ADAPTATION BY THE TYPE III CRISPR-CAS SYSTEM OF

*STREPTOCOCCUS THERMOPHILUS*

by

Xinfu Zhang

(Under the Direction of Michael P. Terns)

ABSTRACT

CRISPR-Cas systems provide their prokaryotic hosts with an adaptive and efficient defense mechanism against invading nucleic acids, and are structurally and functionally classified into six types (Type I-VI). Type III CRISPR-Cas systems are distinguished from the other types of systems by their special requirements and features of target interference. However, our knowledge about adaptation by the Type III systems is very limited. To provide a detailed study to determine the specific properties and patterns of adaptation by the Type III systems, we examined the adaptation by the Type III-A system of *Streptococcus thermophilus*, and the adaptation by the Type II-A system in the same host as a comparison. Unlike Type II systems and some Type I systems, deletion of genes involved in Type III crRNA biogenesis or interference did not disrupt adaptation nor detectably change spacer uptake patterns except those related to counter-selection. No PAM was observed with the Type III system. The lengths of Type III-A spacers were on average longer than the Type II-A spacers: 36 bp. Interestingly, certain regions of plasmids and the host genome were particularly

well-sampled during Type III-A, but not Type II-A, spacer uptake. These regions included the single-strand origins of rolling-circle replicating plasmids, rRNA and tRNA encoding clusters, and promoter regions of expressed genes. We also found that the Type III-A system could adapt and protect the cell from a lytic phage. Collectively, this work indicates that the Type III adaptation machinery preferentially targets DNA secondary structures including imperfect hairpins and other partially double-stranded DNAs.

Primed adaptation provides CRISPR-Cas systems with an important co-evolutionary strategy to minimize the escaping invading nucleic acids. Primed adaptation by a Type III system had never been reported. In this dissertation, preliminary evidence of Type III primed adaptation was obtained. Moreover, by studying primed adaptation of the Type I-E CRISPR-Cas system in *Streptococcus thermophilus*, we established a novel method for the future primed adaptation studies, based on natural CRISPR-Cas escaping viruses.

INDEX WORDS: CRISPR; Cas; Type III; adaptation; primed adaptation; *Streptococcus thermophilus*; phage; plasmid.

# ADAPTATION BY THE TYPE III CRISPR-CAS SYSTEM OF *STREPTOCOCCUS THERMOPHILUS*

by

Xinfu Zhang

B.S., Beijing Forestry University, China, 2011

M.S., Clemson University, 2014

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2021

# ADAPTATION BY THE TYPE III CRISPR-CAS SYSTEM OF *STREPTOCOCCUS THERMOPHILUS*

by

Xinfu Zhang

Major Professor: Michael P. Terns
Committee: Robert Sabatini
Christopher West
Robert J. Schmitz

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
August 2021

## DEDICATION

This dissertation is dedicated to every member of my family, every of my friends, my teachers, my universities, and my country, China.

ACKNOWLEDGEMENTS

I thank my country, China, my friends, and all my family members for love and support. I thank Beijing Forestry University, Clemson University, and The University of Georgia for my education.

I thank my committee members, Dr. Michael Terns, Dr. Robert Sabatini, Dr. Robert J. Schmitz, and Dr. Christopher West, and the previous committee members, Dr. Jianfu Chen, Dr. Rebecca Terns, and Dr. Claiborne Glover. I am grateful to their insightful guidance and nice encouragements, without which the work would not be finished.

I thank Dr. Sandra Garrett at Dr. Brenton Graveley's lab at Department of Genetics and Genome Sciences, Institute for Systems Genomics, University of Connecticut Health Center. Dr. Sandra Garrett is a distinguished researcher and an outstanding bioinformatic technician. She nicely helped with the analyses of high-throughput sequencing data in this dissertation.

## ABBREVIATIONS

1. CRISPR: Clustered regularly interspaced short palindromic repeats.
2. Cas: CRISPR-associated.
3. *Sth*: *Streptococcus thermophilus*.
4. crRNA: CRISPR RNA.
5. crRNP complex: crRNA and Cas protein complex.
6. PAM: protospacer adjacent motif.
7. PFS: protospacer flanking region.
8. dsDNA: double-stranded DNA.
9. ssDNA: single-stranded DNA.
10. RCR: rolling-circle replication.
11. *dso*: double-strand origin.
12. *sso*: single-strand origin.
13. $RS_B$: recombination site in *sso*.
14. CS-6: conserved 6 nucleotide sequence in *sso*.
15. RT: reverse transcriptase.
16. HTS: high-throughput sequencing.
17. LTR: long terminal repeat.
18. LAB: lactic acid bacteria.
19. GRAS: generally recognized as safe.
20. CDM: chemically defined medium.

21. orf: open reading frame.

22. WT: wildtype.

23. LM17: M17 medium supplemented with 0.5% lactose.

24. M.O.I.: multiplicity of infection.

25. PTS: phosphotransferase system.

26. BIM: bacteriophage insensitive mutant.

27. CEM: CRISPR-Cas escaping mutant.

28. cAn: cyclic oligoadenylates.

29. PPS: priming protospacer.

30. CFU: colony forming unit.

TABLE OF CONTENTS

# CHAPTER 1

# INTRODUCTION AND LITERATURE REVIEW

Eubacteria and archaea are always threatened by viruses (1). They are also constantly exposed to environmental plasmids, and the non-beneficial plasmids result in fitness cost for their hosts (2). As a consequence, prokaryotic cells evolved multiple strategies to defend themselves against foreign nucleic acids, including abortive infection, restriction-modification systems (3), and more recently discovered and extensively studied CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas (CRISPR-associated gene) systems (4-6).

CRISPR-Cas systems provide the hosts with an adaptive mechanism against invading nucleic acids (7-12). The specific defense activity of a CRISPR-Cas system is acquired by adaptation step, during which a short fragment (protospacer) of the foreign DNA is captured and integrated into the CRISPR locus at the leader proximal end as a spacer, simultaneously with a duplication of the first repeat (7,13-16). A spacer in a CRISPR array can be expressed as a small CRISPR RNA (crRNA) and then guide an interference protein or a protein complex (crRNP) to destroy the previously encountered foreign nucleic acids from which it was originally derived (7,17,18) (Figure 1.1). CRISPR-Cas systems

are structurally and functionally diverse, and have been classified into six types (Type I-VI) and multiple subtypes (5,6,8,19-21).

Functional studies of CRISPR-Cas systems, especially those regarding target interference, have inspired researchers to develop many unprecedented, convenient and powerful tools for genome editing, gene expression control, disease detection and cures, and many other purposes (22). Adaptation abilities of CRISPR-Cas systems and the dynamic CRISPR arrays they generated have been used for bacterial strain typing (23), bacterial virome detection (24), and even digital movie encoding and data storage (25). The tremendous contribution of CRISPR-Cas systems to biotechnology makes their fundamental studies invaluable, especially those investigating adaptation, since it is the least understood process of CRISPR-Cas functions.

## Adaptation by CRISPR-Cas systems

All CRISPR-Cas systems function in three major steps: adaptation, crRNA biogenesis, and target interference (Figure 1.1). The heritable defense memories against foreign nucleic acids are acquired during adaptation, which includes protospacer recognition, processing, and integration into a CRISPR array at the leader-proximal end (15,26,27).

### *Integration by Cas1-Cas2 complex*

A CRISPR array is usually associated with *cas* genes, and each Cas protein participates in one or more major steps of the CRISPR-Cas system-mediated

defense, including adaptation, crRNA biogenesis, and target interference (6,14). Among the *cas* genes, *cas1* and *cas2* are highly conserved in nearly all the CRISPR loci, and have been shown not to be involved in crRNA biogenesis or the downstream target interference (7,14). Instead, Cas1 and Cas2 have been shown to be essential for adaptation of all tested CRISPR-Cas systems, in that knockout of the two genes or mutations of the endonuclease domain of Cas1 results in the studied systems losing their adaptation abilities (7,28-33). Mutations of the enzymatic domain of *E. coli* Cas2 does not influence the adaptation, indicating that Cas2 plays a structural role in this process, at least in that Type I system of *E. coli* (29). Cas1 and Cas2 proteins form a hexamer (four Cas1 monomers centered by two Cas2 monomers) both *in vivo* and *in vitro* (29,34-36) (Figure 1.2). Through the aid of this complex, the 3'-OH groups of the two strands of the prespacer (protospacer processed for integration) successively attack the junctions between the leader and the first repeat, and between the first repeat and the first pre-existing spacer (37,38) (Figure 1.2). By transesterification reactions, Cas1-Cas2 complex integrates the double-stranded prespacer into the CRISPR array, splitting the plus and the minus strand of the first repeat, and leaving two gaps (37,38). DNA polymerase(s) and ligase(s) are thought to be required to fill the gap and finish the whole process (Figure 1.2). Since DNA polymerase I has been shown required for the Type I adaptation in *E. coli* (39), it is proposed to be the polymerase that fills the integration gap.

For some of CRISPR-Cas systems, e.g., the Type I-E system in *E. coli* K12, Cas1 and Cas2 are the only two Cas proteins required for adaptation

(28,29,40,41); while for some other systems of different types, other Cas proteins are also required for protospacer recognition or processing (7,30,31,42). Besides Cas proteins, the leader sequence and at least one repeat unit (28,43-45), and some other host factors are also required to ensure the integration to happen at the correct position (32,36,46-49).

***The recognition, selection, and processing of the proper protospacers***

For well-studied Type I and Type II CRISPR-Cas systems, the protospacers are selected along foreign DNAs by system-specific protospacer adjacent motifs (PAMs) (Figure 1.1) (50-53). PAM recognition is also required for the authentication of the interference process (50,51,54,55), by which the crRNP complexes of Type I and Type II systems can protect the CRISPR loci (containing the same sequence as the target) within its own genome from interference (Figure 1.1). The Cas1-Cas2 complex of the Type I-E system in *E. coli* K12 is sufficient to recognize the ATG PAM upstream of protospacers (28,29,40,41); while some other Type I systems require Cas4 to recognize PAM sequences (42,56,57). The Cas9 protein of the Type II system of *Streptococcus pyogenes* contains a PAM binding motif and performs PAM recognition to select the proper protospacers (31).

Besides PAMs, some other specific sequences also facilitate or discourage adaptation, and influence the patterns of adaptation, although at different levels. In 2019, Heler et al. found that PAM-proximal sequences of the protospacers determined the frequency of adaptation by the Type II systems in *Streptococcus*

*pyogenes* (58). Additionally, Li et al. found that in the Type I-B system of *Haloarcula hispanica*, the 3rd nucleotide at 3'- end of the spacers tended to be cytosine, and the presence/absence of the cytosine influenced the sizes of the new spacers (59). Results such as these suggest that specific features of protospacers, or protospacer adjacent sequences play important roles in the selection of new spacers.

RecBCD complexes and their homologous protein complexes in prokaryotic cells bind to double-stranded DNA (dsDNA) breaks, and then use their helicase and nuclease activities to repair the broken DNAs by degradation and homologous recombination (60). RecBCD complexes have been shown required for adaptation of some tested Type I systems, and the helicase activity of the enzymes was essential, while the nuclease activities and the homologous recombination they mediated were redundant (39,61,62). Since dsDNA breaks largely happen during DNA replication, the plasmids with high copy numbers and all extensively replicating invaders become more sensitive than the cellular genome to adaptation (39,61). Moreover, RecBCD can be hampered by *chi* sequences (60), and the enrichment of the *chi* sites around the replication termini of the prokaryotic genomes helps the adaptation machineries to more specifically recognize foreign DNAs (39,61).

Besides DNA replication, other parts of the life cycles of invaders also establish the patterns of adaptation of the CRISPR-Cas systems. In 2017, Modell et al. found that the Type II CRISPR-Cas system of *Staphylococcus aureus* immediately acquired new spacers after phage injection, and most of the new

spacers were acquired within the *cos* site, which was the first part of the phage to be injected into the cells (63). In my work here, I found that the Type III CRISPR-Cas system of *Streptococcus thermophilus* efficiently adapts against the *single-strand origin*s (*sso*s) of rolling-circle replicating (RCR) plasmids, the beginning of the highly transcribed genes of both the invader and the host, the rRNA and tRNA encoding strands, and the early transcribed genes of a lytic phage, to ensure efficient defenses (see Chapter 2).

Cas4 is a RecB-like nuclease (64,65), and has been shown to recognize PAM and determine the length and the orientation of the new spacers for some of the Type I CRISPR-Cas systems (42,56,57,66,67). A structural study of the adaptation machinery of a Type I-D system revealed a two-step procedure of prespacer capture by the adaptation complex (68). In the absence of a prespacer, Cas4 but not Cas2 assembles with Cas1 to recognize and process the protospacers; and in the presence of a mature prespacer, Cas2 but not Cas4 binds to Cas1 with a strikingly higher affinity to form the complex for the following integration (68). However, another *in vitro* investigation of a Type I-C system suggested that Cas4 formed a complex with both Cas1 and Cas2, and then processed the protospacers into mature prespacers for the integration (66).

The processing of the protospacers from the long substrates to the short and mature prespacers is a prerequisite of adaptation, but it is the least understood step of the adaptation process. The existence of 3'- single-stranded DNA (ssDNA) tails of the prespacers substantially facilitate adaptation (16,37,38,47,69,70). Cas1 is a non-specific exonuclease *in vitro* when associated

with Cas2 in the adaptation complex (62,71-73), and it trims 5' ends of the protospacers, leaving 3'- ssDNA tails for the following integration (32,53). In *Streptococcus thermophilus* (*Sth*), Cas2 of the Type I-E system possesses a DnaQ-like 3'-5' exonuclease domain, which has been proposed to process the 3'-overhangs of the prespacers to promote integration (74). Some other non-Cas exonucleases have also been shown to be involved in the 3'- ssDNA tail generation of the prespacers (75).

***Primed adaptation***

Target nucleic acids can escape from CRISPR-Cas-mediated interference by mutation(s) at pivotal positions within the protospacers/targets or the PAMs (76,77) (Figure 1.3). However, a pre-existing spacer in a CRISPR array, which is partially or totally complementary to a fragment of a molecule, can greatly stimulate adaptation against the same molecule (41,78). To acquire new spacers from a molecule that the system has never processed before is termed 'naïve adaptation', whereas adaptation triggered by a pre-existing spacer (priming spacer) is termed 'primed adaptation' (Figure 1.3). Primed adaptation is substantially more efficient than naïve adaptation (79), and directs the adaptation machinery to the invader DNA instead of self-genome (41), thus providing the hosts with a co-evolutionary strategy to minimize the amount of CRISPR-Cas escapers. Different studies have revealed that primed adaptation has strand or position biases. For the Type I-E system in *E. coli*, a priming spacer stimulates adaptation against the primed strand (41,78); while for the other tested Type I

systems, the secondarily acquired spacers were distributed across both primed and non-primed strands with an obvious gradient centered at the target/protospacer of the pre-existing spacer (41,78-83). While Cas1 and Cas2 are sufficient for Cas elements in naïve adaptation of Type I-E system (28), all tested primed adaptation processes of different systems require all the Cas proteins involved in crRNA biogenesis and target interference (41,78-80,83-85).

Although primed adaptation has been extensively studied only for Type I systems, bioinformatic analyses of the clustering of spacers acquired from viruses suggests that primed adaptation and the consequential host-viral coevolution probably exists in some Type II systems as well (86,87). In 2019, Nussenzweig et al. reported primed adaptation by a Type II-A system of *Streptococcus pyogenes*, which increased the efficiency of the secondary adaptation against the phage by Cas9-mediated DNA breaks, and the new protospacers acquired during priming located to the immediate vicinity of the target (88). In 2021, a second report regarding primed adaptation by the Type II-A system of *Streptococcus mutans* revealed the same distribution pattern of the secondary protospacers (89).

While the outcome of primed adaptation (extensive spacer uptake) is easily observed, the mechanism(s) of primed adaptation are still under research and debate. Several very possible functional models have been proposed, including: 1) The DNA breaks or short DNA fragments caused by target interference (even if DNA cleavage is very inefficient due to mutations of the escaping targets) facilitate adaptation (90-94); 2) Single molecule studies have revealed that the

DNA interference complex can recruit Cas1-Cas2 complex to the targeted invaders, which is thought to facilitate adaptation (95); 3) The crRNPs and the critical subunits acquire different conformations when binding to a *bona fide* target compared to an escaping target, and consequently authenticate target interference or primed adaptation, separately (96,97). It is worth noting that the models do not conflict with each other, and they may all be real mechanisms of primed adaptation. In *E. coli*, Cas3 and RecG, a helicase that dissipates R-loops, are required for primed adaptation to resolve invader DNA blocks by the crRNP, in order to enable adaptation (39).

### *Adaptation by Type III CRISPR-Cas systems*

While Type I, Type II and Type V systems target DNAs (7,54,98-101), and Type VI systems target single-strand RNAs (ssRNAs) (102-105), Type III systems have been shown to have both DNA and RNA cleavage abilities both *in vivo* and *in vitro* (106-114) (Figure 1.4). DNA target interference by Type III systems requires the directional transcription of the target, as the DNase activity of the crRNPs is stimulated by base pairing between the guiding crRNAs and the transcript of the target DNAs (107,112,115-118) (Figure 1.4). Additionally, the Palm domain of Cas10 (Csm1 for Type III-A) synthesizes cyclic oligoadenylates (cAns) as secondary messengers, which bind to CARF domain of Csx1 (Csm6 for Type III-A) and activates the RNase activity of HEPN domain of Csx1 to non-specifically cleave the foreign DNA transcripts (119-121) (Figure 1.4). In *Staphylococcus epidermidis*, the non-specific RNase activity of Csm6 is

redundant for defense when the target is extensively transcribed, and Csm6 hampers the growth of the bacteria when active; however, when the target is weakly transcribed, the DNase activity of Csm1 appears insufficient to cleave the invader, and consequentially, Csm6 becomes necessary for invader clearance in this situation (122). The DNase activity of Csm1 is sequence-non-specific as well (107,112,115-118), which not only cleaves the invading DNAs, but also promotes the mutagenesis of the host genome and increases the diversity of the host strains (123). Since the RNase activity of Csx1 and the DNase activity of Cas10 are non-specific, there is a risk to the prokaryotic cells that Type III systems may degrade host nucleic acids as well. Type III crRNPs specifically cleave the transcribed triggering RNAs of targets (110,114), which is reasoned to be a regulating mechanism of the ribonuclease activities of Cas10 and Csx1 (124,125). Moreover, the CARF domain and the HEPN domain of Csx1 degrade the cAns, which has been shown a regulatory mechanism of Csx1 when the defense against the invader is complete (126). While PAM recognition is required for authentication of the interference process of Type I and Type II systems (50,51,54,55), target interference by Type III systems tolerates a broad range of protospacer flanking region (PFS) sequences (112,127,128). Type III systems have been used for genomic editing of the hyperthermophiles, in which the CRISPR-Cas9-based modules might not function (129), and the target RNA cleavage abilities of Type III systems have been used for gene expression control (113,130).

Although Type III systems are intriguing, our knowledge about their adaptation is very limited. In 2016, Silas et al. reported the Type III-B system of *Marinomonas mediterranea* (MMB-1), revealing a novel reverse transcriptase (RT)-fused-Cas1 protein (33). While the reported RT-free systems can only adapt DNAs as CRISPR spacers, the Type III-B system can use both RNAs and DNAs as substrates, and adaptation against RNAs is dependent on the RT. This additional adaptation against RNAs makes the system preferentially acquire new spacers from highly transcribed regions versus weakly transcribed regions, which is beneficial for the function of the system, since target interference by Type III systems requires transcription of the targets (33). Soon after this exciting finding, a similar RT-Cas1-Cas2 complex of *Fusicatenibacter saccharivorans* was used as a novel and efficient tool to record transcription event in *E. coli* (131). A similar RT-mediated Type III adaptation against highly transcribed regions was reported by Gonzalez-Delgado et al. in 2019, and moreover, they observed a preference against the coding strand of the rRNA genes (132). They speculated that the rRNA-encoding strand preference was also caused by RT and there was a correlation between gene transcription and new spacer orientation (132). However, since adaptation against the other genes had no strand bias, it appears less likely that the bias was caused by transcription and RT activity. The results of my work presented in this dissertation consider the RT-free Type III-A CRISPR-Cas system of *Sth*, and show a similar trend toward rRNA and other highly transcribed genes. However, my findings indicate that it is more likely that the coding strand of the rRNA genes forms secondary structures when in their

single-stranded forms (e.g., when the template strand is being processed by RNA polymerase), which serves as additional substrates for CRISPR-Cas adaptation (See Chapter 2).

The reported RT-encoding Type III systems are not representative, because less than 10% of Type III systems have RT activity (33). Very recently, Artamonoka et al. observed and reported adaptation against a lytic phage by a RT-free Type III system of *Thermus thermophilus* (133). They found that the system performed robust adaptive defense against a lytic phage, phiFa. The protospacers detected by high-throughput sequencing (HTS) had a strand bias in that the template strands of the phage were adapted more extensively than the encoding strands, which was caused by counter-selection, since the crRNAs of the Type III system needed to bind to the mRNAs of the phages to be functional (Figure 1.4). More interestingly, they found that the long terminal repeat (LTR), as the firstly invading part and early transcribed region of the phage, was adapted substantially more efficiently than the other parts of the phage, and they reasoned that maybe the LTR region encoded an anti-CRISPR element that blocked the functions of the CRISPR-Cas system (133). While not inconsistent with the data, it is more likely that the LTR formed secondary structures since it was a repeat-rich region, including palindromic, direct, and inverted repeats, and such structures could be recognized by Type III CRISPR-Cas system (see my findings in Chapter 2); or only adaptation against the early transcribed genes could perform timely defense against the phage (see my findings in Chapter 2).

As to well-studied Type I and Type II systems, the protospacers are selected along foreign DNAs by system-specific PAMs, and are inserted at the leader proximal end of the array in a PAM directed orientation (50-53); while the adaptation by the tested Type III systems are PAM-independent (33,132,133) (See also my findings in Chapter 2).

## Genetics of *Streptococcus thermophilus*

Lactic acid bacteria (LAB) have been used for thousands of years in human history to produce dairy and other fermented products, and they are regarded as at 'generally recognized as safe' (GRAS) status of worldwide food safety administrations. *Streptococcus thermophilus* (*Sth*), a low-GC gram-positive bacterial species, is the second most used LAB (second only to *Lactococcus lactis*), as a starter for the production of yogurt and a large variety of cheeses (134,135). *Sth* produces not only lactic acids, but also bio-functional molecules, which increase the nutrition and immune values of milk products (136). As many as $10^{21}$ *Sth* cells are ingested by human beings annually, with a market value of 40 billion dollars (137). Besides its tremendous industrial and commercial values, *Sth* is also a well-studied and extensively used model gram-positive species to study protein function, vaccine activity, and cellular metabolisms (138).

### *Metabolism of Streptococcus thermophilus*

To adapt its fitness to growth in milk, *Sth* has experienced reductive evolutions of its genome, primarily in the decay of its sugar metabolism, protein

transport, and pathogenic genes (139,140). It has also acquired genes for stress tolerance, phage immunity, and acetaldehyde and exopolysaccharide production which benefits the flavor and nutrition of the milk products (141). *Sth* strains have lost the abilities to catabolize a variety of carbohydrates, and preferentially use milk-enriched lactose for glycolysis (140). *Sth* strains encode PrtS or other cell wall associated serine proteases to catabolize casein, which is the major nitrogen source in milk, into oligopeptides (142-144). *Sth* uses ABC transporters, Ami and OTS systems to uptake the oligopeptides (145), and digests them into free amino acids (although *Sth* has few amino acid requirement) (140). This proteolysis is valuable for mild acidification and flavor improvement of milk products. *Sth* has a symbiotic relationship with another LAB, *Lactococcus bulgaricus*, in that *Sth* uptakes amino acids and oligopeptides produced by *Lactococcus bulgaricus*, while *Lactococcus bulgaricus* uptakes the formic, folic and pyruvic acids, carbon dioxide, glutathione and long chain fatty acids produced by *Sth* (143,146). This is termed as protocooperation, which contributes to fast growth of the cells and the acidification of milk.

### *Natural transformation of Streptococcus thermophilus*

To use *Sth* for better industrial production and biological studies, the development of genetic tools has been highly desired. Traditionally, the approach of genome editing for *Sth* was plasmid-based homologous recombination (147,148). However, this traditional way suffered from low efficiency of recombination, reverse mutation to wild type (WT) by secondary recombination

events, and huge laboratory efforts for plasmid construction and curing. Natural competence is a transient physiological status, during which prokaryotic cells can uptake environmental DNAs. Natural competence of *Sth* can be induced to a very high level, allowing genome editing by antibiotic selection-free and non-replicating DNA fragments of the homologous recombinant templates (149,150), thus avoiding the difficulties of plasmid construction and curing.

Natural competence status of most organisms is usually transient and tightly regulated (151), to optimize the benefits and the costs of lateral gene transfer. Natural competence of *Sth* is initiated by competence-related late *com* genes, which are not expressed under standard growth conditions. ComX is a sigma factor ($\sigma^X$) that transiently associates with the RNA polymerase, and directs the later to specifically target ComX-box. ComX-box is a protein binding motif located in the promoter of *com* genes, and ComX-directed RNA polymerase binding to it triggers the expression of *com* genes (151). The expression of *comX* gene is regulated by ComRS signaling system (152). In chemically defined medium (CDM), which lacks oligopeptides, the ComS precursor is produced, matured, and secreted outside the cell membranes by an unknown mechanism (153,154). When the mature and short ComS with its C-terminal pheromone domain is accumulated to a certain concentration in the extracellular environment, the cells uptake the mature ComS oligopeptides back to the cytoplasm (150,154-156). The re-imported ComS forms a dimer with a transcription activator, ComR, and ComRS complex facilitates the transcription of *comX* and induces natural competence status of the *Sth* cells (152). The transcription of *comR* is

downregulated by cell density (152,157,158), and the transcription of *comX* is downregulated by the accumulation of the late *com* gene products it stimulated (159), thus providing negative feedback regulation of natural competence.

Under laboratory conditions, this system can be artificially induced by adding extracellular ComS. If transforming DNAs such as substrates for homologous recombination, efficient genome editing can be achieved without antibiotic selection or any plasmid (See Chapter 2 and 3).

### *Plasmids of Streptococcus thermophilus*

*Sth* must balance the cost of harboring plasmids with the selective advantages they can bestow, e.g., antibiotic resistance and phage immunity conferred by natural environmental plasmids. Most natural plasmids described in gram-positive bacteria, including *Sth*, and many of those of gram-negative bacteria, replicate via a rolling-circle replication (RCR) mechanism (160). Here, a Rep protein encoded by a RCR plasmid recognizes the short and partially palindromic double-stranded origin (*dso*) of the plasmid, generates a dsDNA nick at the *dso*, and displaces the 5'- end of the plus strand. The plus strand is then continuously replicated, a process which is initiated by the cognate 3'-OH end of the parental plus strand. The plus strand is displaced and re-ligated after the complete replication of the minus strand (Figure 1.5). The single-stranded origin (*sso*) is partially palindromic and comprises a long hairpin structure when the plus strand is in the single-stranded and circular DNA form. This hairpin structure acts as a signal to initiate replication of the minus strand (161-163) (Figure 1.5).

This replication initiation of the minus DNA strand requires the synthesis of short RNA primers by RNA polymerase III, similar to lagging strand replication of prokaryotic and eukaryotic genomes (163). Single stranded origins are diverse in sequence, and are classified into 5 types according to their secondary structures and consensus motifs, including *ssoA*, *ssoL*, *ssoT*, *ssoU*, and *ssoW* (163). The loop distal regions of the stems of the *sso*s are termed recombination sites ($RS_{BS}$), and are required for RNA polymerase binding, together with promoters within the *sso*s (163-165). There is usually a conserved 6 nucleotide sequence (CS-6) existing in the loop of a *sso*, which terminates the short RNA primer synthesis and is important for the minus DNA strand replication (163,166,167).

pWV01 is a RCR plasmid originally isolated from *Streptococcus cremoris* (168), and has been developed as a shuttle vector between *E. coli* and many *Lactococcus* and *Streptococcus* species (169). pWV01 has been observed to generate ssDNA forms in *Bacillus subtilis* and *Lactococcus lactis*, which is the major difference between RCR plasmids and theta-replicating plasmids, indicating that it replicates via a RCR mechanism (169). pWV01 has 4 open reading frames (*orf*s): *orfA* (*repA* gene) encodes the Rep protein, *orfC* product regulates the expression of *repA* and is required for replication of pWV01 (169), *orfB* and *orfD* do not appear necessary for any known plasmid function and may not be expressed (169).The partially palindromic *dso* and the DNA nicking site within pWV01 were identified by a simulation assay with a well-studied RCR plasmid, pLS1 (169). The long and partially palindromic *sso* was identified by the studies with a series of deletion derivatives, and the stem-loop structure was

predicted *in silico* (169,170). The *sso* of pWV01 belongs to the *ssoW* family and shows similarities with the conversion signal of phage φX174 (170). The $RS_B$ site and the promoter for RNA primer synthesis, as well as the CS-6 sequence for the termination of the short RNA oligo synthesis were localized within the *ssoW* by sequence alignment (166). pWAR (171), pTRK882 (172), pNZ123 (173) are derivatives of pWV01 and maintain its replication elements, and they are used in my work to determine the adaptation features of the Type III-A CRISPR-Cas system (see Chapter 2).

pSMQ172 is a RCR plasmid first isolated from *Sth* (174). The partially palindromic *dso* and the DNA nicking site within pSMQ172 were identified by a simulation assay with well-studied RCR plasmids, including pMV158, pFX2, and pE194 (174). The *sso* of pSMQ172 belongs to the *ssoA* family, encoding a well conserved $RS_B$ site and the promoter for RNA primer synthesis, but not any known CS-6 sequence for the termination of short RNA oligo synthesis (174). pSMQ172 has 4 *orf*s, and putative promoters are recognized upstream of *orf*1, *orf*3, and *orf*4 (174). *orf*1 and *orf*2 products are a plasmid copy number control protein and a Rep protein, respectively (174). *orf*3 is homologous to *mob* genes, and the associating *oriT* is present closely upstream of it (174). During conjugal transfer, MOB acts as a relaxase that binds to the transfer origin, *oriT*, and generates a dsDNA nick within it. The nicked DNA strand then dissociates from the circular ssDNA, and is injected into the recipient cell by a Type IV secretion system of gram-positive bacteria (175,176). Like other *oriT*s, the *oriT* of pSMQ172 has inverted repeats, and its nicking site is predicted by sequence

similarity (174). Although no secretion system has been identified in the plasmid or *Sth* genome (176), the *oriT* of pSMQ172 is predicted to form a crucial structure, and MOB may be able to generate the dsDNA break at *oriT* even in the absence of other conjugation machineries. *orf4* is not homologous to any known gene (174). pNT1 was constructed by inserting the gene encoding chloramphenicol acetyltransferase into pSMQ172 (177), and is used in my work to determine the adaptation features of the Type III-A CRISPR-Cas system (see Chapter 2).

In contrast to RCR, plasmid theta replication does not require any DNA break or ssDNA formation. Theta replication continuously synthesizes the leading strand of the plasmid, starting from the replication origins and discontinuously synthesizes the lagging strand in a manner similar to that of genome replication. There are 4 different classes of the theta replication, from Class A to D (178). Theta replicating plasmids are rare in gram-positive bacteria, however, Class D theta replicating plasmids, including pAMβ1 plasmid isolated from *Enterococcus faecalis*, are able to replicate in a broad range of gram-positive bacteria. pAMβ1 plasmid encodes a RepE protein, which specifically binds to an AT-rich region of the *pAMβ1 ori* immediately downstream of the *repE* gene, and processes the RNA primer for DNA replication (179-181). The replication of the leading strand and the lagging strand are unidirectional, and require DNA polymerase I and PriA as the initiating replisomes (181,182). pIB184 (183), pRSNPed (184), and pG+off (185) are derivatives of pAMβ1 and maintain its replication elements, and are

used in my work to determine the adaptation features of the Type III-A CRISPR-Cas system (see Chapter 2).

### *Virulent phage 2972 of Streptococcus thermophilus*

Phages of LAB are some of the most extensively investigated bacteriophages as they cause lactic acid fermentation failures and tremendous economic loss in human history. All Streptococcal phages belong to Siphoviridae family of *Caudovirales* order, and the majority can be divided into *cos*-type and *pac*-type according to their dsDNA packaging mechanisms and structural proteins (186). The other several Streptococcal phages identified recently are classified into three smaller independent groups, because of some differences in their morphologies, although they share significant homology with the other Streptococcal phages in the two major groups (187-189).

Phage 2972 is a *pac*-type virulent *Sth* phage isolated from yogurt. It has an isometric head capsid with the diameter of 55 nm, and a noncontractile tail of 260 nm (190). It has the smallest genome among *Sth* phages (37,704 bp), with 44 putative *orf*s involved in packaging, head morphogenesis, tail morphogenesis, host lysis (functional)/lysogeny (not functional), replication, or transcription regulation (190). Phage 2972 has a 34 mins latent period in *Sth* hosts cultured in LM17 broth at 42$^{o}$C, and starts releasing new progeny by 30 mins (191). The maximum burst happens at 40 mins, indicating that it is able to adsorb on the host efficiently and complete the life cycle rapidly (191). The 44 genes are divided into early, middle, and late groups, according to their expression starting

times, and the genes of each group are clustered together. Genes 30-44 along with Gene 1 are early genes, which are involved in phage DNA replication and transcription control; Genes 2-12 are middle genes, which are involved in packaging and head morphogenesis; and Genes 13-29 are late genes, which are involved in tail morphogenesis and host lysis (191).

Phage 2972 was used extensively in this dissertation as a means to challenge the Type III CRISPR-Cas system of Sth JIM8232, and to analyze spacer acquisition by the Type III CRISPR-Cas system.

### CRISPR-Cas systems of *Streptococcus thermophilus*

The adaptive defense function of CRISPR-Cas systems was first observed and reported in *Sth* (7), one of the most extensively used model organisms for CRISPR-Cas-related studies. *Sth* has 4 independent CRISPR-Cas systems: CRISPR1 and CRISPR3 belong to subtype Type II-A, and CRISPR2 and CRISPR4 belong to subtypes Type III-A and Type I-E, respectively (192-194). These 4 loci express mature crRNAs (195), and my work showed that all 4 systems were active in target interference (see Chapter 2, 3 and 4). It is worth noting that like several other *Sth* strains, *Sth* DGCC7710 has a degraded Type III-A system, in that *csm6* genes were disrupted by a nonsense mutation introducing multiple premature stop codons. As a result, target interference by CRISPR2 of DGCC7710 was not observed. In contrast, *Sth* JIM8232, which was the major model organism used in this dissertation, has an intact Type III-A system and performed efficient CRISPR2-mediated defense. CRISPR1 and

CRISPR3 had been shown to be active in adaptation (7), with CRISPR1 to be dominant and CRISPR3 to be active but less efficient in adaptation (50,192). Adaptation by the Type I system was not reported when my dissertation work was initiated and subsequently found to be a highly rare event (196), and the Type III system was thought to be inactive in adaptation before my work. Although CRISPR-Cas systems protect the hosts by eliminating foreign nucleic acids, the expression of *cas* genes is costly to *Sth*, and the *cas* deleted *Sth* mutants are much more competitive than WT in an invader-free environment (197).

### *Type II CRISPR-Cas systems of Streptococcus thermophilus*

*In silico* analyses revealed the existence of CRISPR-Cas systems and spacers with sequence identity to extrachromosomal elements in 24 different *Sth* strains (198). The sensitivities of *Sth* strains to phages were correlated to the spacer numbers, suggesting a phage immunity role of CRISPR-Cas systems in *Sth* (198). In 2007, excitingly, Barrangou et al. detected and reported CRISPR-Cas-mediated defense against two phage species by adaptation in a Type II-A system (CRISPR1) of *Sth* DGCC7710, which was the first report of CRISPR-Cas spacer acquisition and target interference (7). All the *cas* genes around CRISPR1 array, including *cas9*, *cas1*, *cas2*, and *csn2* (7), as well as tracrRNA, the leader sequence and at least one repeat (43,45), were required for adaptation, while only *cas9* was involved in target interference (7). In 2008, Deveau et al. reported for the first time that protospacers were selected along

foreign DNAs by consensus PAMs (NNAGAAW for CRISPR1), which was critical for the authentication of target interference (50). Phages were found to be able to escape from CRISPR-Cas-mediated defense of the BIMs (bacteriophage insensitive mutants) by deletion(s) or mutation(s) within the PAMs or the protospacers (50). Soon after, CRISPR1 was shown to be able to perform adaptive defense not only against phages, but also against plasmids, and to be able to generate a cut within the protospacer of the dsDNA target (199). Although Cas1 and Cas2 of CRISPR1 are sufficient for the integration of a prespacer (43), adaptation against an unprocessed invading DNA required Cas9 and Csn2 (7). In 2012, Lee et al. purified Csn2 of the system, and reported its crystal structure (200). Csn2 forms a homotetramer centered by a channel, and binds to linear dsDNAs but not circular ones (200). Csn2 is proposed to determine the lengths of the new spacers (30 bp for CRISPR1) (200). Cas9 has been shown to recognize the PAM in another Type II system of a different organism (31).

The other Type II-A system, CRISPR3-Cas, is able to generate a cut within the dsDNA target to interfere with the invading phage (201). When transplanted into a distant gram-negative host, *E. coli*, the phage and plasmid interference ability of the system was maintained (202). Moreover, mutagenesis studies in *E. coli* revealed that target cleavage relied upon the HNH and the RuvC domains of Cas9 (202).

Adaptation by the Type II-A systems in *Sth* happens either against foreign nucleic acids (leading to immunity) or against the host genome (leading to autoimmunity and cell death) (30,203). Moreover, when PAMs are nearly evenly

distributed among phage 2972 genome, both of the Type II systems present a strong and reproducible bias regarding protospacer selection across the phage genomes, suggesting that there is an unknown selection mechanism during their adaptation (203).

CRISPR3-Cas system has been expressed in *E. coli* to perform multiplex DNA interference (204), and both the Type II systems have been re-purposed for genome editing and gene regulation (205,206).

### *Type I CRISPR-Cas system of Streptococcus thermophilus*

Target interference ability of Type I systems was first discovered in *E. coli* (99), in a system homologous to the Type I-E system of *Sth*. CasA, CasB, CasC, CasD, and CasE proteins are sufficient to maturate a crRNA and form a crRNP with the later (99). The HD nuclease domain of the trans-acting protein, Cas3, is essential for target cleavage (99). In 2011, Sinkunas et al. purified Cas3 protein of the Type I CRISPR-Cas system of *Sth*, and determined that both the ATP-dependent HD nuclease domain and a helicase domain of Cas3 were necessary for ssDNA cleavage (207). Soon after, the entire interference machinery of the Type I system was reconstituted *in vitro* (208). While the PAM of the system was predicted to be AA immediately upstream by the pre-existing spacers, the *in vitro* work showed that only a single A or T nucleotide at the -1 position of the protospacer was required for authenticating the correct binding between the target and the crRNP, indicating that target interference had a broader range of tolerated PAMs than adaptation (208). Moreover, the *in vitro* work further

revealed that crRNP bound to the target strand of the invader, forming an R-loop and displacing the protospacer strand. The displaced ssDNA was a signal for Cas3 recruitment *in trans*, and the ATP-dependent nuclease of Cas3 generated a cut at the ssDNA, and further degraded the ssDNA in a 3' to 5' unidirectional way (208). Single-molecule supercoiling experiments revealed that the R-loop started from the PAM site, and zipped toward the target direction (209). The R-loop will stall and collapse upon encountering a mismatch, and will activate DNA cleavage by Cas3 upon reaching the end of the target (209).

In contrast to Type II systems, adaptation by the Type I-E system was very inefficient, being barely detectable by high-throughput sequencing after tens and even hundreds of days of co-cultivation with phage (196), despite Cas1-Cas2 complex being capable of integrating a prespacer into the CRISPR array *in vitro* (74). Interestingly, Cas2 of the system is unique, because it possesses a DnaQ-like 3'-5' exonuclease domain, which processes the 3'- overhangs of the prespacers to promote the integration (74).

### *Type III CRISPR-Cas system of Streptococcus thermophilus*

Following characterization of Type I and Type II CRISPR-Cas systems, it was assumed that all the CRISPR-Cas systems bound to and degraded DNA targets. However, works by Tamulaitis et al. revealed that the Type III-A CRISPR-Cas system of *Sth* defended against a RNA phage in a Csm3 endonuclease-dependent way (110). Moreover, they reconstituted the crRNP of the system *in vitro*, and found that it indeed bound to RNA targets rather than

DNA targets, and generated multiple cuts at 6 nt intervals (110). Later, the system was also found to interfere with plasmid DNAs and their transcripts in *E. coli* host by the non-specific nuclease activities of Csm1 and Csm6 (108,121). As stated above, Smalakyte et al. showed that the CARF domain and the HEPN domain of Csx1 of this system degraded the cAns, which was a regulatory mechanism of Csx1 when interference against the invader is complete (126). Adaptation by the Type III system of *Sth* had never been observed before my work.

## Dissertation overview

The chapters of this dissertation describe the scientific studies and contributions I have made investigating naïve and primed adaptation by the Type III-A CRISPR-Cas system, as well as adaptation by the Type I-E CRISPR-Cas system of *Sth*.

Chapter 2 investigates the properties of naïve adaptation by the Type III-A CRISPR-Cas system of *Sth*. Type III CRISPR-Cas systems are distinguished from other types of the systems by their interference activity, however, our knowledge about their adaptation is very limited. In this chapter, I for the first time provide detailed studies about spacer uptake properties of Type III systems. I examined adaptation by the Type III-A system of *Sth*, and adaptation by the Type II-A system in the same host as a comparison. Unlike Type II systems and some Type I systems, deletion of Type III crRNA biogenesis-related gene (*cas6*) or interference-related gene (*csm1-6*) did not disrupt adaptation nor detectably

change spacer uptake patterns (except those altered by counter-selection). No PAM was observed with the Type III system. The lengths of the Type III-A spacers averaged 36 bp. Interestingly, certain regions of plasmids and the host genome were particularly well-sampled during Type III-A, but not Type II-A, spacer uptake. These regions included the single-stranded origins (*sso*s) of RCR plasmids, rRNA and tRNA encoding clusters, and promoter regions of expressed genes. I also found that the Type III-A system could adapt and protect the cell from a lytic phage. Taken together, this chapter indicates that the Type III adaptation machinery preferentially targets DNA secondary structures including imperfect hairpins and other partially double-stranded DNAs.

Chapter 3 investigates primed adaptation by the Type III-A CRISPR-Cas system of *Sth*. Our knowledge about adaptation by the Type III CRISPR-Cas systems was very limited before my work, and in particular, no evidence of primed adaptation had been reported. In this chapter, my work provided preliminary evidence of Type III primed adaptation. Here, primed adaptation increased the frequency of secondary adaptation by the Type III-A system against the invading plasmid, when compared to host genome. The DNA cleavage sites of the targeted plasmid were recognized and preferred by primed adaptation, and moreover, directional transcription of the DNA target and DNase activity of the system were both required for primed adaptation, implying that it is the initial DNA cleavage that facilitates secondary adaptation. This chapter provides preliminary information about primed adaptation by the mysterious Type

III system, and paves the way for further studies about primed adaptation by Type III CRISPR-Cas systems.

