Implications of self-targeting by type I CRISPR-Cas systems

Auswirkungen des Selbst-targetings durch Typ I CRISPR-Cas Systeme

Doctoral thesis for a doctoral degree

at the Graduate School of Life Sciences,

Julius-Maximilians-Universität Würzburg,

Section Infection and Immunity

submitted by

Franziska Wimmer

from

Traunstein

Würzburg, 2022

This document is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0): http://creativecommons.org/licenses/by-nc-nd/4.0 This CC license does not apply to third party material (attributed to another source) in this publication.

Submitted on: ……

Members of the Thesis Committee

Chairperson: Prof. Dr. Manfred Gessler

Primary Supervisor: Prof. Dr. Chase Beisel

Supervisor (Second): Prof. Dr. Cynthia Sharma

Supervisor (Third): Jun. Prof. Dr. Neva Caliskan

Supervisor (Fourth): Prof. Dr. Frederik Börnke

Date of Public Defence: …….……

Date of Receipt of Certificates: …….

The presented work has been carried out as a member of the Graduate School of Life Science under the supervision of Prof. Dr. Chase L. Beisel at the Helmholtz Institute for RNA-based Infection Research (HIRI) in Würzburg between April 2018 and April 2022.

List of Publication

Wimmer, F., & Beisel, C. L. (**2020**). CRISPR-Cas systems and the paradox of self-targeting spacers. *Frontiers in Microbiology*, 3078.

Wimmer, F., Englert, F., & Beisel, C. L. (**2022**). A TXTL-Based Assay to Rapidly Identify PAMs for CRISPR-Cas Systems with Multi-Protein Effector Complexes. In *Cell-Free Gene Expression* (pp. 391-411). Humana, New York, NY**.**

Wimmer, F.*, Mougiakos, I.*, Englert, F., & Beisel, C. L. (**2022**). Rapid cell-free characterization of multi-subunit CRISPR effectors and transposons. *Molecular Cell*.

* These authors contributed equally

Copyrights

Parts of this work have been published previously and are reproduced, adapted, and/or modified with the permission of:

Wimmer, F., & Beisel, C. L. (**2020**). CRISPR-Cas systems and the paradox of self-targeting spacers. *Frontiers in Microbiology*, 3078.

Copyright: 2020 Wimmer and Beisel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

DOI: 10.3389/fmicb.2019.03078

Wimmer, F., Englert, F., & Beisel, C. L. (**2022**). A TXTL-Based Assay to Rapidly Identify PAMs for CRISPR-Cas Systems with Multi-Protein Effector Complexes. In *Cell-Free Gene Expression* (pp. 391-411). Humana, New York, NY**.**

Copyright: 2022, The Author(s), under exclusive license to Springer Science Business Media, LLC, part of Springer Nature

DOI: 10.1007/978-1-0716-1998-8_24

Wimmer, F., Mougiakos, I., Englert, F., & Beisel, C. L. (**2022**). Rapid cell-free characterization of multi-subunit CRISPR effectors and transposons. *Molecular Cell*.

Copyright: 2022 Elsevier Inc.

DOI: 10.1016/j.molcel.2022.01.026

Declaration of Authorship

Publication (Review): Wimmer, F., & Beisel, C. L. (2020). CRISPR-Cas systems and the paradox of self-targeting spacers. *Frontiers in Microbiology*, 3078.					
Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	FW	CB			
Data Collection	FW	CB			
Data Analysis and Interpretation	FW	CB			
Manuscript Writing					
Writing of Introduction	FW	CB			
Writing of Materials & Methods	-				
Writing of Discussion	FW	CB			
Writing of First Draft	FW	CB			

Explanations: exclusive first authorship

Publication: Wimmer, F., Englert, F., & Beisel, C. L. (2022). A TXTL-Based Assay to Rapidly Identify PAMs for CRISPR-Cas Systems with Multi-Protein Effector Complexes. In *Cell-Free Gene Expression* (pp. 391-411). Humana, New York, NY.					
Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	FW	CB			
Data Collection	FW	FE			
Data Analysis and Interpretation	FW	CB			

Manuscript Writing					
Writing of Introduction	FW	CB			
Writing of Materials & Methods	FW	CB			
Writing of Discussion	FW	CB			
Writing of First Draft	FW	CB			

