78 | TGCGCTGGTTTTTTTCGCGTTCTTCTTTAGGGGTATTGGGGAATTCATTA |
| JM79 | CTTTGAGCTTCCGAGACTGGTCTC |

| | |
|---|---|
| JM80 | CAGCATAGCTCTAAAACCTCGTAGACTATTTTTGTC |
| JM81 | *ctcgtagactatttttgtctaaaaaatttCgtaatcgcac* |
| JM82 | agatgaagattatttcttaataactaaaaatatggtataaTGAGACAAATAGTGCGATTA |
| JM83 | TAATCGCACTATTTGTCTCAttataccatatttttagttattaagaaataatcttcatct |
| JM84 | AGACTGAAGTCTAGCgttttagagctatgctgttttgaatggtcccaaaactgagaccag |
| JM85 | ctggtctcagttttgggaccattcaaaacagcatagctctaaaacGCTAGACTTCAGTCT |
| JM86 | AGATGAAGATTATTTCTTAATAACTAAAAATATGGTATAAggggcttttcaagactgaag |
| JM87 | cttcagtcttgaaaagccccTTATACCATATTTTTAGTTATTAAGAAATAATCTTCATCT |
| JM88 | GTGCGATTACGAAATgttttagagctatgctgttttgaatggtcccaaaactgagaccag |
| JM89 | ctggtctcagttttgggaccattcaaaacagcatagctctaaaacATTTCGTAATCGCAC |
| JM90 | GAGAAGATTGAAAAAATCTTGACTTTTCG |
| JM91 | CGAAAAGTCAAGATTTTTTCAATCTTCTC |
| JM92 | caaaagggtttctaatcccattcgaaagcg |
| JM93 | ctctaaaacactcatgagctaagtaagattttccttttc |
| JM94 | GCTAGCA**ATGAACTTTAATAAAATTGATTT**AGACAATTGGAAGAGAAAAGAG |
| JM95 | **TTATAAAAGCCAGTCATTAGGCCTATCTGAC** |
| JM96 | GTCACTCGAAGATGAAGACTTGATTGAAGAAG |
| JM97 | **AAATCAATTTTATTAAAGTTCAT**TGCTAGCTACTCATTTTCATTATTATTTAGTGAAGTG |
| JM98 | **GTCAGATAGGCCTAATGACTGGCTTTTATAA**AACAAGAAACTCGCTCTAACGAGTTTC |
| JM99 | ACAGATGAATATAACTGCAAACTCGGATCATC |
| JM100 | tgagaccagtctcggaagctcaaaggtctcAGGGATAAATATGAACATgatgagtgatcg |
| JM101 | CCTgagacctttgagcttccgagactggtctcaAGAGCTCGTGCTATAATTATACTAATT |
| JM102 | ggtataaGactcttaataaatgcagtaatacag |
| JM103 | ctgtattactgcatttattaagagtCttatacc |
| JM104 | **ggtataa***tgggcttttcaagactg* |
| JM105 | cagtcttgaaaagcccattatacc |
| JM106 | tgagaccagtctcggaagctcaaaggtctcAGGAGTGATCGTTAAATTTATACTGC |
| JM107 | CCTgagacctttgagcttccgagactggtctcaTGTTCATATTTATCAGAGCTCGTG |
| JM108 | ctcggaagctcaaaggtctcAGGgatgagtgatcgttaaatttatactgcaatcggatgc |
| JM109 | CCTgagacctttgagcttccgaggttcatatttatcagagctcgtgctataattatacta |
| JM110 | **Gttttagagctatgctgttttgaatggtcccaaaactgagaccagtctcggaagctcaaa** |
| JM111 | **ggaccattcaaaacagcatagctctaaaac**gctagacttcagtcttgaaaagcccctgta |
| JM112 | **ggaccattcaaaacagcatagctctaaaac**attttgtaatcgcactatttgtctcagcta |
| JM113 | **tatttcttaataactaaaaatatggtataa**atttttagacaaaaatagtctacgag**gtt** |
| JM114 | **ttataccatatttttagttattaagaaataatcttcatctaaaatatacTgcggaaaatc** |
| JM115 | attcaaaacagcatagctctaaaaC**ttataccatatttttagttattaagaaataatctt** |
| JM116 | **gtgaactatatgattttccgcagtatatttagatgaagattatttcttaataac** |
| JM117 | GTTTGAACTCAACAAGTCTCAGTGTGCTG |
| JM118 | GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC**TGCGCTGGTTGATTTCTTCTTGCG** |
| JM119 | CCCTCTTTCTCAAGTTATCATCGGCAATG |
| JM120 | AAAAAACGTCAGGAGAAGATTCGTAATTTGATGAGTGATCGTTAAATTTATACTGCAATC |

| | |
|---|---|
| JM121 | GACTTCTGGAAGGATCCAAATCACAGAATCGTTCATATTTATCAGAGCTCGTGCTATAAT |
| JM122 | ATTATAGCACGAGCTCTGATAAATATGAACGATTCTGTGATTTGGATCCTTCCAGAAGTC |
| JM123 | GATTGCAGTATAAATTTAACGATCACTCATCAAATTACGAATCTTCTCCTGACGTTTTTT |
| JM124 | ATGGCAGCGCCTAAATGTAGAATGATAAAATACTCTTAATAAATGCAGTAATACAGGGGC |
| JM125 | GCCCCTGTATTACTGCATTTATTAAGAGTATTTTATCATTCTACATTTAGGCGCTGCCAT |
| JM126 | ATGAAGATTATTTCTTAATAACTAAAAATATGGTATAATTAATACCAGCAGTCGGATACC |
| JM127 | GGTATCCGACTGCTGGTATTAATTATACCATATTTTTAGTTATTAAGAAATAATCTTCAT |
| JM128 | GGAGGCTTTTGATGAATCTTAATTTTTCCt**a**aCTAGATGAACCGATTCCATTAAGAGGCG |
| JM129 | CGCCTCTTAATGGAATCGGTTCATCTAGt**t**aGGAAAAATTAAGATTCATCAAAAGCCTCC |
| JM130 | TGGGACCATTCAAAACAGCATAGCTCTAAAACATTTTGTAATCGCACTATTTGTCTCAGC |
| JM131 | ACCATTCAAAACAGCATAGCTCTAAAACAAAAAATTTTGTAATCGCACTATTTGTCTCAG |
| JM132 | CCATTCAAAACAGCATAGCTCTAAAACTGTCTAAAAAATTTTGTAATCGCACTATTTGTC |
| JM133 | GGACCATTCAAAACAGCATAGCTCTAAAACATTTTTGTCTAAAAAATTTTGTAATCGCAC |
| JM134 | CCATTCAAAACAGCATAGCTCTAAAACAGACTATTTTTGTCTAAAAAATTTTGTAATCGC |
| JM135 | ACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGCGCTGGTTGATTTCTTC |
| JM136 | AGCATAGCTCTAAAACCTCGTTTATACCATATTTTTAGTTATTAAGAAATAATCTTCATC |
| JM137 | AGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGCGCTGGTTGATT |
| JM138 | AGCTCTAAAACCTCGTAGACTTTATACCATATTTTTAGTTATTAAGAAATAATCTTCATC |
| JM139 | CATCTGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGCGCTGGTTGATTTCTTC |
| JM140 | CAAAACAGCATAGCTCTAAAACAGATGAGACTATTTTTGTCTAAAAAATTTTGTAATCGC |
| JM141 | CTGAGACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGCGCTGGTTGATT |
| JM142 | AAACAGCATAGCTCTAAAACCTCGTCTCAGATTTTTGTCTAAAAAATTTTGTAATCGCAC |
| JM143 | CCCCGAGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGCGCTGGT |
| JM144 | CATAGCTCTAAAACCTCGTAGACTCGGGGTGTCTAAAAAATTTTGTAATCGCACTATTTG |
| JM145 | CTCACAAAATAGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGCG |
| JM146 | GCTCTAAAACCTCGTAGACTATTTTGTGAGAAAAAATTTTGTAATCGCACTATTTGTCTC |
| JM147 | GAGACAAATAGTGCGATTACAAAATGGGGGAGACAAAAATAGTCTACGAGGTTTTAGAGC |
| JM148 | ACCTCGTAGACTATTTTTGTCTCCCCCATTTTGTAATCGCACTATTTGTCTCAGCTAGAC |
| JM149 | attaatacttcgatacgacaaaacaatggttccaacaagattattttataacttttataa |
| JM150 | gttttgtcgtatcgaagtattaataaggctagtccgttatcaacttgaaaaagtgg |
| JM151 | TGCGCTGGTTGATTTCTTCTTGCGCTTTTTgttaatgtcgatacgacaaaacaatggtcccaaaacttcagcac |
| JM152 | AAAAAGCGCAAGAAGAAATCAACCAGCGCAgttttgggaccattgttttgtcgtatcgacattaacCTCGTAGACTATTTTTGTCTAAAAAATTTTG |
| JM153 | AAAATTTTTTAGACAAAAATAGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGCGCTGGT |
| JM154 | ACCAGCGCAGTTTTGGGACCATTCAAAACAGCATAGCTCTAAAACCTCGTAGACTATTTTTGTCTAAAAAATTTT |
| JM155 | TAGCTGAGACAAATAGTGCGATTACAAAATGTTTTAGAGCTATGCTGTTTTGAATGGTCC |
| JM156 | GGACCATTCAAAACAGCATAGCTCTAAAACATTTTGTAATCGCACTATTTGTCTCAGCTA |
| JM157 | ATAGTGCGATTACAAAATTTTTTCTACACCAAGAGTCTAAGAGGTTTTAGAGCTATGCTG |
| JM158 | CAGCATAGCTCTAAAACCTCTTAGACTCTTGGTGTAGAAAAAATTTTGTAATCGCACTAT |
| JM159 | TTTGAATGGTCCCAAAACTAAATTTGTTTAGCAGGTAAACCGTGCTTT |
| JM160 | ACAGCATAGCTCTAAAACAGAGCGCAATTAATTATTGCGGATATTCCT |
| JM161 | aaacAAAATTTTTTAGACAAAAATAGTCTACGAGg |