Chapter 4 investigates both naïve and primed adaptation by Type I-E CRISPR-Cas system of *Sth*. Previously to the work of this dissertation, adaptation by the Type I-E system could only be very rarely detected. To better understand this process, in this chapter, I increased the frequency of adaptation of the Type I system to a high level by overexpression of *cas1* and *cas2* genes, allowing for a thorough investigation into protospacer selection and adaptation events. This work reveals that the Type I-E system of *Sth* preferentially selects protospacers with an upstream AA PAM. Studies regarding primed adaptation have been hampered by the difficulties in the design of escaping targets. To study primed adaptation of the Type I system, I developed a novel method to naturally obtain virulent phages that escaped from CRISPR-Cas-mediated defense and acquire the escaping target sequences that triggered primed adaptation. I showed that a priming spacer was able to increase the frequency of adaptation by this Type I-E system in the absence of overexpression of any *cas* gene. The secondarily protospacers acquired during primed adaptation accumulated at both primed and non-primed strands with an obvious gradient centered at the target/protospacer of the pre-existing spacer. This chapter not only characterized naïve and primed adaptation by the Type I-E system of *Sth*, but also provided a new method enable future work on understanding the detailed mechanism of primed adaptation.

## References


1. Breitbart, M. and Rohwer, F. (2005) Here a virus, there a virus, everywhere the same virus? *Trends in microbiology*, **13**, 278-284.
2. Bahl, M.I., Hansen, L.H. and Sorensen, S.J. (2009) Persistence mechanisms of conjugative plasmids. *Methods in molecular biology*, **532**, 73-102.
3. Sturino, J.M. and Klaenhammer, T.R. (2006) Engineered bacteriophage-defence systems in bioprocessing. *Nature reviews. Microbiology*, **4**, 395-404.
4. Hille, F., Richter, H., Wong, S.P., Bratovic, M., Ressel, S. and Charpentier, E. (2018) The Biology of CRISPR-Cas: Backward and Forward. *Cell*, **172**, 1239-1259.
5. Haft, D.H., Selengut, J., Mongodin, E.F. and Nelson, K.E. (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS computational biology*, **1**, e60.
6. Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F. *et al.* (2011) Evolution and classification of the CRISPR-Cas systems. *Nature reviews. Microbiology*, **9**, 467-477.
7. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. (2007) CRISPR provides

acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709-1712.

8. Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H. *et al.* (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nature reviews. Microbiology*, **13**, 722-736.
9. Terns, M.P. and Terns, R.M. (2011) CRISPR-based adaptive immune systems. *Current opinion in microbiology*, **14**, 321-327.
10. Terns, R.M. and Terns, M.P. (2013) The RNA- and DNA-targeting CRISPR-Cas immune systems of Pyrococcus furiosus. *Biochemical Society transactions*, **41**, 1416-1421.
11. van der Oost, J., Westra, E.R., Jackson, R.N. and Wiedenheft, B. (2014) Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nature reviews. Microbiology*, **12**, 479-492.
12. Tsui, T.K. and Li, H. (2015) Structure Principles of CRISPR-Cas Surveillance and Effector Complexes. *Annu Rev Biophys*, **44**, 229-255.
13. Agari, Y., Sakamoto, K., Tamakoshi, M., Oshima, T., Kuramitsu, S. and Shinkai, A. (2010) Transcription profile of Thermus thermophilus CRISPR systems after phage infection. *Journal of molecular biology*, **395**, 270-281.
14. Bhaya, D., Davison, M. and Barrangou, R. (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annual review of genetics*, **45**, 273-297.

15. Heler, R., Marraffini, L.A. and Bikard, D. (2014) Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. *Molecular microbiology*, **93**, 1-9.

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

17. Pougach, K., Semenova, E., Bogdanova, E., Datsenko, K.A., Djordjevic, M., Wanner, B.L. and Severinov, K. (2010) Transcription, processing and function of CRISPR cassettes in Escherichia coli. *Molecular microbiology*, **77**, 1367-1379.

18. Sorek, R., Lawrence, C.M. and Wiedenheft, B. (2013) CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annual review of biochemistry*, **82**, 237-266.

19. Kunin, V., Sorek, R. and Hugenholtz, P. (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome biology*, **8**, R61.

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

21. Shmakov, S., Abudayyeh, O.O., Makarova, K.S., Wolf, Y.I., Gootenberg, J.S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K.

*et al.* (2015) Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Molecular cell*, **60**, 385-397.

22. Pickar-Oliver, A. and Gersbach, C.A. (2019) The next generation of CRISPR-Cas technologies and applications. *Nature reviews. Molecular cell biology*, **20**, 490-507.
23. Barrangou, R. and Dudley, E.G. (2016) CRISPR-Based Typing and Next-Generation Tracking Technologies. *Annual review of food science and technology*, **7**, 395-411.
24. Choi, K.R. and Lee, S.Y. (2016) CRISPR technologies for bacterial systems: Current achievements and future directions. *Biotechnol Adv*, **34**, 1180-1209.
25. Shipman, S.L., Nivala, J., Macklis, J.D. and Church, G.M. (2017) CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature*, **547**, 345-349.
26. Isaev, A.B., Musharova, O.S. and Severinov, K.V. (2021) Microbial Arsenal of Antiviral Defenses. Part II. *Biochemistry (Mosc)*, **86**, 449-470.
27. Nussenzweig, P.M. and Marraffini, L.A. (2020) Molecular Mechanisms of CRISPR-Cas Immunity in Bacteria. *Annual review of genetics*, **54**, 93-120.
28. Yosef, I., Goren, M.G. and Qimron, U. (2012) Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic acids research*, **40**, 5569-5576.
29. Nunez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W. and Doudna, J.A. (2014) Cas1-Cas2 complex formation mediates spacer

acquisition during CRISPR-Cas adaptive immunity. *Nature structural & molecular biology*, **21**, 528-534.

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

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

32. Fagerlund, R.D., Wilkinson, M.E., Klykov, O., Barendregt, A., Pearce, F.G., Kieper, S.N., Maxwell, H.W.R., Capolupo, A., Heck, A.J.R., Krause, K.L. *et al.* (2017) Spacer capture and integration by a type I-F Cas1-Cas2-3 CRISPR adaptation complex. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, E5122-E5128.

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

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

35. Wan, H., Li, J., Chang, S., Lin, S., Tian, Y., Tian, X., Wang, M. and Hu, J. (2019) Probing the Behaviour of Cas1-Cas2 upon Protospacer Binding in

CRISPR-Cas Systems using Molecular Dynamics Simulations. *Scientific reports*, **9**, 3188.

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

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

38. Rollie, C., Schneider, S., Brinkmann, A.S., Bolt, E.L. and White, M.F. (2015) Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *Elife*, **4**.

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

40. Diez-Villasenor, C., Guzman, N.M., Almendros, C., Garcia-Martinez, J. and Mojica, F.J. (2013) CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of Escherichia coli. *RNA biology*, **10**, 792-802.

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

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

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

44. Grainy, J., Garrett, S., Graveley, B.R. and M, P.T. (2019) CRISPR repeat sequences and relative spacing specify DNA integration by Pyrococcus furiosus Cas1 and Cas2. *Nucleic acids research*, **47**, 7518-7531.

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

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

47. Rollie, C., Graham, S., Rouillon, C. and White, M.F. (2018) Prespacer processing and specific integration in a Type I-A CRISPR system. *Nucleic acids research*, **46**, 1007-1020.

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

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

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

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

52. Shah, S.A., Erdmann, S., Mojica, F.J. and Garrett, R.A. (2013) Protospacer recognition motifs: mixed identities and functional diversity. *RNA biology*, **10**, 891-899.

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

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

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

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

57. Kieper, S.N., Almendros, C., Behler, J., McKenzie, R.E., Nobrega, F.L., Haagsma, A.C., Vink, J.N.A., Hess, W.R. and Brouns, S.J.J. (2018) Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation. *Cell reports*, **22**, 3377-3384.

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

59. Li, M., Gong, L., Zhao, D., Zhou, J. and Xiang, H. (2017) The spacer size of I-B CRISPR is modulated by the terminal sequence of the protospacer. *Nucleic acids research*, **45**, 4642-4654.

60. Dillingham, M.S. and Kowalczykowski, S.C. (2008) RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiology and molecular biology reviews : MMBR*, **72**, 642-671, Table of Contents.

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

62. Radovcic, M., Killelea, T., Savitskaya, E., Wettstein, L., Bolt, E.L. and Ivancic-Bace, I. (2018) CRISPR-Cas adaptation in Escherichia coli requires RecBCD helicase but not nuclease activity, is independent of

homologous recombination, and is antagonized by 5' ssDNA exonucleases. *Nucleic acids research*, **46**, 10173-10183.

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

64. Zhang, J., Kasciukovic, T. and White, M.F. (2012) The CRISPR associated protein Cas4 Is a 5' to 3' DNA exonuclease with an iron-sulfur cluster. *PloS one*, **7**, e47232.

65. Lemak, S., Nocek, B., Beloglazova, N., Skarina, T., Flick, R., Brown, G., Joachimiak, A., Savchenko, A. and Yakunin, A.F. (2014) The CRISPR-associated Cas4 protein Pcal_0546 from Pyrobaculum calidifontis contains a [2Fe-2S] cluster: crystal structure and nuclease activity. *Nucleic acids research*, **42**, 11144-11155.

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

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

68. Wu, C., Tang, D., Cheng, J., Hu, D., Yang, Z., Ma, X., He, H., Yao, S., Fu, T.M., Yu, Y. *et al.* (2021) Mechanisms of spacer acquisition by sequential assembly of the adaptation module in Synechocystis. *Nucleic acids research*, **49**, 2973-2984.

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

70. Van Orden, M.J., Newsom, S. and Rajan, R. (2020) CRISPR type II-A subgroups exhibit phylogenetically distinct mechanisms for prespacer insertion. *The Journal of biological chemistry*, **295**, 10956-10968.

71. Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., Binkowski, A. *et al.* (2011) A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. *Molecular microbiology*, **79**, 484-502.

72. Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S.M., Ma, W. and Doudna, J.A. (2009) Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure*, **17**, 904-912.

73. He, Y., Wang, M., Liu, M., Huang, L., Liu, C., Zhang, X., Yi, H., Cheng, A., Zhu, D., Yang, Q. *et al.* (2018) Cas1 and Cas2 From the Type II-C CRISPR-Cas System of Riemerella anatipestifer Are Required for Spacer Acquisition. *Frontiers in cellular and infection microbiology*, **8**, 195.

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

75. Ramachandran, A., Summerville, L., Learn, B.A., DeBell, L. and Bailey, S. (2020) Processing and integration of functionally oriented prespacers in

the Escherichia coli CRISPR system depends on bacterial host exonucleases. *The Journal of biological chemistry*, **295**, 3403-3414.

76. Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J. and Severinov, K. (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 10098-10103.

77. Westra, E.R., Semenova, E., Datsenko, K.A., Jackson, R.N., Wiedenheft, B., Severinov, K. and Brouns, S.J.J. (2013) Type I-E CRISPR-Cas Systems Discriminate Target from Non-Target DNA through Base Pairing-Independent PAM Recognition. *PLoS genetics*, **9**.

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

79. Fineran, P.C., Gerritzen, M.J., Suarez-Diez, M., Kunne, T., Boekhorst, J., van Hijum, S.A., Staals, R.H. and Brouns, S.J. (2014) Degenerate target sites mediate rapid primed CRISPR adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, E1629-1638.

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

81. Richter, C., Dy, R.L., McKenzie, R.E., Watson, B.N., Taylor, C., Chang, J.T., McNeil, M.B., Staals, R.H. and Fineran, P.C. (2014) Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic acids research*, **42**, 8516-8526.

82. Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A. and Severinov, K. (2013) High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in E. coli. *RNA biology*, **10**, 716-725.

83. Garrett, S., Shiimori, M., Watts, E.A., Clark, L., Graveley, B.R. and Terns, M.P. (2020) Primed CRISPR DNA uptake in Pyrococcus furiosus. *Nucleic acids research*, **48**, 6120-6135.

84. Hoikkala, V., Ravantti, J., Diez-Villasenor, C., Tiirola, M., Conrad, R.A., McBride, M.J., Moineau, S. and Sundberg, L.R. (2021) Cooperation between Different CRISPR-Cas Types Enables Adaptation in an RNA-Targeting System. *mBio*, **12**.

85. Wiegand, T., Semenova, E., Shiriaeva, A., Fedorov, I., Datsenko, K., Severinov, K. and Wiedenheft, B. (2020) Reproducible Antigen Recognition by the Type I-F CRISPR-Cas System. *CRISPR J*, **3**, 378-387.

86. Nicholson, T.J., Jackson, S.A., Croft, B.I., Staals, R.H.J., Fineran, P.C. and Brown, C.M. (2019) Bioinformatic evidence of widespread priming in type I and II CRISPR-Cas systems. *RNA biology*, **16**, 566-576.

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

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

89. Mosterd, C. and Moineau, S. (2021) Primed CRISPR-Cas Adaptation and Impaired Phage Adsorption in Streptococcus mutans. *mSphere*, **6**.

90. Stachler, A.E., Wortz, J., Alkhnbashi, O.S., Turgeman-Grott, I., Smith, R., Allers, T., Backofen, R., Gophna, U. and Marchfelder, A. (2020) Adaptation induced by self-targeting in a type I-B CRISPR-Cas system. *The Journal of biological chemistry*, **295**, 13502-13515.

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

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

93. Staals, R.H., Jackson, S.A., Biswas, A., Brouns, S.J., Brown, C.M. and Fineran, P.C. (2016) Interference-driven spacer acquisition is dominant

over naive and primed adaptation in a native CRISPR-Cas system. *Nature communications*, **7**, 12853.

94. Kunne, T., Kieper, S.N., Bannenberg, J.W., Vogel, A.I., Miellet, W.R., Klein, M., Depken, M., Suarez-Diez, M. and Brouns, S.J. (2016) Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation. *Molecular cell*, **63**, 852-864.

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

96. Xue, C., Whitis, N.R. and Sashital, D.G. (2016) Conformational Control of Cascade Interference and Priming Activities in CRISPR Immunity. *Molecular cell*, **64**, 826-834.

97. Blosser, T.R., Loeff, L., Westra, E.R., Vlot, M., Kunne, T., Sobota, M., Dekker, C., Brouns, S.J.J. and Joo, C. (2015) Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. *Molecular cell*, **58**, 60-70.

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

99. Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V. and van der

Oost, J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960-964.

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

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

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

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

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

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

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

107. Liu, T.Y., Iavarone, A.T. and Doudna, J.A. (2017) RNA and DNA Targeting by a Reconstituted Thermus thermophilus Type III-A CRISPR-Cas System. *PloS one*, **12**, e0170552.

108. Ichikawa, H.T., Cooper, J.C., Lo, L., Potter, J., Terns, R.M. and Terns, M.P. (2017) Programmable type III-A CRISPR-Cas DNA targeting modules. *PloS one*, **12**, e0176221.

109. Staals, R.H., Zhu, Y., Taylor, D.W., Kornfeld, J.E., Sharma, K., Barendregt, A., Koehorst, J.J., Vlot, M., Neupane, N., Varossieau, K. *et al.* (2014) RNA targeting by the type III-A CRISPR-Cas Csm complex of Thermus thermophilus. *Molecular cell*, **56**, 518-530.

110. Tamulaitis, G., Kazlauskiene, M., Manakova, E., Venclovas, C., Nwokeoji, A.O., Dickman, M.J., Horvath, P. and Siksnys, V. (2014) Programmable RNA shredding by the type III-A CRISPR-Cas system of Streptococcus thermophilus. *Molecular cell*, **56**, 506-517.

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

112. Elmore, J.R., Sheppard, N.F., Ramia, N., Deighan, T., Li, H., Terns, R.M. and Terns, M.P. (2016) Bipartite recognition of target RNAs activates DNA

cleavage by the Type III-B CRISPR-Cas system. *Genes & development*, **30**, 447-459.

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

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

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

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

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

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

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

120. Niewoehner, O., Garcia-Doval, C., Rostol, J.T., Berk, C., Schwede, F., Bigler, L., Hall, J., Marraffini, L.A. and Jinek, M. (2017) Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature*, **548**, 543-548.

121. Foster, K., Kalter, J., Woodside, W., Terns, R.M. and Terns, M.P. (2019) The ribonuclease activity of Csm6 is required for anti-plasmid immunity by Type III-A CRISPR-Cas systems. *RNA biology*, **16**, 449-460.

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

123. Mo, C.Y., Mathai, J., Rostol, J.T., Varble, A., Banh, D.V. and Marraffini, L.A. (2021) Type III-A CRISPR immunity promotes mutagenesis of staphylococci. *Nature*, **592**, 611-615.

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

125. Nasef, M., Muffly, M.C., Beckman, A.B., Rowe, S.J., Walker, F.C., Hatoum-Aslan, A. and Dunkle, J.A. (2019) Regulation of cyclic

oligoadenylate synthesis by the Staphylococcus epidermidis Cas10-Csm complex. *Rna*, **25**, 948-962.

126. Smalakyte, D., Kazlauskiene, M., J, F.H., Ruksenaite, A., Rimaite, A., Tamulaitiene, G., Faergeman, N.J., Tamulaitis, G. and Siksnys, V. (2020) Type III-A CRISPR-associated protein Csm6 degrades cyclic hexa-adenylate activator using both CARF and HEPN domains. *Nucleic acids research*, **48**, 9204-9217.

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

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

129. Li, Y., Pan, S., Zhang, Y., Ren, M., Feng, M., Peng, N., Chen, L., Liang, Y.X. and She, Q. (2015) Harnessing Type I and Type III CRISPR-Cas systems for genome editing. *Nucleic acids research*.

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

131. Schmidt, F., Cherepkova, M.Y. and Platt, R.J. (2018) Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature*, **562**, 380-385.

132. Gonzalez-Delgado, A., Mestre, M.R., Martinez-Abarca, F. and Toro, N. (2019) Spacer acquisition from RNA mediated by a natural reverse

transcriptase-Cas1 fusion protein associated with a type III-D CRISPR-Cas system in Vibrio vulnificus. *Nucleic acids research*, **47**, 10202-10211.

133. Artamonova, D., Karneyeva, K., Medvedeva, S., Klimuk, E., Kolesnik, M., Yasinskaya, A., Samolygo, A. and Severinov, K. (2020) Spacer acquisition by Type III CRISPR-Cas system during bacteriophage infection of Thermus thermophilus. *Nucleic acids research*, **48**, 9787-9803.

134. Wasilewska, E., Zlotkowska, D. and Wroblewska, B. (2019) Yogurt starter cultures of Streptococcus thermophilus and Lactobacillus bulgaricus ameliorate symptoms and modulate the immune response in a mouse model of dextran sulfate sodium-induced colitis. *J Dairy Sci*, **102**, 37-53.

135. Evivie, S.E., Abdelazez, A., Li, B., Bian, X., Li, W., Du, J., Huo, G. and Liu, F. (2019) In vitro Organic Acid Production and In Vivo Food Pathogen Suppression by Probiotic S. thermophilus and L. bulgaricus. *Frontiers in microbiology*, **10**, 782.

136. Miclo, L., Roux, E., Genay, M., Brusseaux, E., Poirson, C., Jameh, N., Perrin, C. and Dary, A. (2012) Variability of hydrolysis of beta-, alphas1-, and alphas2-caseins by 10 strains of Streptococcus thermophilus and resulting bioactive peptides. *J Agric Food Chem*, **60**, 554-565.

137. Mitchell, T.J. (2003) The pathogenesis of streptococcal infections: from tooth decay to meningitis. *Nature reviews. Microbiology*, **1**, 219-230.

138. Lecomte, X., Gagnaire, V., Lortal, S., Dary, A. and Genay, M. (2016) Streptococcus thermophilus, an emerging and promising tool for

heterologous expression: Advantages and future trends. *Food microbiology*, **53**, 2-9.

139. Bolotin, A., Quinquis, B., Renault, P., Sorokin, A., Ehrlich, S.D., Kulakauskas, S., Lapidus, A., Goltsman, E., Mazur, M., Pusch, G.D. *et al.* (2004) Complete sequence and comparative genome analysis of the dairy bacterium Streptococcus thermophilus. *Nature biotechnology*, **22**, 1554-1558.

140. Hols, P., Hancy, F., Fontaine, L., Grossiord, B., Prozzi, D., Leblond-Bourget, N., Decaris, B., Bolotin, A., Delorme, C., Dusko Ehrlich, S. *et al.* (2005) New insights in the molecular biology and physiology of Streptococcus thermophilus revealed by comparative genomics. *FEMS microbiology reviews*, **29**, 435-463.

141. Eng, C., Thibessard, A., Danielsen, M., Rasmussen, T.B., Mari, J.F. and Leblond, P. (2011) In silico prediction of horizontal gene transfer in Streptococcus thermophilus. *Arch Microbiol*, **193**, 287-297.

142. Delorme, C., Bartholini, C., Bolotine, A., Ehrlich, S.D. and Renault, P. (2010) Emergence of a cell wall protease in the Streptococcus thermophilus population. *Applied and environmental microbiology*, **76**, 451-460.

143. Sieuwerts, S., Molenaar, D., van Hijum, S.A., Beerthuyzen, M., Stevens, M.J., Janssen, P.W., Ingham, C.J., de Bok, F.A., de Vos, W.M. and van Hylckama Vlieg, J.E. (2010) Mixed-culture transcriptome analysis reveals the molecular basis of mixed-culture growth in Streptococcus thermophilus

and Lactobacillus bulgaricus. *Applied and environmental microbiology*, **76**, 7775-7784.

144. Courtin, P., Monnet, V. and Rul, F. (2002) Cell-wall proteinases PrtS and PrtB have a different role in Streptococcus thermophilus/Lactobacillus bulgaricus mixed cultures in milk. *Microbiology (Reading)*, **148**, 3413-3421.

145. Garault, P., Le Bars, D., Besset, C. and Monnet, V. (2002) Three oligopeptide-binding proteins are involved in the oligopeptide transport of Streptococcus thermophilus. *The Journal of biological chemistry*, **277**, 32-39.

146. Sieuwerts, S., de Bok, F.A., Hugenholtz, J. and van Hylckama Vlieg, J.E. (2008) Unraveling microbial interactions in food fermentations: from classical to genomics approaches. *Applied and environmental microbiology*, **74**, 4997-5007.

147. Lu, W.W., Wang, Y., Wang, T. and Kong, J. (2015) The global regulator CodY in Streptococcus thermophilus controls the metabolic network for escalating growth in the milk environment. *Applied and environmental microbiology*, **81**, 2349-2358.

148. Renye, J.A., Jr. and Somkuti, G.A. (2009) Insertion of a heterologous gene construct into a non-functional ORF of the Streptococcus thermophilus chromosome. *Biotechnol Lett*, **31**, 759-764.

149. Blomqvist, T., Steinmoen, H. and Havarstein, L.S. (2006) Natural genetic transformation: A novel tool for efficient genetic engineering of the dairy

bacterium Streptococcus thermophilus. *Applied and environmental microbiology*, **72**, 6751-6756.

150. Fontaine, L., Dandoy, D., Boutry, C., Delplace, B., de Frahan, M.H., Fremaux, C., Horvath, P., Boyaval, P. and Hols, P. (2010) Development of a versatile procedure based on natural transformation for marker-free targeted genetic modification in Streptococcus thermophilus. *Applied and environmental microbiology*, **76**, 7870-7877.

151. Johnsborg, O., Eldholm, V. and Havarstein, L.S. (2007) Natural genetic transformation: prevalence, mechanisms and function. *Research in microbiology*, **158**, 767-778.

152. Fontaine, L., Goffin, P., Dubout, H., Delplace, B., Baulard, A., Lecat-Guillet, N., Chambellon, E., Gardan, R. and Hols, P. (2013) Mechanism of competence activation by the ComRS signalling system in streptococci. *Molecular microbiology*, **87**, 1113-1132.

153. Khan, R., Rukke, H.V., Ricomini Filho, A.P., Fimland, G., Arntzen, M.O., Thiede, B. and Petersen, F.C. (2012) Extracellular identification of a processed type II ComR/ComS pheromone of Streptococcus mutans. *Journal of bacteriology*, **194**, 3781-3788.

154. Gardan, R., Besset, C., Gitton, C., Guillot, A., Fontaine, L., Hols, P. and Monnet, V. (2013) Extracellular life cycle of ComS, the competence-stimulating peptide of Streptococcus thermophilus. *Journal of bacteriology*, **195**, 1845-1855.

155. Fontaine, L., Boutry, C., de Frahan, M.H., Delplace, B., Fremaux, C., Horvath, P., Boyaval, P. and Hols, P. (2010) A novel pheromone quorum-sensing system controls the development of natural competence in Streptococcus thermophilus and Streptococcus salivarius. *Journal of bacteriology*, **192**, 1444-1454.

156. Gardan, R., Besset, C., Guillot, A., Gitton, C. and Monnet, V. (2009) The oligopeptide transport system is essential for the development of natural competence in Streptococcus thermophilus strain LMD-9. *Journal of bacteriology*, **191**, 4647-4655.

157. Mashburn-Warren, L., Morrison, D.A. and Federle, M.J. (2010) A novel double-tryptophan peptide pheromone controls competence in Streptococcus spp. via an Rgg regulator. *Molecular microbiology*, **78**, 589-606.

158. Fontaine, L., Wahl, A., Flechard, M., Mignolet, J. and Hols, P. (2015) Regulation of competence for natural transformation in streptococci. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, **33**, 343-360.

159. Haustenne, L., Bastin, G., Hols, P. and Fontaine, L. (2015) Modeling of the ComRS Signaling Pathway Reveals the Limiting Factors Controlling Competence in Streptococcus thermophilus. *Frontiers in microbiology*, **6**, 1413.

160. Khan, S.A. (1997) Rolling-circle replication of bacterial plasmids. *Microbiology and molecular biology reviews : MMBR*, **61**, 442-455.

161. Khan, S.A. (2000) Plasmid rolling-circle replication: recent developments. *Molecular microbiology*, **37**, 477-484.

162. del Solar, G., Giraldo, R., Ruiz-Echevarria, M.J., Espinosa, M. and Diaz-Orejas, R. (1998) Replication and control of circular bacterial plasmids. *Microbiology and molecular biology reviews : MMBR*, **62**, 434-464.

163. Ruiz-Maso, J.A., Macho, N.C., Bordanaba-Ruiseco, L., Espinosa, M., Coll, M. and Del Solar, G. (2015) Plasmid Rolling-Circle Replication. *Microbiol Spectr*, **3**, PLAS-0035-2014.

164. del Solar, G.H., Puyet, A. and Espinosa, M. (1987) Initiation signals for the conversion of single stranded to double stranded DNA forms in the streptococcal plasmid pLS1. *Nucleic acids research*, **15**, 5561-5580.

165. Kramer, M.G., Khan, S.A. and Espinosa, M. (1997) Plasmid rolling circle replication: identification of the RNA polymerase-directed primer RNA and requirement for DNA polymerase I for lagging strand synthesis. *The EMBO journal*, **16**, 5784-5795.

166. Kramer, M.G., Espinosa, M., Misra, T.K. and Khan, S.A. (1999) Characterization of a single-strand origin, ssoU, required for broad host range replication of rolling-circle plasmids. *Molecular microbiology*, **33**, 466-475.

167. Kramer, M.G., Espinosa, M., Misra, T.K. and Khan, S.A. (1998) Lagging strand replication of rolling-circle plasmids: specific recognition of the ssoA-type origins in different gram-positive bacteria. *Proceedings of the*

*National Academy of Sciences of the United States of America*, **95**, 10505-10510.

168. Otto, R., de Vos, W.M. and Gavrieli, J. (1982) Plasmid DNA in Streptococcus cremoris Wg2: Influence of pH on Selection in Chemostats of a Variant Lacking a Protease Plasmid. *Applied and environmental microbiology*, **43**, 1272-1277.

169. Leenhouts, K.J., Tolner, B., Bron, S., Kok, J., Venema, G. and Seegers, J.F. (1991) Nucleotide sequence and characterization of the broad-host-range lactococcal plasmid pWVO1. *Plasmid*, **26**, 55-66.

170. Seegers, J.F., Zhao, A.C., Meijer, W.J., Khan, S.A., Venema, G. and Bron, S. (1995) Structural and functional analysis of the single-strand origin of replication from the lactococcal plasmid pWV01. *Mol Gen Genet*, **249**, 43-50.

171. Mashburn-Warren, L., Morrison, D.A. and Federle, M.J. (2012) The cryptic competence pathway in Streptococcus pyogenes is controlled by a peptide pheromone. *Journal of bacteriology*, **194**, 4589-4600.

172. Duong, T., Miller, M.J., Barrangou, R., Azcarate-Peril, M.A. and Klaenhammer, T.R. (2011) Construction of vectors for inducible and constitutive gene expression in Lactobacillus. *Microb Biotechnol*, **4**, 357-367.

173. Moineau, S., Walker, S.A., Holler, B.J., Vedamuthu, E.R. and Vandenbergh, P.A. (1995) Expression of a Lactococcus lactis Phage

Resistance Mechanism by Streptococcus thermophilus. *Applied and environmental microbiology*, **61**, 2461-2466.

174. Turgeon, N. and Moineau, S. (2001) Isolation and characterization of a Streptococcus thermophilus plasmid closely related to the pMV158 family. *Plasmid*, **45**, 171-183.

175. Smillie, C., Garcillan-Barcia, M.P., Francia, M.V., Rocha, E.P. and de la Cruz, F. (2010) Mobility of plasmids. *Microbiology and molecular biology reviews : MMBR*, **74**, 434-452.

176. Zhang, W., Rong, C., Chen, C. and Gao, G.F. (2012) Type-IVC secretion system: a novel subclass of type IV secretion system (T4SS) common existing in gram-positive genus Streptococcus. *PloS one*, **7**, e46390.

177. Vaillancourt, K., Bedard, N., Bart, C., Tessier, M., Robitaille, G., Turgeon, N., Frenette, M., Moineau, S. and Vadeboncoeur, C. (2008) Role of galK and galM in galactose metabolism by Streptococcus thermophilus. *Applied and environmental microbiology*, **74**, 1264-1267.

178. Lilly, J. and Camps, M. (2015) Mechanisms of Theta Plasmid Replication. *Microbiol Spectr*, **3**, PLAS-0029-2014.

179. Brantl, S. (2015) Antisense-RNA mediated control of plasmid replication - pIP501 revisited. *Plasmid*, **78**, 4-16.

180. Bruand, C. and Ehrlich, S.D. (1998) Transcription-driven DNA replication of plasmid pAMbeta1 in Bacillus subtilis. *Molecular microbiology*, **30**, 135-145.

181. Bruand, C., Le Chatelier, E., Ehrlich, S.D. and Janniere, L. (1993) A fourth class of theta-replicating plasmids: the pAM beta 1 family from gram-positive bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, **90**, 11668-11672.

182. Bidnenko, V., Ehrlich, S.D. and Janniere, L. (1998) In vivo relations between pAMbeta1-encoded type I topoisomerase and plasmid replication. *Molecular microbiology*, **28**, 1005-1016.

183. Biswas, I., Jha, J.K. and Fromm, N. (2008) Shuttle expression plasmids for genetic studies in Streptococcus mutans. *Microbiology (Reading)*, **154**, 2275-2282.

184. Renye, J.A., Jr. and Somkuti, G.A. (2010) Nisin-induced expression of pediocin in dairy lactic acid bacteria. *Journal of applied microbiology*, **108**, 2142-2151.

185. Lartigue, M.F. and Bouloc, P. (2014) A tetracycline-inducible expression vector for Streptococcus agalactiae allowing controllable gene expression. *J Microbiol Methods*, **96**, 16-18.

186. Quiberoni, A., Moineau, S., Rousseau, G.M., Reinheimer, J. and Ackermann, H.-W. (2010) Streptococcus thermophilus bacteriophages. *Int Dairy J*, **20**, 657-664.

187. Philippe, C., Levesque, S., Dion, M.B., Tremblay, D.M., Horvath, P., Luth, N., Cambillau, C., Franz, C., Neve, H., Fremaux, C. *et al.* (2020) Novel Genus of Phages Infecting Streptococcus thermophilus: Genomic and

Morphological Characterization. *Applied and environmental microbiology*, **86**.

188. Mills, S., Griffin, C., O'Sullivan, O., Coffey, A., McAuliffe, O.E., Meijer, W.C., Serrano, L.M. and Ross, R.P. (2011) A new phage on the 'Mozzarella' block: Bacteriophage 5093 shares a low level of homology with other Streptococcus thermophilus phages. *Int Dairy J*, **21**, 963-969.
189. McDonnell, B., Mahony, J., Neve, H., Hanemaaijer, L., Noben, J.P., Kouwen, T. and van Sinderen, D. (2016) Identification and Analysis of a Novel Group of Bacteriophages Infecting the Lactic Acid Bacterium Streptococcus thermophilus. *Applied and environmental microbiology*, **82**, 5153-5165.
190. Levesque, C., Duplessis, M., Labonte, J., Labrie, S., Fremaux, C., Tremblay, D. and Moineau, S. (2005) Genomic organization and molecular analysis of virulent bacteriophage 2972 infecting an exopolysaccharide-producing Streptococcus thermophilus strain. *Applied and environmental microbiology*, **71**, 4057-4068.
191. Duplessis, M., Russell, W.M., Romero, D.A. and Moineau, S. (2005) Global gene expression analysis of two Streptococcus thermophilus bacteriophages using DNA microarray. *Virology*, **340**, 192-208.
192. Horvath, P., Romero, D.A., Coute-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C. and Barrangou, R. (2008) Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. *Journal of bacteriology*, **190**, 1401-1412.

193. Horvath, P., Coute-Monvoisin, A.C., Romero, D.A., Boyaval, P., Fremaux, C. and Barrangou, R. (2009) Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *International journal of food microbiology*, **131**, 62-70.

194. Horvath, P. and Barrangou, R. (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science*, **327**, 167-170.

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

196. Paez-Espino, D., Sharon, I., Morovic, W., Stahl, B., Thomas, B.C., Barrangou, R. and Banfield, J.F. (2015) CRISPR immunity drives rapid phage genome evolution in Streptococcus thermophilus. *mBio*, **6**.

197. Vale, P.F., Lafforgue, G., Gatchitch, F., Gardan, R., Moineau, S. and Gandon, S. (2015) Costs of CRISPR-Cas-mediated resistance in Streptococcus thermophilus. *Proceedings. Biological sciences / The Royal Society*, **282**, 20151270.

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

199. Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H. and Moineau, S.

(2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, **468**, 67-71.

200. Lee, K.H., Lee, S.G., Eun Lee, K., Jeon, H., Robinson, H. and Oh, B.H. (2012) Identification, structural, and biochemical characterization of a group of large Csn2 proteins involved in CRISPR-mediated bacterial immunity. *Proteins*, **80**, 2573-2582.

201. Magadan, A.H., Dupuis, M.E., Villion, M. and Moineau, S. (2012) Cleavage of phage DNA by the Streptococcus thermophilus CRISPR3-Cas system. *PloS one*, **7**, e40913.

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

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

204. Guo, L., Xu, K., Liu, Z., Zhang, C., Xin, Y. and Zhang, Z. (2015) Assembling the Streptococcus thermophilus clustered regularly interspaced short palindromic repeats (CRISPR) array for multiplex DNA targeting. *Analytical biochemistry*, **478**, 131-133.

205. Hao, M., Cui, Y. and Qu, X. (2018) Analysis of CRISPR-Cas System in Streptococcus thermophilus and Its Application. *Frontiers in microbiology*, **9**, 257.

206. Agudelo, D., Carter, S., Velimirovic, M., Duringer, A., Rivest, J.F., Levesque, S., Loehr, J., Mouchiroud, M., Cyr, D., Waters, P.J. *et al.* (2020) Versatile and robust genome editing with Streptococcus thermophilus CRISPR1-Cas9. *Genome research*, **30**, 107-117.

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

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

209. Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M.S., Siksnys, V. and Seidel, R. (2015) Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection. *Cell reports*, **10**, 1534-1543.

**Figure 1.1. Schematic of the process of CRISPR-Cas-mediated invading nucleic acid immunity.**

All the CRISPR-Cas systems function in three major steps: adaptation, crRNA biogenesis, and target interference. The heritable defense memories against the foreign nucleic acids are acquired during adaptation, where protospacers from foreign nucleic acids are selected (by PAM for the Type I and Type II systems), processed, and integrated into a CRISPR locus at its leader-proximal end as new spacers. The CRISPR locus is transcribed into a long and pre-mature RNA, which is further processed into short and mature crRNAs. crRNAs direct the formation of crRNPs, and the cleavage of invading nucleic acids.

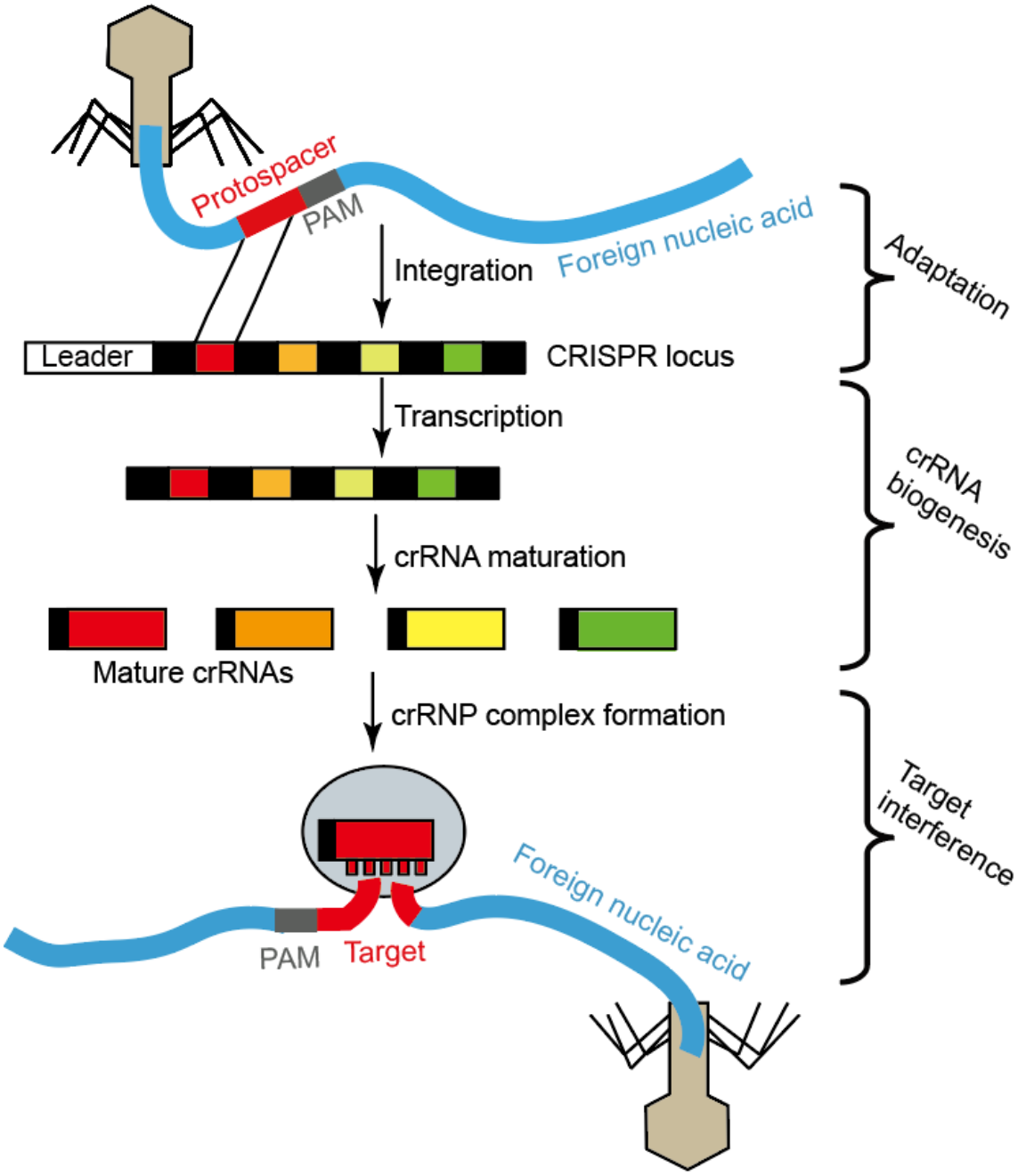

Protospacer
PAM
Foreign nucleic acid
Integration
Leader
CRISPR locus
Transcription
crRNA maturation
Mature crRNAs
crRNP complex formation
Foreign nucleic acid
PAM
Target
Adaptation
crRNA biogenesis
Target interference

**Figure 1.2. Schematic of Cas1-Cas2 complex-mediated prespacer integration.**

The Cas1-Cas2 complex binds to a dsDNA prespacer substrate and directs the 3’- hydroxyl groups of the prespacer to attack the leader-repeat junction and the repeat-Spacer1 junction. The resulting single-stranded repeats will be filled in by host DNA polymerase and ligase.

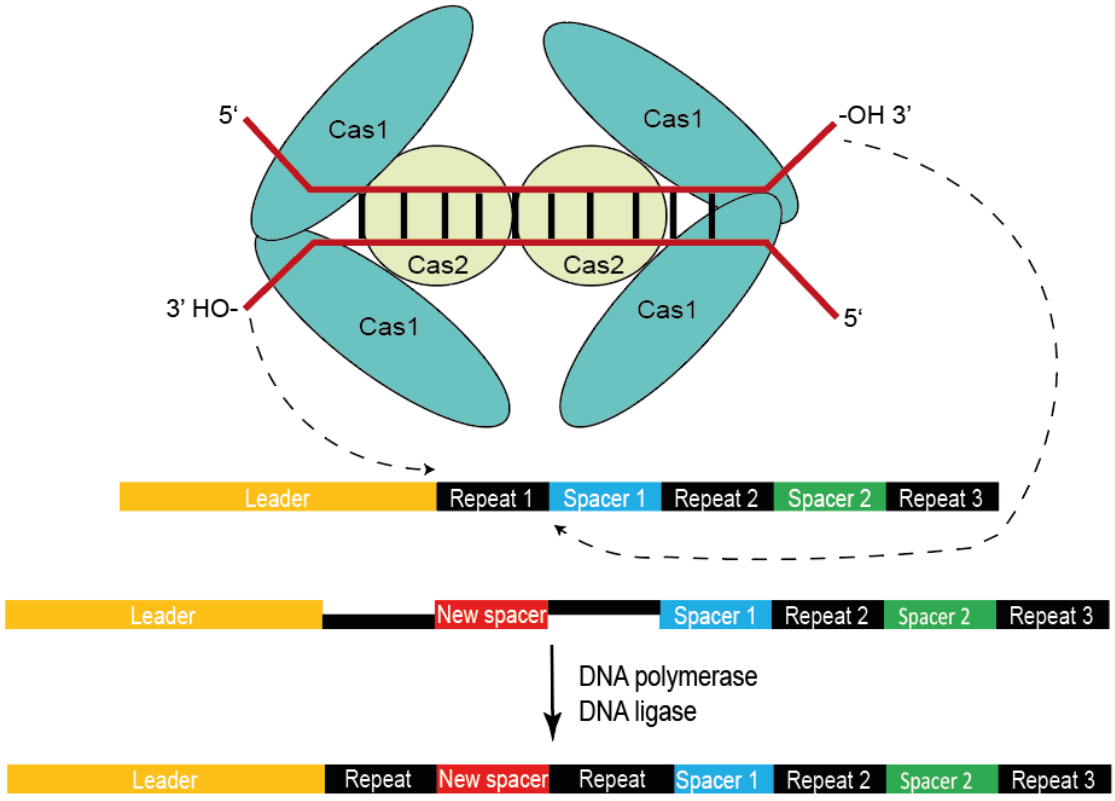

5‘
Cas1
Cas1
-OH 3’
Cas2
Cas2
3’ HO-
Cas1
Cas1
5‘
Leader
Repeat 1
Spacer 1
Repeat 2
Spacer 2
Repeat 3
Leader
New spacer
Spacer 1
Repeat 2
Spacer 2
Repeat 3
DNA polymerase
DNA ligase
Leader
Repeat
New spacer
Repeat
Spacer 1
Repeat 2
Spacer 2
Repeat 3

**Figure 1.3. Schematic of the process of primed adaptation.**

A mature crRNA generated from a CRISPR locus directs a crRNP to recognize and degrade the foreign nucleic acid by base pairing. The invaders may be able to escape from CRISPR-Cas-mediated defense by mutation(s) within the target or the PAM. However, a crRNA which is totally or partially reverse complementary to a fragment of the invader, may be able to trigger the highly efficient secondary adaptation by Cas1-2 complex against the same invading molecule, thus minimizing the amount of the escaping invaders.