Explanations: exclusive first authorship, not peer reviewed

Publication: Wimmer, F.*, Mougiakos, I.*, Englert, F., & Beisel, C. L. (2022). Rapid cell-free characterization of multi-subunit CRISPR effectors and transposons. *Molecular Cell*.					
Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	FW	IM	CB		
Data Collection	FW	IM	FE		
Data Analysis and Interpretation	FW	IM	FE		
Manuscript Writing					
Writing of Introduction	FW/CB	IM			
Writing of Materials & Methods	FW	IM			
Writing of Discussion	FW/CB	IM			
Writing of First Draft	FW/CB	IM			

Explanations: shared first authorship (*)

Unpublished manuscript: Wimmer, F., Collins, S. P., Alkhnbashi, O. S., Beisel, C. L. Host-encoded anti-CRISPR proteins block DNA degradation by two extensively self-targeting CRISPR-Cas systems in *Xanthomonas albilineans*.					
Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	FW	CB	SC/OA		

Data Collection	FW	SC/OA			
Data Analysis and Interpretation	FW	SC/OA			
Manuscript Writing					
Writing of Introduction	FW	CB			
Writing of Materials & Methods	FW	CB			
	FW				
Writing of Discussion	FW	CB			
Writing of First Draft	FW	CB			

Explanations: Unpublished manuscript, exclusive first authorship

The doctoral researcher confirms that she has obtained permission from both the publishers and the co-authors for legal second publication.

The doctoral researcher and the primary supervisor confirm the correctness of the above mentioned assessment.

Franziska Wimmer	12.04.2022	Würzburg	
Doctoral Researcher's Name	Date	Place	Signature
Prof. Dr. Chase Beisel	12.04.2022	Würzburg	
Primary Supervisor's Name	Date	Place	Signature

Publication (Review): Wimmer, F., & Beisel, C. L. (2020). CRISPR-Cas systems and the paradox of self-targeting spacers. *Frontiers in Microbiology*, 3078.					
Figure	**Author Initials,** Responsibility decreasing from left to right				
1	FW	CB			
2	FW	CB			
3	FW	CB			

Explanations: exclusive first authorship

Publication: Wimmer, F., Englert, F., & Beisel, C. L. (2022). A TXTL-Based Assay to Rapidly Identify PAMs for CRISPR-Cas Systems with Multi-Protein Effector Complexes. In *Cell-Free Gene Expression* (pp. 391-411). Humana, New York, NY.					
Figure	**Author Initials,** Responsibility decreasing from left to right				
1	FW	CB			
2	FW	CB			
3	FW	CB			
4	FW	CB			
5	FW	CB			
6	FW	CB			
7	FW	CB			
Tables	**Author Initials,** Responsibility decreasing from left to right				
1	FW				
2	FW				
3	FW				
4	FW				
5	FW				
6	FW				
7	FW				

Explanations: exclusive first authorship, not peer reviewed

Publication: Wimmer, F.*, Mougiakos, I.*, Englert, F., & Beisel, C. L. (2022). Rapid cell-free characterization of multi-subunit CRISPR effectors and transposons. *Molecular Cell*.					
Figure	**Author Initials,** Responsibility decreasing from left to right				
1	FW	CB			
2	FW	FE	CB		
3	FW	FE	IM	CB	
4	FW	FE	CB		
5	FW	IM	CB		
6	FW	IM	CB		
7	FW	IM	CB		
S1	FW	FE	CB		
S2	FW	FE	CB		
S3	FW/IM	CB			
S4	FW/IM	CB			
S5	FW	CB			
S6	FW	IM	CB		
Tables	**Author Initials,** Responsibility decreasing from left to right				
S1	FW				
S2	FW				
S3	IM				
S4	FW	CB			
S5	IM				
S6	FW	IM			

Explanations: shared first authorship (*)

Unpublished manuscript: Wimmer, F., Collins, S. P., Alkhnbashi, O. S., Beisel, C. L. Host-encoded anti-CRISPR proteins block DNA degradation by two extensively self-targeting CRISPR-Cas systems in *Xanthomonas albilineans*.					
Figure	**Author Initials,** Responsibility decreasing from left to right				
1	FW	CB			
2	FW	CB			
3	FW	SC/OA	CB		
4	FW	CB			
S1	FW	CB			
S2	FW	CB			
Tables	**Author Initials,** Responsibility decreasing from left to right				
S1	FW				
S2	FW	SC/OA			
S3	FW				

Explanations: Unpublished manuscript, exclusive first authorship

I also confirm my primary supervisor's acceptance.