| | |
|---|---|
| JM162 | aaaacCTCGTAGACTATTTTTGTCTAAAAAATTTT |
| JM163 | AAGATTATTTCTTAATAACTAAAAATATGGTATAAAAAATAGTCTACGAGGTTTTAGAGC |
| JM164 | TTATACCATATTTTTAGTTATTAAGAAATAATCTTCATCTAAAATATACTGC |
| JM169 | CGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCC |
| JM170 | GGAATTCGAAAAGTCAAGATTTTTTCAATCTTCTCACG |
| JM171 | TGCGCTGGTTGATTTCTTCTTGCGCTTTTT |
| JM172 | ggaattttgaagaagttgtcgataaaggtgcttcagc |
| JM173 | gctgaagcacctttatcgacaacttcttcaaaattcc |
| JM174 | CATCTGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACAACATTGCCGATGATAACT |
| JM175 | GTCTAGCTGAGACAAATAGTGCGATTACAAAATGGGGGAGACAAAAATAGTCTACGAGG |
| JM176 | AGACACCCCGAGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGCG |
| JM177 | CTCGTAGACTCGGGGTGTCTAAAAAATTTTGTAATCGCACTATTTGTCTC |
| JM178 | CTCGTAGACTATTTTGTGAGAAAAAATTTTGTAATCGCACTATTTGTCTC |
| JM179 | aaacAAAATTTTTTAGACAAAAATAGTCTCATCTg |
| JM180 | aaaacAGATGAGACTATTTTTGTCTAAAAAATTTT |
| JM181 | aaacAAAATTTTTTAGACAAAAATCTGAGACGAGg |
| JM182 | aaaacCTCGTCTCAGATTTTTGTCTAAAAAATTTT |
| JM183 | aaacAAAATTTTTTAGACACCCCGAGTCTACGAGg |
| JM184 | aaaacCTCGTAGACTCGGGGTGTCTAAAAAATTTT |
| JM185 | aaacAAAATTTTTTCTCACAAAATAGTCTACGAGg |
| JM186 | aaaacCTCGTAGACTATTTTGTGAGAAAAAATTTT |
| JM187 | aaacAAAATGGGGGAGACAAAAATAGTCTACGAGg |
| JM188 | aaaacCTCGTAGACTATTTTTGTCTCCCCCATTTT |
| JM189 | aaacCCCCGTTTTTAGACAAAAATAGTCTACGAGg |
| JM190 | aaaacCTCGTAGACTATTTTTGTCTAAAAACGGGG |
| JM191 | ctcggaagctcaaaggtctcGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGCG |
| JM192 | gagacctttgagcttccgagactggtctcaGTAATCGCACTATTTGTCTCAGCTAGACTT |
| JM193 | ttacAAAATTTTTTAGACAAAAATAGTCTCATCTg |
| JM194 | ttacAAAATTTTTTAGACAAAAATCTGAGACGAGg |
| JM195 | ttacAAAATTTTTTAGACACCCCGAGTCTACGAGg |
| JM196 | ttacAAAATTTTTTCTCACAAAATAGTCTACGAGg |
| JM197 | ttacAAAATGGGGGAGACAAAAATAGTCTACGAGg |
| JM198 | ttacCCCCGTTTTTAGACAAAAATAGTCTACGAGg |
| JM199 | aaacAAAATTTTTTAGACAAAAATAGTCTACGATg |
| JM200 | aaaacATCGTAGACTATTTTTGTCTAAAAAATTTT |
| JM201 | aaacAAAATTTTTTAGACAAAAATAGTCTACGCGg |
| JM202 | aaaacCGCGTAGACTATTTTTGTCTAAAAAATTTT |
| JM203 | aaacAAAATTTTTTAGACAAAAATAGTCTACTAGg |
| JM204 | aaaacCTAGTAGACTATTTTTGTCTAAAAAATTTT |
| JM205 | aaacAAAATTTTTTAGACAAAAATAGTCTAAGAGg |
| JM206 | aaaacCTCTTAGACTATTTTTGTCTAAAAAATTTT |

| | |
|---|---|
| JM207 | aaacAAAATTTTTTAGACAAAAATAGTCTCCGAGg |
| JM208 | aaaacCTCGGAGACTATTTTTGTCTAAAAAATTTT |
| JM209 | tgagaccagtctcggaagctcaaaggtctcGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGCGCTGGTTGATTTCTTCTTGCG |
| JM210 | tgagaccagtctcggaagctcaaaggtctcGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACAACATTGCCGATGATAACTTGAGAAAGAGG |
| JM211 | gagacctttgagcttccgagactggtctcaGTTTTGGGACCATTCAAAACAGCATAGCTCTAAAACCTCGTAGACTATTTTTGTCTAAAAAATTTTGTAATCGCAC |
| JM212 | CATCTgttttagagctatgctgttttgaatggtcccaaaactgagaccagtctcggaagc |
| JM213 | AAAAATATGGTATAACTGAGACAAATAGTGCGATTACAAAATTTTTTAGACAAAAATAG |
| JM214 | CACTATTTGTCTCAGTTATACCATATTTTTAGTTATTAAGAAATAATCTTCATCTAAAAT |
| JM215 | GTGCGATTACAAAATGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGCGCTGGT |
| JM216 | GCATAGCTCTAAAACATTTTGTAATCGCACTATTTGTCTCAGCTAGACTTCAGTCTTGAA |
| JM217 | GACTGAAGTCTAGCTGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGCGCTGGT |
| JM218 | GCATAGCTCTAAAACAGCTAGACTTCAGTCTTGAAAAGCCCCTGTATTACTGCATTTATT |
| JM219 | ATGCAGTAATACAGGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGCGCTGGT |
| JM220 | GCATAGCTCTAAAACCCTGTATTACTGCATTTATTAAGAGTATTATACCATATTTTTAGT |
| JM221 | AAAATTTTTTAGACAAAAATAGTCTNNNNNGTTTTAGAGCTATGCTGTTTTGAATGGTCC |
| JM222 | AGACTATTTTTGTCTAAAAAATTTTGTAATCGCACTATTTGTC |
| JM223 | tattcttcaactaaaaaacactaaattaataagaaagagccaaacctcgaaag |
| JM224 | tgttgAACTAATGGGTGCgttccaattgaaaccaaagtgctttctcttatctattattc |
| JM225 | ttcaattggaacGCACCCATTAGTTCAACAAACGAAAATTGGATaaagtg |
| JM226 | attaatttagtgtttttttagttgaagaataaagaccacattaaaaaatgtggtcttttg |
| JM227 | CAAACGctcgtgTAGAGATAAAAGAGTCCTTTGGATGATTCCAAGG |
| JM228 | TTATTAAGAGTACTAGTTATTATAGGGCATCAAAATATAATCCTTGTCTAAAATAAATTG |
| JM229 | CTATAATAACTAGTACTCTTAATAAATGCAGTAATACAGGGGCTTTTCAAGAC |
| JM230 | TCTTTTATCTCTAcacgagCGTTTGTTGAACTAATGGgtgctttagttg |
| JM231 | GAAGATTCGTAATTTctctgataaatatgaacatgatgagtgatcgttaaatttatact |
| JM232 | CCAAATCACAGAATCctcgtgctataattatactaattttataaggaggaaaaaatatgg |
| JM233 | taattatagcacgagGATTCTGTGATTTGGATCCTTCCAGAAGTC |
| JM234 | tcatatttatcagagAAATTACGAATCTTCTCCTGACGTTTTTTTAAATCTTG |
| JM235 | GATTCGTAATTTcgttagcaaaaacaggtttaagcctcgc |
| JM236 | AAATCACAGAATCcaaaaaatgagtggcaaaatgctagccactc |
| JM237 | actcatttttttgGATTCTGTGATTTGGATCCTTCCAGAAGTC |
| JM238 | gtttttgctaacgAAATTACGAATCTTCTCCTGACGTTTTTTTAAATCTtgactg |
| JM239 | aaacGTTTTATATCCAGAATCAGCAAAAGCCGAGg |
| JM240 | aaaacCTCGGCTTTTGCTGATTCTGGATATAAAAC |
| JM241 | aaacAATCTTACTAAAAGAGTTAGAACGAGTTTTg |
| JM242 | aaaacAAAACTCGTTCTAACTCTTTTAGTAAGATT |
| JM243 | aaacTGAGATAATGCCGACTGTACTTTTTACAGTg |
| JM244 | aaaacACTGTAAAAAGTACAGTCGGCATTATCTCA |
| JM245 | ctgaatttggcggtggtacgcaagg |
| JM246 | ccttgcgtaccaccgccaaattcag |
| JM247 | AAATCACAGAATCcaaaaaatgagtggcaaaatgctagcca |