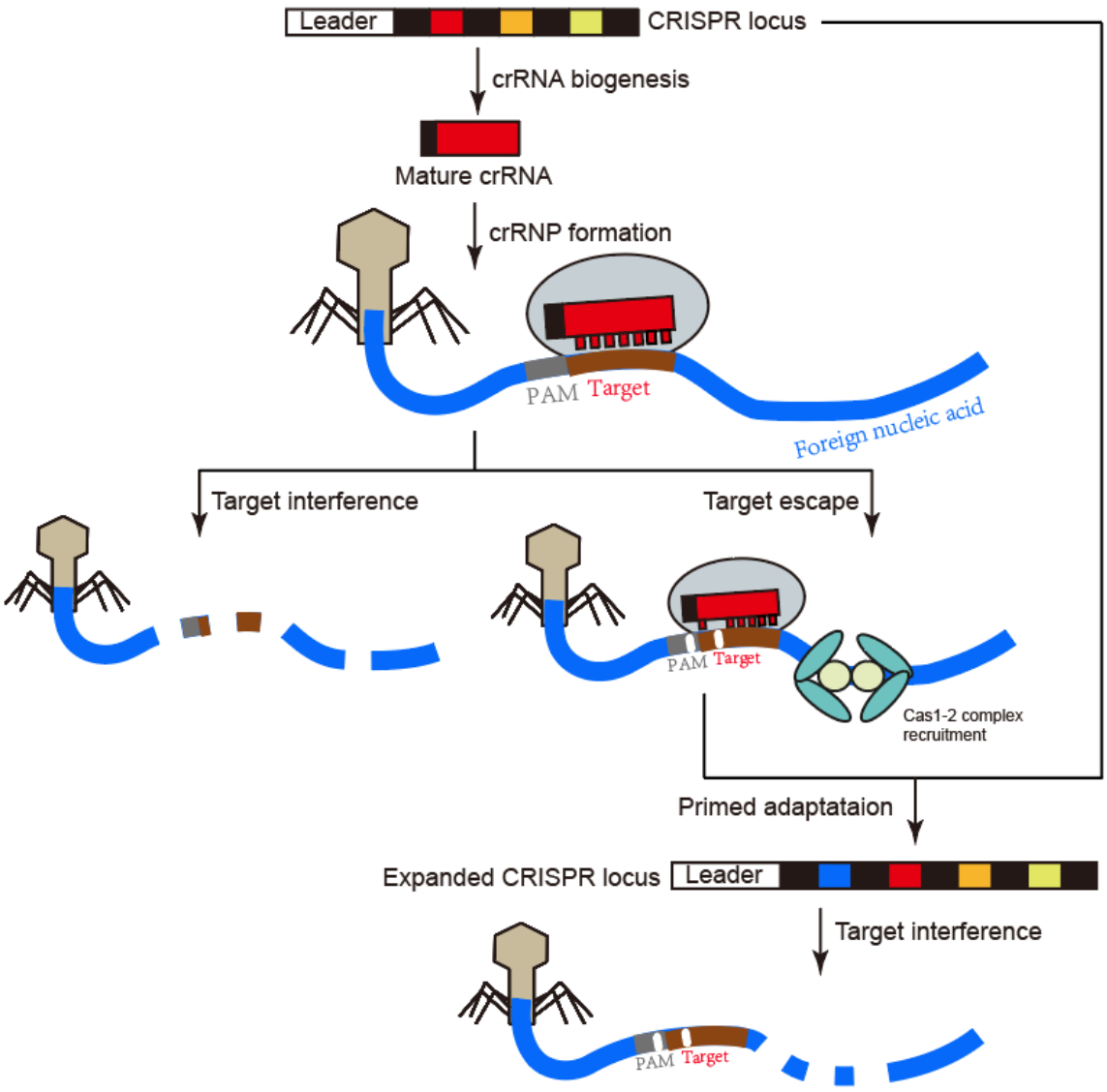

Leader
CRISPR locus
crRNA biogenesis
Mature crRNA
crRNP formation
PAM
Target
Foreign nucleic acid
Target interference
Target escape
PAM
Target
Cas1-2 complex
recruitment
Primed adaptataion
Expanded CRISPR locus
Leader
Target interference
PAM
Target

**Figure 1.4. Schematic of target interference by Type III systems.**

A mature Type III crRNA directs the crRNP (Csm complex for Type III-A) to bind to the transcript of the target by base pairing. The DNase activity (HD domain) of the crRNPs is stimulated by base pairing, and non-specifically cleaves the ssDNAs. Additionally, the Palm domain of Cas10 (Csm1 for Type III-A) synthesizes cAns, which stimulate the RNase activity of Csx1 (Csm6 for Type III-A) to non-specifically cleave the foreign DNA transcripts.

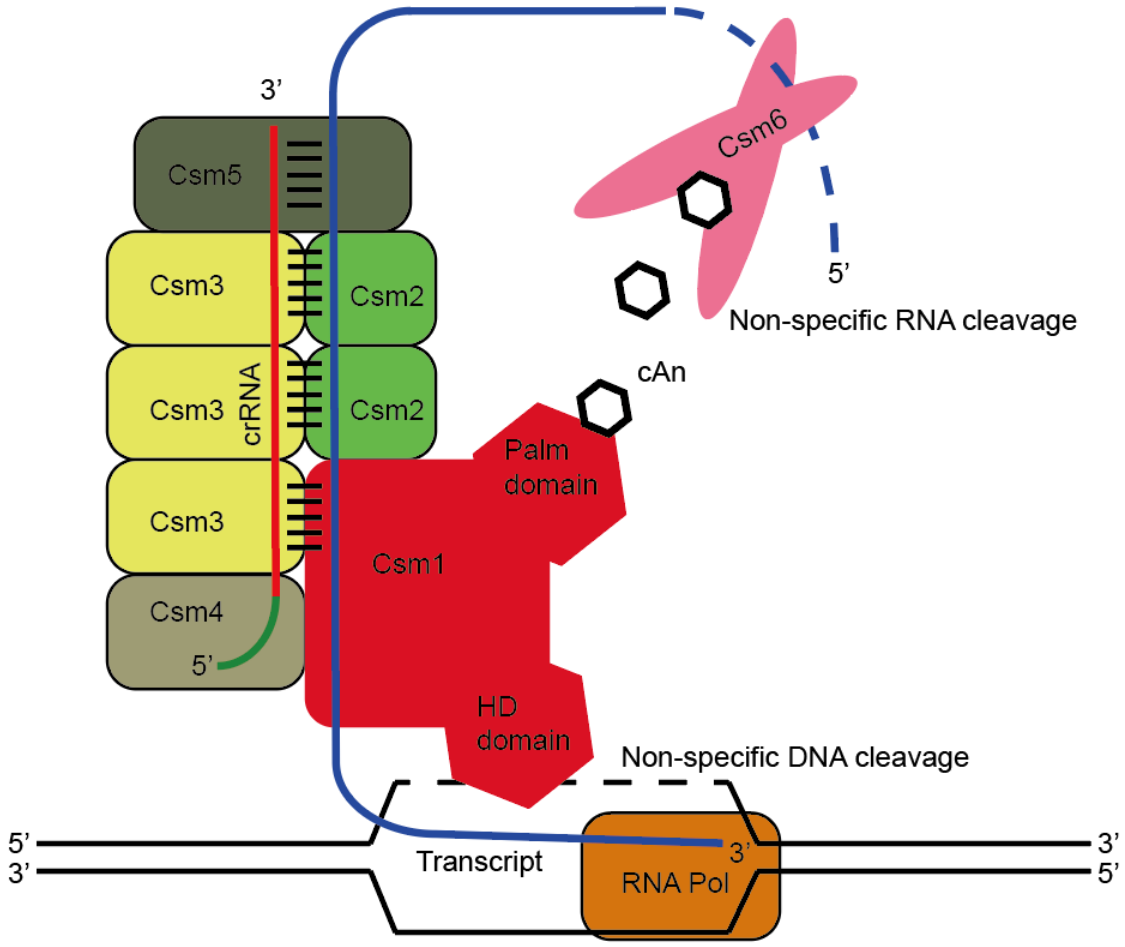
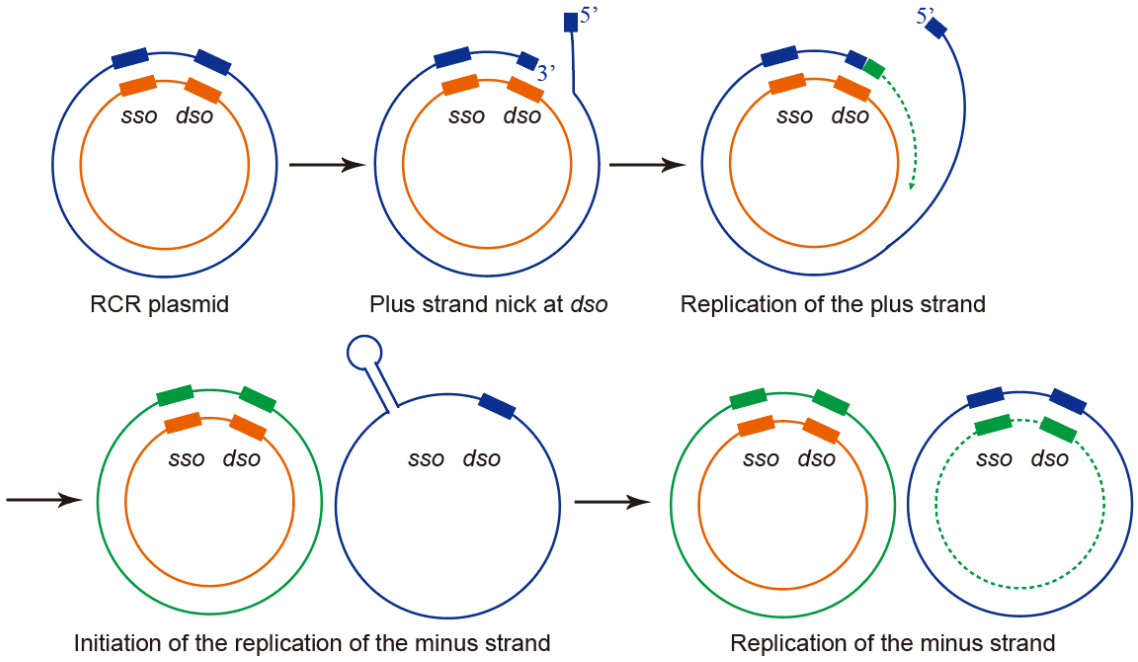

3'
Csm5
Csm3
Csm2
Csm3
crRNA
Csm2
Csm3
Csm4
5'
Csm6
5'
Non-specific RNA cleavage
cAn
Palm
domain
Csm1
HD
domain
Non-specific DNA cleavage
5'
3'
Transcript
3'
RNA Pol
3'
5'

**Figure 1.5. Schematic of the process of rolling-circle replication.**

The parental plus strand (nicked during the replication) is illustrated in blue, the parental minus strand is illustrated in orange, and the nascent strands are illustrated in green. The rectangles represent single-stranded origin (*sso*) and double-stranded origin (*dso*).


5'
3'
sso dso
sso dso
5'
sso dso
RCR plasmid
Plus strand nick at dso
Replication of the plus strand
sso dso
sso dso
sso dso
sso dso
Initiation of the replication of the minus strand
Replication of the minus strand

# CHAPTER 2

# PROPERTIES OF SPACER ACQUISITION BY THE TYPE III-A CRISPR-CAS SYSTEM OF *STREPTOCOCCUS THERMOPHILUS*.

[1]Xinfu Zhang, Sandra Garrett, Brenton G. Graveley, Michael P. Terns. To be submitted to Nucleic Acids Research.

## Abstract

The requirements for interference in Type III CRISPR-Cas systems are broadly different from those in well-studied Type I or Type II systems, with no apparent PAM (protospacer adjacent motif) and the crRNA binding nascent RNA rather than DNA. In this chapter, I examined the process of adaptation to determine if CRISPR spacer uptake properties and patterns also differ for a Type III system. *Streptococcus thermophilus* JIM8232 strain contains interference-proficient Type II-A and III-A systems, as assessed by target plasmid transformation efficiency. Both systems were also competent for adaptation, enabling direct comparison of spacer uptake patterns between Type II and III CRISPR systems in the same organism against a variety of plasmid or phage invaders. Several novel Type III characteristics were observed. First, unlike Type II systems that require the effector Cas9 nuclease for adaptation, deletion of Type III crRNP effector complex genes (*csm1-csm6*) or *cas6* did not disrupt adaptation and did not detectably change spacer uptake patterns except those related to counter-selection. Moreover, in contrast to Type II adaptation, no PAM was observed with the Type III system for either the wildtype or the interference-null (*csm1-csm6* deletion) strain. Spacer lengths were also different between the systems, with the Type III-A spacers averaging 36 bp, consistent with the lengths observed for existing spacers in the native array. Interestingly, certain regions of plasmids and the host chromosome were particularly well-sampled during Type III-A, but not Type II-A, spacer uptake. These regions included the single-strand origins of rolling circle plasmids, rRNA and tRNA gene clusters, and promoter

regions of expressed genes. I also found that the Type III-A system could adapt and protect the cell from a lytic phage. Collectively, my findings indicate that the Type III adaptation machinery preferentially targets DNA secondary structures including imperfect hairpins and other partially double-stranded DNAs. I will discuss how this Type III spacer selection strategy may skew adaption toward the expressed genes of mobile genetic elements which can be recognized and targeted by III-A crRNAs.

## Introduction

CRISPR-Cas systems are diverse, adaptive and heritable immune systems that protect many bacteria and archaea against potentially harmful viruses, plasmids and other invasive mobile genetic elements (1-6). The specific defense activity of a CRISPR-Cas system is acquired by the adaptation phase, during which a short fragment (protospacer) of the foreign DNA can be captured and integrated into the CRISPR locus at the leader proximal end as a spacer, simultaneously with a duplication of the first repeat (2,7-10). Subsequent expression and processing of the repeat-spacer units that comprise the CRISPR arrays, generates a pool of individual CRISPR RNAs (crRNAs) that each have the potential to direct Cas effector nucleases to destroy complementary nucleic acids of the invasive mobile genetic elements (2,11,12). Of the six broad types of CRISPR-Cas systems identified (I-VI) (3,13,14), Types I, II and VI systems recognize and destroy invasive DNAs (15-20), while Type VI systems target RNA destruction (15,21-23) (the mechanisms of action for Type IV systems are yet not

understood). Type III systems are unique in that they recognize complementary target RNAs but are capable of destroying both transcriptionally active DNAs and the RNA products of viruses and plasmids (Figure S1) (24-30).

Recent progress has offered key insight into the detailed molecular mechanisms governing the ability of various CRISPR-Cas systems to recognize, excise, process and accurately integrate spacers at CRISPR arrays (a process known as CRISPR adaptation). Much of the work performed to date has centered on how distinct Type I and II (DNA targeting) systems operate. The available evidence indicates a key and universal role for Cas1-Cas2 integrase complexes in capturing spacer DNA and catalyzing integration of the DNA fragments at CRISPR loci (2,8,31-34). Moreover, an important role for specific sequence elements within the leader (a variably sized DNA region positioned immediately upstream of the CRISPR array) were determined to be critical in directing spacer integration at the leader-proximal repeats of both Type I and Type II systems (31,35-37). The polarized addition of new spacers specifically at the leader end of the CRISPR array generates a chronological record of past invasions (38). The key leader sequences have been shown to function either by providing recognition sites for direct binding of Cas1-Cas2 complexes (39) or for binding of additional non-Cas factors (such as integration host factor (IHF) for Type I-E and I-F systems) needed to recruit Cas1-Cas2 complexes to the leader-proximal CRISPR repeat (40-42). Biochemical and structural studies have revealed that spacer DNA integration proceeds via a two-step transesterification reaction whereby the 3’ hydroxyl groups of the incoming spacer DNA each attack the

terminal nucleotides of the CRISPR repeat element, followed by host cell repair of the gapped DNA intermediates (43,44). While all investigated systems were found to rely on Cas1 and Cas2, the various steps of spacer generation and integration have been shown to require additional Cas (e.g. Cas4, Cas9, Csn2) (2,33,34,45) and non-Cas factors (e.g. IHF, RecBCD, DNA polymerase) (40,42,46-50) in different systems and in different organisms. Indeed, in some systems but not others, efficient spacer acquisition by the adaptation machinery (including Cas1, Cas2 and any additional adaptation factors) depends upon the presence of the effector crRNP interference machinery required also for sequence-specific destruction of foreign nucleic acids (51). Adaptation is a highly infrequent event that only occurs in few cells of a population for all tested CRISPR-Cas systems. Moreover, the protospacer sources of spacers can be either foreign nucleic acids (leading to immunity) or the host chromosome (leading to autoimmunity and cell death) (33,52,53). Additional work with less characterized CRISPR-Cas systems is required to more fully comprehend both the common and unique mechanistic features associated with each diverse CRISPR-Cas system as well as the contribution of host cell environment to spacer acquisition process.

We are only now just beginning to gain initial insights into mechanisms directing adaptation by Type III systems. Type III systems are among the most abundant and widespread CRISPR systems (second only to Type I systems) and occur in both bacteria and archaea (3). Type III systems have been classified into six distinct subtypes (Type III-A to F) based on the properties of the interference

proteins including Cas10 signature protein and other distinguishing features (3,54). Interestingly, most Type III-A systems include genes encoding their own Cas1 and Cas2 adaptation proteins, CRISPR arrays and crRNA processing (Cas6) enzymes, however, the majority of other Type III systems (III-B, C, D, E and F) lack these features and are predicted to be unable to adapt (1,3,54). Instead, there is evidence that these Type III subtypes co-occur with Type I systems and appear be functionally dependent upon the Type I systems to provide them with processed crRNAs (55). The picture that is emerging is one where most Type III systems (besides Type III-A) serve primarily as “back-up” interference modules that are acquired by horizontal gene transmission and provide major benefits to the organism including the ability to target at the RNA level and to resist phage escape by Type I systems (55). A notable exception to this trend is the rare reverse-transcriptase (RT)-Cas1 fusion systems that enable some Type III systems to adapt (55-58). While the other reported systems can only adapt DNAs as CRISPR spacers, the Type III systems with RT activity can use both RNAs and DNAs as the substrates, and the adaptation against RNAs is dependent on the RT (55-58). Work to understand how the RT-free Type III systems adapt has just begun. The best characterized Type III-A system of *S. epidermidis* (59), has yet to be observed to carry out adaptation. In contrast, a Type III-A system of *Thermus thermophilus* was very recently shown to be active against a phage and documented that new spacers were preferentially acquired from a region of the phage predicted to harbor genes that are expressed early in phage infection (60). Further work is necessary to understand in greater detail

the properties used by Type III-A systems to recognize and integrate foreign nucleic acid spacer sequences into their affiliated CRISPR arrays.

Type III systems are particularly interesting systems to explore adaptation properties since they differ in several important ways to the more well-studied Type I and II systems and are therefore expected to have unique attributes. For example, the DNA targeting systems rely on a short (2-5 bp) PAM (protospacer adjacent motif) both as a signal recognized during adaption and interference (61-64). During adaption, the PAM plays a critical role in directing spacers integration in an orientation that leads to a functional CRISPR RNA. At the interference stage, PAMs play a key role in self- vs. non-self, recognition as invader DNA but not captured spacer DNA in CRISPR arrays, contain PAMs (Figure S1) (16,61,62,65). In contrast, Type III systems appear to lack PAM sequences but like Types I and II systems, newly acquired spacers must be integrated in one of two possible orientations to give rise to functional crRNAs capable of recognizing the target RNAs by complementary base-pairing (Figure S1) (56,58,60). Analyses of native Type III-A CRISPR arrays reveal a major bias for spacers inserted in the sense orientation relative to their viral or plasmid RNA targets (66). However, it is not yet fully understood if Type III-A systems utilize a novel (PAM-independent) mechanism to ensure capture of spacers in a particular orientation or if as early work suggests, spacers are initially acquired in both orientations and downstream processes ultimately preserve spacers that impart a selective advantage to the organism and purge those that do not confer immunity or induce autoimmunity (60). It is also unclear if the Cas1 and Cas2 proteins are

sufficient for executing adaptation or if (like other Type I and II systems) the interference components including crRNA, Csm1-6 proteins are needed as well. Given that Type III systems ultimately recognize target RNAs from expressed mobile genetic elements, an important question is whether mechanisms governing spacer choice have evolved to enhance the probability of leading to functional spacers or if instead spacers are chosen on a random basis.

In this chapter, I demonstrate that the native Type III-A system of *Streptococcus thermophilus* (*Sth*) JIM 8232 strain is active for executing adaptation against plasmid and phage invaders. By comparing the adaptation properties of Type III-A and a Type II system that co-exists in the strain, I show that Type III-A systems differ with regard to size range of spacers integrated, lack of PAM, and apparent lack of integration orientation bias. I show that Cas1 and Cas2 are necessary and sufficient for Type III-A adaptation and performing assays in the absence of an interference pathway revealed that spacers are normally integrated in both possible orientations but a strand bias found in wild-type (WT) strains is due to negative selection against cells that incorporate self-targeting spacers capable of triggering lethal autoimmunity or plasmid loss under selective growth conditions. Through challenging *Sth* with a variety of different plasmids and phage invaders, I found that spacer choice is non-random and that DNA secondary structures (e.g. hairpins formed on single-stranded rolling-circle replication intermediates or as a results of predicted R-loop formation during transcription or replication) are preferentially targeted by the Type III-A adaptation machinery. I discuss how this spacer selection strategy may promote

recognition of mobile genetic elements vs. host genomes and skew adaption toward the expressed genes of mobile genetic elements capable of being recognized by RNA-targeting Type III-A acquired crRNAs.

## Materials and methods

### *Strain and plasmid manipulation*

*Sth* JIM8232 was kindly provided by Dr. Pierre Renault. *Sth* DGCC7710 and phage 2972 were kindly provided by Dr. Sylvain Moineau. *Sth* strains were inoculated in M17 medium supplemented with 0.5% lactose (LM17) (Oxoid or HiMedia), and the cultures were incubated at 37$^{o}$C overnight, or at 42$^{o}$C during the day. *E. coli* Top10 was used for plasmid construction and maintenance. *E. coli* Stellar (*dcm*-/*dam*-) was used to generate unmethylated plasmids for the target interference assay. pWAR, pTRK882, pNT1, pNZ123, pIB184, pG+Off, and pRSNPed plasmids were kindly provided by Drs. Michael Federle, Todd Klaenhammer, Sylvain Moineau, Indranil Biswas, Marie-Frédérique Lartigue, and John Renye. M13mp18 single-stranded DNA (ssM13) and RF DNA (dsM13) were purchased from New England BioLabs. When needed, chloramphenicol was supplemented at 2 μg/mL in LM17 liquid broth, and at 5 μg/mL in LM17 plates (with 1% agar) for *Sth*; erythromycin and kanamycin were supplemented at 15 μg/ml and 150 μg/ml for *Sth*, separately. The construction of the *Sth* mutant strains were achieved by a well-developed natural transformation procedure (67). The primers used for PCR amplification of the recombination templates are listed in Table 2.1.

### *Target interference assay*

The *Sth* strains were inoculated in 5 mL LM17 and the cultures were incubated at 37°C overnight. 100 mL fresh LM17 was inoculated with 1% of an overnight culture, and incubated at 42°C until $OD_{600}$ value reached 0.5. The culture was then placed on ice for 15 minutes and then centrifuged for 10 minutes at 5000 RCF at 4°C. Supernatant was decanted and the pellet was resuspended in 2 mL ice-cold wash solution (10% glycerol + 0.4 M sorbitol). The resuspended cells were washed three times by centrifuge for 1 minute at 15000 RPM at 4°C, followed by the resuspension in the wash solution. After the final wash, the electroporation-competent cells were resuspended in 500 µL wash solution and aliquoted. One µg unmethylated target or control plasmid DNA (generated in the *E. coli* Stellar strain), was mixed into 40 µL competent cells, and electroporated into the cells by Gene Pulser (BioRad) at 25 µF, 200 Ω, and 1.8 KV. The transformants were incubated in 1 mL recuperation solution (LM17 + 0.4 M sorbitol + 20 mM $MgCl_2$ + 2 mM $CaCl_2$) for 2 hours at 42°C, and then plated onto an LM17 plate (1% agar) with the appropriate antibiotic.

### *Adaptation assay*

The foreign plasmids were transformed by the well-developed natural transformation procedure (67).Twenty to thirty colonies of the transformed *Sth* strains were inoculated into 10 mL LM17 with appropriate antibiotics, and the cultures were incubated at 37°C for overnight. The total DNA of the overnight

cultures was extracted by Quick-DNA Fungal/Bacterial Miniprep kit (Zymo Research).

To monitor adaptation, the leader-proximal end of the CRISPR array was amplified by CAPTURE PCR or multiple round PCR from the extracted DNAs (Figure 1B). PCR primers are listed in Table 2.1. Expanded amplicons from the first round PCRs were separated from unexpanded products by gel electrophoresis, bands of the correct size were cut, and DNAs were isolated by a gel recovery kit (Zymo Research). When expanded amplicons were too faint to visualize, the region of a gel lane corresponding to amplicons in the expanded size range was cut. Illumina high-throughput sequencing (HTS) overhangs and Illumina HTS index barcodes were added to the expanded array amplicons by PCR. Purified PCR products were ranked by PCR intensity and then pooled, concentrated by ethanol precipitation, quantitated, and diluted to a suitable concentration for Illumina platform sequencing.

Array libraries were sequenced on an Illumina MiSeq, set to yield 250 by 50 paired end reads; the 250 base read 1 sequences were used in this study. After sequencing, samples were de-multiplexed by index, and the sequence corresponding to a new (expanded) spacer was extracted from each read. New spacers were aligned to reference sequences (bacterial chromosome and appropriate plasmids) using Bowtie (68) to identify the protospacer sequence. Protospacer sequences were then characterized with respect to length, PAM, and position on the genome or plasmid. To detect PAMs, protospacer adjacent

upstream and downstream sequences were extracted using bedtools (69) and a consensus sequence logo was made using weblogo (70).

### *RNAseq assay*

To evaluate RNA expression patterns, RNA sequencing was done on cultures grown to either exponential or stationary phase. Briefly, cultures were pelleted and decanted, then frozen at -80°C. Pellets were thawed and resuspended directly in lysis buffer and RNA was isolated using the PowerBiofilm RNA Isolation kit (Qiagen). Stranded, total RNA libraries (with no rRNA or tRNA depletion) were prepared using the Illumina TruSeq kit and were sequenced on an Illumina NextSeq instrument, generating paired 2 by 150bp reads. Reads were demultiplexed by index, adapter trimmed, and aligned to the appropriate reference sequences (bacterial chromosome, plasmid) by bowtie2 (71).

For both the adaptation and RNAseq assays, alignment outputs were processed (69,72) and custom genome browser tracks were generated using tools available from the University of California Santa Cruz Genome Browser (https://genome.ucsc.edu). We noted that the distribution of protospacers across the genome was not even, so regions with significant protospacer enrichment of were identified using the findPeaks software in the HOMER analysis package (73). For spacers that aligned within the boundaries of annotated protein-coding genes, we determined the percent that matched the coding versus template strand of those annotated genes using a custom python script.

***Phage infection and BIM isolation and analyses.***

The *Sth* strains were inoculated in 5 mL LM17 and the cultures were incubated at 37°C overnight. Five mL fresh LM17 with 10 mM $CaCl_2$ was inoculated with 1% of an overnight culture, and incubated at 42°C until $OD_{600}$ value reached 0.3. Phage infection was performed with phage 2972 at multiplicity of infection (M.O.I.) of 0.1, 1, and 10, separately. After phage addition, Cells were kept incubated at 42°C for 60 minutes, 90 minutes, or overnight. At the end of each of the three time points, 100 µL of each culture was plated onto an LM17 plate with 10 mM $CaCl_2$. The plates were incubated at 37°C overnight, and then individual colonies were ramdomly picked and evaluated for CRISPR array expansion by colony PCR, using the primers listed in Table 2.1. The remaining volume of each of the cultures was subjected to total DNA extraction, CAPTURE PCR, and HTS to evaluate adaptation as described above.

## Results

***Type III-A system of Streptococcus thermophilus actively acquires new spacers at the CRISPR locus.***

*Sth* JIM8232 has an intact Type III-A CRISPR-Cas system in its genome, containing 17 pre-existing spacers, as well as a Type II-A system, and another Type II CRISPR array without *cas* genes adjacent to it (Figure 2.1A) (74). Interestingly, the sequence of Spacers 3-8 of the Type III-A system is identical to those of Spacers 9-14.

To test adaptation ability of the Type III-A CRISPR-Cas system, I transformed *Sth* cells with pWAR plasmid (75) as a heterologous DNA, and examined the leader proximal ends of the CRISPR arrays by PCR and HTS (Figure 2.1B). As a comparator to the Type III system, we examined the adaptation by the Type II-A system in the same host in parallel. The homologous Type II system (CRISPR1-Cas) of another well-studied *Sth* strain, DGCC7710, has been shown very efficient in adaptation (2,33,37). However, adaptation by the Type II system of *Sth* JIM8232 strain was inefficient as a result of polymorphisms within the leader sequence (Figure S2A). As a consequence, we recovered the high efficient Type II adaptation phenotype by inserting the leader sequence and the following 4 repeat-spacer units of the DGCC7710 CRISPR1 array into a pseudogene of JIM8232 genome that encoded non-functional components of the glucose phosphoenolpyruvate-dependent phosphotransferase system (PTS locus). The short DGCC7710 CRISPR1 array at PTS locus kept the adaptation features (Figure S2B and C) and mediated defense against the tested target plasmid (Figure S2D). We then monitored adaptation by the Type II system at PTS locus in the new strain.

The sizes of the pre-existing spacers of Type III-A CRISPR locus ranged from 35 to 39 bp, with 36 bp to be the most frequent one. The lengths of about 99% of the unique reads of new spacers acquired by the Type III-A system fell into a roughly normal distribution from 32 to 42 bp, with 36 bp as the peak of the curve (Figure 2.1C). As a comparison, more than 80% of the new spacers acquired by the Type II-A system were 30 bp, with the others being 29 or 31 bp

(Figure 2.1C). As to well-studied Type I and Type II systems, the protospacers are selected along foreign DNAs by system-specific PAMs, and are inserted at the leader proximal ends of the arrays in a PAM directed orientation (61-64), while the adaptation by the tested Type III systems are PAM-independent (56,58,60) (Figure S1). We examined the particular DNA bases of 5'- and 3'-protospacer flanking regions (PFSs) of the protospacers selected by the Type III-A system, and no consensus sequence motif was identified (Figure 2.1D), which was expected since target interference by the Type III systems tolerates a broad range of PFSs (76). To rule out the possibility of survivor bias, since CRISPR-Cas systems diminish the adapted nucleic acids, we examined the interference-inert strains (Figure 2.2B), and still did not find any consensus PFS. In contrast, Type II-A system in the same host selected protospacers with a downstream NNAGAAW PAM (Figure 2.1D). The majority of new spacers were mapped to self-genome, which was not surprising, since the genome is about 464-fold larger than the plasmid (Figure 2.1E).

### *Cas1 and Cas2 are the only two Cas proteins required for adaptation by the Type III-A system*

Little was known about which Cas proteins were involved in adaptation by the Type III CRISPR-Cas systems, since it had been tested only in the three studies with Type III systems employing RT-Cas1 fusion protein (55-58). In this work, new spacer acquisitions were observed by four round PCR for wild-type (WT) strain, the *cas6* knockout (KO) strain, and the *csm* genes KO strains, but

not for the *cas1* or *cas2* KO strains (Figure 2.2A). These results indicate that Cas1 and Cas2 are essential for adaptation by the Type III systems and sufficient for Cas protein requirement.

### *Adaptation-independent replication of the repeat-spacer units in both Type III-A and Type II-A systems*

A large number of the new spacers appeared to be derived from the pre-existing spacer1 of the Type III-A system, and moreover, all the pre-existing spacers were ‘adapted’ at the leader proximal end of the CRISPR array (Figure 2.2C). Interestingly, the same ‘adaptation’ against the pre-existing spacers also occurred in the arrays of the *cas1* and *cas2* KO strains, while the adaptation against the plasmid and genome was not detected except for several rare unique spacers which might be caused by contamination (Figure 2.2C). This indicated that the insertion of the repeat-spacer units at the leader proximal end were likely caused by homologous recombination or other DNA replication procedures at the repeat-rich regions during genomic replication, instead of adaptation events of the CRISPR-Cas system. The extensive repeat-spacer replications were also detected in the Type II CRISPR-Cas locus (Figure 2.2D), indicating that adaptation-independent replication of the repeat-spacer units happens in both Class 1 and Class 2 CRISPR-Cas systems (Figure 2.2E).

### *Type III spacers are integrated in both possible orientations*

Target interference by the Type III CRISPR-Cas systems requires the directional transcription of the target, as the DNase activity of the crRNPs and the RNase activity of Csx1 (Csm6 for Type III-A) are stimulated by base pairing between the guiding crRNAs and the transcript of the target DNAs (28,77-84). As a consequence, only when the template strands of the double-stranded spacers are integrated into the top strand of the Type III CRISPR array can defense be triggered; in contrast, spacers with the encoding strand on the top strand are not bona fide spacers (Figure 2.3A).

Nearly all the pre-existing spacers of Type III-A systems of *Sth* strains were from the template strands of the heterologous DNAs (Figure 2.3B). Surprisingly, in this work, most of the new protospacers detected were from the coding strands of the plasmid or self-genome. However, after knocking out the interference-related *csm* genes, approximately half of the spacers were acquired from the coding strands while the other half were from the template strands (Figure 2.3C, E, and F). These findings indicate that self-targeting spacers were subject to counter selection, as they are capable of triggering ‘defense’ by the Type III crRNPs. Interestingly, the most highly transcribed regions of the rRNA and tRNA encoding clusters were most extensively sampled by the Type III-A adaptation machinery but not by the Type II-A system (Figure 2.3D and S3). Moreover, the Type III adaptation machinery was revealed a coding strand preference in the rRNA and tRNA clusters even in the absence of *csm* genes (i.e., without the influence of the counter selection) (Figure 2.3D and S3). This phenomenon is

discussed further below (See Impact of transcription on protospacer selection and Discussion).

### *Selective targeting of partially palindromic sequences by Type III adaptation*

pWAR is a rolling-circle replicating (RCR) plasmid (85,86). The Rep protein encoded by a RCR plasmid recognizes the *double-strand origin* (*dso*) of the plasmid, generates a double-stranded DNA (dsDNA) nick at the *dso*, and peels off the 5'-end of the plus strand. The plus strand is continuously replicated, which is initiated by the cognate 3'-OH end of the parental plus strand. The plus strand is displaced and re-ligated after the replication. The s*ingle-strand origin* (*sso*) is partially palindromic and forms a hairpin structure when the plus strand is displaced in the circular single-stranded DNA (ssDNA) form, and the hairpin structure can trigger the replication of the minus strand (Figure 2.4A) (87,88).

The new spacers acquired by the Type III-A system which target pWAR formed consistent peaks covering the *sso*. This was further amplified when the interference-required *csm* genes were absent, with two extraordinary peaks of mapped spacers appeared at the end of *sso* (Figure 2.4B). Since *sso* formed a hairpin structure during its ssDNA stage, we reasoned that this imperfect dsDNA structure could be a preferred substrate of Type III adaptation, and the adaptation peaks of WT system were not as pronounced due to plasmid elimination by downstream target interference.

To further test this finding, we monitored adaptation by both Type III-A and Type II-A CRISPR systems against 3 other RCR plasmids, pTRK882 (89), pNT1 (90), and pNZ123 (91); 3 theta replicating plasmids, pIB184 (92), pRSNPed (93), and pG+Off (94); and non-replicating ssM13 and dsM13 DNAs, separately. For the Type III-A system, all the *sso*s of RCR plasmids were recognized by the adaptation machinery and covered by new spacer mapping peaks, while the replication origins of the theta replicating plasmids were not preferred DNA substrates for adaptation (Figure 2.4C and S4). pNT1 contains a *mob* gene for conjugation, as well as a putative *oriT* with its $RS_A$ recombination site, inverted repeats and nick site (95). Interestingly, both the *sso* and *oriT* of pNT1 were highly sampled by the Type III-A system (Figure 2.4D). Unlike the other plasmids, M13 DNAs are not able to replicate in *Sth* cells and do not contain any selective marker, hence detection of spacers targeting these DNAs is difficult. Despite this, we still successfully detected 5 unique spacers from dsM13 DNA, as well as 3 unique spacers from ssM13 DNA over 6 independent experiments (Table 2.2). All the protospacers of ssM13 were located within *lacI* gene, which was a hairpin structure enriched region (96), and they were located at the partially palindromic sequences. These findings support our hypothesis that stem-loop structures or other secondary structures formed by a ssDNA could serve as additional substrates of adaptation by the Type III-A CRISPR system.

Although the *sso* and *dso* sequences, as well as the *oriT* of pNT1 plasmid were adapted by the Type II-A system, they were not preferred over the other

parts of the plasmid (Figure S4), revealing the specificity of the features of the Type III-A adaptation.

Besides the functions in the life cycles of plasmids, partially palindromic sequences also play an important role in expression control of some genes. For example, the replication rate-limiting gene *repD* of pIB184 is regulated by antisense RNA III and *cis*-acting sequences, which are enriched in partially palindromic structures (Figure 2.5A) (97,98). The upstream sequences of ribosomal protein L10 and amino acid synthetase genes are enriched of stem-loop structures (Figure 2.5B and C), which play important roles in expression control of the genes (99-101). These partially palindromic sequences in the plasmids and self-genome were well-sampled by the Type III-A CRISPR-Cas systems but not by the Type II-A systems (Figure 5D-I).

***Impact of transcription on protospacer selection***

We noticed strong adaptation against several highly transcribed genes of the plasmids and self-genome, as well as the extensive adaptation against the 5 rRNA and tRNA encoding regions, by the Type III-A system, but not by the Type II-A system (Figure 2.3, 2.5, S3, and S4). More specifically, strong adaptation around the transcription start sites of some highly transcribed genes were observed for the Type III-A system (Figure 2.6A-C). This raised the question of whether transcription of the substrates impacted protospacer selection during adaptation by the Type III-A system of *Sth*, despite the system being RT-free.

To test this hypothesis, we inverted the p23 promoter in the multiple cloning site of pIB184 (which would not influence the gene expression or the stability of the plasmid), and compared between the adaptation against the new and the original pIB184 plasmid. For the Type III-A system, the accumulation of the protospacers following the transcription start site was diminished by the p23 inversion, which in turn facilitated the adaptation against the upstream region, a lowly adapted region in the original pIB184 plasmid (Figure 2.6B). In contrast, the strong promoter had no influence on the adaptation by the Type II-A system (Figure 2.6B). Moreover, when correlating adaptation strength to transcription level, we observed that the first ~50 bp of the transcribed regions were apparently preferred by the adaptation of the Type III-A system (Figure 2.6D). Except for the most highly transcribed rRNA and tRNA encoding regions (Figure 2.6E), no direct correlation was found between the adaptation strength and the transcription level of the substrates (Figure 2.6E and F). Since the Type III-A system had no RT activity, this feature was not caused by direct adaptation against RNAs; instead, since the system appears to recognize secondary structures of the DNA substrates, we concluded that when the DNA substrates being transcribed, the encoding strand was displaced and had the potential to form the secondary structures to serve as the additional adaptation substrates for the Type III-A system (Figure 2.8).

### *Type III systems mediate adaptive defense against the lytic phage*

No lytic phage has been identified for *Sth* JIM8232 strain, while phage 2972 had been shown lytic for *Sth* DGCC7710 strain (with a degraded Type III-A system) (102). To test whether the Type III-A CRISPR-Cas system is able to mediate defense against lytic phage, we replaced the native Type III-A CRISPR-Cas system of DGCC7710 by the one of JIM8232, and co-cultivated the new strain with phage 2972 at different multiplicities of infection (M.O.I.) and for different durations. By HTS, we detected 13 unique spacers acquired from the phage, and they were roughly equally distributed at the coding and the template strands (Figure 2.7A). Interestingly, nearly all the adaptation events were detected at the early transcribed regions of the phage (Figure 2.7A). Since the phage was lytic to the hosts, we reasoned that cells had a limited time period to adapt against the phage after infection, during which only the early genes were transcribed and became more sensitive to the adaptation by the Type III-A CRISPR-Cas system. This is theoretically beneficial for the defense mediated by the system, because targeting the transcripts of the early genes allows timely authentication of the Type III-mediated defense.

We successfully isolated two bacteriophage insensitive mutants (BIMs), which acquired a new spacer from the template strands of the phage into their Type III CRISPR-Cas locus and did not acquire any new spacer in their other CRISPR loci (Figure 2.7B and S5). These findings show that the Type III-A CRISPR-Cas system is able to mediate the defense against the lytic phage by adaptation and subsequent target interference.

## Discussion

Type III CRISPR-Cas systems are distinguished from the other types of systems by their unique and diverse mechanisms of target interference (Figure 2.2A and S1) (24). Type III systems have been used for genomic editing of the hyperthermophiles, in which the CRISPR-Cas9-based modules might not function (103), and their target RNA cleavage abilities have been used for gene expression control (29,104). Nevertheless, our knowledge about adaptation by the Type III systems is very limited. After 2016, three Type III systems with RT activity were reported to acquire both RNA and DNA substrates during adaptation (55-58), however, since less than 10% of Type III systems have RT activity, these systems were not representative. Only very recently, a RT-free Type III-A system of *Thermus thermophilus* was shown to actively acquire new spacers from phage phiFa to perform the defense (60). In this chapter, we for the first time, provided a detailed analyses of the properties of adaptation by the Type III CRISPR-Cas system.