Franziska Wimmer 12.04.2022 Würzburg

Doctoral Researcher's Name Date Place Signature

Other Scientific Contributions

Wandera, K. G., Collins, S. P., **Wimmer, F.**, Marshall, R., Noireaux, V., & Beisel, C. L. (**2020**). An enhanced assay to characterize anti-CRISPR proteins using a cell-free transcription-translation system. *Methods*, *172*, 42-50.

Jiao, C., Sharma, S., Dugar, G., Peeck, N. L., Bischler, T., **Wimmer, F.**, Yu, Y., Barquist, L., Schoen, C., Kurzai O., Sharma, C. M. & Beisel, C. L. (**2021**). Noncanonical crRNAs derived from host transcripts enable multiplexable RNA detection by Cas9. *Science*, *372*(6545), 941-948.

Affidavit

I hereby confirm that my thesis entitled “Implications of self-targeting by type I CRISPR-Cas systems” is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Würzburg,	12.04.2022	
Place	Date	Signature

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation „Auswirkungen des Selbst-targetings durch Typ I CRISPR-Cas Systeme“ eigenständig, d.h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mit angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegt hat.

Würzburg,	12.04.2022	
Place	Date	Signature

Table of Content

LIST OF PUBLICATION V

COPYRIGHTS VI

DECLARATION OF AUTHORSHIP VII

OTHER SCIENTIFIC CONTRIBUTIONS XIII

AFFIDAVIT XIV

EIDESSTATTLICHE ERKLÄRUNG XIV

TABLE OF CONTENT XV

SUMMARY 1

ZUSAMMENFASSUNG 2

INTRODUCTION 3

BACTERIAL IMMUNE SYSTEMS 3

CRISPR-CAS SYSTEMS 4

ANTI-CRISPR PROTEINS CAN COUNTERACT CRISPR-CAS IMMUNITY 7

SELF-TARGETING CRISPR-CAS SYSTEMS CAN HARBOR ALTERNATIVE FUNCTIONS 8

APPLICATIONS OF TYPE I CRISPR-CAS SYSTEMS 10

OVERCOMING CHALLENGES IN STUDYING TYPE I CRISPR-CAS SYSTEMS BY THE USE OF CELL-FREE TRANSCRIPTION-TRANSLATION SYSTEMS 12

CHAPTER 1: CRISPR-CAS SYSTEMS AND THE PARADOX OF SELF-TARGETING SPACERS 14

ABSTRACT 14

INTRODUCTION 14

NATURAL OCCURRENCE OF SELF-TARGETING SPACERS 17

INCORPORATION OF SELF-TARGETING SPACERS 18

NAÏVE ACQUISITION 18

PROTOSPACER WITHIN TRANSFERRED MOBILE GENETIC ELEMENTS 19

PRIMED ADAPTATION 20

NAÏVE ACQUISITION OF RNA-DERIVED SPACERS 20

ACQUISITION OF SELF-TARGETING SPACERS TRIGGERED BY FOREIGN INVADERS 21

SURVIVING SELF-TARGETING BY CRISPR-CAS SYSTEMS 22

ACTIVE DNA REPAIR 22

MUTATIONS DISRUPTING CRISPR-BASED TARGETING 24

PARTIALLY COMPLEMENTARY SPACERS DIRECTING TARGET BINDING BUT NOT CLEAVAGE 26