| | |
|---|---|
| JM248 | NNNNNtagCAAAATTTTTTAGACAAAAATAGTC |
| JM249 | NNNNNcagCAAAATTTTTTAGACAAAAATAGTC |
| JM250 | NNNNNctcCAAAATTTTTTAGACAAAAATAGTC |
| JM251 | NNNNNgagCAAAATTTTTTAGACAAAAATAGTC |
| JM252 | NNNNNtccCAAAATTTTTTAGACAAAAATAGTC |
| JM253 | NNNNNcctCAAAATTTTTTAGACAAAAATAGTC |
| JM254 | NNNNNattCAAAATTTTTTAGACAAAAATAGTC |
| JM255 | NNNNNgatCAAAATTTTTTAGACAAAAATAGTC |
| JM256 | NNNNNaaaCAAAATTTTTTAGACAAAAATAGTC |
| JM257 | NNNNNagtCTTTCTCAAGTTATCATCGGC |
| JM258 | NNNNNtagCTTTCTCAAGTTATCATCGGC |
| JM259 | NNNNNcagCTTTCTCAAGTTATCATCGGC |
| JM260 | NNNNNctcCTTTCTCAAGTTATCATCGGC |
| JM261 | NNNNNgagCTTTCTCAAGTTATCATCGGC |
| JM262 | NNNNNtccCTTTCTCAAGTTATCATCGGC |
| JM263 | NNNNNcctCTTTCTCAAGTTATCATCGGC |
| JM264 | NNNNNattCTTTCTCAAGTTATCATCGGC |
| JM265 | NNNNNgatCTTTCTCAAGTTATCATCGGC |
| JM266 | NNNNNaaaCTTTCTCAAGTTATCATCGGC |
| JM267 | CATCTGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACAGTGCCGAGGAAAAATTAG |
| JM268 | AGACAAAAATAGTCTNNNNNGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACAGTG |
| JM269 | AGACTATTTTTGTCTAAAAAATTTTGTAATCGCACTATTTGTCTCAGCTAGACTTCAG |
| JM270 | CATCTGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACCTCGTAGACTATTTTTGTCTAAAAAATTTTGTTAATACCAGCAGTCGGATACCTTCC |
| JM271 | CAAAACAGCATAGCTCTAAAACAGATGAGACTATTTTTGTCTAAAAAATTTTGTAATCGC |
| JM272 | GATTATTTCTTAATAACTAAAAATATGGTATAAGTTAATACCAGCAGTCGGATACCTTCC |
| JM273 | TTATACCATATTTTTAGTTATTAAGAAATAATCTTCATCTAAAATATACTGCGG |
| JM274 | TTGGGACCATTCAAAACAGCATAGCTCTAAAACAACATTGCCGATGATAACTTGAGAAAG |
| JM275 | GCTGTTTTGAATGGTCCCAAAACCTCGTAGACTATTTTTGTCTAAAAAATTTTGTAATCG |
| JM276 | GGAGAAGATTCGTAATTTtttatactgcaatcggatgcgat |
| JM277 | CCAAATCACAGAATCatgttcatatttatcagagctcg |
| JM278 | gataaatatgaacatGATTCTGTGATTTGGATCCTTCCAGAAGTC |
| JM279 | cgattgcagtataaaAAATTACGAATCTTCTCCTGACGTTTTTTTAAATCTTG |
| JM280 | CTTGGTATTTCACAAATTGTAACGCTTTAATACGC |
| JM281 | ccttaacaatcccaaaacttgtcgaatctcactatcaatcatttctccaccattag |
| JM282 | ctaatggtggagaaatgattgatagtgagattcgacaagttttgggattgttaagg |
| JM283 | gtttgataaaatattctgcaaaaaaggggacggtcggcttaatagctcacgctatgcc |
| JM284 | ggcatagcgtgagctattaagccgaccgtccccttttttgcagaatattttatcaaac |
| JM285 | CCCCGAGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGCGCTGGTTGATT |
| JM286 | GCATAGCTCTAAAACCTCGTAGACTCGGGGTGTCTAAAAAATTTTGTAATCGCAC |
| JM287 | CTCACAAAATAGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGCGCTGGTTGATT |
| JM288 | GCTCTAAAACCTCGTAGACTATTTTGTGAGAAAAAATTTTGTAATCGCACTATTTGTCTC |

| | |
|---|---|
| JM289 | GGGGGAGACAAAAATAGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGCGCTGGTTGATT |
| JM290 | AAAACCTCGTAGACTATTTTTGTCTCCCCCATTTTGTAATCGCACTATTTGTCTCAGCTA |
| JM291 | CCCCGTTTTTAGACAAAAATAGTCTACGAGGTTTTAGAGC |
| JM292 | CTCGTAGACTATTTTTGTCTAAAAACGGGGGTAATCGCACTATTTGTCTCAGCTAGAC |
| JM293 | GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGCGCTGGTTGATTTCTTCTTGCG |
| JM294 | GGACCATTCAAAACAGCATAGCTCTAAAACATTTTGTAATCGCACTATTTGTCTCAGCTA |
| JM295 | GGACCATTCAAAACAGCATAGCTCTAAAACAGCTAGACTTCAGTCTTGAAAAGCCCCTGT |
| JM296 | GGACCATTCAAAACAGCATAGCTCTAAAACCCTGTATTACTGCATTTATTAAGAGTATTA |
| JM297 | TATTTCTTAATAACTAAAAATATGGTATAAGATTAAGCCATATGGGGTTAAGCAAGG |
| JM298 | GCTTAATCTTATACCATATTTTTAGTTATTAAGAAATAATCTTCATCTAAAATATACTGC |
| JM299 | GGCAAAAGAAGAGTAGTGTGATGGAACAAACggtcataacctgaaggaagatctggatcc |
| JM300 | GATATATACTCTTGTAAACTAGCCACTTCATCCatctgtgccagttcgtaatgtctggtc |
| JM301 | acattacgaactggcacagatGGATGAAGTGGCTAGTTTACAAGAGTATATATCATTGAC |
| JM302 | ggatccagatcttccttcaggttatgaccGTTTGTTCCATCACACTACTCTTCTTTTGCC |
| JM303 | caggtatgaaaacaagcttaggcatggg |
| JM304 | gctgaaatttaattgtcaaaaaatgtgacattcggtcggcttaatagctcacgctatgcc |
| JM305 | ggcatagcgtgagctattaagccgaccgaatgtcacatttttttgacaattaaatttcagc |
| JM306 | aatcccaaaacttgtcgaattaaatttaattacactctaaaattcgtgaatttttaatgg |
| JM307 | aaattcacgaattttagagtgtaattaaatttaattcgacaagttttgggattgttaagg |
| JM308 | gaacGTATTTTAGTTAATGGTGATTCAATGGAACCTATG |
| JM309 | tttaCATAGGTTCCATTGAATCACCATTAACTAAAATAC |
| JM310 | CCATTAGTTCAACAAACGCTCGTGAGCTTACTATGCCATTATTaataacttagcc |
| JM311 | CCAATTTTCGTTTTGTCACTTTGCTTGATATATGAGAATTATTTAACC |
| JM312 | CAAGCAAAGTGACAAAACGAAAATTGGATAAAGTGGGATATTTTTAAAATATAT |
| JM313 | ggctaagttattAATAATGGCATAGTAAGCTCACGAGCGTTTGTTGAACTAATGG |
| JM314 | CGGATTTAGAGATACCAAATTAATATAGTCTTCGC |
| JM315 | TAGCAAAGGCTTTAAACGTTAACGAAGC |
| JM316 | GCGCTCTGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTAAATTTGTTTAGCAGG |
| JM317 | ATAGCTCTAAAACAGAGCGCAATTAATTATTGCGGATATTCCT |
| JM318 | TTTGAATGGTCCCAAAACTAAATTTGTTTAGCAGGTAAACCGTGCTTT |
| JM319 | ACAGCATAGCTCTAAAACAGAGCGCAATTAATTATTGCGGATATTCCT |
| JM320 | aaacGTGTTCTCTTCAATCCATTCATCTATTGCTg |
| JM321 | aaaacAGCAATAGATGAATGGATTGAAGAGAACAC |
| JM322 | aaacCGAATATACAGGAAGAATACACGATGTTGTg |
| JM323 | aaaacACAACATCGTGTATTCTTCCTGTATATTCG |
| JM324 | aaacGTTTGTGTATCAGCCATATTTTGCTTTAATg |
| JM325 | aaaacATTAAAGCAAAATATGGCTGATACACAAAC |
| JM326 | GTGCGATTACAAAATTTTTTAGACAAAAATAGTCTNNNNNGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACAGTGCCGAGG |
| JM327 | TTTTTAGACAAAAATAGTCTNNNNNGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACAGTGCCGAGGAAAAATTAGGTGCGCTTGGC |
| JM328 | cgagctctgataaatatgaacatGGATGAAGTGGCTAGTTTACAAGAGTATATATCATTG |
| JM329 | CACTTCATCCatgttcatatttatcagagctcgtgctataattatactaattttataagg |

| | |
|---|---|
| JM330 | TGTCATGATAATAATGGTTTCTTAGGATTCTGTGATTTGGATCCTTCCAGAAGTC |
| JM331 | CGAAAAGTGCCACCTGACGTAAATTACGAATCTTCTCCTGACGTTTTTTTAAATCTTG |
| JM332 | CAAGATTTAAAAAAACGTCAGGAGAAGATTCGTAATTTACGTCAGGTGGCACTTTTCG |
| JM333 | CTGGAAGGATCCAAATCACAGAATCCTAAGAAACCATTATTATCATGACATTAACC |
| JM334 | GGTTTCTTAGGATTCTGTGATTTGGATCCTTCCAGAAGTCTTTTTCTTGCACTG |
| JM335 | AAATTACGAATCTTCTCCTGACGTTTTTTTAAATCTTGACTGTTCATTC |
| JM336 | GAAGGAGATATACCATGGGCGCCAGCGTTAATAGATCTTGTACAGCAAGG |
| JM337 | CAGCCGGATCAAGTATACATTTGTTCTTAAAGCATTTTTTCCCATAAAAACAACC |
| JM338 | TTAAGAACAAATGTATACTTGATCCGGCTGCTAACAAAGCCC |
| JM339 | TAACGCTGGCGCCCATGGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGG |
| JM340 | GCGTTGCGCCATACTTGCACATTATCAAACAAAATCAATTATTAAAATAGATGTTTAAG |
| JM341 | GCGGTGGCAGCAGCCTAGGCCATCCCCTAGAAATTAATCAATGCG |
| JM342 | CGCATTGATTAATTTCTAGGGGATGGCCTAGGCTGCTGCCACCGC |
| JM343 | GATTTTGTTTGATAATGTGCAAGTATGGCGCAACGCAATTAATGTAAGTTAGCTC |
| JM344 | AAAATTTTTTAGACAAAAATctgagcatctGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGCGCTGGTTG |
| JM345 | agatgctcagATTTTTGTCTAAAAAATTTTGTAATCGCACTATTTGTCTCAGC |
| JM346 | AAAATTTTTTctcacccccgAGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGCGCTGGTTG |
| JM347 | CTCGTAGACTcgggggtgagAAAAAATTTTGTAATCGCACTATTTGTCTCAGCTAGACTT |
| JM348 | CccccgggggggAGACAAAAATAGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGCGCTGGTTG |
| JM349 | TCGTAGACTATTTTTGTCTccccccggggGTAATCGCACTATTTGTCTCAGCTAGACTTC |
| JM350 | gaccagacattacgaactggcacagatGATTCTGTGATTTGGATCCTTCCAGAAGTC |
| JM351 | gatcttccttcaggttatgaccAAATTACGAATCTTCTCCTGACGTTTTTTTAAATCTTG |
| JM352 | TAAAAAAACGTCAGGAGAAGATTCGTAATTTggtcataacctgaaggaagatctggatcc |
| JM353 | GACTTCTGGAAGGATCCAAATCACAGAATCatctgtgccagttcgtaatgtctggtc |
| JM354 | AAAATTTTTTAGACAAAAATctgagcatctGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGCGCTGGTTGATTTCTTCTTGCGCTTTTT |
| JM355 | GAGCTAACTTACATTAATTGCGTTGCGCacctctggtgaaggagttggc |
| JM356 | GCGGTGGCAGCAGCCTAGGgtttaaggttggtgtctttttacctgtttgaaaacaaag |
| JM357 | ctttgttttcaaacaggtaaaaaagacaccaaccttaaacCCTAGGCTGCTGCCACCGC |
| JM358 | gccaactccttcaccagaggtGCGCAACGCAATTAATGTAAGTTAGCTC |
| JM359 | aaacGTGTTCTCTTCAATCCATTCATCTATTGCTg |
| JM360 | aaaacAGCAATAGATGAATGGATTGAAGAGAACAC |
| JM361 | aaacGTTTGTGTATCAGCCATATTTTGCTTTAATg |
| JM362 | aaaacATTAAAGCAAAATATGGCTGATACACAAAC |
| JM363 | aaactcgcttcttctttactctcTAGGGATAACAg |
| JM364 | aaaacTGTTATCCCTAgagagtaaagaagaagcga |
| JM365 | aaacagaagctatatcctaaaatTAGGGATAACAg |
| JM366 | aaaacTGTTATCCCTAattttaggatatagcttct |
| JM367 | GTTAAGTATAAGAAGGAGATATACATatgtatctgctggtgagctttgatct |
| JM368 | CAGCAGttaaataaaaatcagtttcggttctttgttcacgg |
| JM369 | gaaccgaaactgatttttatttaaCTGCTGCCACCGCTGA |
| JM370 | acatATGTATATCTCCTTCTTATACTTAACTAATATACTAAGATGGGG |