We compared the patterns of adaptation by the Type III-A and the Type II-A CRISPR-Cas systems of *Sth* JIM8232 against different RCR plasmids and theta-replicating plasmids, as well as the host genome. A prominent and intriguing feature of the adaptation by the Type III system was the apparent recognition of the *sso*s of the RCR plasmids, contrasting with that of the Type II system. RCR plasmids produce ssDNA intermediate forms during their replication, and the long and partially palindromic *sso*s form stem-loop structures to trigger the synthesis

of the minus strand. We reasoned that the stem-loop structures served as additional dsDNA substrates for adaptation of the Type III system, making the *sso*s preferentially adaptation substrates. Similarly, the partially palindromic *oriT* sequence of pNT1, and the stem-loop structures enriched regulatory regions of pIB184 *rep* gene and the genomic genes, as well as the clover structure enriched rRNA and tRNA encoding regions of self-genome, were also enriched in Type III adaptation, but not in Type II adaptation. Most of the natural plasmids of gram-positive bacteria, including *Sth*, and many of those of gram-negative bacteria are RCR plasmids (105). Moreover, the crucial structure of *oriT* and other DNA secondary structures are important for the conjugative transfer and other functions of environmental mobile genetic elements (106). As a consequence, it seems likely that secondary structure recognition by the Type III CRISPR-Cas system is beneficial for the system to specifically and efficiently eliminate the invaders (Figure 2.8).

Cas1 and Cas2 have been shown to be essential for adaptation by all the tested CRISPR-Cas systems (2,31-34,41,56). The Type III-A CRISPR-Cas system of *Sth* requirs Cas1 and Cas2 proteins, but not Cas6 or any interference-related Cas proteins for adaptation. After knocking out all the *csm* genes, the system retained the ability to adapt, and the intrinsic features of adaptation. Intriguingly, although the adaptation was inert after knocking out *cas1* or *cas2* genes, I observed the replication of the repeat and the pre-existing spacer units, indicating that such repeat-spacer duplication is an adaptation-independent cellular event. Such replication was observed in both the Type III and Type II

systems, suggesting that it is a universal feature of all or many of the CRISPR-Cas systems. We hypothesize that homologous recombination or DNA replication errors in the repeat-rich region can help the CRISPR-Cas systems to replicate recently acquired spacers to enhance the expression of the crRNAs, as well as to lose the old spacers to keep a compact CRISPR array. This may explain why the sequences of Spacers3-8 of the Type III-A system of *Sth* JIM8232 are identical to those of Spacers9-14. While the analyses presented here would be unable to detect spacer loss, such loss has been observed in a study regarding a Type I CRISPR-Cas system (107).

In 2016, Silas et al. reported adaptation by the Type III-B system of *Marinomonas mediterranea* (MMB-1), which expresses a RT-fused-Cas1 protein (56). While RT-free CRISPR-Cas systems can only adapt DNAs as new spacers, this Type III-B system can use both RNAs and DNAs as the substrates, and the adaptation against RNAs is dependent on RT. This additional adaptation against RNAs makes the system preferentially acquire new spacers from highly transcribed regions versus lowly transcribed regions, which is beneficial for the function of the system, since target interference by the Type III systems requires transcription of the targets (56). Soon after this exciting finding, a similar RT-Cas1-Cas2 complex of *Fusicatenibacter saccharivorans* was used as a novel and efficient tool to record transcription events in *E. coli* (57). A similar RT-mediated Type III adaptation against highly transcribed regions was then reported by Gonzalez-Delgado et al. in 2019, and moreover, they observed a preference toward the encoding strand of the rRNA genes (58). They speculated

that the rRNA encoding strand preference was also caused by RT and there was a correlation between the gene transcription and the new spacer orientation (58). In this work about the Type III CRISPR-Cas system of *Sth* JIM8232, we found no strong correlation between transcription level of the DNA substrates and adaptation efficiency against them. However, we observed that the first ~50 bp regions following strong transcription start sites were preferentially adapted by the RT-free system. This feature was further confirmed by our investigation into the influence of the p23 promoter on adaptation. The extensive adaptation against rRNA as well as tRNA encoding sequences was also observed in adaptation by the Type III system of *Sth*, indicating that this Type III-specific preference was RT-independent and was not caused by direct adaptation against the RNA substrates. We reasoned that during the transcription of the highly expressed regions, the coding strand was displaced in a ssDNA form, which can form secondary structures and be recognized by the Type III system. In particular, the rRNA and tRNA coding sequences are the most highly transcribed regions of a genome and are enriched in the clover-leaf structures in their ssDNA forms. Adaptation against the coding strands of highly transcribed regions were more efficiently detected over the template strands in our analyses, but such strand bias was abolished after knocking out the interference related *csm* genes, indicating that the prespacers were integrated into the CRISPR array in both possible orientations and the bias in the WT system was caused by counter-selection.

Very recently, Artamonoka et al. observed and reported adaptation against a lytic phage by a RT-free Type III system of *Thermus thermophilus* (60). They found that the system performed robust adaptive defense against a lytic phage, phiFa. The protospacers detected by HTS had a strand bias in that the template strand of the phage was adapted more extensively than the coding strand, which was caused by survivor bias, since the crRNAs of the Type III system needed to bind to the mRNAs of the phages to be functional. More interestingly, they found that the long terminal repeat (LTR), the firstly invading region and early transcribed part of the phage, was adapted substantially more efficiently than the other parts of the phage, and they reasoned that maybe the LTR region encoded an anti-CRISPR element that blocked the functions of the CRISPR-Cas system (60). In light of our results, it seems plausible that the LTR formed secondary structures since it was a repeat-rich region, including palindromic, direct, and inverted repeats, and such structures could be recognized by the Type III CRISPR-Cas system; or only the adaptation against the early transcribed genes could perform the timely defense against the phage. In this chapter, I also isolated two *Sth* DGCC7710 BIMs that separately acquired one new spacer from the template strand from phage 2972 in the transplanted Type III-A CRISPR-Cas system, indicating that the Type III-A system of *Sth* performed robust adaptive defense against the phages as well. Moreover, nearly all the new spacers detected by HTS and the two new spacers of the isolated BIMs came from the early transcribed genes of the phage, suggesting that only the crRNAs targeting

early transcribed genes could mediate timely defense against lytic viruses for Type III CRISPR-Cas systems.

## References


1. Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F. *et al.* (2011) Evolution and classification of the CRISPR-Cas systems. *Nature reviews. Microbiology*, **9**, 467-477.
2. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709-1712.
3. Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H. *et al.* (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nature reviews. Microbiology*, **13**, 722-736.
4. Terns, M.P. and Terns, R.M. (2011) CRISPR-based adaptive immune systems. *Current opinion in microbiology*, **14**, 321-327.
5. Terns, R.M. and Terns, M.P. (2013) The RNA- and DNA-targeting CRISPR-Cas immune systems of Pyrococcus furiosus. *Biochemical Society transactions*, **41**, 1416-1421.

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

7. Agari, Y., Sakamoto, K., Tamakoshi, M., Oshima, T., Kuramitsu, S. and Shinkai, A. (2010) Transcription profile of Thermus thermophilus CRISPR systems after phage infection. *Journal of molecular biology*, **395**, 270-281.

8. Bhaya, D., Davison, M. and Barrangou, R. (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annual review of genetics*, **45**, 273-297.

9. Heler, R., Marraffini, L.A. and Bikard, D. (2014) Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. *Molecular microbiology*, **93**, 1-9.

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

11. Pougach, K., Semenova, E., Bogdanova, E., Datsenko, K.A., Djordjevic, M., Wanner, B.L. and Severinov, K. (2010) Transcription, processing and function of CRISPR cassettes in Escherichia coli. *Molecular microbiology*, **77**, 1367-1379.

12. Sorek, R., Lawrence, C.M. and Wiedenheft, B. (2013) CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annual review of biochemistry*, **82**, 237-266.

13. Kunin, V., Sorek, R. and Hugenholtz, P. (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome biology*, **8**, R61.

14. Shmakov, S., Abudayyeh, O.O., Makarova, K.S., Wolf, Y.I., Gootenberg, J.S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K. *et al.* (2015) Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Molecular cell*, **60**, 385-397.

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

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

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

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

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

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

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

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

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

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

25. Staals, R.H., Zhu, Y., Taylor, D.W., Kornfeld, J.E., Sharma, K., Barendregt, A., Koehorst, J.J., Vlot, M., Neupane, N., Varossieau, K. *et al.* (2014) RNA targeting by the type III-A CRISPR-Cas Csm complex of Thermus thermophilus. *Molecular cell*, **56**, 518-530.

26. Tamulaitis, G., Kazlauskiene, M., Manakova, E., Venclovas, C., Nwokeoji, A.O., Dickman, M.J., Horvath, P. and Siksnys, V. (2014) Programmable RNA shredding by the type III-A CRISPR-Cas system of Streptococcus thermophilus. *Molecular cell*, **56**, 506-517.

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

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

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

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

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

32. Nunez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W. and Doudna, J.A. (2014) Cas1-Cas2 complex formation mediates spacer

acquisition during CRISPR-Cas adaptive immunity. *Nature structural & molecular biology*, **21**, 528-534.

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

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

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

36. Grainy, J., Garrett, S., Graveley, B.R. and M, P.T. (2019) CRISPR repeat sequences and relative spacing specify DNA integration by Pyrococcus furiosus Cas1 and Cas2. *Nucleic acids research*, **47**, 7518-7531.

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

38. Barrangou, R. and Dudley, E.G. (2016) CRISPR-Based Typing and Next-Generation Tracking Technologies. *Annual review of food science and technology*, **7**, 395-411.

39. Van Orden, M.J., Newsom, S. and Rajan, R. (2020) CRISPR type II-A subgroups exhibit phylogenetically distinct mechanisms for prespacer insertion. *The Journal of biological chemistry*, **295**, 10956-10968.

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

41. Fagerlund, R.D., Wilkinson, M.E., Klykov, O., Barendregt, A., Pearce, F.G., Kieper, S.N., Maxwell, H.W.R., Capolupo, A., Heck, A.J.R., Krause, K.L. *et al.* (2017) Spacer capture and integration by a type I-F Cas1-Cas2-3 CRISPR adaptation complex. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, E5122-E5128.

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

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

44. Rollie, C., Schneider, S., Brinkmann, A.S., Bolt, E.L. and White, M.F. (2015) Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *Elife*, **4**.

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

46. Wu, C., Tang, D., Cheng, J., Hu, D., Yang, Z., Ma, X., He, H., Yao, S., Fu, T.M., Yu, Y. *et al.* (2021) Mechanisms of spacer acquisition by sequential assembly of the adaptation module in Synechocystis. *Nucleic acids research*, **49**, 2973-2984.

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

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

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

50. Radovcic, M., Killelea, T., Savitskaya, E., Wettstein, L., Bolt, E.L. and Ivancic-Bace, I. (2018) CRISPR-Cas adaptation in Escherichia coli requires RecBCD helicase but not nuclease activity, is independent of homologous recombination, and is antagonized by 5' ssDNA exonucleases. *Nucleic acids research*, **46**, 10173-10183.

51. Wiegand, T., Semenova, E., Shiriaeva, A., Fedorov, I., Datsenko, K., Severinov, K. and Wiedenheft, B. (2020) Reproducible Antigen Recognition by the Type I-F CRISPR-Cas System. *CRISPR J*, **3**, 378-387.

52. Stern, A., Keren, L., Wurtzel, O., Amitai, G. and Sorek, R. (2010) Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends in genetics : TIG*, **26**, 335-340.

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

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

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

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

57. Schmidt, F., Cherepkova, M.Y. and Platt, R.J. (2018) Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature*, **562**, 380-385.

58. Gonzalez-Delgado, A., Mestre, M.R., Martinez-Abarca, F. and Toro, N. (2019) Spacer acquisition from RNA mediated by a natural reverse transcriptase-Cas1 fusion protein associated with a type III-D CRISPR-Cas system in Vibrio vulnificus. *Nucleic acids research*, **47**, 10202-10211.

59. Nasef, M., Muffly, M.C., Beckman, A.B., Rowe, S.J., Walker, F.C., Hatoum-Aslan, A. and Dunkle, J.A. (2019) Regulation of cyclic oligoadenylate synthesis by the Staphylococcus epidermidis Cas10-Csm complex. *Rna*, **25**, 948-962.

60. Artamonova, D., Karneyeva, K., Medvedeva, S., Klimuk, E., Kolesnik, M., Yasinskaya, A., Samolygo, A. and Severinov, K. (2020) Spacer acquisition by Type III CRISPR-Cas system during bacteriophage infection of Thermus thermophilus. *Nucleic acids research*, **48**, 9787-9803.

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

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

63. Shah, S.A., Erdmann, S., Mojica, F.J. and Garrett, R.A. (2013) Protospacer recognition motifs: mixed identities and functional diversity. *RNA biology*, **10**, 891-899.

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

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

66. Millen, A.M., Samson, J.E., Tremblay, D.M., Magadan, A.H., Rousseau, G.M., Moineau, S. and Romero, D.A. (2019) Lactococcus lactis type III-A CRISPR-Cas system cleaves bacteriophage RNA. *RNA biology*, **16**, 461-468.

67. Fontaine, L., Dandoy, D., Boutry, C., Delplace, B., de Frahan, M.H., Fremaux, C., Horvath, P., Boyaval, P. and Hols, P. (2010) Development of a versatile procedure based on natural transformation for marker-free targeted genetic modification in Streptococcus thermophilus. *Applied and environmental microbiology*, **76**, 7870-7877.

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

69. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841-842.

70. Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome research*, **14**, 1188-1190.

71. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods*, **9**, 357-359.

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

73. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell*, **38**, 576-589.

74. Delorme, C., Bartholini, C., Luraschi, M., Pons, N., Loux, V., Almeida, M., Guedon, E., Gibrat, J.F. and Renault, P. (2011) Complete genome sequence of the pigmented Streptococcus thermophilus strain JIM8232. *Journal of bacteriology*, **193**, 5581-5582.

75. Otto, R., de Vos, W.M. and Gavrieli, J. (1982) Plasmid DNA in Streptococcus cremoris Wg2: Influence of pH on Selection in Chemostats of a Variant Lacking a Protease Plasmid. *Applied and environmental microbiology*, **43**, 1272-1277.

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

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

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

79. Liu, T.Y., Iavarone, A.T. and Doudna, J.A. (2017) RNA and DNA Targeting by a Reconstituted Thermus thermophilus Type III-A CRISPR-Cas System. *PloS one*, **12**, e0170552.

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

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

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

83. Niewoehner, O., Garcia-Doval, C., Rostol, J.T., Berk, C., Schwede, F., Bigler, L., Hall, J., Marraffini, L.A. and Jinek, M. (2017) Type III CRISPR-

Cas systems produce cyclic oligoadenylate second messengers. *Nature*, **548**, 543-548.

84. Foster, K., Kalter, J., Woodside, W., Terns, R.M. and Terns, M.P. (2019) The ribonuclease activity of Csm6 is required for anti-plasmid immunity by Type III-A CRISPR-Cas systems. *RNA biology*, **16**, 449-460.

85. Leenhouts, K.J., Tolner, B., Bron, S., Kok, J., Venema, G. and Seegers, J.F. (1991) Nucleotide sequence and characterization of the broad-host-range lactococcal plasmid pWVO1. *Plasmid*, **26**, 55-66.

86. Mashburn-Warren, L., Morrison, D.A. and Federle, M.J. (2012) The cryptic competence pathway in Streptococcus pyogenes is controlled by a peptide pheromone. *Journal of bacteriology*, **194**, 4589-4600.

87. Khan, S.A. (2000) Plasmid rolling-circle replication: recent developments. *Molecular microbiology*, **37**, 477-484.

88. del Solar, G., Giraldo, R., Ruiz-Echevarria, M.J., Espinosa, M. and Diaz-Orejas, R. (1998) Replication and control of circular bacterial plasmids. *Microbiology and molecular biology reviews : MMBR*, **62**, 434-464.

89. Duong, T., Miller, M.J., Barrangou, R., Azcarate-Peril, M.A. and Klaenhammer, T.R. (2011) Construction of vectors for inducible and constitutive gene expression in Lactobacillus. *Microb Biotechnol*, **4**, 357-367.

90. Vaillancourt, K., Bedard, N., Bart, C., Tessier, M., Robitaille, G., Turgeon, N., Frenette, M., Moineau, S. and Vadeboncoeur, C. (2008) Role of galK

and galM in galactose metabolism by Streptococcus thermophilus. *Applied and environmental microbiology*, **74**, 1264-1267.

91. Moineau, S., Walker, S.A., Holler, B.J., Vedamuthu, E.R. and Vandenbergh, P.A. (1995) Expression of a Lactococcus lactis Phage Resistance Mechanism by Streptococcus thermophilus. *Applied and environmental microbiology*, **61**, 2461-2466.

92. Biswas, I., Jha, J.K. and Fromm, N. (2008) Shuttle expression plasmids for genetic studies in Streptococcus mutans. *Microbiology (Reading)*, **154**, 2275-2282.

93. Renye, J.A., Jr. and Somkuti, G.A. (2010) Nisin-induced expression of pediocin in dairy lactic acid bacteria. *Journal of applied microbiology*, **108**, 2142-2151.

94. Lartigue, M.F. and Bouloc, P. (2014) A tetracycline-inducible expression vector for Streptococcus agalactiae allowing controllable gene expression. *J Microbiol Methods*, **96**, 16-18.

95. Turgeon, N. and Moineau, S. (2001) Isolation and characterization of a Streptococcus thermophilus plasmid closely related to the pMV158 family. *Plasmid*, **45**, 171-183.

96. Reckmann, B., Grosse, F., Urbanke, C., Frank, R., Blocker, H. and Krauss, G. (1985) Analysis of secondary structures in M13mp8 (+) single-stranded DNA by the pausing of DNA polymerase alpha. *Eur J Biochem*, **152**, 633-643.

97. Le Chatelier, E., Ehrlich, S.D. and Janniere, L. (1996) Countertranscript-driven attenuation system of the pAM beta 1 repE gene. *Molecular microbiology*, **20**, 1099-1112.

98. Brantl, S. (2015) Antisense-RNA mediated control of plasmid replication - pIP501 revisited. *Plasmid*, **78**, 4-16.

99. Gutierrez-Preciado, A., Henkin, T.M., Grundy, F.J., Yanofsky, C. and Merino, E. (2009) Biochemical features and functional implications of the RNA-based T-box regulatory mechanism. *Microbiology and molecular biology reviews : MMBR*, **73**, 36-61.

100. Novichkov, P.S., Kazakov, A.E., Ravcheev, D.A., Leyn, S.A., Kovaleva, G.Y., Sutormin, R.A., Kazanov, M.D., Riehl, W., Arkin, A.P., Dubchak, I. *et al.* (2013) RegPrecise 3.0--a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC genomics*, **14**, 745.

101. Yakhnin, H., Yakhnin, A.V. and Babitzke, P. (2015) Ribosomal protein L10(L12)4 autoregulates expression of the Bacillus subtilis rplJL operon by a transcription attenuation mechanism. *Nucleic acids research*, **43**, 7032-7043.

102. Duplessis, M., Russell, W.M., Romero, D.A. and Moineau, S. (2005) Global gene expression analysis of two Streptococcus thermophilus bacteriophages using DNA microarray. *Virology*, **340**, 192-208.

103. Li, Y., Pan, S., Zhang, Y., Ren, M., Feng, M., Peng, N., Chen, L., Liang, Y.X. and She, Q. (2015) Harnessing Type I and Type III CRISPR-Cas systems for genome editing. *Nucleic acids research*.

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

105. Khan, S.A. (1997) Rolling-circle replication of bacterial plasmids. *Microbiology and molecular biology reviews : MMBR*, **61**, 442-455.

106. Bikard, D., Loot, C., Baharoglu, Z. and Mazel, D. (2010) Folded DNA in action: hairpin formation and biological functions in prokaryotes. *Microbiology and molecular biology reviews : MMBR*, **74**, 570-588.

107. Rao, C., Chin, D. and Ensminger, A.W. (2017) Priming in a permissive type I-C CRISPR-Cas system reveals distinct dynamics of spacer acquisition and loss. *Rna*, **23**, 1525-1538.

**Figure 2.1. Adaptation by Type III-A and Type II-A CRISPR-Cas systems.**

(A) The schematic of CRISPR-Cas systems of *Sth* JIM8232. The *cas* genes are represented by colored arrows. 'L' represents the leader sequences of the systems. The fence structures following the leader sequences represent the CRISPR arrays. (B) The schematic of the process of adaptation assay. Un, unexpanded PCR product that reflects WT CRISPR array. Ex, expanded PCR product that reflects adaptation. (C) The column graph shows the length distribution of the new spacers of the Type II-A system (cyan) and the Type III-A system (yellow). The x axis indicates the spacer lengths, and the y axis indicates the percentages of the unique spacers observed. (D) Analysis of PFS sequences in adaptation. (E) Proportion of spacers derived from the plasmid (blue) and genome (green). Pooled data of at least three independent experiments are presented in (C), (D), and (E).

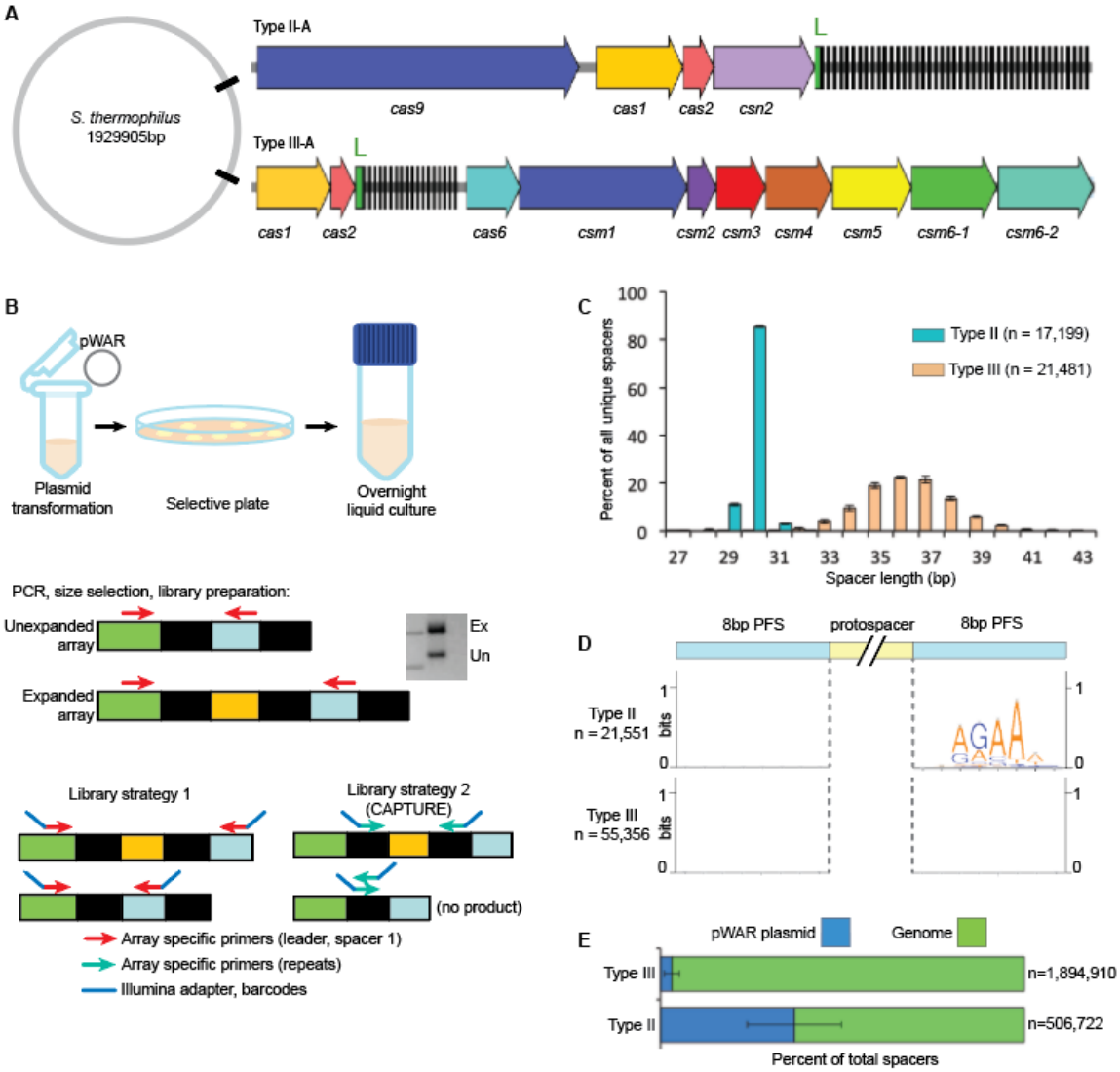

A
S. thermophilus
1929905bp
Type II-A
cas9
cas1
cas2
csn2
L
Type III-A
L
cas1
cas2
cas6
csm1
csm2
csm3
csm4
csm5
csm6-1
csm6-2
B
pWAR
Plasmid transformation
Selective plate
Overnight liquid culture
PCR, size selection, library preparation:
Unexpanded array
Expanded array
Ex
Un
Library strategy 1
Library strategy 2 (CAPTURE)
(no product)
Array specific primers (leader, spacer 1)
Array specific primers (repeats)
Illumina adapter, barcodes
C
Percent of all unique spacers
Type II (n = 17,199)
Type III (n = 21,481)
Spacer length (bp)
D
8bp PFS
protospacer
8bp PFS
Type II
n = 21,551
Type III
n = 55,356
bits
E
pWAR plasmid
Genome
Type III
n=1,894,910
Type II
n=506,722
Percent of total spacers

**Figure 2.2. Cas1-Cas2-dependent adaptation and -independent repeat-spacer replication.**

(A) Four round PCR results of the Type III-A CRISPR-Cas systems. Δ, gene deleted. Un, unexpanded PCR product that reflects WT CRISPR array. Ex, expanded PCR product that reflects adaptation. The picture represents more than 3 independent experiments. (B) The transformation efficiencies of the non-target plasmid (pNo-target), transcribed target plasmid (pTarget), and the non-transcribed target plasmid (pTarget-NT). n = 3. (C) The adaptation-independent replications of the repeat-spacer units of the *cas1* KO and WT Type III-A system, as well as (D) the Type II-A system. Pooled data of at least three independent experiments are presented. (E) Schematic of adaptation-independent replications of the repeat-spacer units (left), and Cas1-Cas2-dependent adaptation (right).

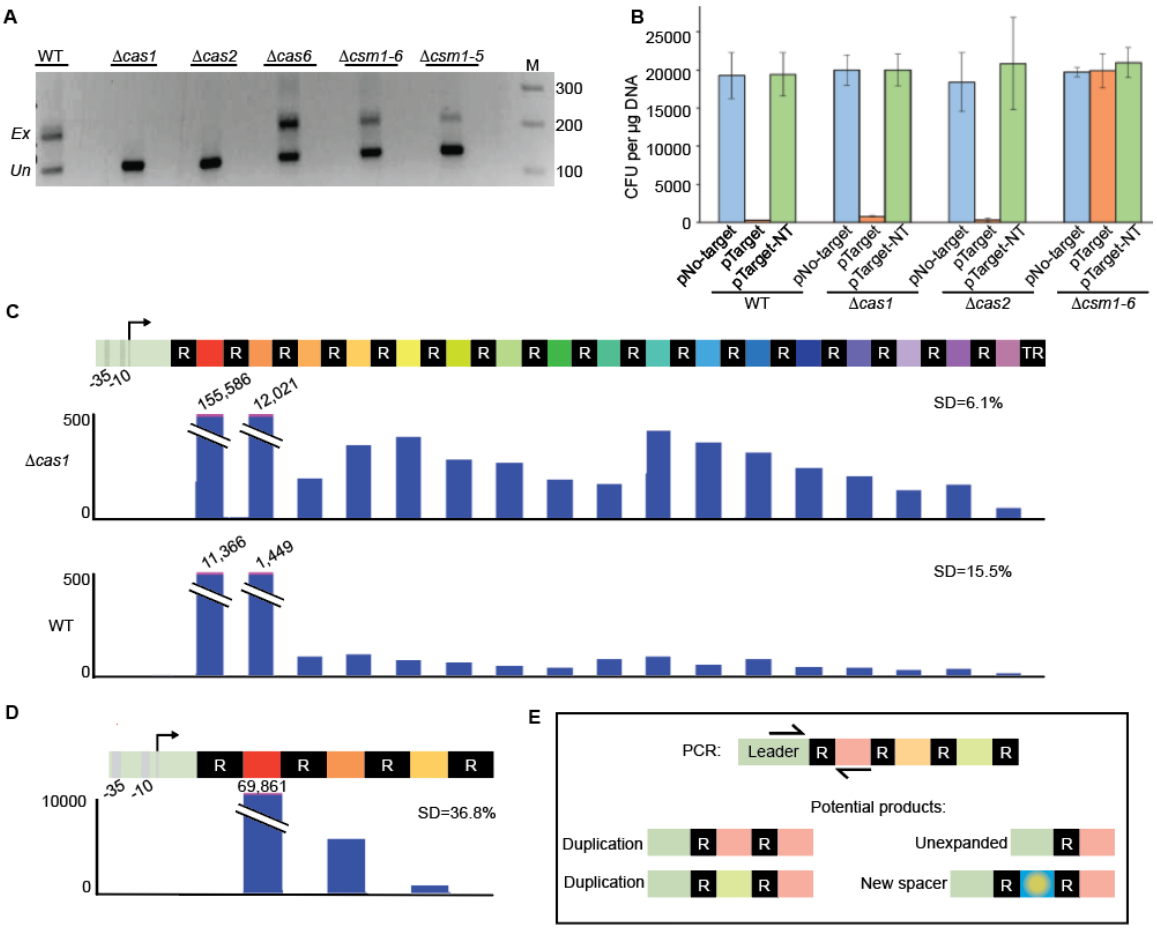

A
WT
Δcas1
Δcas2
Δcas6
Δcsm1-6
Δcsm1-5
M
300
200
100
Ex
Un
B
CFU per µg DNA
25000
20000
15000
10000
5000
0
pNo-target
pTarget
pTarget-NT
WT
Δcas1
Δcas2
Δcsm1-6
C
-35
-10
155,586
12,021
SD=6.1%
500
Δcas1
0
11,366
1,449
SD=15.5%
WT
D
69,861
SD=36.8%
10000
E
PCR:
Leader
Potential products:
Duplication
Unexpanded
New spacer

**Figure 2.3. Type III spacers are integrated in both possible orientations.** (A) The effects of the two possible spacer integration orientations on Type III target interference. When the coding strand of the spacer is integrated into the top strand of the CRISPR array, the expressed Type III crRNA will not be able to perform defense (left). In contrast, only when the template strand of the spacer is integrated into the top strand of the CRISPR array, the expressed Type III crRNA can perform defense (right). (B) Proportion of Type III spacers derived from the coding strand (blue) and template strand (orange). Pooled data of at least three independent experiments are presented for the new spacers. (C-F) Adaptation by WT and the *csm1-6* KO Type III systems, and the expression level (revealed by RNAseq) of the highly transcribed regions of the plasmid or the genome. The protospacers matched the plus strand of the plasmid are shown as cyan bars, and the other protospacers matched the minus strand are shown as pink bars. The transcripts matched the plus strand of the plasmid are shown as purple bars, and the other transcripts matched the minus strand are shown as yellow bars. The height of the bars is at the Y-axis, which reflects the numbers of the reads corresponding to a particular position. The pictures represent at least three independent experiments.

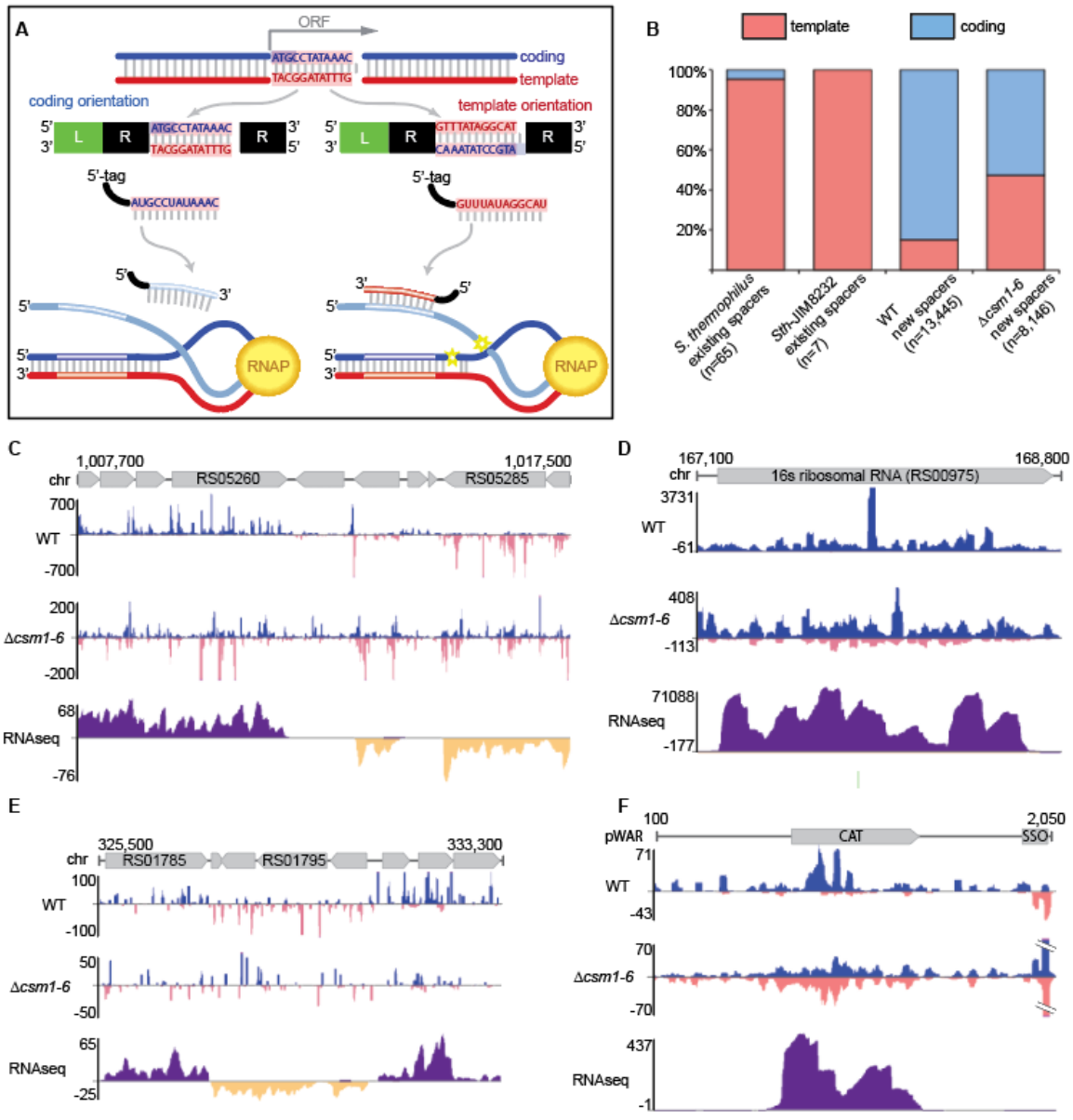

A
ORF
ATGCCTATAAAC
TACGGATATTTG
coding
template
coding orientation
template orientation
5'
3'
L
R
ATGCCTATAAAC
TACGGATATTTG
GTTTATAGGCAT
CAAATATCCGTA
5'-tag
AUGCCUAUAAAC
GUUUAUAGGCAU
RNAP
B
template
coding
100%
80%
60%
40%
20%
S. thermophilus existing spacers (n=65)
Sth-JIM8232 existing spacers (n=7)
WT new spacers (n=13,445)
Δcsm1-6 new spacers (n=8,146)
C
1,007,700
1,017,500
chr
RS05260
RS05285
700
WT
-700
200
Δcsm1-6
-200
68
RNAseq
-76
D
167,100
168,800
chr
16s ribosomal RNA (RS00975)
3731
WT
-61
408
Δcsm1-6
-113
71088
RNAseq
-177
E
325,500
333,300
chr
RS01785
RS01795
100
WT
-100
50
Δcsm1-6
-50
65
RNAseq
-25
F
100
2,050
pWAR
CAT
SSO
71
WT
-43
70
Δcsm1-6
-70
437
RNAseq
-1

**Figure 2.4. Selective targeting of the origins of RCR plasmids.**

(A) Schematic of the process of rolling-circle replication. The parental plus strand (nicked during the replication) is illustrated by blue, the parental minus strand is illustrated by orange, and the nascent strands are illustrated by dashed lines. The rectangles represent *sso* and *dso*. The RNA primer synthesized to trigger the replication of the minus strand is illustrated by yellow arrow. RSB, recombination site for RNA polymerase binding. CS-6, conserved 6 nt sequence to hamper the elongation of the RNA primer. Adaptation by the Type III CRISPR-Cas systems against (B) *sso* of pWAR, and (C) *sso* of pTRK882, as well as (D) *oriT* and *sso* of pNT1. The pictures represent at least three independent experiments.

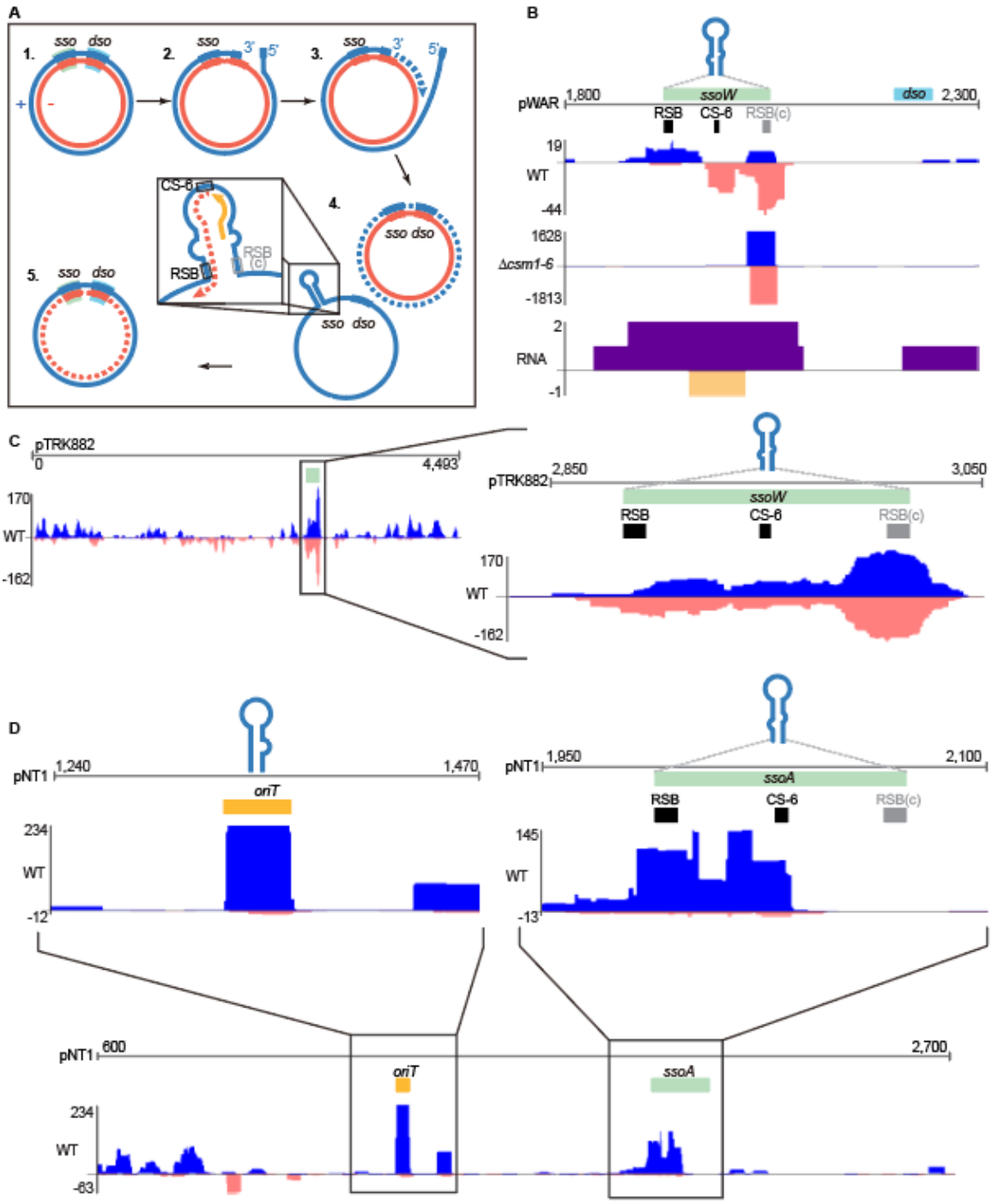

A
1.
sso
dso
2.
3'
5'
3.
4.
sso dso
5.
CS-6
RSB
RSB (c)
B
pWAR
1,800
ssoW
dso
2,300
RSB
CS-6
RSB(c)
19
WT
-44
1628
Δcsm1-6
-1813
2
RNA
-1
C
pTRK882
0
4,493
170
WT
-162
2,850
3,050
D
pNT1
1,240
1,470
oriT
234
-12
1,950
2,100
ssoA
145
-13
600
2,700
-63

**Figure 2.5. Impact of partially palindromic sequences on protospacer selection.**

(A-C) Schematics of the regulation of the expression of *repD* gene of plasmid pIB184 and ribosomal protein L10 and amino acid synthetase genes in self-genome. (D-I) Adaptation by the Type III-A system and the transcription level (revealed by RNAseq) of the partially palindromic sequences. The pictures represent at least three independent experiments.