RNA TARGETING 26

GENOME-ENCODED ANTI-CRISPR PROTEINS 28

SELF-TARGETING SPACERS UNDERLYING ALTERNATIVE FUNCTIONS OF CRISPR-CAS SYSTEMS 28

GENOME EVOLUTION 29

CRISPR-CAS INDUCED MRNA CLEAVAGE 30

CRISPR-CAS INDUCED DNA DAMAGE RESPONSE 32

TRANSCRIPTIONAL REGULATION 32

INVADERS CO-OPTING CRISPR-CAS SELF-TARGETING 33

CONCLUSION AND FUTURE PERSPECTIVES 34

REFERENCES 36

CHAPTER 2: A TXTL-BASED ASSAY TO RAPIDLY IDENTIFY PAMS FOR CRISPR-CAS SYSTEMS WITH MULTI-PROTEIN EFFECTOR COMPLEXES 44

ABSTRACT 44

INTRODUCTION 44

MATERIALS 47

REAGENTS AND KITS 47

EQUIPMENT 47

METHODS 47

PAM LIBRARY CONSTRUCTION 49

CRISPR-CAS PLASMID DESIGN AND PREPARATION 50

PAM ASSAY 51

DNA EXTRACTION 52

QUALITY CHECK 52

NGS LIBRARY PREPARATION 54

NGS DATA ANALYSIS 55

DATA VALIDATION 56

NOTES 59

REFERENCES 60

CHAPTER 3: RAPID CELL-FREE CHARACTERIZATION OF MULTI-SUBUNIT CRISPR EFFECTORS AND TRANSPOSONS 62

SUMMARY 63

INTRODUCTION 63

DESIGN – PAM-DETECT: A TXTL-BASED ENRICHMENT ASSAY FOR PAM DETERMINATION 64

RESULTS 66

PAM-DETECT VALIDATED WITH THE CANONICAL TYPE I-E CRISPR-CAS SYSTEM FROM *ESCHERICHIA COLI* 66

DISTINCT PAM PROFILES PERVADE I-E CRISPR-CAS SYSTEMS 68

EXTENSIVE SELF-TARGETING I-C AND I-F1 CRISPR-CAS SYSTEMS IN *XANTHOMONAS ALBILINEANS* ARE FUNCTIONALLY ENCODED 70

THE I-F CRISPR TRANSPOSON FROM *VIBRIO CHOLERA* RECOGNIZES AN EXTREMELY FLEXIBLE PAM PROFILE 72

THE I-B2 CRISPR TRANSPOSON FROM *RIPPKAEA ORIENTALIS* RECOGNIZES A LESS FLEXIBLE PAM PROFILE 74

DNA TRANSPOSITION BY CRISPR TRANSPOSONS CAN BE RECAPITULATED IN TXTL 75

DNA TRANSPOSITION BY CRISPR TRANSPOSONS CAN BE RECAPITULATED IN TXTL 76

DNA TRANSPOSITION IN TXTL WITH THE *RIPPKAEA ORIENTALIS* CAST ESTABLISHES A DISTINCT BRANCH WITHIN I-B2 CRISPR TRANSPOSONS 76

DISCUSSION 78

LIMITATIONS OF THE STUDY 80

STAR METHODS 81

KEY RESOURCE TABLE 81

METHOD DETAILS 82

QUANTIFICATION AND STATISTICAL ANALYSIS 89

REFERENCES 92

SUPPLEMENTARY INFORMATION 96

SUPPLEMENTARY FIGURES 96

SUPPLEMENTARY TABLES 101

SUPPLEMENTARY METHODS S1 154

CHAPTER 4: HOST-ENCODED ANTI-CRISPR PROTEINS BLOCK DNA DEGRADATION BY TWO EXTENSIVELY SELF-TARGETING CRISPR-CAS SYSTEMS IN *XANTHOMONAS ALBILINEANS* 157

ABSTRACT 157

INTRODUCTION 157

RESULTS 159

THE TWO SELF-TARGETING CRISPR-CAS SYSTEMS IN *XANTHOMONAS ALBILINEANS* DO NOT AVOID SELF-TARGETING THROUGH GENE REPRESSION 159