| | |
|---|---|
| JM371 | TCACCACAGCatgaaagatgtgatctatgtggagaatcattattttgtg |
| JM372 | TTCTGTTCGAttaattgccaacttcaatcatcttcggc |
| JM373 | tgattgaagttggcaattaaTCGAACAGAAAGTAATCGTATTGTACACGG |
| JM374 | acatagatcacatctttcatGCTGTGGTGATGATGGTGATGG |
| JM375 | TAAGAAGGAGATATACCATGaaagatgtgatctatgtggagaatcattattttgtg |
| JM376 | tcacatctttCATGGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGG |
| JM377 | NNNNNatgAGTGCCGAGGAAAAATTAGGTGCGCTTGGC |
| JM378 | NNNNNatgAAAGCACGGTTTACCTGCTAAACAAATTTA |
| JM379 | NNNNNtacAGTGCCGAGGAAAAATTAGGTGCGCTTGGC |
| JM380 | NNNNNtacAAAGCACGGTTTACCTGCTAAACAAATTTA |
| JM381 | NNNNNctaAGTGCCGAGGAAAAATTAGGTGCGCTTGGC |
| JM382 | NNNNNctaAAAGCACGGTTTACCTGCTAAACAAATTTA |
| JM383 | CGCAGTATATTTTAGATGAAGATTATTTCTTAATAACTAAAAATATGGTATAA |
| JM384 | AAAAAGCGCAAGAAGAAATCAACCAGCGCA |
| JM385 | AGAAATAATCTTCATCTAAAATATACTGCGGAAAATCATATAGTTC |
| JM386 | aaaacctacagaaaacactaaattaataagaaagagccaaacctcgaaag |
| JM387 | ttaaaagctgttccaattgaaaccaaagtgctttctcttatctattattc |
| JM388 | taagagaaagcactttggtttcaattggaacagcttttaaaaagcaaatatgagcc |
| JM389 | ggctctttcttattaatttagtgttttctgtaggtttttaggcataaaactatatg |
| JM390 | aaacTgtggagaggCAATCCATTCATCTATTGCTg |
| JM391 | aaaacAGCAATAGATGAATGGATTGcctctccacA |
| JM392 | aaacTgggtgtgcgCAGCCATATTTTGCTTTAATg |
| JM393 | aaaacATTAAAGCAAAATATGGCTGcgcacacccA |
| JM394 | ATGCGTCCGGCGTAGA |
| JM395 | GATTATGCGGCCGTGTACAA |
| JM396 | cctttgatattaatgagaaaaaactgctgctgaatgtg |
| JM397 | gaacATGTCACCTAAGTCAACACCATCATTTTTTATTCTg |
| JM398 | cgatcAGAATAAAAAATGATGGTGTTGACTTAGGTGACAT |
| JM399 | aaacGTCACGAGGTTATTATGTACGAGTTAATCAg |
| JM400 | aaaacTGATTAACTCGTACATAATAACCTCGTGAC |
| JM401 | aaacTTTGACGCAATAGGTGCTATAACTAAAATGg |
| JM402 | aaaacCATTTTAGTTATAGCACCTATTGCGTCAAA |
| JM403 | aaacATGTAAACTTGTAATTAATCCGTAAGCCATg |
| JM404 | aaaacATGGCTTACGGATTAATTACAAGTTTACAT |
| JM405 | aaacTTATACTCTTGGTAAATCCTAGAATTGTTAg |
| JM406 | aaaacTAACAATTCTAGGATTTACCAAGAGTATAA |
| JM407 | aaacGTACATAATAACCTCGTGACCTTTTTCGTTg |
| JM408 | aaaacAACGAAAAAGGTCACGAGGTTATTATGTAC |
| JM409 | catatagttttatgcctaaaaacctacagatagtttaaaatcatttgttcaaaaat |
| JM410 | ggctcatatttgctttttaaaagctcactgtgaattattatcaaaacc |
| JM411 | ggctcatatttgctttttaaaagcttttagtccagctctaagatttg |

| | |
|---|---|
| JM412 | CAAAATTTTTTAGACAAAAATAGTCTNNNNNGTTTTAGAGCTATGCTGTTTTGAATGGTC |
| JM413 | AGTGCCGAGGAAAAATTAGGTGCGCTTGGCGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTGCGCTGGTTGATTTCTTCTTGCGCTTTTT |
| JM414 | GCCAAGCGCACCTAATTTTTCCTCGGCACTTTATACCATATTTTTAGTTATTAAGAAATA |
| JM415 | TCAATGAGTGGTATCCAAGACGAAAACTTATTTTGTTATCACAATTTTCGGTTGACATC |
| JM416 | GATGTCAACCGAAAATTGTGATAACAAAATAAGTTTTCGTCTTGGATACCACTCATTGA |
| JM417 | cgagctctgataaatatgaacatgaatatcctaaatcagttgtttcattttagttacc |
| JM418 | atcgcatccgattgcagtataaatttagtccagctctaagatttgaggtgaag |
| JM419 | cttcacctcaaatcttagagctggactaaatttatactgcaatcggatgcgat |
| JM420 | ggtaactaaaatgaaacaactgatttaggatattcatgttcatatttatcagagctcg |
| JM421 | catatagttttatgcctaaaaacctacagaATCACAAGCTGATGTATTATTTTTG |
| JM422 | ggctcatatttgctttttaaaagctGTCCCTTCTCACATTTCAATACTAG |
| JM423 | cgagctctgataaatatgaacatATCACAAGCTGATGTATTATTTTTG |
| JM424 | CTCGTAGGATATCTTTTCTACCAGA |
| JM425 | TCTGGTAGAAAAGATATCCTACGAGgtttttgtactctcaagatttaagtaactgtacaacatgatgatgaagtatcgtcatctactaac |
| JM426 | CAAAAATAATACATCAGCTTGTGATatgttcatatttatcagagctcg |
| JM427 | agttttatgcctaaaaacctacagagaatatcctaaatcagttgtttcattttagttacc |
| JM428 | ctcaaatgatttttgagattttcatacc |
| JM429 | ggtatgaaaatctcaaaaatcatttgagGTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAACGGTGAAAAAGGTTCACTGTACGAGTACTTA |
| JM430 | ggtttcggaatgaagtgaataacgttgtttag |
| JM431 | cagcttatcgttcagaagccttaagg |
| JM432 | gcaaagcaaatagttccatttttggaaaagg |
| JM433 | ggatgatgcgtggtctttccg |
| JM434 | gtcaaaattcctcatgtatcgtcacgac |
| JM435 | ggtaagaacagatgtcctaggaaatcagc |
| JM436 | gagagggacatgctgccaag |
| JM437 | gtgagacttttggcggttactgaaaag |
| JM438 | gaccatcttcttttacaaacaacagagaagtatc |
| JM439 | ctgatgaaaaactttacagaaatttttagaaagtaagg |
| JM440 | gaaaattttaatttgcatgtaactgggcagtg |
| JM441 | gtcgttaaaccgtgtgctctacgacc |
| JM442 | GCTATAACTAAAATGgttttagagctatgctgttttgaatggtcccaaaacttcagcaca |
| JM443 | ACCTATTGCGTCAAAgttttgggaccattcaaaacagcatagctctaaaacctcgtagac |
| JM444 | aaacTGTAATTGTCATACTTGATAAGTCTACTATg |
| JM445 | aaaacATAGTAGACTTATCAAGTATGACAATTACA |
| JM446 | aaaaatcatacgaggtttttgtactctcaagatttaagtaactgtacaacatgatgatga |
| JM447 | aaaaacctcgtatgatttttgagattttcatacccacatagtgattataa |
| JM448 | AGATATCCTttGAGgtttttgtactctcaagatttaagtaactgtacaacatgatgatga |
| JM449 | aaaaacCTCaaAGGATATCTTTTCTACCAGATTTCGTAATCGCACTATTAATC |
| JM450 | gtttttgtactctcaagatttaagtaactgtacaacagcttttaaaaagcaaatatgagc |
| JM451 | gttgtacagttacttaaatcttgagagtacaaaaacctcaaatgatttttgagattttca |
| JM452 | AGAGCTGTGTTGTTTCGAATGGTTCCAAAACTTTTGTTATCACAATTTTCGGTTGACATC |