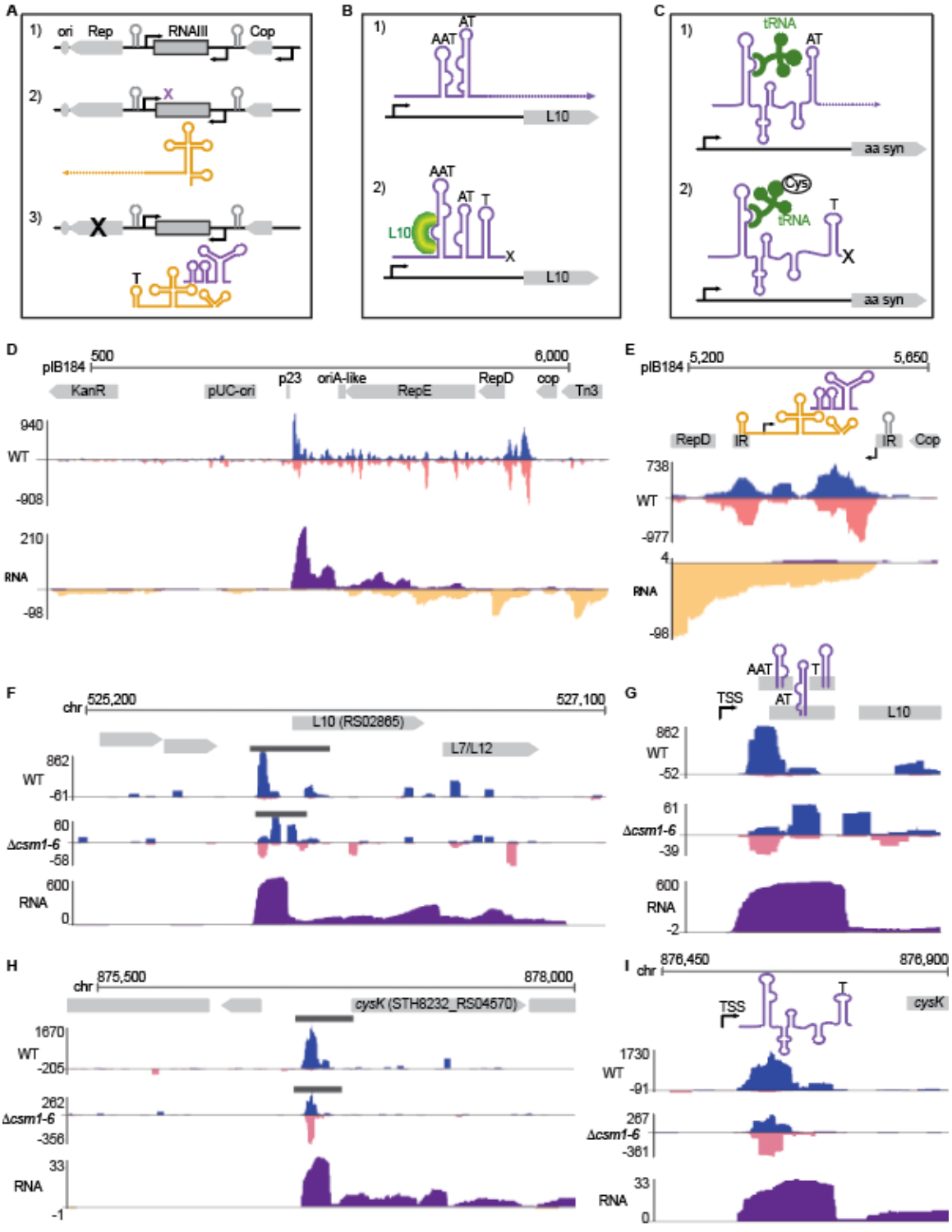

A
1) ori Rep RNAIII Cop
2) X
3) X T
B
1) AT AAT L10
2) AAT AT T L10 X L10
C
1) tRNA AT aa syn
2) Cys tRNA T X aa syn
D
pIB184 500 6,000
KanR pUC-ori p23 oriA-like RepE RepD cop Tn3
940 WT -908
210 RNA -98
E
pIB184 5,200 5,650
RepD IR IR Cop
738 WT -977
4 RNA -98
F
chr 525,200 527,100
L10 (RS02865) L7/L12
862 WT -61
60 Δcsm1-6 -58
600 RNA 0
G
AAT T TSS AT L10
862 WT -52
61 Δcsm1-6 -39
600 RNA -2
H
chr 875,500 878,000
cysK (STH8232_RS04570)
1670 WT -205
262 Δcsm1-6 -356
33 RNA -1
I
chr 876,450 876,900
TSS T cysK
1730 WT -91
267 Δcsm1-6 -361
33 RNA 0

**Figure 2.6. Impact of transcription on protospacer selection.**

(A and C) Adaptation by the Type III-A system against two highly transcribed region in the genome. (B) Adaptation by the Type III-A and the Type II systems against the original (p23-F) and the p23 inverted (p23-R) pIB184 plasmid. The pictures of (A-C) represent at least three independent experiments. (D) Cumulative distribution of protospacers acquired by the Type III-A system (blue) among the *Sth* JIM8232 genes sorted by RNAseq (yellow) with the promoter listed on the left. (E) Cumulative distribution of protospacers acquired by the Type III-A system (red) and the Type II-A system (black) among *Sth* JIM8232 genes sorted by RNAseq with the highly expressed genes listed on the left. (F) RNAseq reads of rRNAs and tRNAs were filtered out from (E). Pooled data of at least three independent experiments are presented in (D-F).

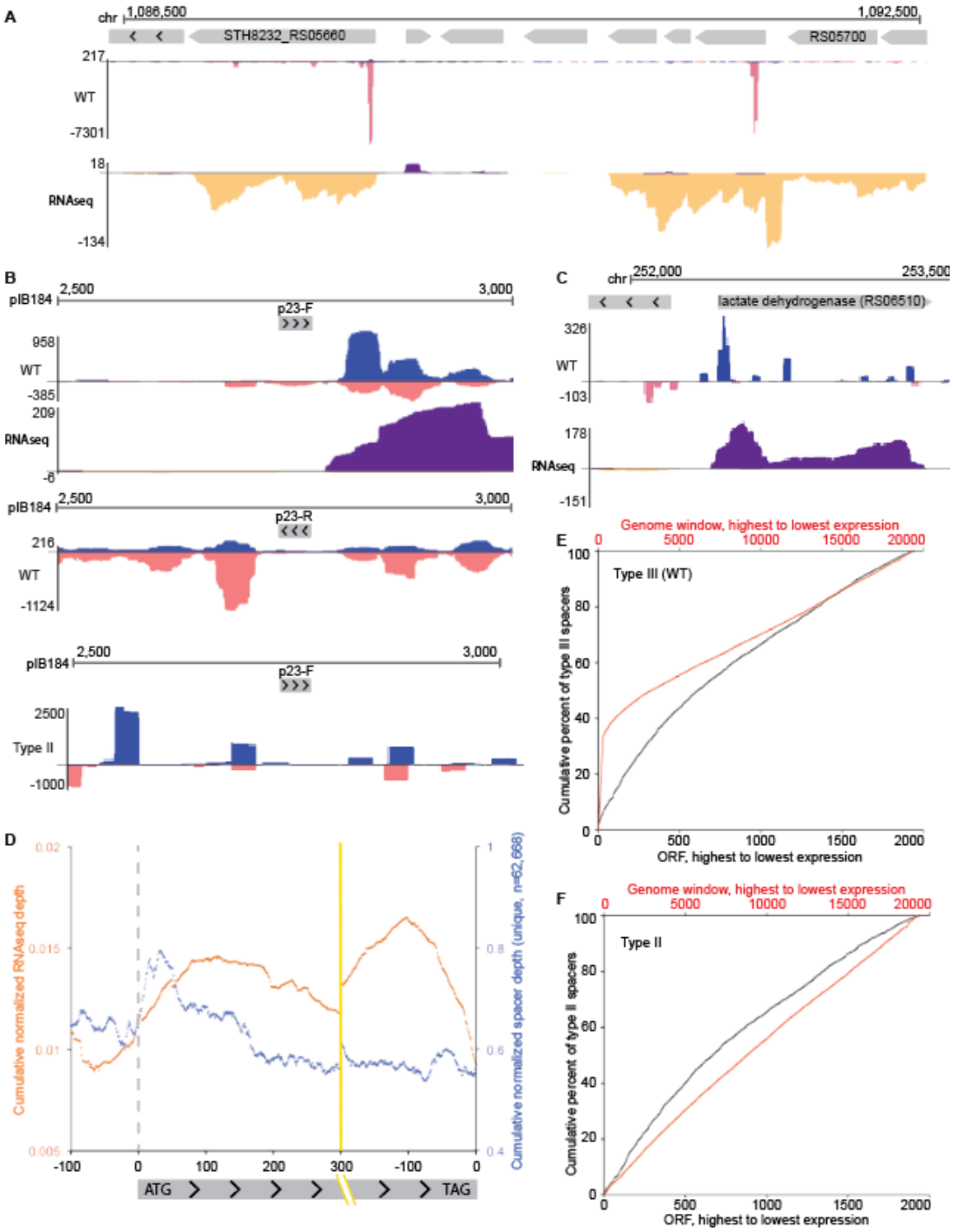

A
chr
1,086,500
1,092,500
STH8232_RS05660
RS05700
217
WT
-7301
18
RNAseq
-134
B
pIB184
2,500
3,000
p23-F
958
WT
-385
209
RNAseq
-8
pIB184
2,500
3,000
p23-R
216
WT
-1124
pIB184
2,500
3,000
p23-F
2500
Type II
-1000
C
chr
252,000
253,500
lactate dehydrogenase (RS06510)
326
WT
-103
178
RNAseq
-151
D
Cumulative normalized RNAseq depth
0.02
0.015
0.01
0.005
Cumulative normalized spacer depth (unique, n=62,668)
1
0.8
0.6
0.4
-100
0
100
200
300
-100
0
ATG
TAG
E
Genome window, highest to lowest expression
0
5000
10000
15000
20000
Type III (WT)
Cumulative percent of type III spacers
100
80
60
40
20
0
0
500
1000
1500
2000
ORF, highest to lowest expression
F
Genome window, highest to lowest expression
0
5000
10000
15000
20000
Type II
Cumulative percent of type II spacers
100
80
60
40
20
0
0
500
1000
1500
2000
ORF, highest to lowest expression

**Figure 2.7. Type III systems mediate defense against the lytic phage.**

(A) Distribution of the protospacers acquired by the Type III-A system among phage 2972. Green, early transcribed genes. Blue, middle transcribed genes. Red, late transcribed genes. Yellow, degraded and late transcribed lysogenic genes. (B) Growth curves of WT *Sth* DGCC7710 strain (left) and the BIMs (middle and right) with (red) and without (blue) phage 2972.

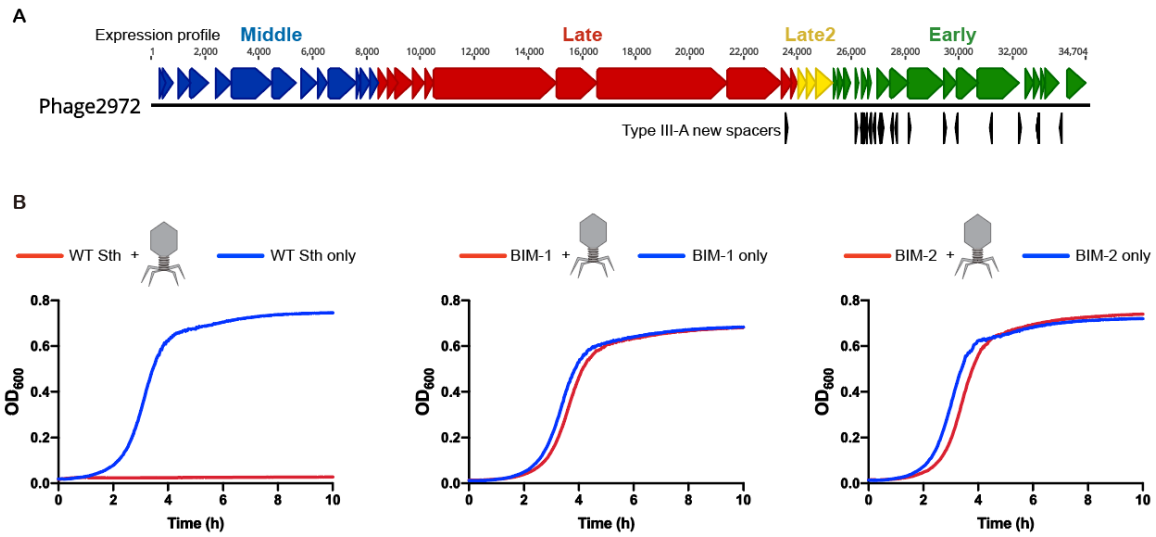

A
Expression profile
Middle
Late
Late2
Early
1
2,000
4,000
6,000
8,000
10,000
12,000
14,000
16,000
18,000
20,000
22,000
24,000
26,000
28,000
30,000
32,000
34,704
Phage2972
Type III-A new spacers
B
WT Sth +
WT Sth only
BIM-1 +
BIM-1 only
BIM-2 +
BIM-2 only
OD600
Time (h)
0.0
0.2
0.4
0.6
0.8
0
2
4
6
8
10

**Figure 2.8. Spacer uptake by Type III systems against the stem-loop structures.**

The mobile genetic elements invade the prokaryotic cells by conjugation, transduction, and transformation. The stem-loop structures present in the ssDNA forms of the invaders during their invasion, replication, and gene expression and are important for the life cycles of the mobile genetic elements. Type III CRISPR-Cas systems recognize the stem-loop structures to efficiently uptake spacers from the mobile genetic elements.

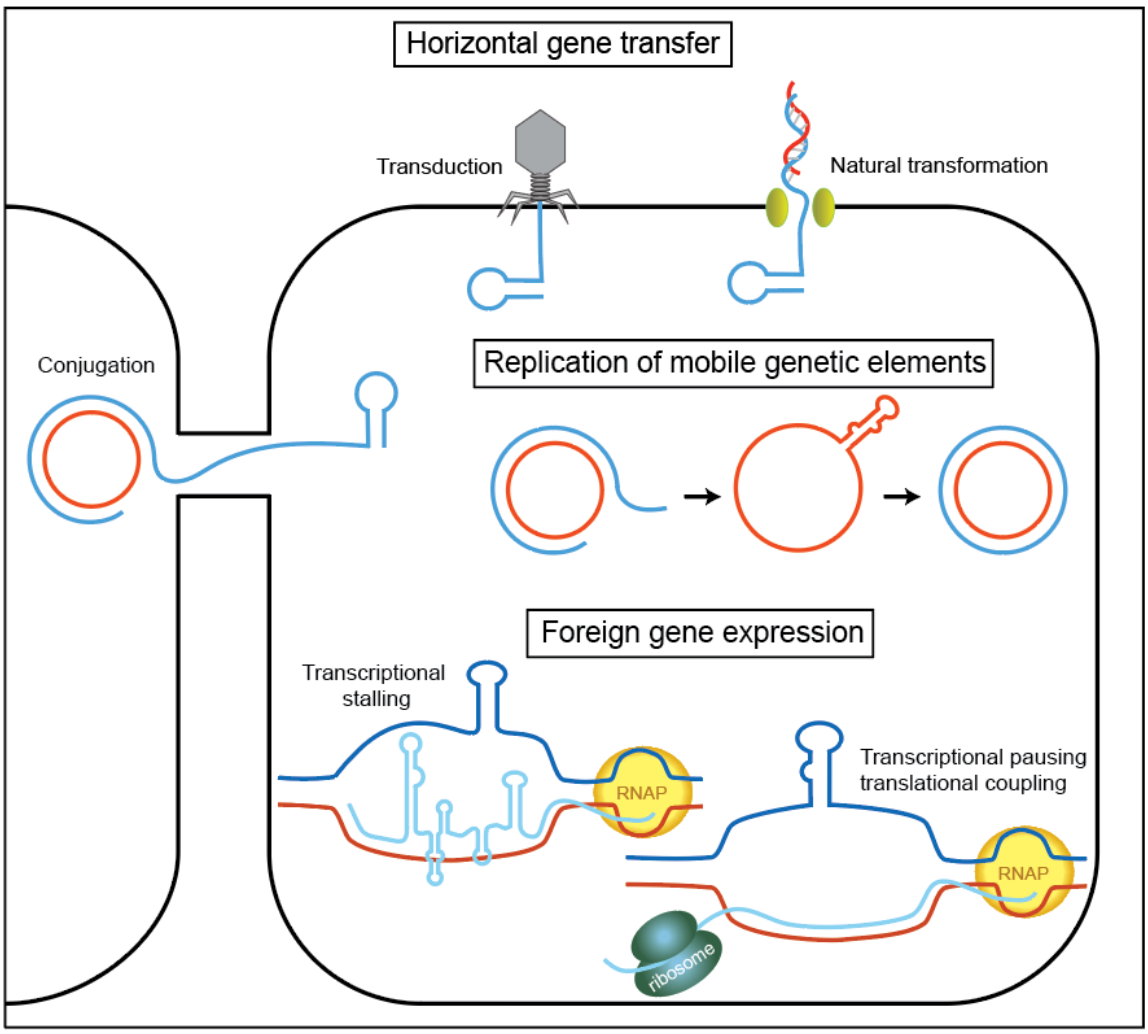

Horizontal gene transfer
Transduction
Natural transformation
Conjugation
Replication of mobile genetic elements
Foreign gene expression
Transcriptional stalling
RNAP
Transcriptional pausing translational coupling
RNAP
ribosome

**Figure S1. Schematic comparing defense by Type III systems and other types of systems.**

Adaptation and the following target interference by well-studied Type I and Type II systems are dependent on system-specific PAMs (e.g., CCT here). The orientation of spacer integration is dependent on PAM, and only the correctly integrated spacers can express bona fide crRNAs to perform defense against the invaders (left). In contrast, the spacers integrated in the reverse orientation cannot perform defense (right). Target interference by the Type III systems is dependent on the directional transcription of the DNA target. When the coding strand of the spacer is integrated into the top strand of the CRISPR array, the expressed Type III crRNA will not be able to perform defense (left). In contrast, only when the template strand of the spacer is integrated into the top strand of the CRISPR array, the expressed Type III crRNA can perform defense (right).

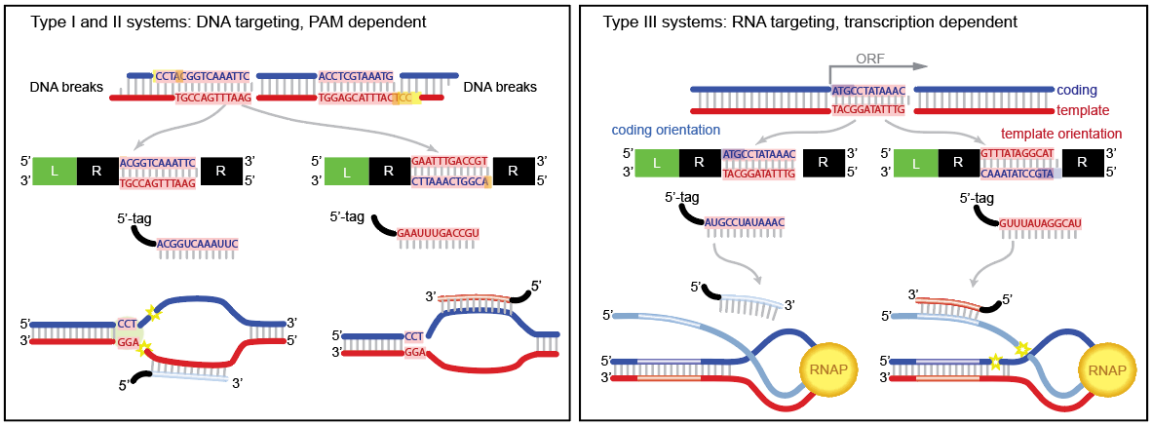

Type I and II systems: DNA targeting, PAM dependent
DNA breaks
DNA breaks
L
R
R
5'-tag
Type III systems: RNA targeting, transcription dependent
ORF
coding
template
coding orientation
template orientation
5'-tag
RNAP

**Figure S2. The recovery of adaptation of the Type II-A system.**

(A) The changes of the leader sequence of the Type II-A system of *Sth* JIM8232 that hampered adaptation. (B) Analysis of PFS sequences in adaptation. (C) The column graph shows the length distribution of the new spacers of integrated into the original Type II locus (red) and into the transplanted Type II array in PTS locus (cyan). (D)The transformation efficiencies of the different target/non-target plasmids to the JIM8232 strain with the DGCC7710 CRISPR array inserted into the PTS locus. n = 3.

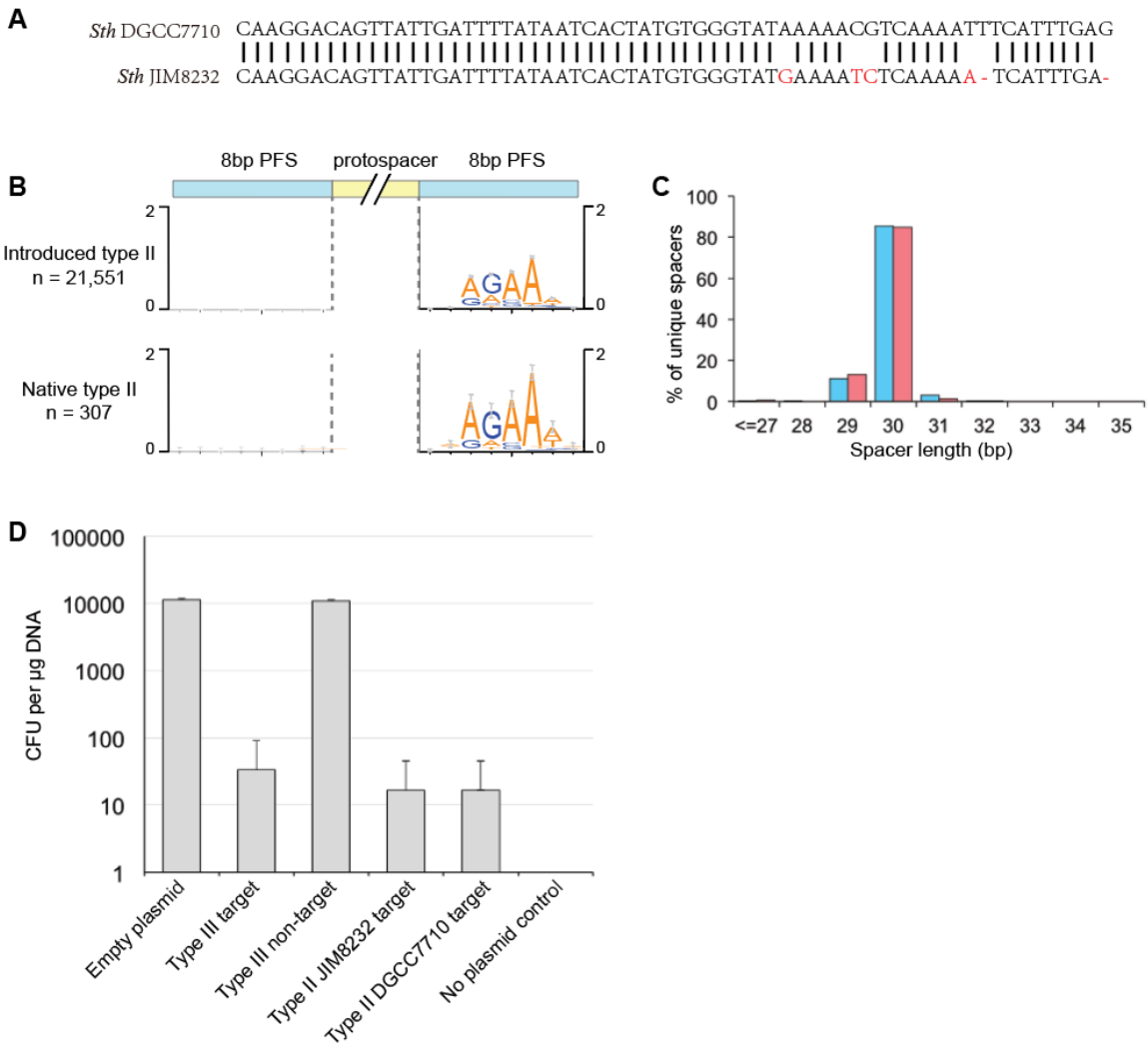

A
Sth DGCC7710 CAAGGACAGTTATTGATTTTATAATCACTATGTGGGTATAAAAACGTCAAAATTTCATTTGAG
Sth JIM8232 CAAGGACAGTTATTGATTTTATAATCACTATGTGGGTATGAAAATCTCAAAAA-TCATTTGA-
B
8bp PFS
protospacer
8bp PFS
Introduced type II
n = 21,551
Native type II
n = 307
C
% of unique spacers
Spacer length (bp)
D
CFU per µg DNA
Empty plasmid
Type III target
Type III non-target
Type II JIM8232 target
Type II DGCC7710 target
No plasmid control

**Figure S3. Adaptation against rRNA and tRNA encoding clusters.**

Adaptation by WT and the *csm1-6* KO Type III systems, and the expression level (revealed by RNAseq) of the 5 clusters of the rRNA and tRNA encoding sequences in the genome. Green triangles represent the tRNA encoding sequences. The pictures represent at least three independent experiments.

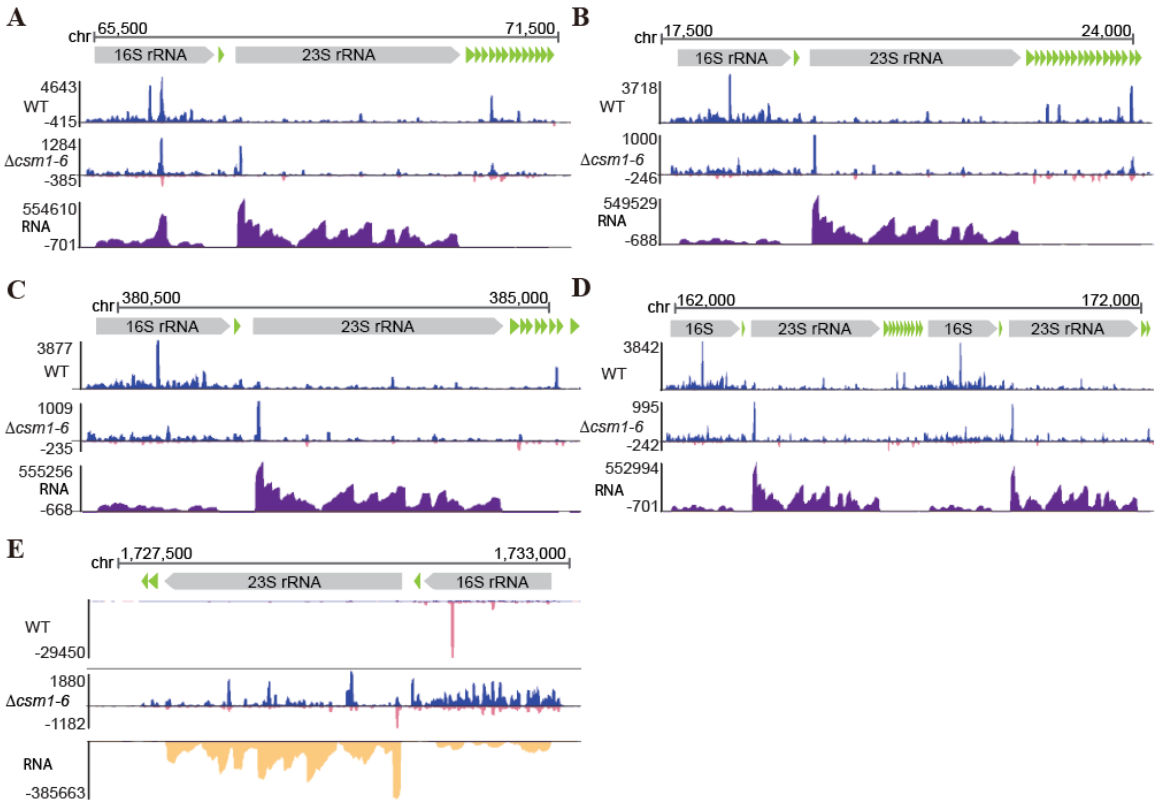

A
chr 65,500
71,500
16S rRNA
23S rRNA
4643
WT
-415
1284
Δcsm1-6
-385
554610
RNA
-701
B
chr 17,500
24,000
16S rRNA
23S rRNA
3718
WT
1000
Δcsm1-6
-246
549529
RNA
-688
C
chr 380,500
385,000
16S rRNA
23S rRNA
3877
WT
1009
Δcsm1-6
-235
555256
RNA
-668
D
chr 162,000
172,000
16S
23S rRNA
16S
23S rRNA
3842
WT
995
Δcsm1-6
-242
552994
RNA
-701
E
chr 1,727,500
1,733,000
23S rRNA
16S rRNA
WT
-29450
1880
Δcsm1-6
-1182
RNA
-385663

Figure S4. Adaptation by the Type III-A and the Type II-A systems against (A) pWAR, (B) pIB184, (C) pTRK882, and (D) pNT1 plasmids. The pictures represent at least three independent experiments.

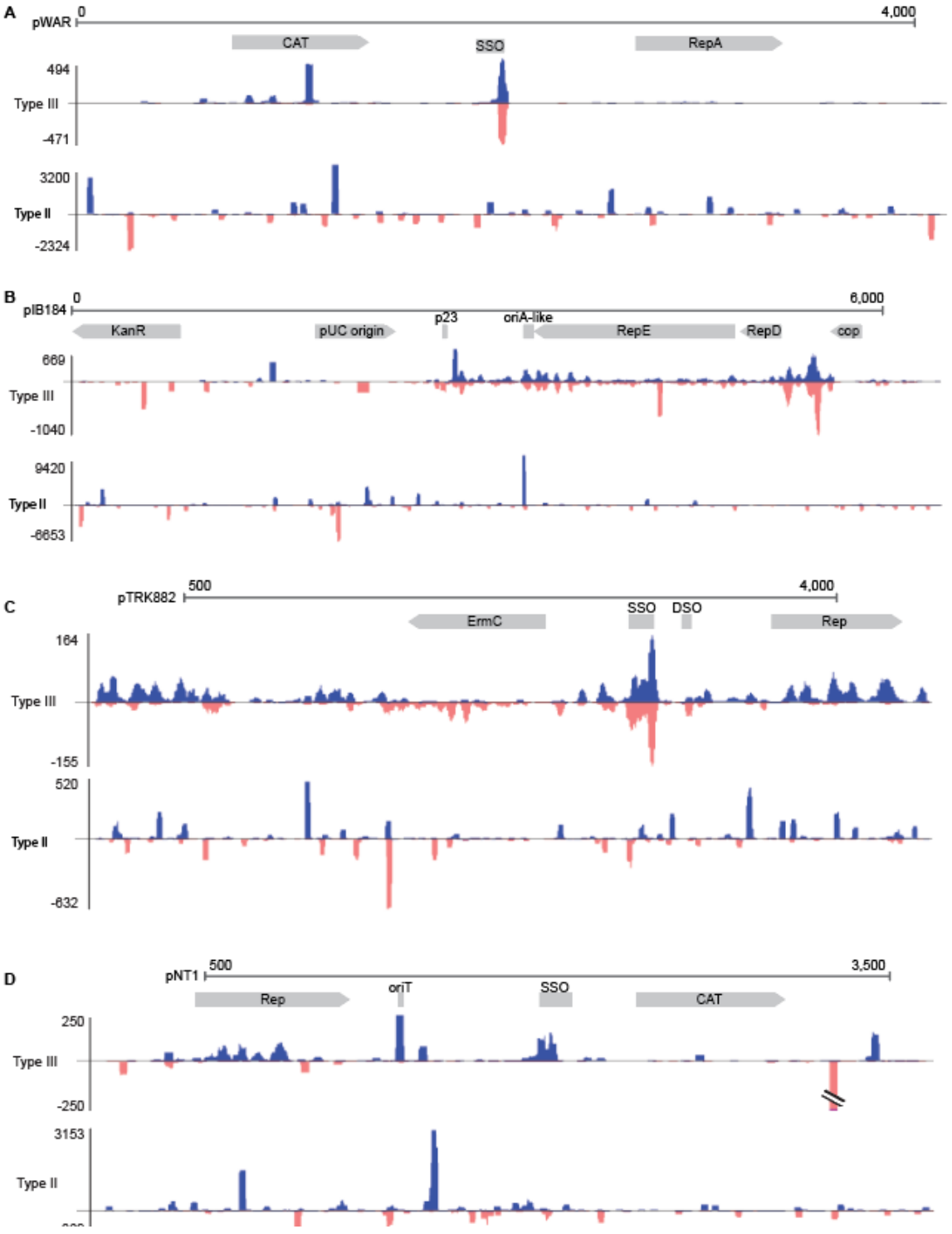

A
pWAR
0
4,000
CAT
SSO
RepA
494
Type III
-471
3200
Type II
-2324
B
pIB184
0
6,000
KanR
pUC origin
p23
oriA-like
RepE
RepD
cop
669
Type III
-1040
9420
Type II
-6653
C
pTRK882
500
4,000
ErmC
SSO
DSO
Rep
164
Type III
-155
520
Type II
-632
D
pNT1
500
3,500
Rep
oriT
SSO
CAT
250
Type III
-250
3153
Type II

**Figure S5. Type III systems mediate defense against the lytic phage.**

(A) The phage tittering assay by phage 2972 at a gradient of amount against WT *Sth* DGCC7710 strain and the BIMs. (B) Growth curves of WT *Sth* DGCC7710 strain (left) and the BIMs (middle and right) with different amounts of phage 2972. (C) PCR results of all the 4 CRISPR loci of WT *Sth* DGCC7710 strain and the BIMs.

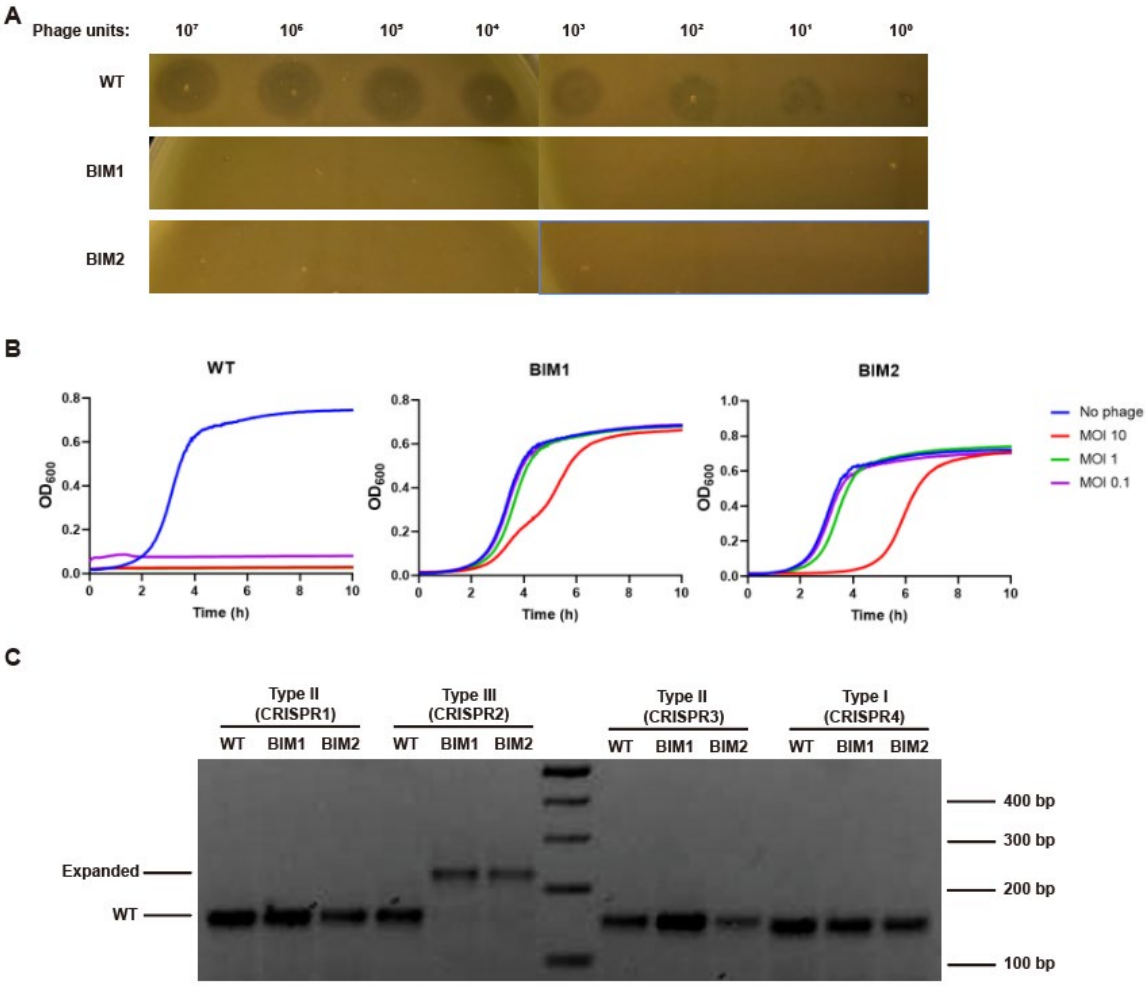

A
Phage units:
10⁷
10⁶
10⁵
10⁴
10³
10²
10¹
10⁰
WT
BIM1
BIM2
B
WT
BIM1
BIM2
OD600
Time (h)
No phage
MOI 10
MOI 1
MOI 0.1
C
Type II (CRISPR1)
Type III (CRISPR2)
Type II (CRISPR3)
Type I (CRISPR4)
WT
BIM1
BIM2
Expanded
WT
400 bp
300 bp
200 bp
100 bp

**Table 2.1. Primers used in this work.**

| **PCR primers used in this study** | | |
|---|---|---|
| | | |
| *Sth* mutant strain construction | | |
| Purpose | Name | Sequence (5'-3') |
| *cas1* KO | sl-JIMCas1+2-2-f | CACCGCTAGCATCTGCAGGCAGAAGTCTCTCACAAGATTTCC |
| | JIMCas1KO-M-r | CTAAAAGCCCTAATAAGCTACCTCAGAGTATTTTATGAC |
| | JIMCas1KO-M-f | GAGGTAGCTTATTAGGGCTTTTAGAGAACTTGACC |
| | sl-JIMCas1KO-r | GCATAAGCTTGCGTCGACGGTAAAACCTCTATGCTGTTCTTTGC |
| *cas2* KO | Sl-JIMCas2KO-f | CACCGCTAGCATCTGCAGGGAGTATTGCTACGACCAAGG |
| | JIMCas2KO-M-r | GTCTATAAAGGCTTAATGCCCTCCTGTGTAACTTG |
| | JIMCas2KO-M-f | GAGGGCATTAAGCCTTTATAGACCTTTAATCATATGGTAC |
| | sl-JIMCas1+2KO-r | GCATAAGCTTGCGTCGACGATTGATAAGAAAGTTATTGAAAAACGCCAACAAG |
| *cas6* KO | Sl-JIMCas6Csm-f | CACCGCTAGCATCTGCAGCCAGTTCTGCAAAGAACAGC |
| | JIMCas6KO-M-r | CATTTTGACATAAGCTCGACTCCTTCATCTTTGTATG |
| | JIMCas6KO-M-f | GGAGTCGAGCTTATGTCAAAATGCTTCTAACATTCG |
| | Sl-JIMCas6KO-r | GCATAAGCTTGCGTCGACGAGCAAAAGCTGCTGTCAG |
| *csm1-5* KO | sl-JIMallCsm-f | CACCGCTAGCATCTGCAGGAAAAAATTAGTATTTACTTTTAAAAGGATCGACC |
| | JIMCsm1-5KO-OL-r | GACTGTATCGGCTCCGTAAAATAAATCAATCTTTTCTTTCTTC |
| | JIMCsm1-5KO-OL-f | CGGAGCCGATACAGTCGAATGAAAACTAAAATGG |
| | Sl-JIMCas6-Csm5-r | GCATAAGCTTGCGTCGACCTGATAATTGGTCTAGAATATCTAG |
| All *csm* KO | sl-JIMallCsm-f | CACCGCTAGCATCTGCAGGAAAAAATTAGTATTTACTTTTAAAAGGATCGACC |
| | JIMAllCsmKO-Mol-r | GACACCCAAATATGCTCCGTAAAATAAATCAATCTTTTCTTTCTTC |
| | JIMAllCsmKO-Mol-f | CGGAGCATATTTGGGTGTCATCCCCCTTTG |

| | | |
|---|---|---|
| | Sl-JIMallCsm-r | GCATAAGCTTGCGTCGACGGTTAGTTGCTTCTTTGATTTTGGCTAC |
| PTS clone | JCPTS-DC1L-U-f | GGTTTGGTTGGAATGCTCTC |
| | JCPTS-DC1L-U-r | GCTGATTTAGGGAGTTGGCAATCCACATACC |
| | JCPTS-DC1L-M-f | GCCAACTCCCTAAATCAGCTGTTTCATTTTAGTTAC |
| | JCPTS-DC1L-M-r | GCTGGTGTATCGCTGTTGAGTAATAAGCCTG |
| | JCPTS-DC1L-D-f | CAACAGCGATACACCAGCTCGTTTCAAAGTTG |
| | JCPTS-DC1L-D-r | CTGGCTAATATTGTTTGGTAACCTTC |
| DGCC7710 Type III replacement | sl-JIMCas1+2-2-f | CACCGCTAGCATCTGCAGGCAGAAGTCTCTCACAAGATTTCC |
| | Sl-JIMallCsm-r | GCATAAGCTTGCGTCGACGGTTAGTTGCTTCTTTGATTTTGGCTAC |
| | | |
| CAPTURE PCR for the Type III system in *Sth* | | |
| Purpose | Name | Sequence (5'-3') |
| 1st PCR | JIMCC2-DS-NA-f | GCAACTGCCTAGAGCACAAGATATGG |
| | JIMCC2-DS-NA-r | CGCTTATTTAGAAGTAGCGTTAGAATCAAGG |
| 2nd PCR | JIMCC2-DS-rpt-f | ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGAGAGGGGACGGAAAC |
| | JIMCC2-DS-rpt-r | GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCTCGAGGTAATTAGGTTTATATC |
| | | |
| CAPTURE PCR for the Type II system at the PTS locus | | |
| Purpose | Name | Sequence (5'-3') |
| 1st PCR | DC1-DS-NA-f | GTGGGTATAAAAACGTCAAAATTTCATTTGAG |
| | DC1-DS-NA-r | CAATTCGAATCTTGATTTGCTGTCAAAC |
| 2nd PCR | JIMCC1-DS-rpt-f | ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTCTCAAGATTTAAGTAACTGTACAAC |
| | JIMCC1-DS-rpt-r | GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGTTACTTAAATCTTGAGAGTACAAAAAC |
| | | |
| Multiple round PCR for the Type III system in *Sth* | | |
| Purpose | Name | Sequence (5'-3') |
| 1st PCR | JC2-leader-1.5f | GCCTTTATAGACCTTTAATCATATGG |

| | | |
|---|---|---|
| | JIM-2.02-1r | CTAAAGAACATAAACATATGATGAATGCTTTAGAAC |
| 2nd PCR | JIMCC2-leader-2f | CAATATAGATAGTGTTTCCAGTAGGTCC |
| | JIMCC2.02-2r | GAACATAAACATATGATGAATGCTTTAGAACTG |
| 3rd PCR | JIMCC2-DS-f | ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCAACTGCCTAGAGCACAAGATATGG |
| | JIMCC2-DS-r | GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGCTTATTTAGAAGTAGCGTTAGAATCAAGG |
| | | |
| CRISPR loci expansion monitor in *Sth* DGCC7710 | | |
| Purpose | Name | Sequence (5'-3') |
| CRISPR 1 | SthCR1-DS-f | ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGGTATAAAAACGTCAAAATTTCATTTGAG |
| | SthCR1-DS-r | GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAATTCGAATCTTGATTTGCTGTCAAAC |
| CRISPR 2 | JIMCC2-DS-f | ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCAACTGCCTAGAGCACAAGATATGG |
| | JIMCC2-DS-r | GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGCTTATTTAGAAGTAGCGTTAGAATCAAGG |
| CRISPR 3 | SthCR3-DS-f | ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGGTAGAAAAGATATCCTACGAG |
| | SthCR3-DS-r | GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCTTCCTCTTTAGCGTTTAG |
| CRISPR 4 | SthCR4-DS-f | ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTCTTAATTCCATTGGGATCTTTTAG |
| | SthCR4-DS-r | GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTGGGAGGCCATTGATATAGG |

**Table 2. The new spacers adapted from ssM13 and dsM13 DNAs by the Type III-A system.**

| | No. | New spacer sequence | Strand | Size | Locus |
|---|---|---|---|---|---|
| **ssM13** | 1 | CGTGCCAGCTGCATTAATGAA TCGGCCAACGCGCGGGGA | minus | 39 | 6024-6062, lacI gene |
| | 2 | GCATTAATGAATCGGCCAACG CGCGGGGAGAGGCGG | minus | 36 | 6017-6052, lacI gene |
| | 3 | AAGCGGGCAGTGAGCGCAAC GCAATTAATGTGAGTTA | plus | 37 | 6081-6117, lacI gene |
| **dsM13** | 1 | TGGCCAACAGAGATAGAACCC TTCTGACCTGAAA | minus | 34 | 5051-5084, g4 assembly gene |
| | 2 | GCTGCGCGTAACCACCACACC CGCCGCGCTTAA | minus | 33 | 5507-5539, replicatio n origin |
| | 3 | TGTGAGCGAGTAACAACCCGT CGGATTCTCCGT | minus | 33 | 6618-6650, lacZ gene |
| | 4 | TTAGTTGTTAGTGCTCCTAAAG ATATTTTAGATAAC | plus | 36 | 4732-4767, g4 assembly gene |
| | 5 | AAGCGGGCAGTGAGCGCAAC GCAATTAATGTGAGTT | plus | 36 | 6081-6116, lacI gene |

# CHAPTER 3

# PRIMED ADAPTATION BY THE TYPE III-A CRISPR-CAS SYSTEM OF *STREPTOCOCCUS THERMOPHILUS*

[1]Xinfu Zhang, Sandra Garrett, Brenton G. Graveley, Michael P. Terns. To be submitted to Nucleic Acids Research.