BOTH CRISPR-CAS SYSTEMS BIND AND DEGRADE TARGET DNA IN *E. COLI* 160

PREDICTED ANTI-CRISPR PROTEINS INHIBIT BOTH CRISPR-CAS SYSTEMS IN TXTL 162

ACR_3 AND ACR_1 INHIBIT DNA DEGRADATION BY THE I-C AND I-F1 CAS3, RESPECTIVELY, IN *E. COLI* 164

DISCUSSION 166

MATERIAL AND METHODS 168

PLASMID CONSTRUCTION 168

RNA-SEQUENCING 168

CASCADE BINDING ASSAY IN *E. COLI* 169

CAS3 DEGRADATION ASSAY IN *E. COLI* 170

PROPHAGE PREDICTION 170

ACR PREDICTION 170

ACR ACTIVITY IN TXTL CASCADE BINDING ASSAY 171

ACR ACTIVITY IN TXTL CAS3 DEGRADATION ASSAY 171

ACR ACTIVITY IN *E. COLI* CASCADE BINDING ASSAY 173

ACR ACTIVITY IN *E. COLI* CAS3 DEGRADATION ASSAY 173

AMINO-ACID SEQUENCE ALIGNMENT 174

REFERENCES 174

SUPPLEMENTARY MATERIAL 178

SUPPLEMENTARY FIGURES 178

SUPPLEMENTARY TABLES 179

DISCUSSION 188

REFERENCES FOR INTRODUCTION AND DISCUSSION 190

ACKNOWLEDGMENTS/DANKSAGUNG 198

APPENDIX: CURRICULUM VITAE 199

Summary

CRISPR-Cas systems are highly diverse and canonically function as prokaryotic adaptive immune systems. The canonical resistance mechanism relies on spacers that are complementary to the invaders' nucleic acids. By accidental incorporation or other mechanisms, prokaryotes can also acquire self-targeting spacers that are complementary to their own genome. As self-targeting commonly leads to lethal autoimmunity, the existence of self-targeting spacers poses a paradox. In Chapter 1, we provide an overview of the prevalence of self-targeting spacers, summarize how they can be incorporated, and which means can be employed by the host to evade lethal self-targeting. In addition, we outline alternative functions of CRISPR-Cas systems that are associated with self-targeting spacers. Whether CRISPR-Cas systems can efficiently target their own genome depends heavily on the presence of protospacer adjacent motifs (PAMs) next to the target region. In Chapter 2, we developed a method to determine PAM requirements. Thereby, we specifically focused on type I systems that engage multi-protein complexes, which are challenging to assess. Using the cell-free transcription-translation (TXTL) system, we developed an enrichment-based binding assay and validated its reliability by examining the well-known PAM requirements of the *E. coli* type I-E system. In Chapter 3, we applied the TXTL-based PAM assay to assess 16 additional CRISPR-Cas systems. These 16 systems included three CRISPR-Cas associated transposons (CASTs). CASTs are recently discovered transposons that employ CRISPR-Cas systems in a non-canonical function for the directed integration of the transposon. To further characterize CASTs in TXTL outside their PAM requirements, we reconstituted the transposition of CASTs in TXTL. In Chapter 4, we turned to non-canonical self-targeting CRISPR-Cas systems, which were already discussed in Chapter 1. While investigating how the plant pathogen *Xanthomonas albilineans* survives self-targeting by its two endogenous CRISPR-Cas systems, we identified multiple putative anti-CRISPR proteins (Acrs) in the genome of *X. albilineans*. Two of the Acrs, named AcrIC11 and AcrIF12$_{Xal}$, inhibited degradation by their respective CRISPR-Cas systems but still retained Cascade-binding ability, and appear responsible for the lack of autoimmunity in *X. albilineans*. In summary, we developed new technologies that eased the investigation of non-canonical multi-component systems and, if applied to additional systems, might reveal unique properties that could be implemented in new CRISPR-Cas based tools.