| | |
|---|---|
| JM453 | ACCATTCGAAACAACACAGCTCTAAAACCTCGTAGGATATCTTTTCTACCAGATTTCGTA |
| JM454 | gataaatatgaacatTTTGGTAATAGGCATTATTACAAATG |
| JM455 | ttgcagtataaaCTACACCCTGTAGACTTTTTAATAAAATTG |
| JM456 | AAAGTCTACAGGGTGTAGtttatactgcaatcggatgcgat |
| JM457 | ATGCCTATTACCAAAatgttcatatttatcagagctcg |
| JM458 | GTTTCAACTTATTATAAGAAGCGGaacttttcactggagttgtcc |
| JM459 | ACCTCCCTTATAACCGCTTCTTATTTGTATAGTTCATCCATGCC |
| JM460 | ATACAAATAAGAAGCGGTTATAAGGGAGGTTTGTTAGTTGATAGC |
| JM461 | aaagttCCGCTTCTTATAATAAGTTGAAACTTCCGTCTATAG |
| JM462 | aaacCGCTATAGACGGAAGTTTCAACTTATTATAg |
| JM463 | aaaacTATAATAAGTTGAAACTTCCGTCTATAGCG |
| JM464 | GTTGGTCACAGATATGGAGTACAATCACTAATTTTG |
| JM465 | TCACGCATTGATGACATGAACGCTATAG |
| JM466 | cctgtccttttaccagacaaccattacc |
| JM467 | CTTTGTCGTTTGAAACAACCGAACTACC |
| JM468 | aaacGCTATAGACGGAAGTTTCAACTTATTATAAg |
| JM469 | aaaacTTATAATAAGTTGAAACTTCCGTCTATAGC |
| JM470 | aaacATAGACGGAAGTTTCAACTTATTATAAGGGg |
| JM471 | aaaacCCCTTATAATAAGTTGAAACTTCCGTCTAT |
| JM472 | tatgtgggtatgaaaatctcaaaa |
| JM473 | gttacttaaatcttgagagtacaaaaacA |
| JM474 | gttacttaaatcttgagagtacaaaaacT |
| JM475 | gttacttaaatcttgagagtacaaaaacG |
| JM476 | TAGTGCGATTACGAAATCTGG |
| JM477 | GAAACAACACAGCTCTAAAACa |
| JM478 | GAAACAACACAGCTCTAAAACt |
| JM479 | GAAACAACACAGCTCTAAAACg |
| JM480 | gggattattaaatgagtgacttagttcatatcatcaaaaatgaggg |
| JM481 | ccctcatttttgatgatatgaactaagtcactcatttaataatccc |
| JM482 | gaagaaatgacttggagagttaaattccactgcccag |
| JM483 | ctgggcagtggaatttaactctccaagtcatttcttc |
| JM484 | atgaaaatctcaaaaatcatttgagGTTTTAgtactctcaagatttaagTGGTTCCAAAACatgatgatgaagtatcgtcatctactaac |
| JM485 | ctcaaatgatttttgagattttcatacccac |
| JM486 | TCTGGTAGAAAAGATATCCTACGAGgtttttGAGCTGTGTTGTTTCGAAtaactgtacaacGGTGAAAAAGGTTCACTGTACGAGTACTTA |
| JM487 | CTCGTAGGATATCTTTTCTACCAGATTTCGTAATC |
| JM488 | atgatgaagtatcgtcatctactaacCCAGAAAACAagtcgttgaggcagagagta |
| JM489 | TCTGGgttagtagatgacgatacttcatcatcatgtttacgtcctgttgaatctttg |
| JM490 | GGTGAAAAAGGTTCACTGTACGAGTACTTATGGCGagtcgttgaggcagagagta |
| JM491 | ATAAGTACTCGTACAGTGAACCTTTTTCACCgtttacgtcctgttgaatctttg |
| JM492 | atgatgatgaagtatcgtcatctactaacGTTTTAgtactctcaagatttaagTGGTTCCAAAACcttcacctcaaatcttagagctggactaaa |
| JM493 | gttagtagatgacgatacttcatcatcatg |

| | |
|---|---|
| JM494 | GGTGAAAAAGGTTCACTGTACGAGTACTTAgtttttGAGCTGTGTTGTTTCGAAtaactgtacaacTCAATGAGTGGTATCCAAGACGAAAACTTA |
| JM495 | TAAGTACTCGTACAGTGAACCTTTTTCACC |
| JM496 | catgatgatgaagtatcgtcatctactaac |
| JM497 | GGAACCacttaaatcttgcagaagctacTaagataaggcttcatgccg |
| JM498 | AgtagcttctgcaagatttaagtGGTTCCgtaaggcgtcccttacactattattatag |
| JM499 | GTTGTTTCGAATaacTTtacACCACAATTATTATTATATAGTTTTATAACTAATATGTC |
| JM500 | GGTgtaAAgttATTCGAAACAACACAGCGAGTaAAAATAAGGCTTAGTCCGTACTCAAC |
| JM501 | GAGCTAACTTACATTAATTGCGTTGCGCTTAAATATTTACCTTTTTAATCAATGG |
| JM502 | GCGGTGGCAGCAGCCTAGGGGGTTATATGGTGGTTTATCC |
| JM503 | GGATAAACCACCATATAACCCCCTAGGCTGCTGCCACCGC |
| JM504 | CCATTGATTAAAAAGGTAAATATTTAAGCGCAACGCAATTAATGTAAGTTAGCTC |
| JM505 | GATTACGAAATCTGGTAGAAAAGATATCCTttGAGGTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAACGGTGAAAAAGGTTCACTGTACGAGTACTTA |
| JM506 | AGGATATCTTTTCTACCAGATTTCGTAATCGC |
| JM507 | ATCATTAATTAATAATAGGTTATGTTTAGAGATCGATACCCACCCCGAAGAAAAGGGGACGAGAACACGTATGCCGAAGTATATAAATCATCAGTACAAAG |
| JM508 | TCTAAACATAACCTATTATTAATTAATGATTTTTTAAGCCAGTCAC |
| JM509 | TGCTGAAGAGAGGAAAGCGTATCG |
| JM510 | CGATACGCTTTCCTCTCTTCAGCA |
| JM511 | GCTAGTAAACCGCCTCGCGCagcttttaaaaagcaaatatgagcc |
| JM512 | ggctcatatttgctttttaaaagctGCGCGAGGCGGTTTACTAGC |
| JM513 | TCTGGTAGAAAAGATATCCTACGAGgtttttGAGCTGTGTTGTTTCGAATGGTTCCAAAACGGTGAAAAAGGTTCACTGTACGAGTACTTA |
| JM514 | TCTGGTAGAAAAGATATCCTACGAGGTTTTAGAGCTGTGTTGTTTCGAAtaactgtacaacGGTGAAAAAGGTTCACTGTACGAGTACTTA |
| JM523 | GCGGATAACAATTAAGCTTGTACAATAAAAGGAGAAATGTTATTTGCC |
| JM524 | CCTAGGAGCTCTCGAGATCTACTTCCTCCTCTATCATAATAAATTTAAA |
| JM525 | TTTAAATTTATTATGATAGAGGAGGAAGTAGATCTCGAGAGCTCCTAGG |
| JM526 | GGCAAATAACATTTCTCCTTTTATTGTACAAGCTTAATTGTTATCCGC |
| JM527 | tgGTCTTCgtctcggaagctcaaGAAGACcGTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAACGATTGTGCCCGCTAGTAAA |
| JM528 | gGTCTTCttgagcttccgagacGAAGACcaGTTTTGGAACCATTCGAAACAACACAGCTCTAAAACCTCGTGGACCAATTTTCG |
| JM529 | GATTGAGAAAATATTGACTTTCCGTATTCCCTA |
| JM530 | TAGGGAATACGGAAAGTCAATATTTTCTCAATC |
| JM531 | atgaaaatctcaaaaatcatACgagGTTTTAgtactctcaagatttaagTGGTTCCAAAACatgatgatgaagtatcgtcatctactaac |
| JM532 | ctcGTatgatttttgagattttcatacccac |
| JM533 | TCTGGTAGAAAAGATATCCTttGAGgtttttGAGCTGTGTTGTTTCGAAtaactgtacaacGGTGAAAAAGGTTCACTGTACGAGTACTTA |
| JM534 | CTCaaAGGATATCTTTTCTACCAGATTTCGTAATC |
| JM535 | gagagggacatgctgccAGAgtctactttaatgagctctttggtaaac |
| JM536 | gtttaccaaagagctcattaaagtagacTCTggcagcatgtccctctc |
| JM537 | caaacatctacaatctgtctaaatg |
| JM538 | GTGAGGGTCACTCTGCCaagATTTATTTTAATACGCTTTTCGGTA |
| JM539 | TACCGAAAAGCGTATTAAAATAAATcttGGCAGAGTGACCCTCAC |
| JM540 | AAACctcgtTGGTTGATTTCTTCTTGCGCTTTTTg |
| JM541 | aaaacAAAAAGCGCAAGAAGAAATCAACCAacgag |
| JM542 | AAACagatgTGGTTGATTTCTTCTTGCGCTTTTTg |

| | |
|---|---|
| JM543 | aaaacAAAAAGCGCAAGAAGAAATCAACCAcatct |
| JM544 | tggttgacgacttgatggCgccatttagacagattgtagatg |
| JM545 | catctacaatctgtctaaatggcGccatcaagtcgtcaacca |
| JM546 | ATTTAACCTTGCTAGTGATATTATGGcGCCATTTAGGCCAATTATAGA |
| JM547 | TCTATAATTGGCCTAAATGGCgCCATAATATCACTAGCAAGGTTAAAT |
| JM548 | aatcatttgaggtttttgtactctcaagatttaagtaactgtacaacatgatgatgaagtatcgtcatctacta |
| JM549 | tagtagatgacgatacttcatcatcatgttgtacagttacttaaatcttgagagtacaaaaacctcaaatgatt |
| JM550 | TATCCTACGAGGTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAACGGTGAAAAAGGTTCACTGTACGAGTAC |
| JM551 | GTACTCGTACAGTGAACCTTTTTCACCGTTTTGGAACCATTCGAAACAACACAGCTCTAAAACCTCGTAGGATA |
| JM552 | tagtataataaagcattttaacattatacttttgataatcgtttatcgtc |
| JM553 | gactaggttgtctcagcatttttgtcaataaataaggtgtcacagt |
| JM554 | actgtgacaccttatttattgacaaaaatgctgagacaacctagtc |
| JM555 | gattatcaaaagtataatgttaaaatgctttattatactaacgttatata |
| JM556 | atttctttttttcttgtaaaaaaagaaagttcttaaaggttttatagtttt |
| JM557 | CTCTATCAAATCCTTATAATCTACTCTCTAACAAAATGATTATCCCATTCAA |
| JM558 | TTGAATGGGATAATCATTTTGTTAGAGAGTAGATTATAAGGATTTGATAGAG |
| JM559 | aaacctttaagaactttctttttttacaagaaaaaagaaattagataaat |
| JM560 | catatagttttatgcctaaaaacctacagaTAGAGATAAAAGAGTCCTTTGGATG |
| JM561 | CAAAAATAATACATCAGCTTGTGATAACTTTGTATATTCCAAGATTTGATAAA |
| JM562 | TTTATCAAATCTTGGAATATACAAAGTTATCACAAGCTGATGTATTATTTTTG |
| JM563 | CATCCAAAGGACTCTTTTATCTCTAtctgtaggtttttaggcataaaactatatg |
| JM564 | cattacgaactggcacagattagtttaaaatcatttgttcaaaaataaaatcc |
| JM565 | ggattttatttttgaacaaatgattttaaactaatctgtgccagttcgtaatg |
| JM566 | ggtccattcactattctcattccctttcagataattttagatttgcttttcta |
| JM567 | tagaaaagcaaatctaaaattatctgaaaagggaatgagaatagtgaatggacc |
| JM568 | CTATCAAATCCTTATAATCTACTCTCTAACAAAATGATTATCCCATTCAACTTCA |
| JM569 | cattacgaactggcacagatTAGAGATAAAAGAGTCCTTTGGATG |
| JM570 | CATCCAAAGGACTCTTTTATCTCTAatctgtgccagttcgtaatg |
| JM571 | ggtccattcactattctcattccctttcagataattttagatttgcttttcta |
| JM572 | tagaaaagcaaatctaaaattatctgaaaagggaatgagaatagtgaatggacc |
| JM573 | CAATCGTATGACAAATTATGATaaGTACTTGCCAAATCAAAAAGTTCTTC |
| JM574 | CAAATAATTCCTTTTTAAAAGATTTTAATTAATTATATTATACC |
| JM575 | GGTATAATATAATTAATTAAAATCTTTTAAAAAGGAATTATTTG |
| JM576 | TTTTGATTTGGCAAGTACttATCATAATTTGTCATACGATTGATAAAAGC |
| JM577 | GATTCTGTGATTTGGATCCTTCCAGAAGTCTTTTTCTTGC |
| JM578 | GCAAGAAAAAGACTTCTGGAAGGATCCAAATCACAGAATC |
| JM579 | actggcacagatAAACACTAAATTAATAAGAAAGAGCCAAACCTCGAAAG |
| JM580 | ggatccagatcttccttcaggttatgaccGCGCGAGGCGGTTTACTAGC |
| JM581 | GCTAGTAAACCGCCTCGCGCggtcataacctgaaggaagatctggatcc |
| JM582 | TCTTTCTTATTAATTTAGTGTTTatctgtgccagttcgtaatgtctggtc |