**Abstract**

CRISPR-Cas systems provide prokaryotic cells with a mechanism of defense against invaders, including virus and plasmids. The heritable and specific defense ability of the systems is acquired during adaptation against a fragment of the invading nucleic acid. However, invaders can escape from CRISPR-Cas mediated interference by mutation(s) within or adjacent to the originally acquired fragments. Primed adaptation provides CRISPR-Cas systems with an efficient and specific secondary adaptation against escaping invaders, thus allowing hosts to minimize escapers and co-evolve with the invaders. Our knowledge about adaptation by the Type III CRISPR-Cas systems is very limited, and in particular, no Type III primed adaptation has been reported yet. Here, we gained preliminary evidence indicating primed adaptation by the Type III CRISPR-Cas system. The priming process increased the proportion of adaptation by the Type III-A system of *Streptococcus thermophilus* against the invading plasmid, rather than host genome. Specifically, the proposed DNA cleavage sites of the targeted plasmid were recognized and preferred by secondary adaptation of the Type III-A CRISPR-Cas system, and directional transcription of the DNA target and the DNase activity of the system were both required for the primed adaptation, implying that it is DNA cleavage that facilitates secondary adaptation. Although Type III CRISPR-Cas systems tolerate more mutations of the invader compared with Type I or Type II systems to constrain the escaping invader numbers, primed adaptation still theoretically benefits the Type III systems, because of the special features of the adaptation

and the target interference of them. This work here provides the first evidence regarding primed adaptation by the mysterious Type III systems, and paves the way for further studies about Type III primed adaptation by CRISPR-Cas systems.

## Introduction

CRISPR-Cas systems exist in half of eubacteria and nearly all the archaea (1-3), and provide hosts with defense against invading nucleic acids by three major procedures: adaptation, crRNA biogenesis, and target interference (4-6). The specific and heritable defense memory of a CRISPR-Cas system is acquired by adaptation, during which a short fragment (protospacer) of the foreign DNA can be captured and integrated into the CRISPR locus as a spacer (4,5,7-9).

During biogenesis of crRNAs, a CRISPR array is transcribed into a long RNA, and then further processed into small and mature crRNAs (10). For Type III systems, the long transcripts of the CRISPR arrays are processed by Cas6 endonuclease, which cut within the repeat sequences (11-14). The resulting intermediate crRNAs contain the spacer-derived sequences, flanked by the last 8 nucleotides of the repeats at their 5' -ends, and the rest of the sequences of the repeats at their 3'-ends (11-14). The repeat-derived ribonucleic acids at the 3'-ends of the crRNA intermediates, as well as several nucleotides at the 3' -ends derived from the spacers, are further cleaved to make the short and mature crRNAs (12,15-17), though the mechanism of this process is unknown. In *Sulfolobus islandicus*, mutations within Type III crRNP subunits as well as within

Cas6, result in the accumulation of long crRNA intermediates and block their 3'-processing (18), suggesting that the crRNP and Cas6 are involved in the uncharacterized secondary processing. The 8 nt repeat-derived sequences at the 5'- ends are maintained with the crRNAs as the Type III crRNA tags, and their existence is critical for the authentication of downstream interference process (12,15,16,19-22).

During CRISPR-Cas-mediated target interference, the mature crRNPs of Type I, Type II and Type V systems directly bind to DNA targets and mediate DNA cleavage (9,23-27), while those of Type VI systems target ssRNAs rather than DNAs (28-31). In contrast, the crRNPs of Type III systems directly bind to ssRNA targets (the transcripts of the DNA targets, triggering RNAs), and such binding activates both the RNase and DNase activities of the crRNPs (18,22,32-37). Cas7 (Csm3 for Type III-A) in Type III crRNPs specifically cleaves the reverse complementary RNA targets (36,37). The DNase activity is undertaken by the HD domain of Cas10 (Csm1 for Type III-A), which cuts ssDNA in a sequence-non-specific way (22,34,38-40). Since Cas10 is a subunit of the crRNP which binds to the transcript of the DNA target, it is hypothesized that because it is close to the ssDNA in the transcription bubble of the DNA target, the active Cas10 is able to find and cut the encoding strand of the invader. After cleavage of the triggering RNAs, the non-specific DNase activity of Cas10 becomes inactive to protect self-ssDNAs from cleavage. Additionally, the Palm domain of Cas10 synthesizes cyclic oligoadenylates (cA*n*s) as a secondary messenger, which binds to the CARF domain of Csx1 (Csm6 for Type III-A) and stimulates

the RNase activity of the HEPN domain of Csx1 to non-specifically cleave foreign DNA transcripts (41-43). In *Staphylococcus epidermidis*, the non-specific RNase activity of Csm6 is not necessary for defense when the target is extensively transcribed, and hamper the growth of the bacteria when activated; however, when the target is weakly transcribed and the DNase activity of Csm1 is insufficient to cleave the invader, Csm6 is necessary for invader clearance (44). PAM recognition is required for the authentication of interference process of Type I and Type II systems, and this is the mechanism to protect host genomes at the CRISPR loci, which contain the same target sequences as spacers (24,45-47). Instead, the discrimination between self- and non-self-DNAs by Type III systems relies on the complementarity between the crRNA and the PFS of the triggering RNA. Base pairing at several critical positions within the 8 nt crRNA tag with the 3'-downstream sequence of the target RNA prevents the activation of Cas10 (20,33,48,49), though it still activates specific cleavage of the triggering RNA by Cas7 (48).

In Type I and Type II systems, target DNAs can escape from CRISPR-Cas-mediated interference by mutation(s) at pivotal positions within the targets or the PAMs (50,51). However, a pre-existing spacer in a CRISPR array, which is partially or totally complementary to a fragment of a molecule, can greatly stimulate adaptation efficiency against the same molecule (52,53). To acquire a new spacer from a molecule that the system never processed before is termed naïve adaptation, and the pre-existing spacer (priming spacer) triggered adaptation is termed primed adaptation. Current evidence suggests a somewhat

inverse relationship between the sensitivity of the escaping targets to CRISPR-Cas-mediated defense and their ability to trigger primed adaptation, in that some targets (even those with mutations) that are still sensitive to the initial interference may not trigger secondary adaptation, while some those which are less sensitive to direct interference become able to trigger priming (54,55). Primed adaptation has been relatively well-studied in Type I systems (52-54,56-59), and has also been reported in Type II systems (60,61). Primed adaptation is substantially more efficient than naïve adaptation (54), and directs the adaptation machinery to the invader DNA instead of self-genome (53), thus providing the hosts with a co-evolutionary strategy to minimize the amount of CRISPR-Cas escapers. Different studies have revealed that primed adaptation has strand or position biases (52-54,56-61). All tested primed adaptation processes of different systems require all the Cas proteins involved in crRNA biogenesis and target interference (52-54,56,59,60,62,63). Several functional models have been proposed for the mechanism of primed adaptation, including: 1. DNA breaks or short DNA fragments generated by inefficient DNA cleavage against the imperfect target become substrates for adaptation (64-68); 2. The DNA interference complex recognizes the escaping target DNAs and recruits Cas1-Cas2 complex to the invaders (69); 3. The crRNPs have different conformations when binding to a bona fide target or an escaping target, and consequentially authenticate target interference or primed adaptation, separately (70,71). It is important to note that the models do not conflict with each other, and thus more than one may be true mechanisms of primed adaptation.

Target interference of Type III systems tolerates a broad range of protospacer flanking sequences (PFS) (20,22,72), and also tolerates apparently more mutations within the targets than Type I and Type II systems (72,73). As a result, Type III systems minimize the potential escapers of the invading nucleic acids (72,73), and consistent with this, the rare phages identified escaping from Type III systems were found to accumulate long deletions within or covering the targets, rather than few nucleotide changes (74). Despite this difficulty of escape, primed adaptation may still be beneficial for Type III CRISPR-Cas-mediated defense. We found that naïve adaptation by the Type III system preferentially takes up the protospacers at the encoding strands of the promoter regions of expressed genes (See Chapter 2). Since the target interference ability of the Type III system requires a reverse complementary RNA, DNA uptake against the encoding strand will not directly contribute to defense. Moreover, as to the *bona fide* protospacers derived from the template strands, if the protospacer region was weakly transcribed or a late transcript in phage infection, the Type III spacer-mediated defense may be insufficient to efficiently clear phage or plasmid nucleic acids (32,44). In these situations, the potential primed adaptation triggered by the 'inefficient' spacers may be able to provide a chance to the system to perform efficient secondary uptake to counter against the invaders.

We previously observed and characterized naïve adaptation by the Type III-A system of *Sth* JIM8232 (See Chapter 2). In the course of this chapter, analyses of adaptation by the Type III-A system against a target plasmid and a non-target plasmid revealed that the target sequence increased the proportions of the

secondary spacers acquired from the plasmid relative to the host genome. The secondary protospacers accumulated at the DNA cleavage sites of the plasmid, and moreover, transcription of the target sequence and the nuclease domain of Csm1 were shown to be required for this process. This work details primed adaptation by Type III CRISPR-Cas systems, and the possible mechanism and benefits of primed adaptation are discussed in this chapter.

## Materials and methods

### *Strain and plasmid manipulation*

*Sth* JIM8232 was kindly provided by Dr. Pierre Renault (AgroParisTech, France). *Sth* strains were inoculated in M17 medium supplemented with 0.5% lactose (LM17) (Oxoid or HiMedia), and the cultures were incubated at 37°C overnight, or at 42°C during the day. *E. coli* Top10 was used for plasmid construction and maintenance. *E. coli* Stellar (*dcm*-/*dam*-) was used to generate unmethylated plasmids for the target interference assay. pWAR plasmids was kindly provided by Dr. Michael Federle (University of Illinois). When needed, chloramphenicol was supplemented at 2 μg/mL in LM17 liquid broth, and at 5 μg/mL in LM17 plates (1% agar) for *Sth*; and was supplemented at 10 μg/mL in LB for *E. coli*.

Construction of the *Sth* mutant strains was achieved by a well-developed natural transformation procedure (75). The primers used for PCR amplification of the recombination templates are listed in Table 1. Linear pWAR plasmid was used as the backbone of the construction of the target/non-target plasmids. The

primers used for PCR amplification and the reverse complimentary oligos to make the insertions are listed in Table 3.1. The insertions were ligated into the vector by GeneArt Seamless Cloning & Assembly kit (Thermo Fisher).

***Plasmid transformation***

The target interference assay in Figure 3.1B was performed by the Glycerol protocol. Briefly, *Sth* strains were inoculated in 5 mL LM17 and the cultures were incubated at 37°C overnight. 100 mL fresh LM17 was inoculated with 1% of an overnight culture, and incubated at 42°C until $OD_{600}$ value reached 0.5. The culture was then placed on ice for 15 minutes and then centrifuged for 10 minutes at 5000x*g* at 4°C. Supernatant was decanted and the pellet was resuspended in 2 mL ice-cold wash solution (10% glycerol + 0.4 M sorbitol). The resuspended cells were washed three times by centrifugation for 1 minute at 15000 RPM at 4°C in the wash solution. The electroporation-competent cells were then resuspended in 500 µL wash solution and aliquoted. 1 µg unmethylated target or control plasmid DNA (generated by *E. coli* Stellar strain) was mixed with 40 µL competent cells, and electroporated into the cells by Gene Pulser (BioRad) at 25 µF, 200 Ω, and 1.8 KV. The transformants were incubated in 1 mL recovery solution (LM17 + 0.4 M sorbitol + 20 mM $MgCl_2$ + 2 mM $CaCl_2$) for 2 hours at 42°C, and then plated onto an LM17 plate (1% agar) with chloramphenicol.

The target interference assay in Figure 3.2B was performed by the Threonine protocol. Briefly, *Sth* strains were inoculated in 5 mL TYL broth (0.5%

tryptone, 0.5% yeast extract, and 0.5% lactose) at 37°C overnight. 5 mL fresh TYL broth supplemented with 40 mM DL-threonine was inoculated with 1% of an overnight culture, and incubated at 42°C until $OD_{600}$ value reached 0.8. The culture was then centrifuged for 10 minutes at 3200 rpm at 4°C. Supernatant was decanted and the pellet was resuspended in 800 µL ice-cold 5 mM $KH_2PO_4$-1 mM $MgCl_2$, pH 4.6. The cells were then centrifuged again for 1 minutes at 14000 rpm at 4°C. Supernatant was decanted and the pellet was resuspended in 800 µL ice-cold electroporation medium (5 mM $KH_2PO_4$ - 0.3 M raffinose- 1 mM $MgCl_2$, pH 4.6). 1 µg unmethylated target or control plasmid DNA generated by *E. coli* Stellar strain, was mixed into the 800 µL competent cells, and electroporated into the cells by Gene Pulser (BioRad) at 25 µF, 200 Ω, and 1.65 KV. The 800 µL transformants were incubated in 200 µL of 5x TYL broth (5% tryptone, 2.5% yeast extract, and 2.5% lactose) for 2 hours at 42°C, and then plated onto an LM17 plate (1% agar) with chloramphenicol.

***Adaptation assay***

20 to 30 colonies of the transformed *Sth* strains were inoculated in 10 mL LM17 with chloramphenicol, and the cultures were incubated at 37°C for overnight. The total DNAs of the overnight cultures were extracted by Quick-DNA Fungal/Bacterial Miniprep kit (Zymo Research). Tthe leader-proximal ends of the CRISPR arrays were amplified from extracted DNAs by multiple round PCR from the extracted DNAs (See Chapter 2, Figure 2.1B). PCR primers are listed in Table S1. Expanded amplicons from the first round PCRs were separated from

unexpanded products by gel electrophoresis, bands of the correct size were cut, and DNAs were isolated by a gel recovery kit (Zymo Research). When expanded amplicons were too faint to visualize, the region of a gel lane corresponding to amplicons in the expanded size range was cut. Illumina high-throughput sequencing (HTS) overhangs and Illumina HTS index barcodes were added to the expanded array amplicons by PCR. Purified PCR products were ranked by PCR intensity and then pooled, concentrated by ethanol precipitation, quantitated, and diluted to a suitable concentration for Illumina platform sequencing. Array libraries were sequenced on an Illumina MiSeq, set to yield 250 by 50 paired end reads; the 250 base read 1 sequences were used in this study. After sequencing, samples were de-multiplexed by index, and the sequence corresponding to a new (expanded) spacer was extracted from each read. New spacers were aligned to reference sequences (bacterial chromosome and appropriate plasmids) using Bowtie (76) to identify the protospacer sequence.

## Results

### *Sth JIM8232 Type III-A CRISPR-Cas system is active in target interference*

*Sth* JIM8322 has an intact Type III-A CRISPR-Cas system in its genome, containing 17 pre-existing spacers (77). Based on the target cleavage features of Type III systems discussed above, to investigate the target interference ability of the Type III-A system of this strain, we used a pWAR-based plasmid containing a transcribed-target of pre-existing spacer1 (Figure 3.1A), and a pWAR plasmid

containing a transcribed-non-target sequence as well as the empty plasmid as the negative controls (Figure 3.1A), to challenge wild-type (WT) *Sth* JIM8232, separately. To test whether directional transcription is required for target interference by this system, we also included another pWAR plasmid containing a non-transcribed-target sequence (Figure 3.1A). Spacer1 of the system matches a fragment of *Sth* bacteriophage 5093 (78). Because the 3'-PFS is important for the authentication of target cleavage by Type III systems (22,32,34), the natural PFS (CTCGCTTG) of bacteriophage 5093 was inserted downstream of the target/non-target sequence for each of the three target/non-target plasmids. There are terminators following the PFSs to prevent differences resulting from run-through transcription. The transformation efficiency of the transcribed-target plasmid was about two orders of magnitudes lower than that of the other three plasmids (Figure 3.1B), indicating that the system is active in defense, which is dependent on the directional transcription of the target sequence. To explore the roles of Cas proteins and their functional domains in target interference, the *csm1* mutated and the *cas* gene(s) deleted strains were challenged by the plasmids as well. While Cas1 and Cas2 were not involved in target interference, all the other Cas proteins were required (Figure 3.1B). Interestingly, mutations within the Csm1 Palm domain (D575A, D576A) abolished the defense, and the ones within the Csm1 HD domain (H15A, D16A) only slightly hampered the activity (Figure 3.1B), indicating that the non-specific RNase activity of Csm6 activated by cA*n*s was essential for interference against the target plasmid, while the DNase activity of Csm1 was less important for that.

***Priming spacers change the proportion and the distribution of protospacers mapped to the plasmid***

The Type III-A CRISPR-Cas system of *Sth* JIM8232 had been shown active in naïve adaptation (See Chapter 2). To investigate whether the system has primed adaptation ability, we transformed WT JIM8232 with a pWAR-based target plasmid with the reverse PFS (CAAAGGCA), and a pWAR-based non-target plasmid as the negative control, separately. The reverse PFS was reverse to the crRNA tag, so that it authenticated target interference by the lack of any complementarity to the crRNA tag. However, compared with the natural PFS adopted from phage 5093, the reverse PFS made the target plasmid sensitive but less sensitive to Type III CRIPSR-Cas-mediated target interference (Figure 3.2A), so that spacer1 target sequence with the reverse PFS was expected to be able to trigger the potential primed adaptation. We performed cell culture, multiple round PCR, and HTS (See Materials and Methods, and Chapter 2 Figure 2.1). If the system had primed adaptation activity, it would be expected that the proportion and the distribution of the new protospacers might change relative to the target/priming protospacer (PPS), and the adaptation against the target plasmid would be much more efficient than the one against the non-target plasmid. The expanded PCR amplicons of the leader-proximal ends of the CRISPR arrays were observed by the second round of PCR for both the experimental and the control group, and the expanded band of the experiment group was not stronger than that of the control group (Figure 3.2B). We reasoned

that since the target plasmid was sensitive to CRISPR-Cas-mediated clearance, *Sth* cells might have limited period to maintain the substrate for CRISPR-Cas adaptation. Despite this, the existence of the transcribed-target elevated the proportion of the protospacers from the plasmid by about 3-fold (Figure 3.2C). Moreover, compared with the non-target plasmid, the protospacers of the target plasmid accumulated at the encoding strand of the ~800 bp region downstream to the transcribed target (Figure 3.2D), which was the DNA cut site of Csm1 according to the current knowledge (34). These findings suggested that the primed adaptation mechanism exists in Type III CRISPR-Cas systems, which might be due to DNA breaks generated by the HD domain of Csm1.

***The transcription-dependent DNA breaks by Csm1 DNase activity triggered primed adaptation***

To test whether it was indeed the DNA breaks by the Csm1 DNase activity that triggered the primed adaptation, we monitored the adaptation by WT, the *csm1* HD mutant (H15A, D16A), and the *csm1* Palm mutant (D575A, D576A) strains, against the empty pWAR plasmid, the transcribed-target plasmid, and the non-target plasmid. Furthermore, to test whether the directional transcription of the target DNA (required for Csm1 activation) was required for primed adaptation, adaptation by the WT strain against the non-transcribed-target plasmid was also monitored (Figure 3.3A). Consistent with the previous findings, the transcribed-target sequence increased the proportion and changed the distribution of the protospacers mapped to the plasmid in the WT strain; in

contrast, the non-target sequence had no influence to the protospacer selection (Figure 3.3B). Interestingly, the mutations within the Csm1 HD DNase domain eliminated the plasmid selection preference during adaptation. Unexpectedly, the mutations within the Csm1 Palm domain also had influence to the selection bias (Figure 3.3B). The terminator in the non-transcribed-target plasmid abolished the position bias for the WT system (Figure 3.3D). These findings support the hypothesis that the DNA breaks by the directional transcription-dependent DNase activity of Csm1 triggered primed adaptation (Figure 3.4).

## Discussion

### *Type III-A CRISPR-Cas system preferentially acquired secondary protospacers at the cleavage sites within the target plasmid.*

Since the first description of primed adaptation in a Type I-E CRISPR-Cas system (52,53), the process has been characterized in at least six Type I systems and two Type II systems (52-54,56-61). For Type I and Type II systems, the target invaders can escape from CRISPR-Cas-mediated cleavage by one or few single nucleotide mutation(s) at critical positions within the target or the PAM (50,51). Primed adaptation provides efficient secondary adaptations against invaders (53,54), and thus minimizes the frequency of CRISPR-Cas escapers.

This work is the first study investigating primed adaptation by Type III systems. Although target interference of Type III systems tolerate a broad range of PFSs and many mutations within the targets to minimize the invader escapers (20,22,72,73), we still observed primed adaptation in the Type III-A system of *Sth*

JIM8232. The existence of a PPS increased the proportion of the plasmid derived protospacers from ~5% to ~20% (Figure 3.2D and 3.3B). The accumulation of the secondary protospacers at the non-template strand downstream to the target sequence was observed (Figure 3.2E and 3.3C), which had been proposed to be the DNA cut sites of Type III systems (34).

Compared with the tested Type I and Type II systems (52-54,56-60), the targeted plasmid recognition during primed adaptation by the Type III system of *Sth* JIM8232 was not as prominent. Moreover, during adaptation assay, the PPS of the Type III system did not apparently increase the adaptation detection rate (Figure 3.2C). Upon first thought, the lack of increased spacer uptake may suggest the absence of primed adaptation, with some other phenomenon resulting in the observed spacer-uptake patterns. However, the transcribed-non-target plasmid was used as a control for the target plasmid, which ruled out the influence of the additional promoter to adaptation (See also Chapter 2). Moreover, the existence of the Type III PPS indeed changed the distribution of the protospacers among the plasmid during secondary adaptation (Figure 3.2E and 3.3C). We propose here 3 possible explanations for the relative deficiencies of the targeted plasmid recognition and adaptation rate improvement by the Type III-A system: 1) The targeted invaders of the Type III systems were found able to escape from the CRISPR-Cas-mediated defense by large or whole target sequence deletion (74), and the antibiotic survivors of the target plasmid transformed *Sth* cells might have mutations within their CRISPR-Cas systems, which could depress their adaptation ability. As a result, more colony collection

may be needed to further test the existence and the patterns of Type III primed adaptation in the future work; 2) It could also possibly be that the target plasmid was not an escaper of the Type III-A CRISPR-Cas-mediated defense. For tested Type I and Type II systems, the PPSs that were used in the primed adaptation assay were designed to make the primed plasmid insensitive to the CRISPR-Cas-mediated defense (52-54,56-60), so that the primed plasmids had roughly the same number and duration with the non-target plasmid controls in the host populations. In contrast, the target plasmid I used was not a CRISPR-Cas escaper and was sensitive to the cleavage (Figure 3.2B). As a result, the plasmid is expected to be cleaved, meaning that the Type III-A system had limited time period to adapt against the plasmid before its clearance. Therefore, while the rate of adaptation is elevated, the time of this elevated adaptation is lessened, giving a 'normal' level of adaptation overall; 3) A third possible reason is that the cloning site of the target/non-target sequence in pWAR was intrinsically highly targeted during naïve adaptation (See Chapter 2), so that the improvement of the adaptation by the PPS was not as prominent as expected. As a result, it would be necessary to clone the target in more regions, which were infrequently sampled during the naïve adaptation, to further test primed adaptation.

***The possible mechanism of primed adaptation by the Type III-A CRISPR-Cas system.***

For the Type I-E system in *E. coli*, a priming spacer stimulates adaptation at the primed strand (52,53). For the other tested Type I systems, the secondary

new spacers distributed at both primed and non-primed strands with an obvious gradient centered at the target/protospacer of the pre-existing spacer; moreover, only the upstream strands of the PPS and the target (reverse complementary to the PPS) were preferentially adapted during priming, rather than the downstream strands (see the Figure 8 of ref. 54 for better view) (54,56-59). The adaptation hotspots during primed adaptation were the cutting sites of Cas3, the *in trans* endonuclease and helicase of the Type I-E system. For the reported Type II-A systems, the new protospacers acquired during priming located at immediate vicinities of the target (60,61). The distribution of the secondary protospacers around the PPS of the Type II-A systems had no strand bias. Correspondingly, Cas9-mediated DNA interference by Type II systems had been shown to cleave each strand at a single position within the target sequence and generate blunt breaks (79,80). The preference of the DNA cleavage sites during secondary adaptation by Type I and Type II systems implies a possible mechanism for the primed adaptation: DNA breaks or short DNA fragments generated by target interference facilitate adaptation (64-68). DNA breaks were shown to promote adaptation in several studies (81-83).

In 2015, Samai et al. showed that the activated Cas10 of the Type III system of *Staphylococcus epidermidis* generated cuts at the non-template strand downstream of the target sequence (34). In this work, during the PPS triggered secondary adaptation by the Type III-A system of *Sth*, the ~800 bp region downstream of the target on the non-templated strand was preferred (Figure 3.2E and 3.3C), which supports the hypothesis that DNA cuts mediated by Csm1

triggered primed adaptation (Figure 3.4). The ssDNAs with free ends generated by DNA cleavage act as substrates for adaptation by Cas1-Cas2 hexamers. How the ssDNAs are integrated into the CRISPR array is unknown. If primed adaptation proceeds similar to naïve adaptation (See Chapter 2), one might predict that released ssDNAs are able to form secondary structures, and can be recognized by the adaptation machinery of the Type III CRISPR-Cas system (See Chapter 2). Supporting the hypothesis that DNA cuts triggered the primed adaptation, the block of the transcription of the target DNA sequence, and the mutations within the DNase domain of Csm1, abolished primed adaptation (Figure 3.3). Unexpectedly, mutations within the Palm domain of Csm1 also made the proportions of plasmid derived protospacers not significantly different between naïve and primed adaptation (Figure 3.3B), implicating a potential role of non-specific RNA cleavage by Csm6 in primed adaptation or more simply, that the introduced mutation disrupts the overall function of the Csm1 protein including its ability to carry out primed adaptation.

***Future directions.***

As discussed above, the relative deficiencies of the targeted plasmid recognition and adaptation rate improvement challenged the reality of the conclusions we made our in primed adaptation study about the Type III-A CRISPR-Cas system. As a result, we need to clone the target sequence in the regions that are normally infrequently sampled during naïve adaptation, and also need to evaluate escaping target plasmids, to further test the findings about

primed adaptation. We tried to clone the target sequence into multiple positions that were lowly adapted during the naïve adaptation. However, because the pWAR plasmid is small, the cloning of the transcribed target/non-target sequences outside its multiple cloning site might influence the expression of the replication related elements and had been found to apparently undermine the stability of the plasmid. As a result, we subcloned the transcribed-target and non-target sequences into a lowly adapted region of another plasmid, pIB184 plasmid (according to naïve adaptation findings in Chapter 2), and the subcloning did not negatively influence the stability of the plasmid. Moreover, we constructed a series of pIB184- and pWAR-based target plasmids with different mutations within the target sequence, and also a series of pIB184- and pWAR-based target plasmids with different PFSs. The new target plasmids had different sensitivities to Type III-A CRISPR-Cas-mediated defense. In the Type I system, it had been shown that the different mutations within the PAM or the target sequence could either trigger primed adaptation, or authenticate interference (54,59). The newly constructed plasmids will be invaluable to further test the existence of primed adaptation in the Type III-A CRISPR-Cas system, and to characterize the relationship between target interference and primed adaptation.

Unlike Type I or Type II systems, Type III CRISPR-Cas systems tolerate a broad range of PFSs and many mutations within the targets during target interference, so few targeted invaders could escape from the defense by mutations (20,22,72,73). However, naïve adaptation by the Type III system preferentially uptakes the protospacers at the coding strands of the promoter

regions of expressed genes, which will not be able to handle target interference (See Chapter 2), since target interference by Type III systems is directional and transcription-dependent. Moreover, the bona fide protospacers acquired from lowly or lately transcribed genes of the lytic phages may be insufficient to perform effective defense (32,44) (See also Chapter 2). Theoretically, primed adaptation provides the hosts with a secondary chance to acquire more spacers and functional spacers to ensure the complete defense. This hypothesis is worth testing.

## References


1. Hille, F., Richter, H., Wong, S.P., Bratovic, M., Ressel, S. and Charpentier, E. (2018) The Biology of CRISPR-Cas: Backward and Forward. *Cell*, **172**, 1239-1259.
2. Haft, D.H., Selengut, J., Mongodin, E.F. and Nelson, K.E. (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS computational biology*, **1**, e60.
3. Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F. *et al.* (2011) Evolution and classification of the CRISPR-Cas systems. *Nature reviews. Microbiology*, **9**, 467-477.

4. Bhaya, D., Davison, M. and Barrangou, R. (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annual review of genetics*, **45**, 273-297.

5. Heler, R., Marraffini, L.A. and Bikard, D. (2014) Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. *Molecular microbiology*, **93**, 1-9.

6. Sorek, R., Lawrence, C.M. and Wiedenheft, B. (2013) CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annual review of biochemistry*, **82**, 237-266.

7. Agari, Y., Sakamoto, K., Tamakoshi, M., Oshima, T., Kuramitsu, S. and Shinkai, A. (2010) Transcription profile of Thermus thermophilus CRISPR systems after phage infection. *Journal of molecular biology*, **395**, 270-281.

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

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

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

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

12. Hatoum-Aslan, A., Maniv, I. and Marraffini, L.A. (2011) Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 21218-21222.

13. Carte, J., Pfister, N.T., Compton, M.M., Terns, R.M. and Terns, M.P. (2010) Binding and cleavage of CRISPR RNA by Cas6. *Rna*, **16**, 2181-2188.

14. Wang, R., Preamplume, G., Terns, M.P., Terns, R.M. and Li, H. (2011) Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. *Structure*, **19**, 257-264.

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

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

17. Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S.V. *et al.* (2012) Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Molecular cell*, **45**, 303-313.

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

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

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

21. Tamulaitis, G., Kazlauskiene, M., Manakova, E., Venclovas, C., Nwokeoji, A.O., Dickman, M.J., Horvath, P. and Siksnys, V. (2014) Programmable RNA shredding by the type III-A CRISPR-Cas system of Streptococcus thermophilus. *Molecular cell*, **56**, 506-517.

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

23. Mulepati, S. and Bailey, S. (2013) In vitro reconstitution of an Escherichia coli RNA-guided immune system reveals unidirectional, ATP-dependent

degradation of DNA target. *The Journal of biological chemistry*, **288**, 22184-22192.

24. Redding, S., Sternberg, S.H., Marshall, M., Gibb, B., Bhat, P., Guegler, C.K., Wiedenheft, B., Doudna, J.A. and Greene, E.C. (2015) Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. *Cell*, **163**, 854-865.
25. Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V. and van der Oost, J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960-964.
26. Jiang, F., Zhou, K., Ma, L., Gressel, S. and Doudna, J.A. (2015) STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science*, **348**, 1477-1481.
27. Anders, C., Niewoehner, O., Duerst, A. and Jinek, M. (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, **513**, 569-573.
28. Liu, L., Li, X., Ma, J., Li, Z., You, L., Wang, J., Wang, M., Zhang, X. and Wang, Y. (2017) The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a. *Cell*, **170**, 714-726 e710.
29. Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I. *et al.* (2017) Diversity and evolution of class 2 CRISPR-Cas systems. *Nature reviews. Microbiology*, **15**, 169-182.

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

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

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

33. Liu, T.Y., Iavarone, A.T. and Doudna, J.A. (2017) RNA and DNA Targeting by a Reconstituted Thermus thermophilus Type III-A CRISPR-Cas System. *PloS one*, **12**, e0170552.

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

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

36. Osawa, T., Inanaga, H., Sato, C. and Numata, T. (2015) Crystal structure of the CRISPR-Cas RNA silencing Cmr complex bound to a target analog. *Molecular cell*, **58**, 418-430.

37. Taylor, D.W., Zhu, Y., Staals, R.H., Kornfeld, J.E., Shinkai, A., van der Oost, J., Nogales, E. and Doudna, J.A. (2015) Structural biology. Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning. *Science*, **348**, 581-585.

38. Jung, T.Y., An, Y., Park, K.H., Lee, M.H., Oh, B.H. and Woo, E. (2015) Crystal structure of the Csm1 subunit of the Csm complex and its single-stranded DNA-specific nuclease activity. *Structure*, **23**, 782-790.

39. Ramia, N.F., Tang, L., Cocozaki, A.I. and Li, H. (2014) Staphylococcus epidermidis Csm1 is a 3'-5' exonuclease. *Nucleic acids research*, **42**, 1129-1138.

40. Hatoum-Aslan, A., Maniv, I., Samai, P. and Marraffini, L.A. (2014) Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system. *Journal of bacteriology*, **196**, 310-317.

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

42. Niewoehner, O., Garcia-Doval, C., Rostol, J.T., Berk, C., Schwede, F., Bigler, L., Hall, J., Marraffini, L.A. and Jinek, M. (2017) Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature*, **548**, 543-548.

43. Foster, K., Kalter, J., Woodside, W., Terns, R.M. and Terns, M.P. (2019) The ribonuclease activity of Csm6 is required for anti-plasmid immunity by Type III-A CRISPR-Cas systems. *RNA biology*, **16**, 449-460.

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

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

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

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

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

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

50. Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J. and Severinov, K. (2011) Interference by clustered regularly interspaced short palindromic repeat

(CRISPR) RNA is governed by a seed sequence. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 10098-10103.

51. Westra, E.R., Semenova, E., Datsenko, K.A., Jackson, R.N., Wiedenheft, B., Severinov, K. and Brouns, S.J.J. (2013) Type I-E CRISPR-Cas Systems Discriminate Target from Non-Target DNA through Base Pairing-Independent PAM Recognition. *PLoS genetics*, **9**.

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

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

54. Fineran, P.C., Gerritzen, M.J., Suarez-Diez, M., Kunne, T., Boekhorst, J., van Hijum, S.A., Staals, R.H. and Brouns, S.J. (2014) Degenerate target sites mediate rapid primed CRISPR adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, E1629-1638.

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

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

57. Richter, C., Dy, R.L., McKenzie, R.E., Watson, B.N., Taylor, C., Chang, J.T., McNeil, M.B., Staals, R.H. and Fineran, P.C. (2014) Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic acids research*, **42**, 8516-8526.

58. Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A. and Severinov, K. (2013) High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in E. coli. *RNA biology*, **10**, 716-725.

59. Garrett, S., Shiimori, M., Watts, E.A., Clark, L., Graveley, B.R. and Terns, M.P. (2020) Primed CRISPR DNA uptake in Pyrococcus furiosus. *Nucleic acids research*, **48**, 6120-6135.

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

61. Mosterd, C. and Moineau, S. (2021) Primed CRISPR-Cas Adaptation and Impaired Phage Adsorption in Streptococcus mutans. *mSphere*, **6**.

62. Hoikkala, V., Ravantti, J., Diez-Villasenor, C., Tiirola, M., Conrad, R.A., McBride, M.J., Moineau, S. and Sundberg, L.R. (2021) Cooperation between Different CRISPR-Cas Types Enables Adaptation in an RNA-Targeting System. *mBio*, **12**.

63. Wiegand, T., Semenova, E., Shiriaeva, A., Fedorov, I., Datsenko, K., Severinov, K. and Wiedenheft, B. (2020) Reproducible Antigen Recognition by the Type I-F CRISPR-Cas System. *CRISPR J*, **3**, 378-387.

64. Stachler, A.E., Wortz, J., Alkhnbashi, O.S., Turgeman-Grott, I., Smith, R., Allers, T., Backofen, R., Gophna, U. and Marchfelder, A. (2020) Adaptation induced by self-targeting in a type I-B CRISPR-Cas system. *The Journal of biological chemistry*, **295**, 13502-13515.

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

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

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

68. Kunne, T., Kieper, S.N., Bannenberg, J.W., Vogel, A.I., Miellet, W.R., Klein, M., Depken, M., Suarez-Diez, M. and Brouns, S.J. (2016) Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation. *Molecular cell*, **63**, 852-864.

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

70. Xue, C., Whitis, N.R. and Sashital, D.G. (2016) Conformational Control of Cascade Interference and Priming Activities in CRISPR Immunity. *Molecular cell*, **64**, 826-834.

71. Blosser, T.R., Loeff, L., Westra, E.R., Vlot, M., Kunne, T., Sobota, M., Dekker, C., Brouns, S.J.J. and Joo, C. (2015) Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. *Molecular cell*, **58**, 60-70.

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

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

74. Artamonova, D., Karneyeva, K., Medvedeva, S., Klimuk, E., Kolesnik, M., Yasinskaya, A., Samolygo, A. and Severinov, K. (2020) Spacer acquisition by Type III CRISPR-Cas system during bacteriophage infection of Thermus thermophilus. *Nucleic acids research*, **48**, 9787-9803.

75. Fontaine, L., Dandoy, D., Boutry, C., Delplace, B., de Frahan, M.H., Fremaux, C., Horvath, P., Boyaval, P. and Hols, P. (2010) Development of

a versatile procedure based on natural transformation for marker-free targeted genetic modification in Streptococcus thermophilus. *Applied and environmental microbiology*, **76**, 7870-7877.

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

77. Delorme, C., Bartholini, C., Luraschi, M., Pons, N., Loux, V., Almeida, M., Guedon, E., Gibrat, J.F. and Renault, P. (2011) Complete genome sequence of the pigmented Streptococcus thermophilus strain JIM8232. *Journal of bacteriology*, **193**, 5581-5582.

78. Mills, S., Griffin, C., O'Sullivan, O., Coffey, A., McAuliffe, O.E., Meijer, W.C., Serrano, L.M. and Ross, R.P. (2011) A new phage on the 'Mozzarella' block: Bacteriophage 5093 shares a low level of homology with other Streptococcus thermophilus phages. *Int Dairy J*, **21**, 963-969.

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

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

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

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

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

**Figure 3.1. Target interference by the Type III CRISPR-Cas system.**

(A) The schematics of the pWAR-based plasmids used into the interference assay. The green arrow represents the promoter, and the red T-shaped symbol represents the terminator. The PFS of all the three target/non-target plasmids is the natural PFS (CTCGCTTG) adopted from the 3'- end of the Spacer1 target of bacteriophage 5093. (B) The transformation assay was done by the Glycerol protocol (See Materials and Methods). ****, $p<0.0001$; ns, no significance; n = 3.

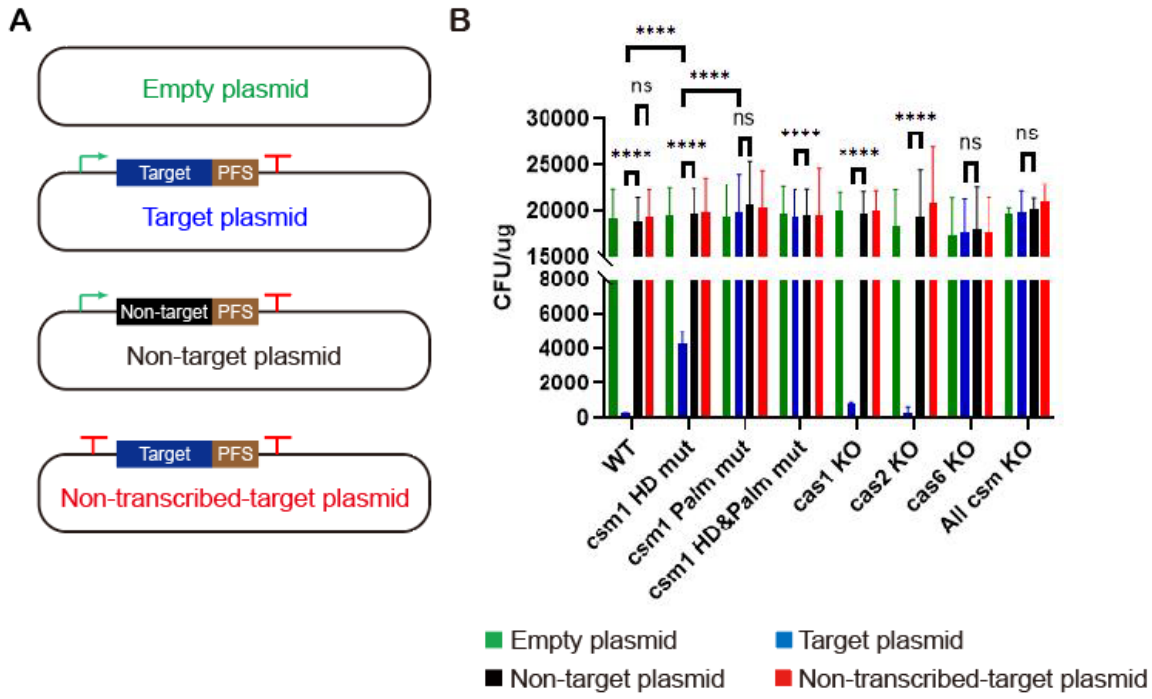

A
Empty plasmid
Target
PFS
Target plasmid
Non-target
PFS
Non-target plasmid
Target
PFS
Non-transcribed-target plasmid
B
CFU/ug
30000
25000
20000
15000
8000
6000
4000
2000
0
****
ns
****
****
ns
****
****
****
ns
ns
WT
csm1 HD mut
csm1 Palm mut
csm1 HD&Palm mut
cas1 KO
cas2 KO
cas6 KO
All csm KO
Empty plasmid
Target plasmid
Non-target plasmid
Non-transcribed-target plasmid

**Figure 3.2. Primed adaptation exists in the Type III-A CRISPR-Cas system.** (A) The schematics of the pWAR-based plasmids used in the adaptation assay. The green arrow represents the promoter. The PFS of both the target and non-target plasmids is the reverse PFS (CAAAGGCA). (B) The transformation assay was done by the Threonine protocol (See Materials and Methods). ****, $p<0.0001$; ***, $p<0.001$; and **, $p<0.01$; ns, no significance; n=3. (C) The gel pictures of the multiple round PCR. *, WT bands of the CRISPR array; +1, the expanded bands of the CRISPR array which reflect adaptation events. (D) The percentages of the total reads of the protospacers came mapped to the plasmid, rather than the self-genome. *, $p=0.0183$; n=3. (E) The protospacer mapping of the target/non-target plasmids. The green arrow represents the promoter in front of the target/non-target sequence. The black rectangle represents the target or the non-target sequence, together with the reverse PFS. The blue arrows and the orange rectangles represent the genes and the replication origins of the plasmid. The unique protospacers matched the plus strand of the plasmid are shown as cyan bars, and the other protospacers matched the minus strand are shown as pink bars. The height of the bars is at the Y-axis, which reflects the numbers of unique reads corresponding to a particular position. (C) and (E) represent at least three independent experiments.