Zusammenfassung

CRISPR-Cas-Systeme sind sehr vielfältig und funktionieren kanonisch als prokaryotische adaptive Immunsysteme. Der kanonische Resistenzmechanismus basiert auf Spacern, die komplementär zu den Nukleinsäuren der Eindringlinge sind. Durch zufällige Inkorporation oder andere Mechanismen können Prokaryoten auch Spacer integrieren, die komplementär zu ihrem eigenen Genom sind. Da Selbst-targeting in der Regel zu letaler Autoimmunität führt, stellt die Existenz von selbst-targeting Spacern ein Paradoxon dar. In Kapitel 1 geben wir einen Überblick über die Verbreitung von selbst-targeting Spacern, fassen zusammen, wie sie eingebaut werden können und welche Mittel der Wirt einsetzen kann, um sich dem letalen Selbst-targeting zu entziehen. Darüber hinaus werden alternative Funktionen von CRISPR-Cas-Systemen skizziert, die mit selbst-targeting Spacern in Verbindung gebracht werden. Ob CRISPR-Cas-Systeme Ziele in ihrem eigenen Genom erkennen können, hängt stark davon ab ob bestimmte Motive neben der Zielregion (protospacer adjacent motifs, PAMs) vorhanden sind. In Kapitel 2 haben wir eine Methode entwickelt, um die Anforderungen an PAMs zu bestimmen. Dabei konzentrierten wir uns speziell auf Typ I Systeme, deren Erforschung durch Nutzung von Multiproteinkomplexen erschwert wird. Unter Verwendung des zellfreien Transkriptions-Translations-Systems (TXTL) entwickelten wir einen Test der zur Anreicherung erkannter PAMs führt. Seine Zuverlässigkeit validierten wir, indem wir die bekannten PAM-Anforderungen des *E. coli* Typ I-E Systems untersuchten. In Kapitel 3 wendeten wir den TXTL-basierten PAM-Assay an, um 16 weitere CRISPR-Cas-Systeme zu untersuchen. Zu diesen 16 Systemen gehörten drei CRISPR-Cas-assoziierte Transposons (CASTs). CASTs sind kürzlich entdeckte Transposons, die CRISPR-Cas-Systeme in einer nicht-kanonischen Funktion für die gerichtete Integration des Transposons einsetzen. Um CASTs in TXTL außerhalb ihrer PAM-Anforderungen weiter zu charakterisieren, haben wir die Transposition von CASTs in TXTL rekonstruiert. In Kapitel 4 wandten wir uns den nicht-kanonischen, selbst-targeting CRISPR-Cas-Systemen zu, die bereits in Kapitel 1 behandelt wurden. Während wir untersuchten, wie das Pflanzenpathogen *Xanthomonas albilinean*s Selbst-targeting durch seine beiden endogenen CRISPR-Cas-Systeme überlebt, identifizierten wir mehrere mutmaßliche Anti-CRISPR-Proteine (Acrs) im Genom von *X. albilineans*. Zwei dieser Acrs, AcrIC11 und AcrIF12$_{Xal}$, hemmten die Degradation durch ihre jeweiligen CRISPR-Cas-Systeme, erlaubten aber dennoch DNA-Bindung durch Cascade. Diese beiden Acrs scheinen für das Fehlen von Autoimmunität bei *X. albilineans* verantwortlich zu sein. Zusammenfassend lässt sich sagen, dass wir neue Technologien entwickelt haben, die die Untersuchung von nicht-kanonischen Mehrkomponentensystemen erleichtert haben und bei Anwendung auf weitere Systeme einzigartige Eigenschaften offenbaren könnten, die in neue CRISPR-Cas-basierte Tools implementiert werden könnten.

Introduction

Bacterial immune systems

Nature is a dynamic environment, and its members are prone to a never-ending evolutionary process. In this rough world everything tries to survive, often at the costs of others. This battle of survival can be described with the “Red Queen Hypothesis” (1). A constant pressure exists that triggers continuous development of ways to escape lethal attack by a hostile organism. This ongoing arms race can once again be found between bacteria and bacteriophages (hereafter called phages) (2). Thereby, bacteria invented various strategies to defend from invading phages.

As an early step within bacterial defense, phage entry can be blocked by surface modifications (3, 4). If a phage still manages to enter the bacterial cell and successfully injects its genome, additional measures come into play that cleave foreign nucleic acids or block phage replication. One example are restriction-modification (R-M) systems that are present in 74% of prokaryotes (5) and cleave phage DNA. R-M systems function by recognizing specific sequence motifs in phages followed by endonucleolytic DNA cleavage (6). Another example of an immune system responsible for bacterial defense upon phage entry is the chemical defense (7). Bacteria produce small molecules that can stop phage replication, which is often achieved by the DNA-intercalating properties of these secondary metabolites.

As a late step in bacterial immunity, abortive infection (Abi) systems are activated if the first lines of defense fail. The infected bacterium commits suicide or enters growth arrest to prevent phage replication. These altruistic defense mechanisms prevent phage spread and protect other members in the bacterial community from phage infection (8). One example of such an Abi system is the lambda T4*rII* exclusion (Rex) system (9–11). Once the system is activated, an ion channel in the bacterial membrane is formed, leading to membrane depolarization and reduced ATP levels (10, 11). Consequently, cell growth is inhibited, and phage infection is stopped. A second example of an Abi system is the recently discovered cyclic oligonucleotide-based anti-phage signaling system (CBASS). Upon phage infection, CBASS produces cyclic oligonucleotides that activate the diverse effector proteins (12). Cell death can be caused via non-specific DNA degradation, membrane disruption by phospholipases or ion channel formation, and other to-date unknown mechanisms (12–14). This list of bacterial defense systems is by far incomplete. Bacteria constantly having to defend from attacking phages forced them to evolve a huge diversity of defense systems that is still not fully explored.