**Supplementary DNA sequences (5'-3')**

>pJM37gw (Swap spacers 2 and 4)
GGTCTCAAAACTGCGCTGGTTGATTTCTTCTTGCGCTTTTTGTTTTAGAGCTA
TGCTGTTTTGAATGGTCCCAAAACAGTGCCGAGGAAAAATTAGGTGCGCTTG
GCGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACAGGAATATCCGCAA
TAATTAATTGCGCTCTGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACT
TATATGAACATAACTCAATTTGTAAAAAAGTTTTAGAGCTATGCTGTTTTGAAT
GGTCCCAAAACTAAATTTGTTTAGCAGGTAAACCGTGCTTTGTTTTAGAGCTA
TGCTGTTTTGAATGGTCCCAAAACTTCAGCACACTGAGACTTGTTGAGTTCC
ATGTTTTAGAGACC

>pJM73gw (LAS Duplication in Spacer 4)
GGTCTCaaaacTGCGCTGGTTGATTTCTTCTTGCGCTTTTTGTTTTAGAGCTAT
GCTGTTTTGAATGGTCCCAAAACTTATATGAACATAACTCAATTTGTAAAAAA
GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACAGGAATATCCGCAATA
ATTAATTGCGCTCTGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACAAA
ATTTTTTAGACAAAAATAGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATG
GTCCCAAAACTAAATTTGTTTAGCAGGTAAACCGTGCTTTGTTTTAGAGCTAT
GCTGTTTTGAATGGTCCCAAAACTTCAGCACACTGAGACTTGTTGAGTTCCA
Tgttttagagacc

>pJM87gw (Position 5 phage targeting spacer)
GGTCTCAAAACTGCGCTGGTTGATTTCTTCTTGCGCTTTTTGTTTTAGAGCTA
TGCTGTTTTGAATGGTCCCAAAACTTATATGAACATAACTCAATTTGTAAAAA
AGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACAGGAATATCCGCAAT
AATTAATTGCGCTCTGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACAG
TGCCGAGGAAAAATTAGGTGCGCTTGGCGTTTTAGAGCTATGCTGTTTTGAA
TGGTCCCAAAACGTGTTCTCTTCAATCCATTCATCTATTGCTGTTTTAGAGCT
ATGCTGTTTTGAATGGTCCCAAAACTAAATTTGTTTAGCAGGTAAACCGTGCT
TTGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTTCAGCACACTGAGA
CTTGTTGAGTTCCATGTTTTAGAGACC

>pJM258
aaaatgctgagacaacctagtctctcacttgttgattcagtaatattgaatatcctaaatcagttgtttcattttagttac
cgtataagatatttacaaaaatctgatgaaaaacttttacagaaatttttagaaagtaaggattgacaagaacagt
tattgattttataatcactatgtgggtatgaaaatctcaaaaatcatttgaggtttttgtactctcaagatttaagtaactg
tacaacATGAAAATCACAGGACAAGCGCAATTTACTgtttttgtactctcaagatttaagtaact
gtacaacTGATATTGTAGAATACTTCTTTAACTATTAgtttttgtactctcaagatttaagtaact
gtacaacggtcataacctgaaggaagatctggatcctaatgaattcatctgcaggggcccatcagtctgacgac
caagagagccataaacaccaatagccttaacatcatccccatatttatccaatattcgttccttaatttcatgaaca
atcttcattctttcttctctagtcattattattggtccattcactattctcattcc

>pJM259

aaaatgctgagacaacctagtctctcacttgttgattcagtaatattgaatatcctaaatcagttgtttcattttagttac
cgtataagatatttacaaaaatctgatgaaaaacttttacagaaatttttagaaagtaaggattgacaagaacagt
tattgattttataatcactatgtgggtatgaaaatctcaaaaatcatttgaggtttttgtactctcaagatttaagtaactg
tacaacACACAATGTATACATCATGGCAGACAAACAgtttttgtactctcaagatttaagtaact
gtacaacTAGGGTCTAATCGAAAAGCAGGCGTCTAGAgtttttgtactctcaagatttaagtaa
ctgtacaacTCTTAAACAAGCCACAATTCGATCAGCCAAgtttttgtactctcaagatttaagta
actgtacaactgagaccagtctcggaagctcaaaggtctcgtttttgtactctcaagatttaagtaactgtacaacA
TGAAAATCACAGGACAAGCGCAATTTACTgtttttgtactctcaagatttaagtaactgtacaac
TGATATTGTAGAATACTTCTTTAACTATTAgtttttgtactctcaagatttaagtaactgtacaac
ggtcataacctgaaggaagatctggatcctaatgaattcatctgcaggggcccatcagtctgacgaccaagaga
gccataaacaccaatagccttaacatcatccccatatttatccaatattcgttccttaatttcatgaacaatcttcattct
ttcttctctagtcattattattggtccattcactattctcattcc

>pJM261
TTAGAGAGTAGATTATAAGGATTTGATAGAGGAGGAATTAAGTTGCTTGACA
TATGATTATTAAGAAATAATCTAATATGGTGACAGTCACATCTTGTCTAAAAC
GTTGATATATAAGGATTTTTAAGGTATAATAAATATAAAAATGGAATTATTTTG
AAGCTGAAGTCATGCTGAGATTAATAGTGCGATTACGAAATCTGGTAGAAAA
GATATCCTACGAGGTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAACTTAA
TACAGGTTTTTACAAAAGCTTTACCATGTTTTAGAGCTGTGTTGTTTCGAATG
GTTCCAAAACCAGAACACGCAAGAGTGCAAACGTTAGCTCGTTTTAGAGCTG
TGTTGTTTCGAATGGTTCCAAAACggtcataacctgaaggaagatctggatcctaatgaattcat
ctgcaggggcccatcagtctgacgaccaagagagccataaacaccaatagccttaacatcatccccatatttat
ccaatattcgttccttaatttcatgaacaatcttcattctttcttctctagtcattattattggtccattcactattctcattcc

>pJM262
TTAGAGAGTAGATTATAAGGATTTGATAGAGGAGGAATTAAGTTGCTTGACA
TATGATTATTAAGAAATAATCTAATATGGTGACAGTCACATCTTGTCTAAAAC
GTTGATATATAAGGATTTTTAAGGTATAATAAATATAAAAATGGAATTATTTTG
AAGCTGAAGTCATGCTGAGATTAATAGTGCGATTACGAAATCTGGTAGAAAA
GATATCCTACGAGGTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAACTGAT
GTTAATGGGCACAAATTTTCTGTCAGGTTTTAGAGCTGTGTTGTTTCGAATGG
TTCCAAAACCCACATGCAATCGCAGCCTTTGTACTCGCAGTTTTAGAGCTGT
GTTGTTTCGAATGGTTCCAAAACTATCCCAGCTTTTGACGCAACGCTCGGAA
AGTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAACtgagaccagtctcggaagctca
aaggtctcGTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAACTTAATACAGGT
TTTTACAAAAGCTTTACCATGTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAA
AACCAGAACACGCAAGAGTGCAAACGTTAGCTCGTTTTAGAGCTGTGTTGTT
TCGAATGGTTCCAAAACggtcataacctgaaggaagatctggatcctaatgaattcatctgcagggg
cccatcagtctgacgaccaagagagccataaacaccaatagccttaacatcatccccatatttatccaatattcgt
tccttaatttcatgaacaatcttcattctttcttctctagtcattattattggtccattcactattctcattcc

## REFERENCES


1 Thomas, C. M. & Nielsen, K. M. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* **3**, 711-721, doi:10.1038/nrmicro1234 (2005).

2 Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712, doi:10.1126/science.1138140 (2007).

3 Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843-1845, doi:10.1126/science.1165771 (2008).

4 Amitai, G. & Sorek, R. CRISPR-Cas adaptation: insights into the mechanism of action. *Nat. Rev. Microbiol.* **14**, 67-76, doi:10.1038/nrmicro.2015.14 (2016).

5 Jackson, S. A. *et al.* CRISPR-Cas: Adapting to change. *Science* **356**, doi:10.1126/science.aal5056 (2017).

6 Sternberg, S. H., Richter, H., Charpentier, E. & Qimron, U. Adaptation in CRISPR-Cas Systems. *Mol. Cell* **61**, 797-808, doi:10.1016/j.molcel.2016.01.030 (2016).