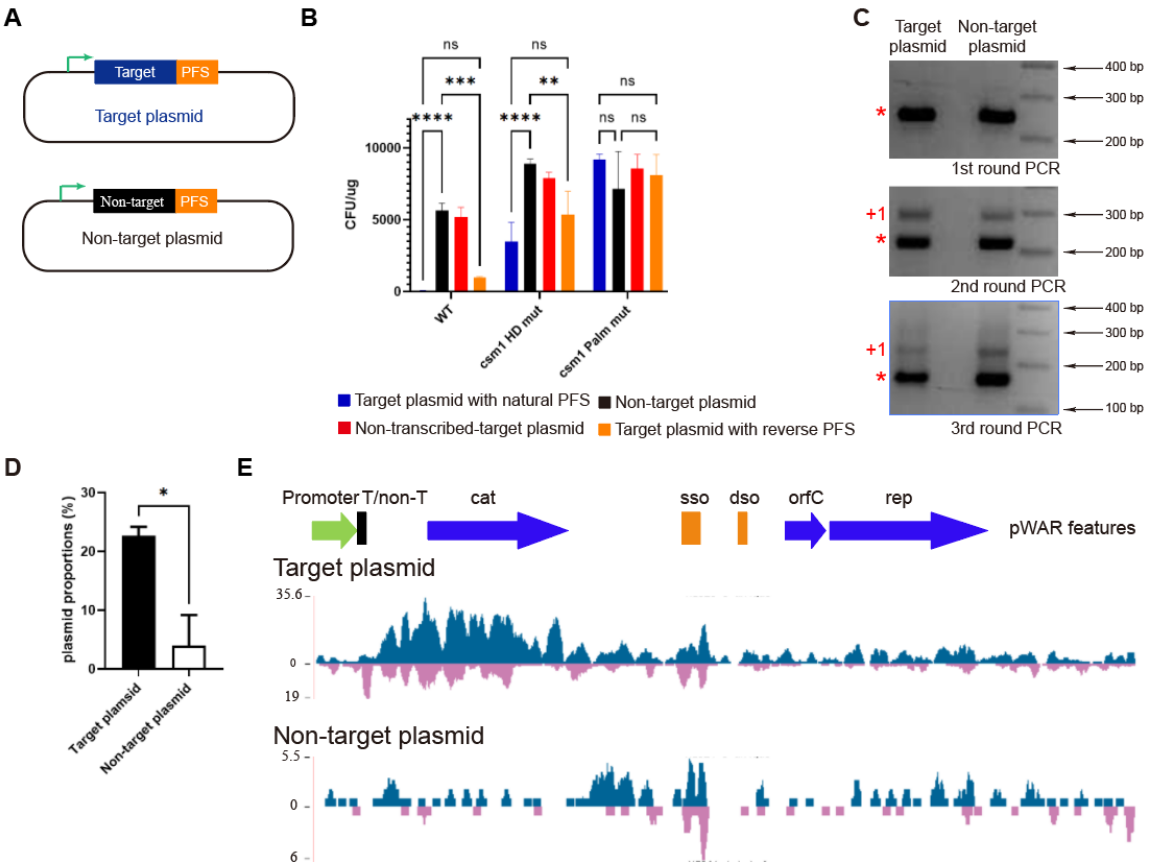

A
Target
PFS
Target plasmid
Non-target
PFS
Non-target plasmid
B
ns
***
****
**
CFU/ug
10000
5000
0
WT
csm1 HD mut
csm1 Palm mut
Target plasmid with natural PFS
Non-target plasmid
Non-transcribed-target plasmid
Target plasmid with reverse PFS
C
Target plasmid
Non-target plasmid
400 bp
300 bp
200 bp
1st round PCR
+1
2nd round PCR
100 bp
3rd round PCR
D
plasmid proportions (%)
30
20
10
0
Target plamsid
Non-target plasmid
E
Promoter
T/non-T
cat
sso
dso
orfC
rep
pWAR features
Target plasmid
35.6
19
Non-target plasmid
5.5
6

**Figure 3.3. The transcription-dependent DNA breaks by Csm1 DNase activity triggered primed adaptation of the Type III-A CRISPR-Cas system.** (A) The schematics of the pWAR-based plasmids used into the interference assay. The green arrow represents the promoter, and the red T-shaped symbol represents the terminator. The PFS of all the three target/non-target plasmids is the reverse PFS (CAAAGGCA) (B) The percentages of the total reads of the protospacers came mapped to the plasmid, rather than the self-genome. **, $p<0.01$; ns, no significance; n=3. (C) and (D) The protospacer mapping of the different plasmids. The green arrow and the red T-shaped symbol represent the promoter and the terminator in front of the target/non-target sequence, separately. The black rectangle represents the target or the non-target sequence, together with the reverse PFS. The blue arrows and the orange rectangles represent the genes and the replication origins of the plasmid. The unique protospacers matched the plus strand of the plasmid are shown as cyan bars, and the other protospacers matched the minus strand are shown as pink bars. The height of the bars is at the Y-axis, which reflects the numbers of unique reads corresponding to a particular position. The pictures represent at least three independent experiments.

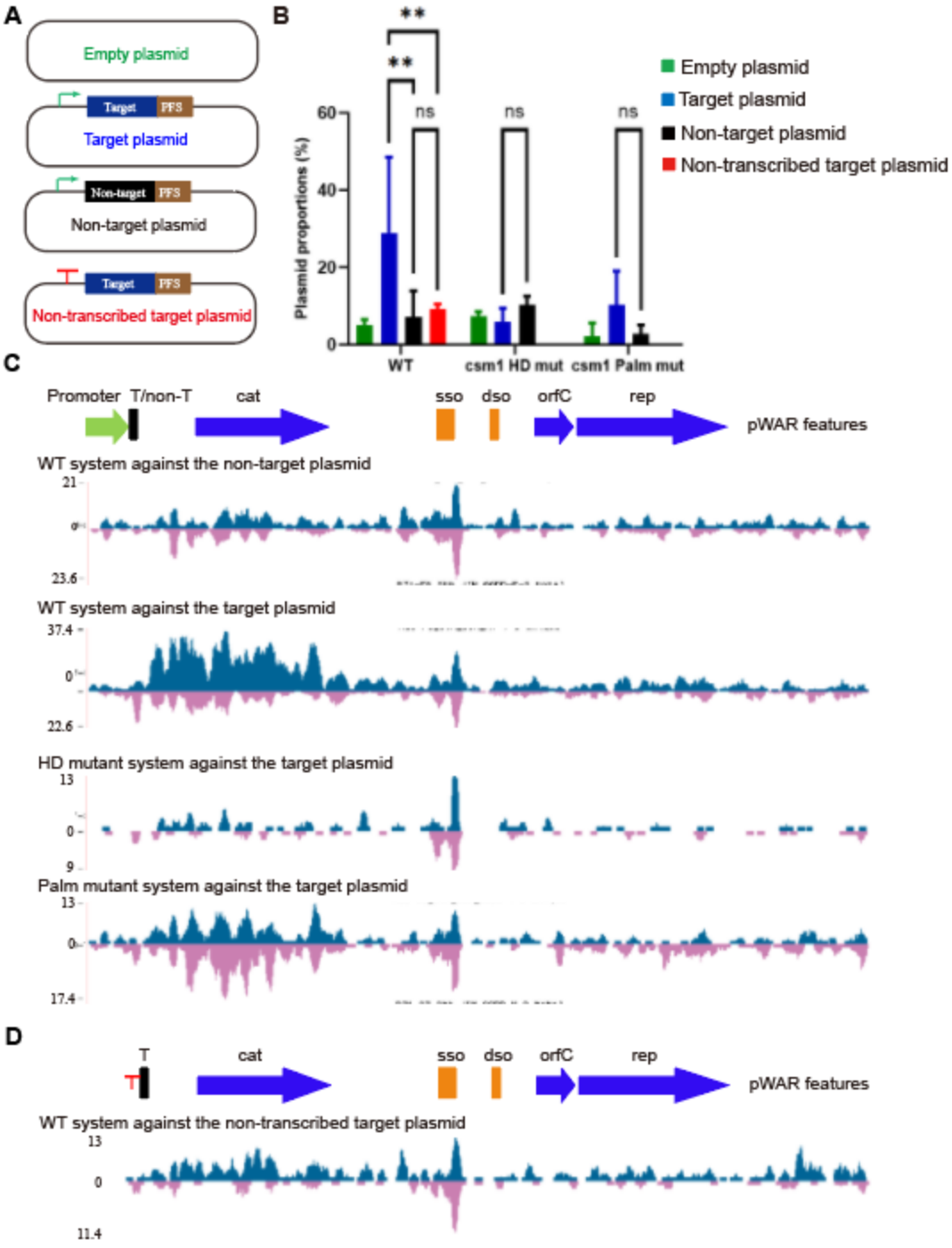

A
Empty plasmid
Target
PFS
Target plasmid
Non-target
PFS
Non-target plasmid
Target
PFS
Non-transcribed target plasmid
B
**
**
ns
ns
ns
60
40
20
0
Plasmid proportions (%)
WT
csm1 HD mut
csm1 Palm mut
Empty plasmid
Target plasmid
Non-target plasmid
Non-transcribed target plasmid
C
Promoter
T/non-T
cat
sso
dso
orfC
rep
pWAR features
WT system against the non-target plasmid
21
0
23.6
WT system against the target plasmid
37.4
0
22.6
HD mutant system against the target plasmid
13
0
9
Palm mutant system against the target plasmid
13
0
17.4
D
T
cat
sso
dso
orfC
rep
pWAR features
WT system against the non-transcribed target plasmid
13
0
11.4

**Figure 3.4. A possible model for primed adaptation by the Type III-A CRISPR-Cas system.** The target sequence and the downstream part of the invader was transcribed by the promoter (green rectangle). The base pairing between the transcript (dashed lines) and the crRNA triggered the dsDNA breaks downstream to the target by the crRNP (black scissors). The ssDNAs with free ends generated by DNA cleavage fuel adaptation by Cas1-Cas2 hexamers. How the ssDNAs were integrated into the CRISPR array is unknown.

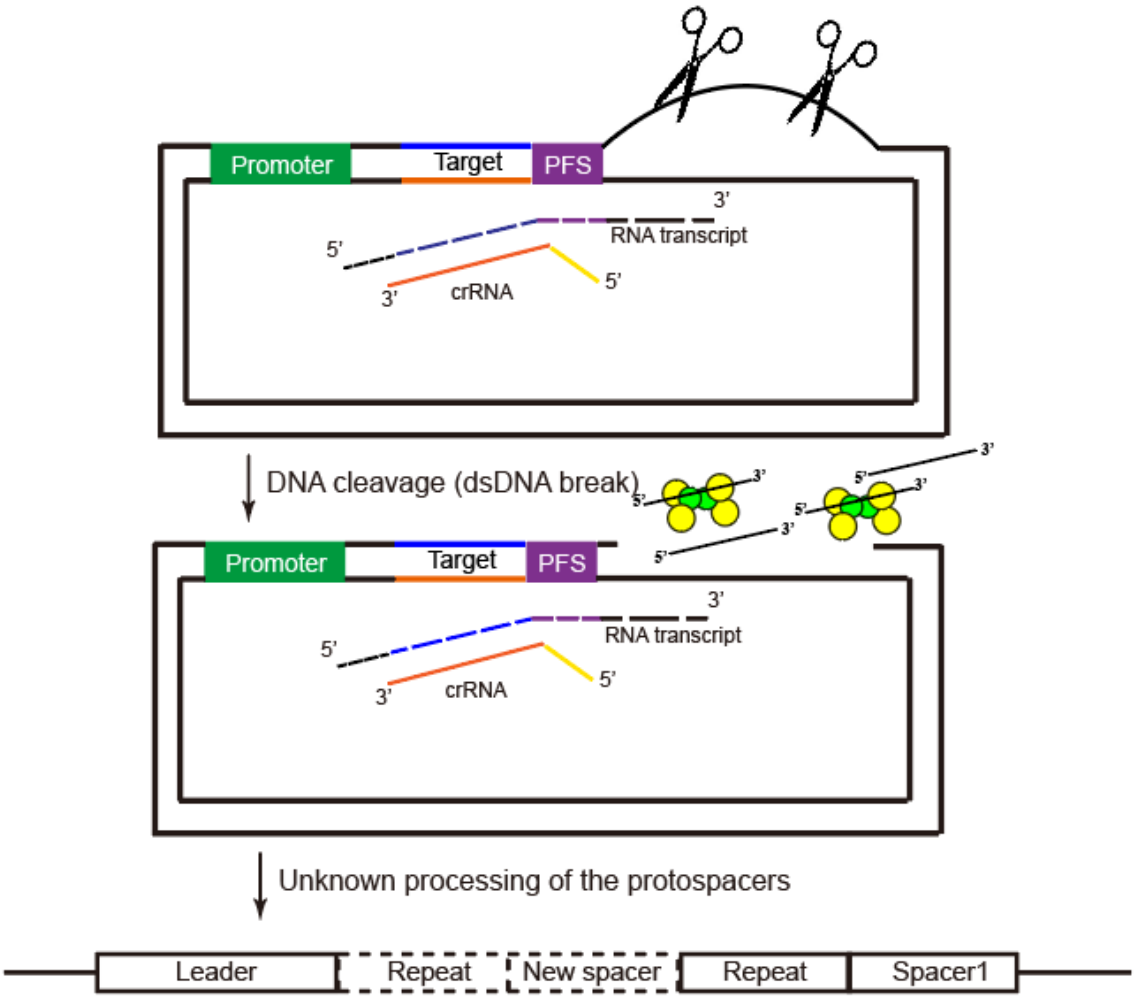

Promoter
Target
PFS
3'
RNA transcript
5'
3'
crRNA
5'
DNA cleavage (dsDNA break)
Promoter
Target
PFS
3'
RNA transcript
5'
3'
crRNA
5'
Unknown processing of the protospacers
Leader
Repeat
New spacer
Repeat
Spacer1

**Table 3.1. The primers used in this work.**

| *Sth* mutant strain construction | | |
|---|---|---|
| Purpose | Name | Sequence (5'-3') |
| *cas1* KO | sl-JIMCas1+2-2-f | CACCGCTAGCATCTGCAGGCAGAAGTCTCTCACAAGATTTCC |
| | JIMCas1KO-M-r | CTAAAAGCCCTAATAAGCTACCTCAGAGTATTTTATGAC |
| | JIMCas1KO-M-f | GAGGTAGCTTATTAGGGCTTTTAGAGAACTTGACC |
| | sl-JIMCas1KO-r | GCATAAGCTTGCGTCGACGGTAAAACCTCTATGCTGTTCTTTGC |
| *cas2* KO | Sl-JIMCas2KO-f | CACCGCTAGCATCTGCAGGGAGTATTGCTACGACCAAGG |
| | JIMCas2KO-M-r | GTCTATAAAGGCTTAATGCCCTCCTGTGTAACTTG |
| | JIMCas2KO-M-f | GAGGGCATTAaGCCTTTATAGACCTTTAATCATATGGTAC |
| | sl-JIMCas1+2KO-r | GCATAAGCTTGCGTCGACGATTGATAAGAAAGTTATTGAAAAACGCCAACAAG |
| *cas6* KO | Sl-JIMCas6Csm-f | CACCGCTAGCATCTGCAGCCAGTTCTGCAAAGAACAGC |
| | JIMCas6KO-M-r | CATTTTGACATAAGCTCGACTCCTTCATCTTTGTATG |
| | JIMCas6KO-M-f | GGAGTCGAGCTTATGTCAAAATGCTTCTAACATTCG |
| | Sl-JIMCas6KO-r | GCATAAGCTTGCGTCGACGAGCAAAAGCTGCTGTCAG |
| All *csm* KO | sl-JIMallCsm-f | CACCGCTAGCATCTGCAGGAAAAAATTAGTATTTACTTTTAAAAGGATCGACC |
| | JIMAllCsmKO-Mol-r | GACACCCAAATATGCTCCGTAAAATAAATCAATCTTTTCTTTCTTC |
| | JIMAllCsmKO-Mol-f | CGGAGCATATTTGGGTGTCATCCCCCTTTG |
| | Sl-JIMallCsm-r | GCATAAGCTTGCGTCGACGGTTAGTTGCTTCTTTGATTTTGGCTAC |
| *csm*1 mutant | sl-dCsm1-WT-f | CACCGCTAGCATCTGCAGGTTAAATTTCATGGCTTCTTGATGG |
| | sl-dCsm1-WT-r | GCATAAGCTTGCGTCGACCTTTACTTCTATCATATAAAGAACTACATAAGG |
| | dCsm1-HD-f | GCGGCGATCGGTAAGGTCATTCAAAGG |
| | dCsm1-HD-r | CAAAAGAGCTCCGTAAAATAAATCAATC |
| | dCsm1-GGDD-f | GCGGCGGTTTTTGCCATCGGCTC |
| | dCsm1-GGDD-r | CCCACCGGCATAGATG |

| | | |
|---|---|---|
| Transcribed target/non-target plasmid construction | | |
| Purpose | Name | Sequence (5'-3') |
| Backbone | endofpWAR-f | GCGGCCGCCACCGCGGTGGG |
| | pPgm-r | GAAATATCTCCTTTTAAATTCAATG |
| Target | 4212 | CATTGCTAACGCTTATTTAGAAGTAGCGTTAGAATCAAGCAAAGGCA |
| | 4213 | TGCCTTTGCTTGATTCTAACGCTACTTCTAAATAAGCGTTAGCAATG |
| Non-target | 4214 | TTTCTAGGAATGGGTAATTATAGCGAGCTAGAAAGCCAAAGGCA |
| | 4215 | TGCCTTTGGCTTTCTAGCTCGCTATAATTACCCATTCCTAGAAA |
| | | |
| Non-transcribed target plasmid construction | | |
| Purpose | Name | Sequence (5'-3') |
| Backbone | linear-pWAR-rev | CTCGAGGGGGGGCCC |
| | XZ160open-f(target) | CATTGCTAACGCTTATTTAGAAGTAGCGTTAG |
| Target | SthTerminator-f | AAATCAACACCCTGTCATTTTATGGCAGGGTGTTTTCG |
| | SthTerminator-r | CGAAAACACCCTGCCATAAAATGACAGGGTGTTGATTT |
| | | |
| Multiple round PCR for the Type III system in *Sth* | | |
| Purpose | Name | Sequence (5'-3') |
| 1st PCR | JC2-leader-1.5f | GCCTTTATAGACCTTTAATCATATGG |
| | JIM-2.02-1r | CTAAAGAACATAAACATATGATGAATGCTTTAGAAC |
| 2nd PCR | JIMCC2-leader-2f | CAATATAGATAGTGTTTCCAGTAGGTCC |
| | JIMCC2.02-2r | GAACATAAACATATGATGAATGCTTTAGAACTG |
| 3rd PCR | JIMCC2-DS-f | ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCAACTGCCTAGAGCACAAGATATGG |
| | JIMCC2-DS-r | GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGCTTATTTAGAAGTAGCGTTAGAATCAAGG |

# CHAPTER 4

# PRIMING SPACERS FACILITATE ADAPTATION BY THE TYPE I-E CRISPR-CAS SYSTEM OF *STREPTOCOCCUS THERMOPHILUS*

---

[1]Xinfu Zhang, Michael P. Terns. To be submitted to PLoS One.

**Abstract**

CRISPR-Cas systems provide prokaryotic cells with a heritable and adaptive mechanism of defense against invaders. *Streptococcus thermophilus* (*Sth*) is one of the most important model organisms for functional studies of CRISPR-Cas systems, with co-existing Type I, II, and III systems. While adaptation by the Type II systems of *Sth* have been well studied, adaptation by the Type I system has only been very rarely detected and not at sufficient levels to thoroughly study adaptation patterns. To better understand Type I adaptation in this important organism, we studied both naïve and primed adaptation by the Type I-E system in this work. Overexpression of *cas1* and *cas2* increased adaptation by the Type I system to a high level. Mapping of the resulting spacers revealed the presence of an upstream AA PAM. Studies regarding primed adaptation have been hampered by the difficulties in the design of escaping target. To study primed adaptation by the Type I system, we developed a novel method to naturally obtain virulent phages that escaped from CRISPR-Cas-mediated defense and acquire the escaping target sequences that triggered primed adaptation. We showed that a priming spacer was able to increase adaptation by this Type I-E system in the absence of overexpression of any *cas* gene. The secondarily adapted protospacers obtained during primed adaptation accumulated at both primed and non-primed strands with an obvious gradient centered at the target/protospacer matching the pre-existing spacer. Taken together, this work characterizes both naïve and primed adaptation by the Type I-E system of *Sth*, and provides a new methodology for future primed adaptation studies.

## Introduction

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins exist in nearly half of eubacteria and almost all archaea (1,2), providing the host with an acquired and highly effective defense mechanism against potential harmful viruses and plasmids in the environment (3-10). A CRISPR array is assembled by identical partially palindromic sequences (repeats), interspaced by variable sequences acquired from invading elements (spacers) with a same or approximate length(s), following an AT rich 'leader' sequence, which promotes the transcription of the array (7,11-16). CRISPR-Cas systems function in three major steps: adaptation, CRISPR RNA (crRNA) biogenesis, and target interference (11,15,16). CRISPR-Cas systems are diverse and classified into six major types (Type I-VI) and multiple subtypes according to the repeat sequence identities of CRISPR arrays and different *cas* genes adjacent to them (1,2,17-20).

The Type I-E system of *E. coli* is one of the best studied CRISPR-Cas subtypes. The maturation of the crRNA of this system is performed by CasE (5). A small and mature crRNA guides the target interference machinery (Cascade) to bind to the target DNA, and then recruits the *trans* acting endonuclease and helicase, Cas3, to cleave the invader (5,21). While such interference activity is readily observed, adaptation by the Type I-E system of *E. coli* could be barely detected (13). The deletion of a global transcription regulator, *hns*, or overexpression of *cas1* and *cas2* of the system, had been shown able to

substantially increase adaptation efficiency (13,22,23). Cas1 and Cas2 are the only two Cas proteins required for adaptation by this system, and the requirement of the leader sequence and at least one repeat unit for CRISPR-Cas adaptation was first discovered in this paradigmatic system (24). This system selects 'suitable' fragments among heterologous DNAs by protospacer adjacent motifs (PAMs), and integrates new spacers into the CRISPR array in a PAM-dependent way (22). Like other Type I systems and Type II systems (25,26), the presence of a PAM is critical in the authentication of target interference by this system (27). It has been shown that one mismatch at the $1^{st}$-$5^{th}$ or $7^{th}$ -$8^{th}$ nucleotide of the crRNA, or a mutation in the PAM region of the target, can eliminate the effect of interference mediated by this system (27).

Although an imperfect target impedes crRNPs from degrading the invading nucleic acids, it can significantly improve the acquisition effectiveness of additional spacers from the same molecule by this system (22,23,28), providing the bacteria a co-evolutionary defense strategy with their invaders (28). This priming spacer-dependent secondary adaptation is termed primed adaptation, distinguished from naïve adaptation. Like primed adaptation of other tested Type I CRISPR-Cas systems (29-32), primed adaptation of this Type I-E system in *E. coli* is substantially more efficient than naïve adaptation (28), and increases the proportion of new spacers adapted from primed invader, rather than host genome (22,23,28). In the other tested Type I systems, primed protospacer (PPS) changed the distribution of the secondarily adapted new protospacers, in that the new protospacers accumulated at both primed and non-primed strands

with an obvious gradient centered at the target/PPS (29-32). In contrast, during primed adaptation by the Type I-E system of *E. coli*, only primed strand with PPS was preferentially adapted, without a great bias of the distance to the PPS (22,23). While Cas1 and Cas2 are sufficient Cas elements for naïve adaptation by this systems, genetic evidences showed that Cas3 and crRNP are also required for primed adaptation (22,28).

Four CRISPR-Cas systems co-exist in *Streptococcus thermophilus* (*Sth*) DGCC7710. CRISPR1 and CRISPR3 belong to subtype Type II-A, CRISPR2 belongs to subtype Type III-A, and CRISPR4 belongs to subtype Type I-E, separately (33-35). CRISPR4-Cas system of *Sth* is homologous to the Type I-E system of *E. coli* (35,36). While CRISPR1 and CRISPR3 have been shown efficient in the adaptation (3,26,33), adaptation by the Type I-E system of *Sth* could only barely be detected by HTS after tens and even hundreds of days of the cocultivation with phage (37). Despite this, it was clear that Cas1-Cas2 complex of this system was capable of integrating a prespacer into the CRISPR array *in vitro* (38). The PAM of this system had been predicted to be AA immediately upstream of the protospacers by analyses against the pre-existing spacers (39), and the system had been observed able to cleave target DNA *in vitro* (36,39).

To better understand adaptation by the Type I-E CRISPR-Cas systems, as well as to fully understand CRISPR-Cas functions in the important model bacterial species, *Sth*, we performed a functional study of both naïve and primed adaptation by the Type I-E system of *Sth*. This work showed that the Type I-E

system of *Sth* was active in target interference *in vivo*, and was very inefficient in adaptation. Overexpression of *cas1* and *cas2* increased adaptation by this system. The protospacers acquired during adaptation were indeed selected from sequences with the upstream AA PAM. Moreover, we developed a novel method to naturally obtain virulent phages that escaped from CRISPR-Cas-mediated defense and acquire the escaping target sequences that triggered primed adaptation. Using this research method, we showed that a priming spacer was able to increase adaptation to readily detectable levels by this Type I-E system in the absence of overexpression of any *cas* gene. In contrast to its homologous system in *E. coli*, the secondarily adapted protospacers during primed adaptation by this Type I-E system in *Sth* accumulated at both primed and non-primed strands with an obvious gradient centered at the target/PPS of the pre-existing spacer. This work fills a large gap of our knowledge about CRISPR-Cas functions in *Sth*, revealed the common secondary protospacer distribution pattern during primed adaptation by the Type I-E systems, and provides a new method for future primed adaptation studies.

## Materials and methods

### *Strain and plasmid manipulation*

*Sth* DGCC7710 and phage 2972 were kindly provided by Dr. Sylvain Moineau (Univerite laval, Canada). *Sth* strains were inoculated in M17 medium supplemented with 0.5% lactose (LM17) (Oxoid), and the cultures were incubated at 37$^{o}$C overnight, or at 42$^{o}$C during the day. *E. coli* Top10 was used

for plasmid construction and maintenance. *E. coli* Stellar (*dcm*-/*dam*-) was used to generate unmethylated plasmids for the target interference assay. pWAR plasmid was kindly provided by Dr. Michael Federle. When needed, chloramphenicol was supplemented at 2 μg/mL in LM17 liquid broth, and at 5 μg/mL in LM17 plates (1% agar) for *Sth*; chloramphenicol was supplemented at 10 μg/mL in LB broth for *E. coli*; erythromycin was supplemented at 15 μg/ml for *Sth* and 150 μg/ml for *E. coli*, separately.

Construction of the *Sth* mutant strains with Type II systems knocked out was achieved by a well-developed method using a temperature sensitive plasmid, pINTRS (40). The overexpression plasmid of *cas1* and *cas2* was constructed using pWAR plasmid as the backbone. The primers used for PCR amplification of the linear vector and the insertions are listed in Table 4.1. The insertions were ligated into the vector by GeneArt Seamless Cloning & Assembly kit (Thermo Fisher).

***Target interference assay***

The *Sth* strains were inoculated in 5 mL LM17 and the cultures were incubated at 37°C for overnight. 100 mL fresh LM17 was inoculated with 1% of an overnight culture, and incubated at 42°C until $OD_{600}$ value reached 0.5. The culture was then placed on ice for 15 minutes and then centrifuged for 10 minutes at 5000x*g* at 4°C. Supernatant was decanted and the pellet was resuspended in 2 mL ice-cold wash solution (10% glycerol + 0.4M sorbitol). The resuspended cells were washed three times by centrifuge for 1 minute at 15000

RPM at 4°C in the wash solution. The electroporation-competent cells were then resuspended in 500 µL wash solution and aliquoted. 1 µg unmethylated target or control plasmid DNA generated by *E. coli* Stellar strain, was mixed into 40 µL competent cells, and electroporated into the cells by Gene Pulser (BioRad) at 25µF, 200Ω, and 1.8KV. The transformants were incubated in 1 mL recovery solution (LM17 + 0.4M sorbitol + 20mM $MgCl_2$ + 2mM $CaCl_2$) for 2 hours at 42°C, and then plated onto an LM17 plate (1% agar) with the according antibiotic.

***Phage infection and mutant analyses.***

The *Sth* strains were inoculated in 5 mL LM17 and the cultures were incubated at 37°C overnight. 5 mL fresh LM17 was inoculated with 1% of an overnight culture, and incubated at 42°C until $OD_{600}$ value reached 0.3. Phage infection was performed with phage 2972 or the mutants by a well-established protocol (41). The primers used to determine the expansion of the CRISPR arrays and the mutations within the protospacer of phage 2972 are listed in Table 4.1.

## Results

***Type I-E CRISPR-Cas system of Streptococcus thermophilus is inefficient in adaptation***

CRISPR4-Cas system of *Sth* DGCC7710 is homologous to a deeply well-studied type I-E system in *E. coli* (35,36), with 12 unique spacers sandwiched by 13 repeats, and all the type I-E specific *cas* genes (Figure 4.1A).

Target interference ability of the system had been observed *in vitro* (36,39), however, it had not been tested *in vivo*. We firstly proposed that the inefficient adaptation detection against the *Sth* population co-cultivated with lytic phages (37) might be a reflection of inability of target interference *in vivo*. Therefore, we challenged *Sth* DGCC7710 with the empty pWAR plasmid as the negative control, and with a pWAR plasmid containing the target sequence of spacer1 of the Type I-E system with the predicted AA PAM, separately. In contrast to the hypothesis, the system was found to be active in defense *in vivo* (Figure 4.1B), indicating that the inefficient adaptation against the lytic phage was not a reflection of target interference ability.

CRISPR1 had been reported dominant for adaptation in *Sth* DGCC7710 and CRISPR3 is also active in spacer acquisition (3,33,42), which suggested another hypothesis that the blindingly high efficiency of *Sth* Type II-A systems in invader defense overshadowed the adaptation events at CRISPR4. To test this assumption, a Type II-A CRISPR-Cas knockout (KO) strain (all the *cas* genes, learder sequences, and CRISPR arrays of the two Type II-A systems were knocked out), and wild-type (WT) *Sth* DGCC7710 strain were separately challenged by lytic phage 2972 (43) under a same condition, at multiplicity of infection (M.O.I.) of 1. The colony-forming units (CFU) of the Type II null strain was ~60 fold lower than the CFU of WT strain. Randomly chosen *Sth* survivors of the two strains were analyzed for adaptation events at all the CRISPR loci that they had, by PCR amplification of the leader proximal region of the individual CRISPR arrays. For WT strain, consistent with the previous findings

(3,33,42,44,45), 14 of 24 survivors (58.3%) that we tested acquired a new spacer at CRISPR1 array, 9 of them (37.5%) acquired a new spacer at CRISPR3 array, and none of them had adaptation at CRISPR2 or CRISPR4 array. The remaining one survivor must have escaped from the phage by a CRISPR-Cas unrelated strategy (Figure 4.2B). In the Type II-A KO strain, none of the 43 tested survivors had any new spacer incorporated (Figure 4.2C). As a conclusion, CRISPR2 and CRISPR4 of *Sth* DGCC7710 are still adaptation inert, even when the two dominant Type II-A systems are removed.

### *Overexpression of cas1 and cas2 improves adaptation by the Type I-E CRISPR-Cas system*

Cas1 and Cas2 are the only two Cas proteins essential for naïve adaptation of the Type I-E system of *E. coli*, and overexpression of the *cas1* and *cas2* genes substantially improved adaptation efficiency of the Type I-E system in *E. coli* (22,24). To test whether a similar amelioration of adaptation frequency would result in *Sth*, we overexpressed *cas1* and *cas2* of *Sth* Type I-E system by a pWAR-based plasmid (46,47) in both WT strain and Type II-A KO strain, separately. After phage 2972 challenge, although the presence of *cas1+2* overexpression plasmid did not greatly increase CFU values of either of the two strain, we observed 12 of 221 tested survivors acquired a new spacer (7/53 from WT strain, 5/168 from the Type II-A KO strain) in CRISPR4, while the Type I-E system of the two strains with the control plasmid remained still inert in adaptation (Figure 4.3A). This result indicated that *cas1+2* overexpression is

sufficient to improve the adaptation ability of Type I-E CRISPR-Cas system in *Sth*, and the low efficiency of the process was caused by low expression of the endogenous genes.

All the 12 new spacers (listed in Figure 4.3B) were from phage 2972 genome, among which 11 were acquired from the coding strand of known or hypothetical genes, and the other one was from the template strand. Among the 12 pre-existing native spacers, the first 9 (those close to the leader proximal end of the array) were 33 nt long, and the last 3 were 34 nt long; while among the 12 newly acquired spacers, 10 were 33 nt, and the other 2 were 34 nt. The PAM sequence of the 12 new protospacers was analyzed by WebLogo (http://weblogo.berkeley.edu/logo.cgi), revealing a conserved 'AA' PAM upstream (Figure 3B), which was consistent with bioinformatic prediction and the immunity assay (39).

***A priming spacer is sufficient to improve adaptation by Type I-E CRISPR-Cas system***

While primed adaptation is a well-studied process in some CRISPR-Cas system subtypes, the ubiquity of priming among CRISPR-Cas systems, and the mechanisms involved remain mysterious. Research has been hampered by the difficulty and uncertainty of the design of PPSs to trigger primed adaptation. To better understand primed adaption by the Type I-E system of *Sth*, I established a novel CRISPR-Cas escaping virus mutants-based method for future research.

The 12 phage 2972 insensitive colonies with a new spacer are referred to as bacteriophage insensitive mutants (BIMs). To further study primed adaptation of the Type I-E CRISPR-Cas system, I infected one of the BIMs of the Type II-A CRISPR-Cas null strain (with the new spacer 5'-TGTTTTAGCATTCATAATTGCACGTTCTGGGTT-3', referred to as XZ2), with WT phage 2972 at a high M.O.I. (=10). Although the plaque-forming units (PFUs) were extremely low, I still attained several phage CRISPR-Cas escaping mutants (CEMs) of the virus, which had point mutation(s) at their target or PAM sequence (Figure 4.4A). Although XZ2 was not resistant to the CEM phages containing mutant targets, its new spacer was still speculated able to pair with the corresponding imperfect protospacers and function as a PPS, which might significantly improve the acquisition effectiveness of additional spacers from the same invading nucleic acid.

To test this, we first selected XZ2 cultures which had lost the *cas1+2* overexpression plasmid by long-term subcultures under no antibiotic selection condition. Plasmid loss was confirmed by PCR amplification, and the new BIM without the plasmid were referred to as XZ2np. We chose three of its CEMs, φxzPh3 (which had a mutation at -1 position in the PAM), φxzPh5 (which had a mutation at +2 position in the protospacer), and φxzPh9 (which had two mutations at +2 and +10 positions in the protospacer), for further study (Figure 4.4A). XZ2np was sensitive to the three CEMs and highly resistant to WT phage 2972, while all the 3 CEMs kept a similar infection effectiveness as WT phage 2972 in control *Sth* group without the newly required spacer (Figure 4.4B). I

infected XZ2, XZ2np, and Type II-A null control strains, with the 3 CEMs under the same condition, at M.O.I. of 1. After the independent infection tests with the three different CEMs, CFU values of XZ2np were 2-8 fold greater than that of the control strain, while the ones of XZ2 strain were ~3 fold greater. The XZ2 strain was grown on chloramphenicol selective plates, which would lower the CFU value of cell cultures.

We randomly chose *Sth* survivors of the three tests to analyze for adaptation events at CRISPR4 systems. While no CRISPR4 adaptation was observed among the 28 survivors tested of the control strain, 43.9% of the survivors of XZ2np acquired a new spacer (21/29, 31/92, 6/11 after the three CEMs challenges, separately), and 98.6% of the survivors of XZ2 acquired a new spacer (23/23, 33/34, 15/15 after the three CEMs challenges, separately) (Figure 4.4C). As a conclusion, a priming spacer is sufficient to improve *Sth* CRISPR4 adaptation ability.

To test the distribution pattern of the secondary protospacers sampled by primed adaptation of the Type I-E system of *Sth*, we sequenced 121 of the survivors that acquired a new spacer of XZ2 or XZ2np after CEM challenges. All of the 121 secondary spacers were from phage 2972 genome with invariable upstream ‘AA’ PAM. Among the 121 protospacers acquired, 40 were from the same strand with the PPS; while the other 81 were from the other strand of the phage. The sequenced secondary protospacers had obvious preference to locate adjacent to the protospacer corresponding to PPS (Figure 4.5). Moreover, there was a clear trend toward 5’-3’ spacer uptake relative to the existing spacer

position, i.e., the top strand provided spacers to the right of the target; the bottom strand provided spacers to the left of the target (Figure 4.5). This suggests the activity of a 5'-3' exonuclease (likely cas3) providing DNA ends for spacer uptake.

## Discussion

### *Naïve adaptation by the Type I-E CRISPR-Cas system of Streptococcus thermophilus*

CRISPR1 is the most dominant CRISPR systems in adaptation in *Sth* DGCC7710, and CRISPR3 is also active for spacer acquisition (3,33,42). Although mature crRNAs of CRISPR4 are also processed (48), and related *cas* genes are expressed and can be induced by phage challenge (48,49), adaptation events in this Type I-E were not able to be detected until hundred days cocultivation with phage (37). In fact, adaptation rarely happens in most of tested bacterial strains, unless *cas1+2* genes were over expressed (22,24), or a priming spacer pre-existed (50).

In this study, we first observed target interference by the Type I-E CRISPR-Cas system in *Sth in vivo*, and excluded the possibility that the extremely low efficiency of the adaptation was a result of the competition by the efficient co-existing Type II systems. We successfully increased the adaptation ability of CRISPR4 by *cas1+2* overexpression, indicating that the low efficiency was likely a reflection of the low expression level of the endogenous *cas* genes. Interestingly, non-CRISPR-Cas based *Sth* survivors were observed after

challenge with the lytic phage 2972 (Figure 4.2B and 4.3A), revealing that there exist more survival strategies in *Sth* to counter invading lysogenic viral challenge. Several restriction-modification protein subunits are expressed in *Sth* and can be induced by phage challenge (49), which may be a candidate strategy.

Paez-Espino et al. (37) cocultured *Sth* with phages for one hundred days, and detected 71 new spacers acquired at the CRISPSR4 locus, however, only one of the 71 new spacers targeted phage, suggesting that this system can cause self-immunity by spacer acquisition from its own genome. As a result, it could be beneficial for the cell to keep *cas1+2* expression at a low level. Under the condition of phage challenge, Cas proteins of this system increases in abundance (49), which probably can provide a naïve adaptation chance for the cells. Once a sequence is acquired by this system from phage, primed adaptation mechanism enables the CRISPR4 system of *Sth* cells to co-evolve with the previously encountered phage.

### *A novel CRISPR-Cas escaping virus mutants-based system to study primed adaptation*

The design of a PPS, or an escaping target sequence, that can trigger primed adaptation had been a difficulty in primed adaptation study. In this work, by studying primed adaptation of the Type I-E system of *Sth*, we developed a novel method based on CRISPR-Cas escaping virus mutants, to acquire the suitable PPS information by natural co-evolution between the prokaryotic host and the lytic virus, rather than by artificial design. To achieve this, prokaryotic

host is first challenged with a lytic virus. Most of the cells will be lysed, but if the strain contains an active CRISPR-Cas system, it is expected to identify several survivors with a new spacer acquired from virus genome, which can be confirmed by colony PCR amplification against the leader-proximal end of the CRISPR array (Figure 4.6A). Stable insensitivity of the survivors to the virus can then be confirmed. The CRISPR-Cas-mediated survivors will then be co-cultivated with WT virus. Most of WT virions units will fail to lyse the CRISPR-Cas-mediated survivors, but it is still expected to identify several plaques formed by CRISPR-Cas-escaping viruses. Most of the escaping viruses will have mutation(s) within or near the protospacer, which prevent the target cleavage by the CRISPR-Cas system. The specific mutation(s) of the CRISPR-Cas-escaping virus mutants (CEMs) can be characterized by PCR and the following sequencing against the protospacers plus their flanking regions. The mutation(s) which recover the lysis ability of the escaping virus against the according prokaryotic survivor (containing one specific new spacer), and whether the escaping viruses retain the ability to lyse WT prokaryotic cells needs to be confirmed (Figure 4.6B). The originally acquired new spacer (derived from WT virus) cached at the locus of the prokaryotic survivor may be able to function as priming spacers of the CEMs, since they are partially or perfectly complementary to their original protospacers in the genomes of the CEMs, but lose the interference initiation ability. Finally, the CEMs can be used to challenge their corresponding prokaryotic survivors (containing a priming spacer), as well as WT

prokaryotic cells (Figure 4.6C). The priming spacer holding strain may end up with primed adaptation ability against the CEMs (Figure 4.6C).

### *Primed adaptation by the Type I-E CRISPR-Cas system of Streptococcus thermophilus*

Using the research system proposed above, we found that a pre-existing priming spacer can greatly improve the acquisition effectiveness of the Type I-E CRISPR-Cas system of *Sth* against the same DNA molecule. We also show that *cas1+2* overexpression increased the ratio of CRISPR4 primed adaptation-mediated survival events among all the survivors from less than 50% to nearly 100% (Figure 4.4C), suggesting a multiplicative effect of these two strategies of spacer acquisition. Different from the well-studied homologous Type I-E system of *E. coli*, secondary protospacers acquired by primed adaptation of this Type I-E system of *Sth* distributed at both the primed and the non-primed strands, centered by the PPS. This finding indicated a universal Type I feature of secondary protospacer distribution during priming, and the specific feature of the Type I-E system of *E. coli* might be caused by some unknown cellular effects.

## References

1. Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F. *et al.* (2011) Evolution and classification of the CRISPR-Cas systems. *Nature reviews. Microbiology*, **9**, 467-477.

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

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

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

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

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

7. Terns, M.P. and Terns, R.M. (2011) CRISPR-based adaptive immune systems. *Current opinion in microbiology*, **14**, 321-327.