7 Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* **37**, 67-78, doi:10.1016/j.mib.2017.05.008 (2017).

8 Krupovic, M., Beguin, P. & Koonin, E. V. Casposons: mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery. *Curr. Opin. Microbiol.* **38**, 36-43, doi:10.1016/j.mib.2017.04.004 (2017).

9 Nunez, J. K. *et al.* Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat. Struct. Mol. Biol.* **21**, 528-534, doi:10.1038/nsmb.2820 (2014).

10 Nunez, J. K., Harrington, L. B., Kranzusch, P. J., Engelman, A. N. & Doudna, J. A. Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature* **527**, 535-538, doi:10.1038/nature15760 (2015).

11 Wang, J. *et al.* Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell* **163**, 840-853, doi:10.1016/j.cell.2015.10.008 (2015).

12 Wright, A. V. *et al.* Structures of the CRISPR genome integration complex. *Science* **357**, 1113-1118, doi:10.1126/science.aao0679 (2017).

13 Xiao, Y., Ng, S., Nam, K. H. & Ke, A. How type II CRISPR-Cas establish immunity through Cas1-Cas2-mediated spacer integration. *Nature* **550**, 137-141, doi:10.1038/nature24020 (2017).

14 Wright, A. V. & Doudna, J. A. Protecting genome integrity during CRISPR immune adaptation. *Nat Struct Mol Biol* **23**, 876-883, doi:10.1038/nsmb.3289 (2016).

15 Nunez, J. K., Lee, A. S., Engelman, A. & Doudna, J. A. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* **519**, 193-198, doi:10.1038/nature14237 (2015).
16 Arslan, Z., Hermanns, V., Wurm, R., Wagner, R. & Pul, U. Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. *Nucleic Acids Res.* **42**, 7884-7893, doi:10.1093/nar/gku510 (2014).
17 Ivancic-Bace, I., Cass, S. D., Wearne, S. J. & Bolt, E. L. Different genome stability proteins underpin primed and naive adaptation in E. coli CRISPR-Cas immunity. *Nucleic Acids Res.* **43**, 10821-10830, doi:10.1093/nar/gkv1213 (2015).
18 Yosef, I., Goren, M. G. & Qimron, U. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res* **40**, 5569-5576, doi:10.1093/nar/gks216 (2012).
19 McGinn, J. & Marraffini, L. A. CRISPR-Cas Systems Optimize Their Immune Response by Specifying the Site of Spacer Integration. *Mol Cell* **64**, 616-623, doi:10.1016/j.molcel.2016.08.038 (2016).
20 Weinberger, A. D. *et al.* Persisting viral sequences shape microbial CRISPR-based immunity. *PLoS Comput Biol* **8**, e1002475, doi:10.1371/journal.pcbi.1002475 (2012).
21 Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602-607, doi:10.1038/nature09886 (2011).
22 Elmore, J. R. *et al.* Programmable plasmid interference by the CRISPR-Cas system in *Thermococcus kodakarensis*. *RNA Biol* **10**, 828-840, doi:10.4161/rna.24084 (2013).
23 Nickel, L. *et al.* Two CRISPR-Cas systems in *Methanosarcina mazei* strain Go1 display common processing features despite belonging to different types I and III. *RNA Biol* **10**, 779-791, doi:10.4161/rna.23928 (2013).
24 Randau, L. RNA processing in the minimal organism *Nanoarchaeum equitans*. *Genome Biol* **13**, R63, doi:10.1186/gb-2012-13-7-r63 (2012).
25 Richter, H. *et al.* Characterization of CRISPR RNA processing in *Clostridium thermocellum* and *Methanococcus maripaludis*. *Nucleic Acids Res.* **40**, 9887-9896, doi:10.1093/nar/gks737 (2012).
26 Jones, D. L. *et al.* Kinetics of dCas9 target search in Escherichia coli. *Science* **357**, 1420-1424, doi:10.1126/science.aah7084 (2017).
27 Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62-67, doi:10.1038/nature13011 (2014).
28 Kupczok, A., Landan, G. & Dagan, T. The contribution of genetic recombination to CRISPR array evolution. *Genome Biol Evol* **7**, 1925-1939, doi:10.1093/gbe/evv113 (2015).

29 Tyson, G. W. & Banfield, J. F. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ Microbiol* **10**, 200-207, doi:10.1111/j.1462-2920.2007.01444.x (2008).
30 Richter, C. *et al.* Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res.* **42**, 8516-8526, doi:10.1093/nar/gku527 (2014).
31 Rollie, C., Schneider, S., Brinkmann, A. S., Bolt, E. L. & White, M. F. Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *Elife* **4**, doi:10.7554/eLife.08716 (2015).
32 Nunez, J. K., Bai, L., Harrington, L. B., Hinder, T. L. & Doudna, J. A. CRISPR Immunological Memory Requires a Host Factor for Specificity. *Mol. Cell* **62**, 824-833, doi:10.1016/j.molcel.2016.04.027 (2016).
33 Rollie, C., Graham, S., Rouillon, C. & White, M. F. Prespacer processing and specific integration in a Type I-A CRISPR system. *Nucleic Acids Res.* **46**, 1007-1020, doi:10.1093/nar/gkx1232 (2018).
34 Chylinski, K., Makarova, K. S., Charpentier, E. & Koonin, E. V. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* **42**, 6091-6105, doi:10.1093/nar/gku241 (2014).
35 Wei, Y., Chesne, M. T., Terns, R. M. & Terns, M. P. Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in Streptococcus thermophilus. *Nucleic Acids Res* **43**, 1749-1758, doi:10.1093/nar/gku1407 (2015).
36 Levy, A. *et al.* CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **520**, 505-510, doi:10.1038/nature14302 (2015).
37 Paez-Espino, D. *et al.* Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nat Commun* **4**, 1430, doi:10.1038/ncomms2440 (2013).
38 Bikard, D., Hatoum-Aslan, A., Mucida, D. & Marraffini, L. A. CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host Microbe* **12**, 177-186, doi:10.1016/j.chom.2012.06.003 (2012).
39 Edgar, R. & Qimron, U. The *Escherichia coli* CRISPR system protects from lambda lysogenization, lysogens, and prophage induction. *J Bacteriol* **192**, 6291-6294, doi:10.1128/JB.00644-10 (2010).
40 Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233-239, doi:10.1038/nbt.2508 (2013).
41 Modell, J. W., Jiang, W. & Marraffini, L. A. CRISPR-Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. *Nature* **544**, 101-104, doi:10.1038/nature21719 (2017).

42 Dillingham, M. S. & Kowalczykowski, S. C. RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol. Mol. Biol. Rev.* **72**, 642-671, Table of Contents, doi:10.1128/MMBR.00020-08 (2008).
43 Yeeles, J. T., van Aelst, K., Dillingham, M. S. & Moreno-Herrero, F. Recombination hotspots and single-stranded DNA binding proteins couple DNA translocation to DNA unwinding by the AddAB helicase-nuclease. *Mol. Cell* **42**, 806-816, doi:10.1016/j.molcel.2011.04.012 (2011).
44 Fagerlund, R. D. *et al.* 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, doi:10.1073/pnas.1618421114 (2017).
45 Künne, T. *et al.* Cas3-derived target DNA degradation fragments fuel primed CRISPR adaptation. *Mol. Cell*, doi:10.1016/j.molcel.2016.07.011 (2016).
46 Musharova, O. *et al.* Spacer-length DNA intermediates are associated with Cas1 in cells undergoing primed CRISPR adaptation. *Nucleic Acids Res.* **45**, 3297-3307, doi:10.1093/nar/gkx097 (2017).
47 Babu, M. *et al.* A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. *Mol Microbiol* **79**, 484-502, doi:10.1111/j.1365-2958.2010.07465.x (2011).
48 Heler, R. *et al.* Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* **519**, 199-202, doi:10.1038/nature14245 (2015).
49 Hynes, A. P., Villion, M. & Moineau, S. Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages. *Nat Commun* **5**, 4399, doi:10.1038/ncomms5399 (2014).
50 Heler, R. *et al.* Mutations in Cas9 Enhance the Rate of Acquisition of Viral Spacer Sequences during the CRISPR-Cas Immune Response. *Mol Cell* **65**, 168-175, doi:10.1016/j.molcel.2016.11.031 (2017).
51 Hoyland-Kroghsbo, N. M. *et al.* Quorum sensing controls the *Pseudomonas aeruginosa* CRISPR-Cas adaptive immune system. *Proc Natl Acad Sci U S A* **114**, 131-135, doi:10.1073/pnas.1617415113 (2017).
52 Patterson, A. G. *et al.* Quorum sensing controls adaptive immunity through the regulation of multiple CRISPR-Cas systems. *Mol. Cell*, doi:10.1016/j.molcel.2016.11.012 (2016).
53 Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol* **190**, 1390-1400, doi:10.1128/JB.01412-07 (2008).
54 Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733-740, doi:10.1099/mic.0.023960-0 (2009).
55 Elmore, J. R. *et al.* Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system. *Genes Dev.* **30**, 447-459, doi:10.1101/gad.272153.115 (2016).

56 Marraffini, L. A. & Sontheimer, E. J. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**, 568-571, doi:10.1038/nature08703 (2010).
57 Pyenson, N. C., Gayvert, K., Varble, A., Elemento, O. & Marraffini, L. A. Broad Targeting Specificity during Bacterial Type III CRISPR-Cas Immunity Constrains Viral Escape. *Cell Host Microbe* **22**, 343-353 e343, doi:10.1016/j.chom.2017.07.016 (2017).
58 Goldberg, G. W., Jiang, W., Bikard, D. & Marraffini, L. A. Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature* **514**, 633-637, doi:10.1038/nature13637 (2014).
59 Samai, P. *et al.* Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell* **161**, 1164-1174, doi:10.1016/j.cell.2015.04.027 (2015).
60 Deng, L., Garrett, R. A., Shah, S. A., Peng, X. & She, Q. A novel interference mechanism by a type IIIB CRISPR-Cmr module in *Sulfolobus*. *Mol. Microbiol.* **87**, 1088-1099, doi:10.1111/mmi.12152 (2013).
61 Peng, W., Feng, M., Feng, X., Liang, Y. X. & She, Q. 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, doi:10.1093/nar/gku1302 (2015).
62 Silas, S. *et al.* Type III CRISPR-Cas systems can provide redundancy to counteract viral escape from type I systems. *elife* **6**, doi:10.7554/eLife.27601 (2017).
63 Silas, S. *et al.* Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science* **351**, aad4234, doi:10.1126/science.aad4234 (2016).
64 Fineran, P. C. *et al.* Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E1629-1638, doi:10.1073/pnas.1400071111 (2014).
65 Rollins, M. F. *et al.* Cas1 and the Csy complex are opposing regulators of Cas2/3 nuclease activity. *Proc Natl Acad Sci U S A* **114**, E5113-E5121, doi:10.1073/pnas.1616395114 (2017).
66 Westra, E. R. *et al.* CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol Cell* **46**, 595-605, doi:10.1016/j.molcel.2012.03.018 (2012).
67 Redding, S. *et al.* Surveillance and Processing of Foreign DNA by the *Escherichia coli* CRISPR-Cas System. *Cell* **163**, 854-865, doi:10.1016/j.cell.2015.10.003 (2015).
68 Blosser, T. R. *et al.* Two Distinct DNA Binding Modes Guide Dual Roles of a CRISPR-Cas Protein Complex. *Mol. Cell* **58**, 60-70, doi:10.1016/j.molcel.2015.01.028 (2015).
69 Xue, C. *et al.* CRISPR interference and priming varies with individual spacer sequences. *Nucleic Acids Res.* **43**, 10831-10847, doi:10.1093/nar/gkv1259 (2015).