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

9. Barrangou, R. and Marraffini, L.A. (2014) CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Molecular cell*, **54**, 234-244.
10. Tsui, T.K. and Li, H. (2015) Structure Principles of CRISPR-Cas Surveillance and Effector Complexes. *Annu Rev Biophys*, **44**, 229-255.
11. Heler, R., Marraffini, L.A. and Bikard, D. (2014) Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. *Molecular microbiology*, **93**, 1-9.
12. Agari, Y., Sakamoto, K., Tamakoshi, M., Oshima, T., Kuramitsu, S. and Shinkai, A. (2010) Transcription profile of Thermus thermophilus CRISPR systems after phage infection. *Journal of molecular biology*, **395**, 270-281.
13. Pougach, K., Semenova, E., Bogdanova, E., Datsenko, K.A., Djordjevic, M., Wanner, B.L. and Severinov, K. (2010) Transcription, processing and function of CRISPR cassettes in Escherichia coli. *Molecular microbiology*, **77**, 1367-1379.
14. Jansen, R., Embden, J.D., Gaastra, W. and Schouls, L.M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology*, **43**, 1565-1575.
15. Bhaya, D., Davison, M. and Barrangou, R. (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annual review of genetics*, **45**, 273-297.
16. Sorek, R., Lawrence, C.M. and Wiedenheft, B. (2013) CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annual review of biochemistry*, **82**, 237-266.

17. Kunin, V., Sorek, R. and Hugenholtz, P. (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome biology*, **8**, R61.

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

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

20. Shmakov, S., Abudayyeh, O.O., Makarova, K.S., Wolf, Y.I., Gootenberg, J.S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K. *et al.* (2015) Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Molecular cell*, **60**, 385-397.

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

22. Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K. and Semenova, E. (2012) Molecular memory of prior infections activates

the CRISPR/Cas adaptive bacterial immunity system. *Nature communications*, **3**, 945.

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

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

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

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

27. Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J. and Severinov, K. (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 10098-10103.

28. Fineran, P.C., Gerritzen, M.J., Suarez-Diez, M., Kunne, T., Boekhorst, J., van Hijum, S.A., Staals, R.H. and Brouns, S.J. (2014) Degenerate target

sites mediate rapid primed CRISPR adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, E1629-1638.

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

30. Richter, C., Dy, R.L., McKenzie, R.E., Watson, B.N., Taylor, C., Chang, J.T., McNeil, M.B., Staals, R.H. and Fineran, P.C. (2014) Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic acids research*, **42**, 8516-8526.

31. Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A. and Severinov, K. (2013) High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in E. coli. *RNA biology*, **10**, 716-725.

32. Garrett, S., Shiimori, M., Watts, E.A., Clark, L., Graveley, B.R. and Terns, M.P. (2020) Primed CRISPR DNA uptake in Pyrococcus furiosus. *Nucleic acids research*, **48**, 6120-6135.

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

34. Horvath, P., Coute-Monvoisin, A.C., Romero, D.A., Boyaval, P., Fremaux, C. and Barrangou, R. (2009) Comparative analysis of CRISPR loci in

lactic acid bacteria genomes. *International journal of food microbiology*, **131**, 62-70.

35. Horvath, P. and Barrangou, R. (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science*, **327**, 167-170.

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

37. Paez-Espino, D., Sharon, I., Morovic, W., Stahl, B., Thomas, B.C., Barrangou, R. and Banfield, J.F. (2015) CRISPR immunity drives rapid phage genome evolution in Streptococcus thermophilus. *mBio*, **6**.

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

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

40. Renye, J.A., Jr. and Somkuti, G.A. (2009) Insertion of a heterologous gene construct into a non-functional ORF of the Streptococcus thermophilus chromosome. *Biotechnol Lett*, **31**, 759-764.

41. Hynes, A.P., Lemay, M.L., Trudel, L., Deveau, H., Frenette, M., Tremblay, D.M. and Moineau, S. (2017) Detecting natural adaptation of the

Streptococcus thermophilus CRISPR-Cas systems in research and classroom settings. *Nature protocols*, **12**, 547-565.

42. Wei, Y., Chesne, M.T., Terns, R.M. and Terns, M.P. (2015) Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in Streptococcus thermophilus. *Nucleic acids research*, **43**, 1749-1758.
43. Duplessis, M., Russell, W.M., Romero, D.A. and Moineau, S. (2005) Global gene expression analysis of two Streptococcus thermophilus bacteriophages using DNA microarray. *Virology*, **340**, 192-208.
44. Paez-Espino, D., Morovic, W., Sun, C.L., Thomas, B.C., Ueda, K., Stahl, B., Barrangou, R. and Banfield, J.F. (2013) Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nature communications*, **4**, 1430.
45. Sun, C.L., Barrangou, R., Thomas, B.C., Horvath, P., Fremaux, C. and Banfield, J.F. (2013) Phage mutations in response to CRISPR diversification in a bacterial population. *Environmental microbiology*, **15**, 463-470.
46. Wei, Y., Terns, R.M. and Terns, M.P. (2015) Cas9 function and host genome sampling in Type II-A CRISPR-Cas adaptation. *Genes & development*, **29**, 356-361.
47. Duong, T., Miller, M.J., Barrangou, R., Azcarate-Peril, M.A. and Klaenhammer, T.R. (2011) Construction of vectors for inducible and constitutive gene expression in Lactobacillus. *Microb Biotechnol*, **4**, 357-367.

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

49. Young, J.C., Dill, B.D., Pan, C., Hettich, R.L., Banfield, J.F., Shah, M., Fremaux, C., Horvath, P., Barrangou, R. and Verberkmoes, N.C. (2012) Phage-induced expression of CRISPR-associated proteins is revealed by shotgun proteomics in Streptococcus thermophilus. *PloS one*, **7**, e38077.

50. Li, M., Wang, R. and Xiang, H. (2014) Haloarcula hispanica CRISPR authenticates PAM of a target sequence to prime discriminative adaptation. *Nucleic acids research*, **42**, 7226-7235.

**Figure 4.1. Type I-E CRISPR-Cas system of *Sth* is active in target interference.**

(A) The schematic of Type I-E CRISPR-Cas system of *Sth*. Purple pentagons, the *cas* genes; dark green rectangle with the letter L, the leader sequence; black rectangles, the identical repeat sequences; variably colored rectangles, unique pre-existing spacers. (B) The transformation efficiencies of the target and the empty plasmid. **, p=0.01; n = 2.

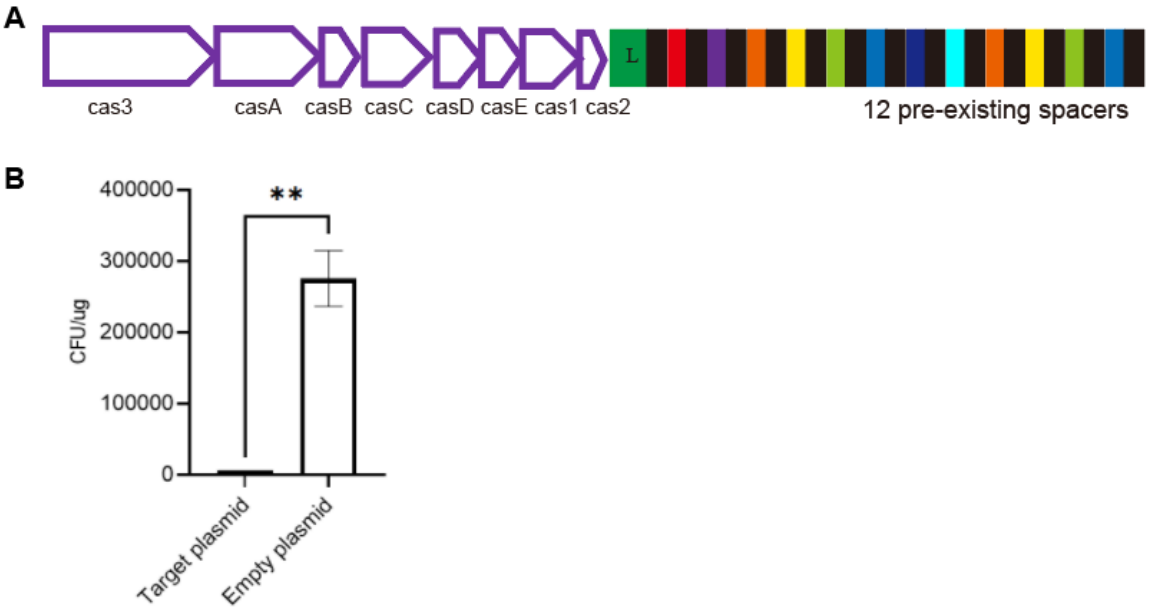
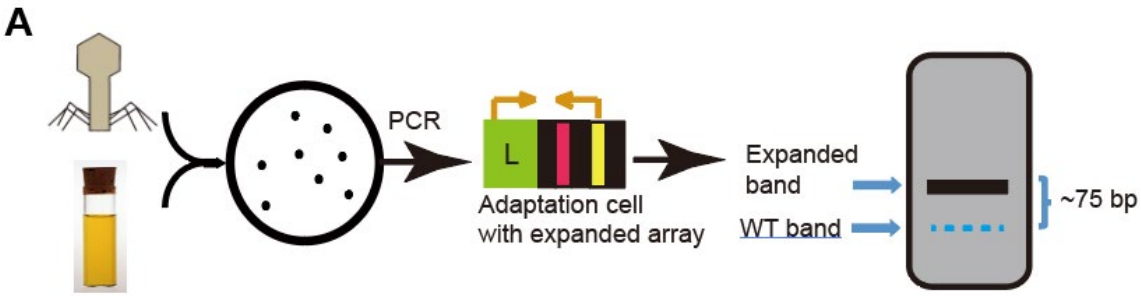
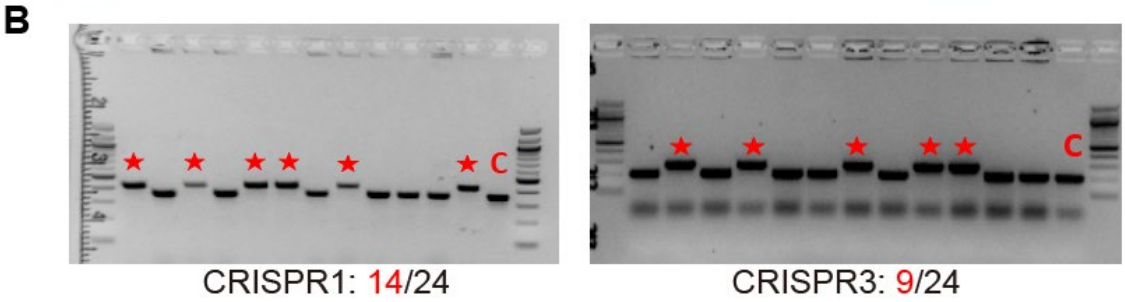
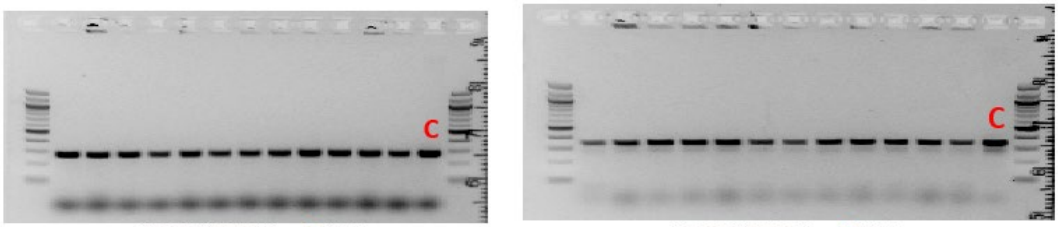

A
L
cas3
casA
casB
casC
casD
casE
cas1
cas2
12 pre-existing spacers
B
400000
300000
200000
100000
0
CFU/ug
**
Target plasmid
Empty plasmid

**Figure 4.2. The Type I-E CRISPR-Cas system is inefficient in adaptation.** (A) The schematic of the phage-based adaptation monitor procedure. *Sth* cells were challenged by lytic phage 2972, and randomly chosen colonies were analyzed by PCR against the leader-proximal ends of the CRISPR arrays. PCR results of the WT strain and the Type II-A CRISPR-Cas null strain were shown in (B) and (C). The last line of each of the gel was no phage challenge control. The bands indicating new protospacer adaptation were indicated by red stars. The numbers below each gel picture indicated the adaptation experienced colonies/total colonies tested.

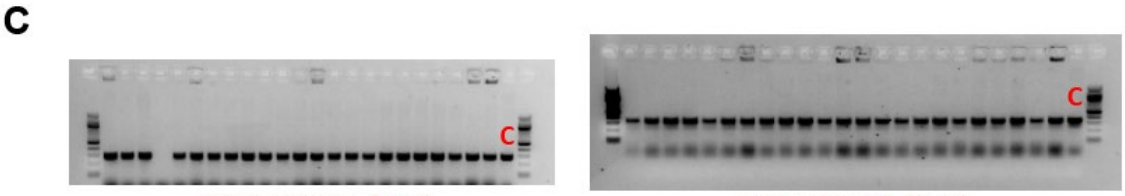



CRISPR1: 14/24 CRISPR3: 9/24




CRISPR2: 0/24 CRISPR4: 0/24

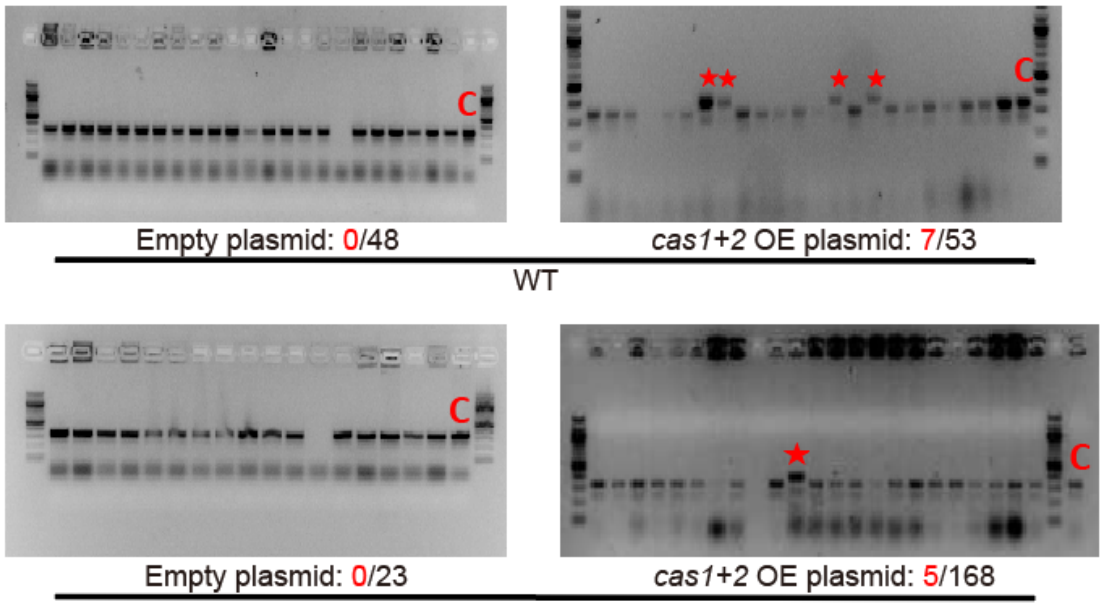



CRISPR2: 0/43 CRISPR4: 0/43

**Figure 4.3. Overexpression of *cas1+2* improves the adaptation ability of Type I-E CRISPR-Cas system.**

(A) The PCR results of the phage survivors of the WT strain (upper) and the Type II-A CRISPR-Cas null strain (lower), with (right) and without (left) the overexpression of *cas1+2*. The last line of each of the gel was no phage challenge control. The bands indicating new protospacer adaptation were indicated by red stars. The numbers below each gel picture indicated the adaptation experienced colonies/total colonies tested. (B) The protospacer sequences (red) together with their upstream and downstream adjacent sequences in phage 2972 (black) were listed. The consensus sequence, AA upstream to the protospacers were identified by WebLogo (http://weblogo.berkeley.edu/logo.cgi).

A




B

```
 1. ACTGATTTGTTAAAAGAAGCACTAGAGGTGATTTACCAAGAAAACTTCCCTCAACGTGCTGAAAA
 2. GCGAATAAGACCGAATGTTTTAGCATTCATAATTGCACGTTCTGGGTT AAGACCAAGTTCACGC
 3. TGGACGTTTGTCAAACTGGAAAGTGCCAGCTAACGTAGAAGACCTTGT AGAAATTGAATCAGA
 4. CATGAATTTGTAGAATAAGCTAATCAATTACCACTACTGGCAGAAATG TCAGTAAGGTTTAAC
 5. ACCGTAGTGCTAAAATATACCGATAAGACAACAAAGCACTATGACCCA GACTTAGGCCGTATG
 6. GTTAAATTTCTATAATGATACTTGGACCCGAAACCTTGAAACAGCATC AAGCACGTTCGAGTT
 7. GCGTATATTCAAAAATACGCTAACAAAAACGGATTAAGCGTTAGTGAT GCTAAACGAAAAGCA
 8. CGCAAAGATGGCAAACAGTAACACATCCGAGCCGTTGGATGATAAGGA ATTTGAAAGGACATT
 9. TCTTTTTGTTGGTAATAAACCCAACCGGGCTTATATCCGTGTTGTTTA GCGAACTCTTTGAGT
10. ATATTAACATTGGAACAGTTACCAGATGAGAAATTATTGAAAGAGATGATAGAACTAGCGAAGA
11. AAATAAGAGGAGGAACATTAAAACATGGGACTTATTTACGATAAAGTA ACAGCATCTAATATTG
12. TGACATCTACGCTAATAACAGTCCGATTCAACAATATCAGCTGACTGA AAACAATGGCAGACCT
```

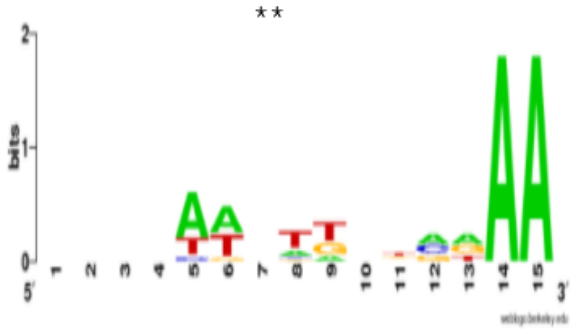


**Figure 4.4. The priming protospacers trigger primed adaptation of the Type I-E CRISPR-Cas system.**

(A) The sequences of the protospacer (red) acquired by XZ2, and the PAM (AA, black) upstream. The mutations of the escaping phages are highlighted by the black stars below. (B) The phage tittering assay by WT phage 2972 and the escaping phages at a gradient of amount against the WT *Sth* and XZ2np. (C) The numbers in the table indicated the adaptation experienced colonies/total colonies tested. Sum, in total.

**A**

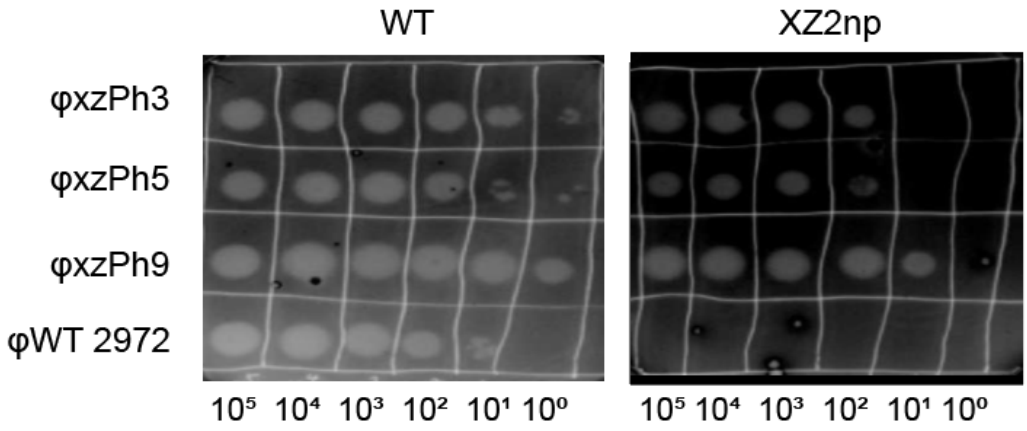
XZ2/XZ2np protospacer 5′-AATGTTTTAGCATTCATAATTGCACGTTCTGGGTT-3′

φxzPh3 5′-AGTGTTTTAGCATTCATAATTGCACGTTCTGGGTT-3′
*

φxzPh5 5′-AATATTTTAGCATTCATAATTGCACGTTCTGGGTT-3′
*

φxzPh9 5′-AATATTTTAGCGTTCATAATTGCACGTTCTGGGTT-3′
* *

**B**




**C**

| | XZ2np | XZ2 | WT |
|---|---|---|---|
| φxzPh3 | 21/29 | 23/23 | 0/3 |
| φxzPh5 | 31/92 | 33/34 | 0/23 |
| φxzPh9 | 6/11 | 15/15 | 0/2 |
| Sum | 58/132 | 71/72 | 0/28 |

**Figure 4.5. The distribution of secondary protospacers acquired by the Type I-E CRISPR-Cas system of *Streptococcus thermophilus*.**

Secondary protospacers acquired from the three phage mutants (xzPh3, xzPh5, and xzPh9) during priming by the Type I-E CRISPR-Cas system of (A) XZ2np and (B) XZ2 are illustrated by orange (primed strand) or blue (non-primed strand) triangles. The black triangle represents the PPS.

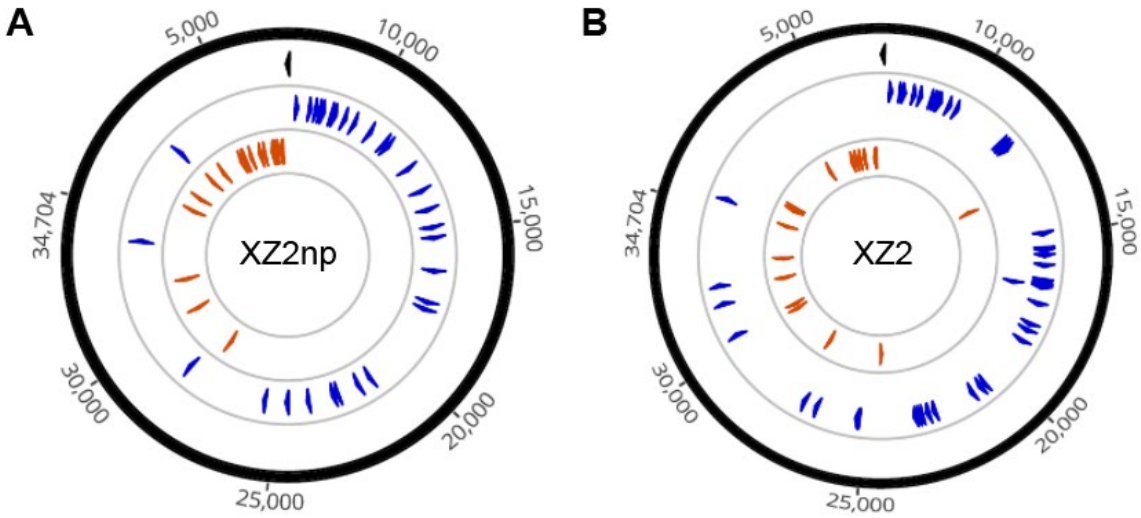

A
5,000
10,000
15,000
20,000
25,000
30,000
34,704
XZ2np
B
5,000
10,000
15,000
20,000
25,000
30,000
34,704
XZ2

**Figure 4.6. Workflow of CRISPR-Cas escaping virus mutants-based system to study primed adaptation.**

The pre-existed spacer in the CRISPR array of the prokaryotic genome is illustrated by the yellow rectangle. Some of the cells acquired a new spacer (from the virus) into the locus, which is illustrated by the red rectangle, and such cells are marked by red star. Virus mutants that escape from the new spacers-mediated immunity are marked by green triangle. The secondary spacer acquired from the virus mutants by primed adaptation is illustrated by a purple rectangle.

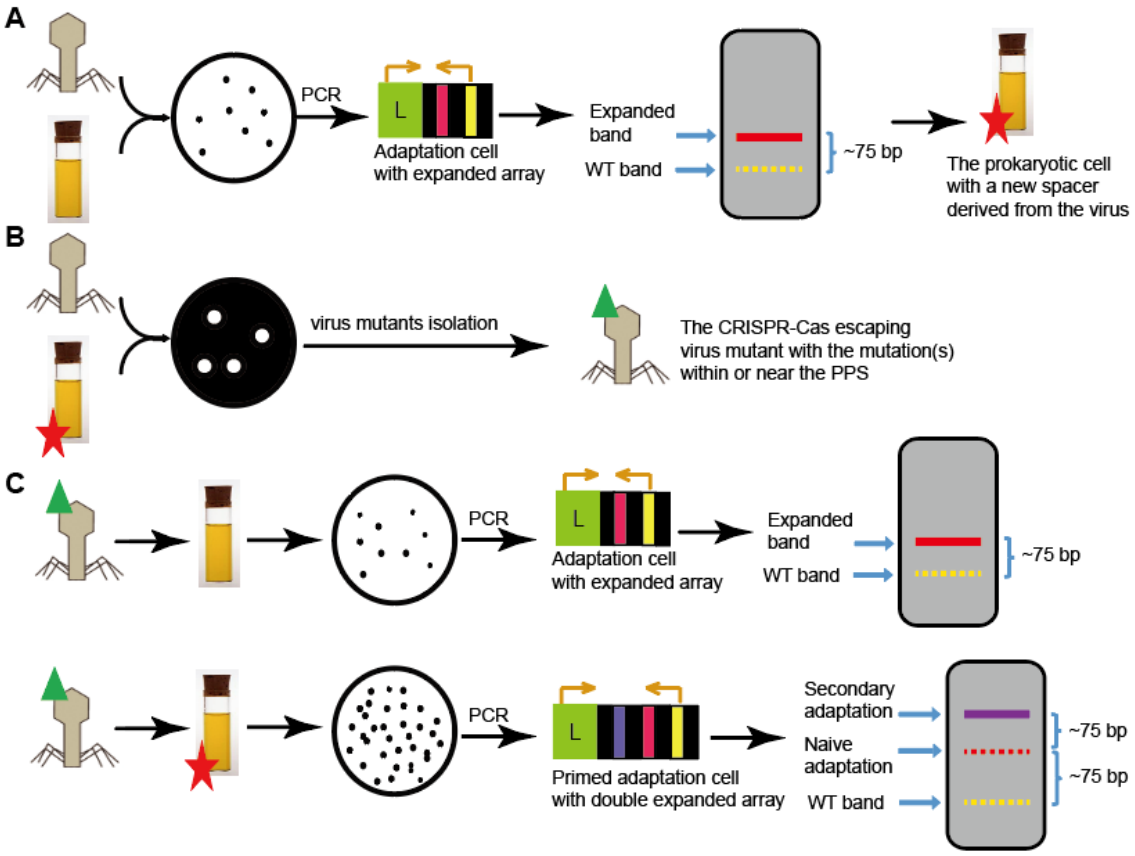

A
PCR
L
Adaptation cell
with expanded array
Expanded
band
WT band
~75 bp
The prokaryotic cell
with a new spacer
derived from the virus
B
virus mutants isolation
The CRISPR-Cas escaping
virus mutant with the mutation(s)
within or near the PPS
C
PCR
L
Adaptation cell
with expanded array
Expanded
band
WT band
~75 bp
PCR
L
Primed adaptation cell
with double expanded array
Secondary
adaptation
Naive
adaptation
WT band
~75 bp
~75 bp

**Table 4.1. The primers used in this work.**

| *Sth* mutant strain construction | | |
|---|---|---|
| Purpose | Name | Sequence (5'-3') |
| Linear pINTRS | Linear-pINTRS-r | CTGCAGATGCTAGCGGTGCCGCGCGCTC |
| | Linear-pINTRS-f | GTCGACGCAAGCTTATGCGGCCGCATAC |
| CRISPR1-Cas KO | sl-StCr1-f | CACCGCTAGCATCTGCAGCTCTACTTATTCCGGACAATATCG |
| | StCr1-OL-r | GGCTTTTTATGTTGAATCAATTAATAATCCCCTTATGCTTTTTTC |
| | StCr1-OL-f | GGGGATTATTAATTGATTCAACATAAAAAGCCAGTTC |
| | sl-cr1-r | GCATAAGCTTGCGTCGACGAGCTCGTCATGTACTTGCAAG |
| CRISPR3-Cas KO | seam-cr3-f | CACCGCTAGCATCTGCAGCACGCGAACAAATTCG |
| | StCr3-OL-r | GTGATAACAAAACTCTTTTATCTCTACTATTTTCCCAC |
| | StCr3-OL-f | GTAGAGATAAAAGAGTTTTGTTATCACAATTTTCGGTTGAC |
| | seam-cr3-r | GCATAAGCTTGCGTCGACCTCCTATGCCCTTTATTG |
| | | |
| *cas1+2* expression vector construction | | |
| Purpose | Name | Sequence (5'-3') |
| Linear pWAR | endofpWAR-f | GCGGCCGCCACCGCGGTGGG |
| | pPgm-r | GAAATATCTCCTTTTAAATTCAATG |
| Insertion | sl-cr4core-f | TTTAAAAGGAGATATTTCATGGTAGAAAAAAATGGAGCTAAG |
| | sl-cr4core-r | CACCGCGGTGGCGGCCGCTTAAGACTTTTCTTTCATTCTAGCTAAG |
| | | |
| CRISPR loci expansion monitor | | |
| Purpose | Name | Sequence (5'-3') |
| CRISPR1 | seqCR1incsn2-f | CCAAGTAATGATTTTCCGACGAG |
| | 1.01rev | CAATTCGAATCTTGATTTGCTGTC |
| CRISPR2 | CR2-lead-f | GGTACACTATAGATAGTGTTTCC |
| | CR2.03-rev | GGATGTACACAGCAAGTGG |
| CRISPR3 | CR3-lead-f | TAAAATTGGAATTATTTTGAAGCTGAAGTC |
| | CR3.02-rev | CAACCAAGCTAATACAGCAGTA |

| | | |
|---|---|---|
| CRISPR4 | CR4-lead-f | GAAAGATGCTAGACTAATCTATC |
| | CR4.01-rev | CTATTCGCCGATAATACAGG |
| | | |
| Protospacer mutation determination | | |
| Purpose | Name | Sequence (5'-3') |
| XZ2 | 1032proto-f | CATGTACGTAATAGCAGATTGG |
| | 1032proto-r | GTAATCCAGCCTAATTCAATTTGC |

# CHAPTER 5

# DISCUSSION

CRISPR-Cas systems exist in eubacteria and archaea (1,2), and provide hosts with a specific defense mechanism against invading nucleic acids (3-8). The functional studies regarding the CRISPR-Cas systems have inspired researchers to develop many unprecedentedly convenient genetic tools for diverse purposes (9). The work presented here for the first time provided detailed studies regarding adaptation by the Type III CRISPR-Cas system, and also contributed to better understanding of primed adaptation by the Type I system.

## Naïve adaptation by the Type III-A CRISPR-Cas system

Type III CRISPR-Cas systems are distinguished from the other types of systems by their unique and diverse mechanisms of target interference (10). Type III systems have been used for genome editing of the hyperthermophiles, in which CRISPR-Cas9-based modules might not function (11), and their target RNA cleavage ability has been used for gene expression control (12,13). Nevertheless, our knowledge about adaptation by the Type III systems is very limited. In this work, we for the first time, provided detailed analyses of the properties of the adaptation by the Type III CRISPR-Cas system.

### *Cas1 and Cas2 proteins mediate adaptation by the Type III CRISPR-Cas system*

We successfully detected adaptation against the plasmids, a lytic phage, and host genome by the Type III-A CRISPR-Cas system of *Streptococcus thermophilus* (*Sth*). Lengths of about 99% of new spacers fell into a roughly normal distribution from 32 to 42 bp, with 36 bp as the peak of the curve. Unlike Type I and Type II systems, adaptation by Type III system is independent on any protospacer flanking sequence.

Cas1 and Cas2 have been shown essential for adaptation by all tested CRISPR-Cas systems (3,14-19). The Type III-A CRISPR-Cas system of *Sth* required Cas1 and Cas2 proteins, but not Cas6 or any interference-related Cas proteins for adaptation. After knocking out all the *csm* genes, the system kept the ability and the intrinsic features of adaptation.

### *Adaptation-independent replication of the repeat-spacer units in the CRISPR arrays*

Intriguingly, although adaptation was impeded after knocking out *cas1* or *cas2* genes, we observed the replication of the pre-existed repeat-spacer units, indicating that it was an adaptation-independent cellular event. Such replication was observed in both the Type III and the Type II systems, indicating that it was a universal feature of all or many of CRISPR-Cas systems. We think that the homologous recombination or the DNA replication errors in the repeat-rich region

can help the CRISPR-Cas systems to replicate the recently acquired spacers to enhance the expression of the crRNAs, as well as to lose the old spacers to keep a compact CRISPR array.

### *Type III CRISPR-Cas system recognizes secondary structures of DNAs during adaptation*

We compared between adaptation by the Type III-A and the Type II-A CRISPR-Cas systems of *Sth* JIM8232 against different rolling-circle replicating (RCR) plasmids and theta replicating plasmids, as well as host genome. A prominent and intriguing feature of the adaptation by the Type III system was the recognition against the *sso*s of the RCR plasmids, distinguished from the Type II system. RCR plasmids have ssDNA forms during their replication, and the long and partially palindromic *sso*s form the stem-loop structures to trigger the synthesis of the minus strand. We reasoned that the stem-loop structures served as additional dsDNA substrates for adaptation by the Type III system, making the *sso*s preferentially adapted by the Type III system. Similarly, the partially palindromic *oriT* sequence of pNT1, and the ones of the regulatory regions of pIB184 *rep* gene and the ones of genomic amino acid synthetase genes, as well as the clover structure enriched rRNA and tRNA encoding regions of the genome, were also recognized by the Type III system but not by the Type II system. Most of natural plasmids of gram-positive bacteria, including *Sth*, and many of those of gram-negative bacteria are RCR plasmids (20). Moreover, the crucial structure of *oriT* and other DNA secondary structures are important for the

conjugative transfer and other life procedures of environmental mobile genetic elements (21). As a consequence, the ability of secondary structure recognition of Type III CRISPR-Cas system is beneficial for the system to specifically and efficiently eliminate invaders.

Unlike the three reported Type III CRISPR-Cas systems with reverse transcriptase (RT) activity (18,22-24), adaptation efficiency of Type III-A system of *Sth* has no strong correlation between transcription level of the DNA substrates. However, the first ~50 bp regions following the strong transcription start sites are preferentially adapted by the RT-free system. Moreover, like the reported Type III-D system of *Vibrio vulnificus* with RT activity, the RT-free Type III-A system of *Sth* extensively acquires new spacers from the rRNA as well as the tRNA sequences, indicating that this Type III-specific preference was RT-independent and was not caused by direct adaptation against the RNA substrates. We reasoned that during transcription of the highly expressed regions, the coding strand was displaced in a ssDNA form, which can form secondary structures and be recognized by the Type III system. Especially, the rRNA and tRNA encoding sequences are the most highly transcribed regions of a genome and are enriched of the clover structures in their ssDNA forms.

***Type III-A CRISPR-Cas system mediates defense against the lytic phage***

We isolated two *Sth* DGCC7710 BIMs that separately acquired one new spacer from the template strand from phage 2972 in the transplanted Type III-A CRISPR-Cas system, indicating that Type III-A system of *Sth* performed robust

adaptive defense against the phages as well. Moreover, nearly all the new spacers detected by HTS and the two new spacers of the isolated BIMs came from the early transcribed genes of the phage, suggesting that only the crRNAs targeting the early transcribed genes could mediate timely defense against the lytic viruses for the Type III CRISPR-Cas systems.

## Primed adaptation by the Type III-A and the Type I-E CRISPR-Cas system

Primed adaptation provides CRISPR-Cas systems with an efficient and specific secondary adaptation against the escaping invaders, thus allowing the hosts to minimize the escapers and co-evolve with invaders. However, primed adaptation had only been studied in few Type I systems and only two Type II systems (25-33). In this work, we for the first time observed primed adaptation by Type III CRISPR-Cas system, and developed a novel CRISPR-Cas escaping virus mutants-based research system for the future study regarding primed adaptation.

### *Primed adaptation by the Type III-A CRISPR-Cas system*

In this work, priming protospacer (PPS) triggered secondary adaptation by the Type III-A system of *Sth*. Primed adaptation by the Type III system preferentially acquired new protospacers against the ~800 bp encoding strand downstream to the target. Adaptation hotspots during primed adaptation had been proposed to be the DNA cleavage sites during target interference by the Type III system (34). The block of transcription of the target DNA sequence, and

the mutations within the DNase domain of Csm1, abolished primed adaptation. These findings supported the hypothesis that DNA cuts mediated by Csm1 triggered primed adaptation. The ssDNAs with free ends generated by DNA cleavage probably fuel adaptation by Cas1-Cas2 hexamers. How the ssDNAs are integrated into the CRISPR array is unknown. Very likely, they are able to form secondary structures, and can be recognized by the adaptation machinery of the Type III CRISPR-Cas system.

### *A novel CRISPR-Cas escaping virus mutants-based system to study primed adaptation*

The design of a PPS, or an escaping target sequence, that can trigger primed adaptation had been a difficulty in the primed adaptation study. In this work, by studying primed adaptation of the Type I-E system of *Sth*, we developed a novel CRISPR-Cas escaping virus mutants-based method, to acquire the suitable PPS information by the natural co-evolution between the prokaryotic host and the lytic virus, rather than by artificial design. Firstly, *Sth* cells were challenged by lytic phage 2972. Type I-E CRISPR-Cas system-mediated bacteriophage insensitive mutants (BIMs) were successfully isolated. One of the BIMs was then co-cultivated with phage 2972 again. Few phage plaques formed by CRISPR-Cas-escaping viruses were isolated. The escaping phages had mutation(s) within or near the protospacer, which prevented target cleavage by the CRISPR-Cas system. The mutation(s) recovered the lysis ability of the escaping phage 2972 against the BIM (containing one specific new spacer). The

originally acquired new spacer (derived from WT phage 2972) functioned as a priming spacer of the escaping phages, and was shown able to trigger primed adaptation by the Type I-E system.

***Future directions***

In this particular study, priming procedure of the Type III-A CRISPR-Cas system of *Sth* did not apparently increase the adaptation efficiency against the targeted plasmid, compared with reported Type I and Type II systems. This finding challenged the reality of the existence of primed adaptation activity of the system. Since the target sequence used here was not an escaping target, I reasoned that the Type III-A CRIPSR-Cas system was able to eliminate the target plasmids and had a short time period to acquire secondary spacers. As a result, the existence and the features of primed adaptation by the Type III system need to be further studied. I constructed a series of target plasmids with different mutations and the protospacer flanking sequences, so that they were differentially sensitive to the target interference. They are useful materials for the further studies. Moreover, the CRISPR-Cas escaping virus mutants-based research system I developed during the Type I-E system study, will be a very useful strategy to generate escaping targets for the study about primed adaptation by the Type III system.

## References


1. Haft, D.H., Selengut, J., Mongodin, E.F. and Nelson, K.E. (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS computational biology*, **1**, e60.
2. Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F. *et al.* (2011) Evolution and classification of the CRISPR-Cas systems. *Nature reviews. Microbiology*, **9**, 467-477.
3. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709-1712.
4. Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H. *et al.* (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nature reviews. Microbiology*, **13**, 722-736.
5. Terns, M.P. and Terns, R.M. (2011) CRISPR-based adaptive immune systems. *Current opinion in microbiology*, **14**, 321-327.
6. Terns, R.M. and Terns, M.P. (2013) The RNA- and DNA-targeting CRISPR-Cas immune systems of Pyrococcus furiosus. *Biochemical Society transactions*, **41**, 1416-1421.

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

8. Tsui, T.K. and Li, H. (2015) Structure Principles of CRISPR-Cas Surveillance and Effector Complexes. *Annu Rev Biophys*, **44**, 229-255.

9. Pickar-Oliver, A. and Gersbach, C.A. (2019) The next generation of CRISPR-Cas technologies and applications. *Nature reviews. Molecular cell biology*, **20**, 490-507.

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

11. Li, Y., Pan, S., Zhang, Y., Ren, M., Feng, M., Peng, N., Chen, L., Liang, Y.X. and She, Q. (2016) Harnessing Type I and Type III CRISPR-Cas systems for genome editing. *Nucleic acids research*, **44**, e34.

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

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

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

15. Nunez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W. and Doudna, J.A. (2014) Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nature structural & molecular biology*, **21**, 528-534.

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

17. Fagerlund, R.D., Wilkinson, M.E., Klykov, O., Barendregt, A., Pearce, F.G., Kieper, S.N., Maxwell, H.W.R., Capolupo, A., Heck, A.J.R., Krause, K.L. *et al.* (2017) Spacer capture and integration by a type I-F Cas1-Cas2-3 CRISPR adaptation complex. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, E5122-E5128.

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

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

20. Khan, S.A. (1997) Rolling-circle replication of bacterial plasmids. *Microbiology and molecular biology reviews : MMBR*, **61**, 442-455.
21. Bikard, D., Loot, C., Baharoglu, Z. and Mazel, D. (2010) Folded DNA in action: hairpin formation and biological functions in prokaryotes. *Microbiology and molecular biology reviews : MMBR*, **74**, 570-588.
22. Schmidt, F., Cherepkova, M.Y. and Platt, R.J. (2018) Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature*, **562**, 380-385.
23. Gonzalez-Delgado, A., Mestre, M.R., Martinez-Abarca, F. and Toro, N. (2019) Spacer acquisition from RNA mediated by a natural reverse transcriptase-Cas1 fusion protein associated with a type III-D CRISPR-Cas system in Vibrio vulnificus. *Nucleic acids research*, **47**, 10202-10211.
24. Silas, S., Lucas-Elio, P., Jackson, S.A., Aroca-Crevillen, A., Hansen, L.L., Fineran, P.C., Fire, A.Z. and Sanchez-Amat, A. (2017) Type III CRISPR-Cas systems can provide redundancy to counteract viral escape from type I systems. *Elife*, **6**.
25. Swarts, D.C., Mosterd, C., van Passel, M.W. and Brouns, S.J. (2012) CRISPR interference directs strand specific spacer acquisition. *PloS one*, **7**, e35888.
26. Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K. and Semenova, E. (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature communications*, **3**, 945.

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

28. Fineran, P.C., Gerritzen, M.J., Suarez-Diez, M., Kunne, T., Boekhorst, J., van Hijum, S.A., Staals, R.H. and Brouns, S.J. (2014) Degenerate target sites mediate rapid primed CRISPR adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, E1629-1638.

29. Richter, C., Dy, R.L., McKenzie, R.E., Watson, B.N., Taylor, C., Chang, J.T., McNeil, M.B., Staals, R.H. and Fineran, P.C. (2014) Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic acids research*, **42**, 8516-8526.

30. Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A. and Severinov, K. (2013) High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in E. coli. *RNA biology*, **10**, 716-725.

31. Garrett, S., Shiimori, M., Watts, E.A., Clark, L., Graveley, B.R. and Terns, M.P. (2020) Primed CRISPR DNA uptake in Pyrococcus furiosus. *Nucleic acids research*, **48**, 6120-6135.

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

33. Mosterd, C. and Moineau, S. (2021) Primed CRISPR-Cas Adaptation and Impaired Phage Adsorption in Streptococcus mutans. *mSphere*, **6**.

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