70 Xue, C., Whitis, N. R. & Sashital, D. G. Conformational control of cascade interference and priming activities in CRISPR immunity. *Mol. Cell* **64**, 826-834, doi:10.1016/j.molcel.2016.09.033 (2016).
71 Semenova, E. *et al.* 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, doi:10.1073/pnas.1602639113 (2016).
72 Staals, R. H. *et al.* Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. *Nat Commun* **7**, 12853, doi:10.1038/ncomms12853 (2016).
73 Severinov, K., Ispolatov, I. & Semenova, E. The influence of copy-number of targeted extrachromosomal genetic elements on the outcome of CRISPR-Cas defense. *Front Mol Biosci* **3**, 45, doi:10.3389/fmolb.2016.00045 (2016).
74 Datsenko, K. A. *et al.* Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat Commun* **3**, 945, doi:10.1038/ncomms1937 (2012).
75 Swarts, D. C., Mosterd, C., van Passel, M. W. & Brouns, S. J. CRISPR interference directs strand specific spacer acquisition. *PLoS One* **7**, e35888, doi:10.1371/journal.pone.0035888 (2012).
76 Shmakov, S. A. *et al.* The CRISPR spacer space is dominated by sequences from species-specific mobilomes. *MBio* **8**, doi:10.1128/mBio.01397-17 (2017).
77 Shmakov, S. *et al.* Diversity and evolution of class 2 CRISPR-Cas systems. *Nat. Rev. Microbiol.* **15**, 169-182, doi:10.1038/nrmicro.2016.184 (2017).
78 Silas, S. *et al.* On the Origin of Reverse Transcriptase-Using CRISPR-Cas Systems and Their Hyperdiverse, Enigmatic Spacer Repertoires. *MBio* **8**, doi:10.1128/mBio.00897-17 (2017).
79 Mohanraju, P. *et al.* Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* **353**, aad5147, doi:10.1126/science.aad5147 (2016).
80 Shipman, S. L., Nivala, J., Macklis, J. D. & Church, G. M. Molecular recordings by directed CRISPR spacer acquisition. *Science*, doi:10.1126/science.aaf1175 (2016).
81 Shipman, S. L., Nivala, J., Macklis, J. D. & Church, G. M. CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature* **547**, 345-349, doi:10.1038/nature23017 (2017).
82 Sheth, R. U., Yim, S. S., Wu, F. L. & Wang, H. H. Multiplex recording of cellular events over time on CRISPR biological tape. *Science*, doi:10.1126/science.aao0958 (2017).
83 Brouns, S. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960-964, doi:10.1126/science.1159689 (2008).

84 Carte, J., Wang, R., Li, H., Terns, R. M. & Terns, M. P. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev* **22**, 3489-3496, doi:10.1101/gad.1742908 (2008).
85 Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* **109**, E2579-2586, doi:10.1073/pnas.1208507109 (2012).
86 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).
87 Jore, M. M. *et al.* Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* **18**, 529-536, doi:10.1038/nsmb.2019 (2011).
88 Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67-71, doi:10.1038/nature09523 (2010).
89 Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* **13**, 722-736, doi:10.1038/nrmicro3569 (2015).
90 Shmakov, S. *et al.* Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol. Cell* **60**, 385-397, doi:10.1016/j.molcel.2015.10.008 (2015).
91 Jansen, R., Embden, J. D., Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43**, 1565-1575 (2002).
92 Kreiswirth, B. N. *et al.* The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage. *Nature* **305**, 709-712 (1983).
93 Chylinski, K., Le Rhun, A. & Charpentier, E. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biol.* **10**, 726-737, doi:10.4161/rna.24321 (2013).
94 Wiedenheft, B. *et al.* Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure* **17**, 904-912, doi:10.1016/j.str.2009.03.019 (2009).
95 Wei, Y. & Terns, M. P. CRISPR Outsourcing: Commissioning IHF for Site-Specific Integration of Foreign DNA at the CRISPR Array. *Mol. Cell* **62**, 803-804, doi:10.1016/j.molcel.2016.06.004 (2016).
96 Murphy, K., Travers, P., Walport, M. & Janeway, C. *Janeway's immunobiology*. (Garland Science, 2012).
97 Andersson, A. F. & Banfield, J. F. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320**, 1047-1050, doi:10.1126/science.1157358 (2008).
98 Sun, C. L. *et al.* Phage mutations in response to CRISPR diversification in a bacterial population. *Environ. Microbiol.* **15**, 463-470, doi:10.1111/j.1462-2920.2012.02879.x (2013).

99 Marraffini, L. A. CRISPR-Cas immunity in prokaryotes. *Nature* **526**, 55-61, doi:10.1038/nature15386 (2015).
100 McGinn, J. & Marraffini, L. A. Molecular mechanisms of CRISPR-Cas spacer acquisition. *Nat Rev Microbiol*, doi:10.1038/s41579-018-0071-7 (2018).
101 Rao, C., Chin, D. & Ensminger, A. W. Priming in a permissive type I-C CRISPR-Cas system reveals distinct dynamics of spacer acquisition and loss. *RNA* **23**, 1525-1538, doi:10.1261/rna.062083.117 (2017).
102 Barrett, R. D. & Schluter, D. Adaptation from standing genetic variation. *Trends Ecol Evol* **23**, 38-44, doi:10.1016/j.tree.2007.09.008 (2008).
103 Van Orden, M. J., Klein, P., Babu, K., Najar, F. Z. & Rajan, R. Conserved DNA motifs in the type II-A CRISPR leader region. *PeerJ* **5**, e3161, doi:10.7717/peerj.3161 (2017).
104 Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935-949, doi:10.1016/j.cell.2014.02.001 (2014).
105 Achigar, R., Magadan, A. H., Tremblay, D. M., Julia Pianzzola, M. & Moineau, S. Phage-host interactions in Streptococcus thermophilus: Genome analysis of phages isolated in Uruguay and ectopic spacer acquisition in CRISPR array. *Sci Rep* **7**, 43438, doi:10.1038/srep43438 (2017).
106 Jiang, W. *et al.* Dealing with the evolutionary downside of CRISPR immunity: bacteria and beneficial plasmids. *PLoS Genet* **9**, e1003844, doi:10.1371/journal.pgen.1003844 (2013).
107 Bergkessel, M., Basta, D. W. & Newman, D. K. The physiology of growth arrest: uniting molecular and environmental microbiology. *Nat Rev Microbiol* **14**, 549-562, doi:10.1038/nrmicro.2016.107 (2016).
108 Pearl, S., Gabay, C., Kishony, R., Oppenheim, A. & Balaban, N. Q. Nongenetic individuality in the host-phage interaction. *PLoS Biol* **6**, e120, doi:10.1371/journal.pbio.0060120 (2008).
109 Kutter, E. & Sulakvelidze, A. *Bacteriophages : biology and applications*. (CRC Press, 2005).
110 Sheth, R. U. & Wang, H. H. DNA-based memory devices for recording cellular events. *Nat Rev Genet* **19**, 718-732, doi:10.1038/s41576-018-0052-8 (2018).
111 Jiang, W. & Marraffini, L. A. CRISPR-Cas: New Tools for Genetic Manipulations from Bacterial Immunity Systems. *Annu Rev Microbiol* **69**, 209-228, doi:10.1146/annurev-micro-091014-104441 (2015).
112 Hatoum-Aslan, A. CRISPR Methods for Nucleic Acid Detection Herald the Future of Molecular Diagnostics. *Clin Chem*, doi:10.1373/clinchem.2018.295485 (2018).
113 Horinouchi, S. & Weisblum, B. Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. *J Bacteriol* **150**, 815-825 (1982).

114 Horinouchi, S. & Weisblum, B. Nucleotide sequence and functional map of pE194, a plasmid that specifies inducible resistance to macrolide, lincosamide, and streptogramin type B antibodies. *J Bacteriol* **150**, 804-814 (1982).
115 Perez-Casal, J., Caparon, M. G. & Scott, J. R. Mry, a trans-acting positive regulator of the M protein gene of *Streptococcus pyogenes* with similarity to the receptor proteins of two-component regulatory systems. *J. Bacteriol.* **173**, 2617-2624 (1991).
116 Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343-345, doi:10.1038/nmeth.1318 (2009).
117 Moore, S. D. & Prevelige, P. E., Jr. A P22 scaffold protein mutation increases the robustness of head assembly in the presence of excess portal protein. *J. Virol.* **76**, 10245-10255 (2002).
118 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
119 Notredame, C., Higgins, D. G. & Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205-217, doi:10.1006/jmbi.2000.4042 (2000).
120 Grissa, I., Vergnaud, G. & Pourcel, C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**, 172, doi:10.1186/1471-2105-8-172 (2007).
121 Rousseau, C., Gonnet, M., Le Romancer, M. & Nicolas, J. CRISPI: a CRISPR interactive database. *Bioinformatics* **25**, 3317-3318, doi:10.1093/bioinformatics/btp586 (2009).
122 Jiang, W., Samai, P. & Marraffini, L. A. Degradation of Phage Transcripts by CRISPR-Associated RNases Enables Type III CRISPR-Cas Immunity. *Cell* **164**, 710-721, doi:10.1016/j.cell.2015.12.053 (2016).