CRISPR-Cas systems

Important members of bacterial immune systems that were not mentioned yet are CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats - CRISPR-associated) systems. These systems are the only adaptive immune systems of bacteria and archaea and are present in about 40% of all bacteria and 85% of all archaea (15). Generally, CRISPR-Cas systems confer immunity in three steps (**Figure 1**). As a first step (acquisition), parts of the invaders' nucleic acids are acquired as spacers in CRISPR arrays separated by repeats (16). These spacers act as memories of past infections. To integrate spacers, Cas1 and Cas2 are required (17, 18) and they can be assisted by accessory factors or effector proteins (19). In the second step of CRISPR-Cas immunity (biogenesis), the CRISPR arrays are transcribed as long precursor CRISPR-RNA (crRNA) and processed in mature crRNAs. Processing is achieved by various Cas proteins whereby RNase III can also be employed (20). During the last step (interference), the crRNA forms a ribonucleoprotein (RNP) complex with the effector protein or the effector protein complex and, upon reinfection, guides the CRISPR-Cas machinery to the invader (21). Foreign nucleic acids are recognized by their complementarity to the spacer portion of the crRNA and are often differentiated from host nucleic acids by protospacer adjacent motifs (PAMs) that are located 5′ or 3′ of the target region (22, 23). A lack of complementarity to the repeat portion within the host can also be used to differentiate between foreigner and host (24, 25). Subsequent DNA or RNA cleavage leads to clearance of the intruder.

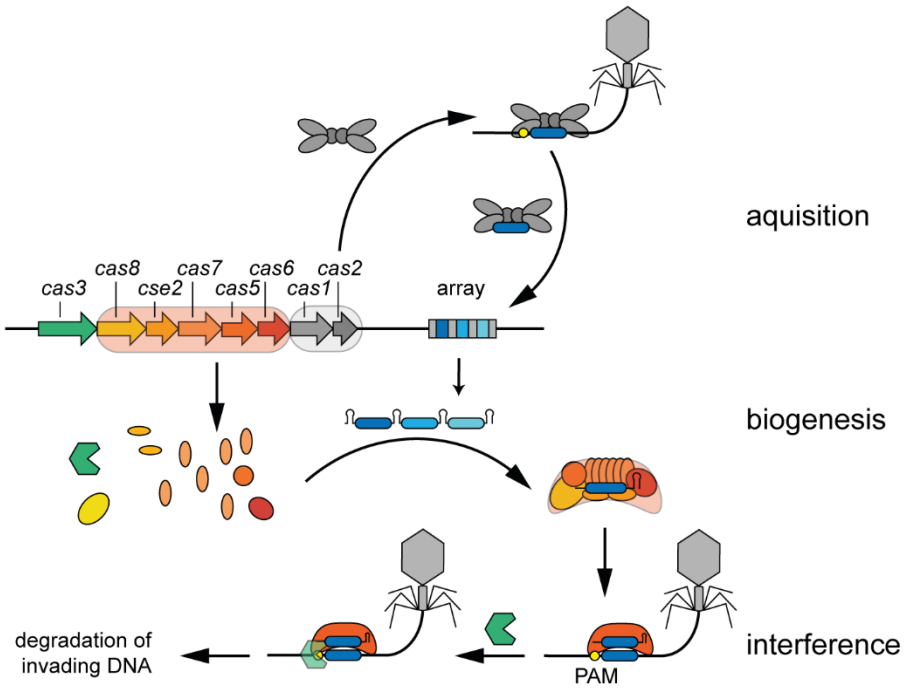


Figure 1: CRISPR-Cas systems act in the steps. The *E. coli* type I-E system is exemplified here**.**

Although this mode of action for CRISPR-Cas immunity generally follows the same principle for all CRISPR-Cas systems, CRISPR-Cas systems are highly diverse. To-date, there are two classes, six types and more than 30 subtypes uncovered (15) (**Figure 2**). High variability can be found in the interference steps, where class I systems (types I, III and IV) utilize multi-protein complexes and class II systems (types II, V and VI) rely on a single multi-domain protein (**Figure 2**) (15). The mode of action differs not only between the two classes but also between the six types.

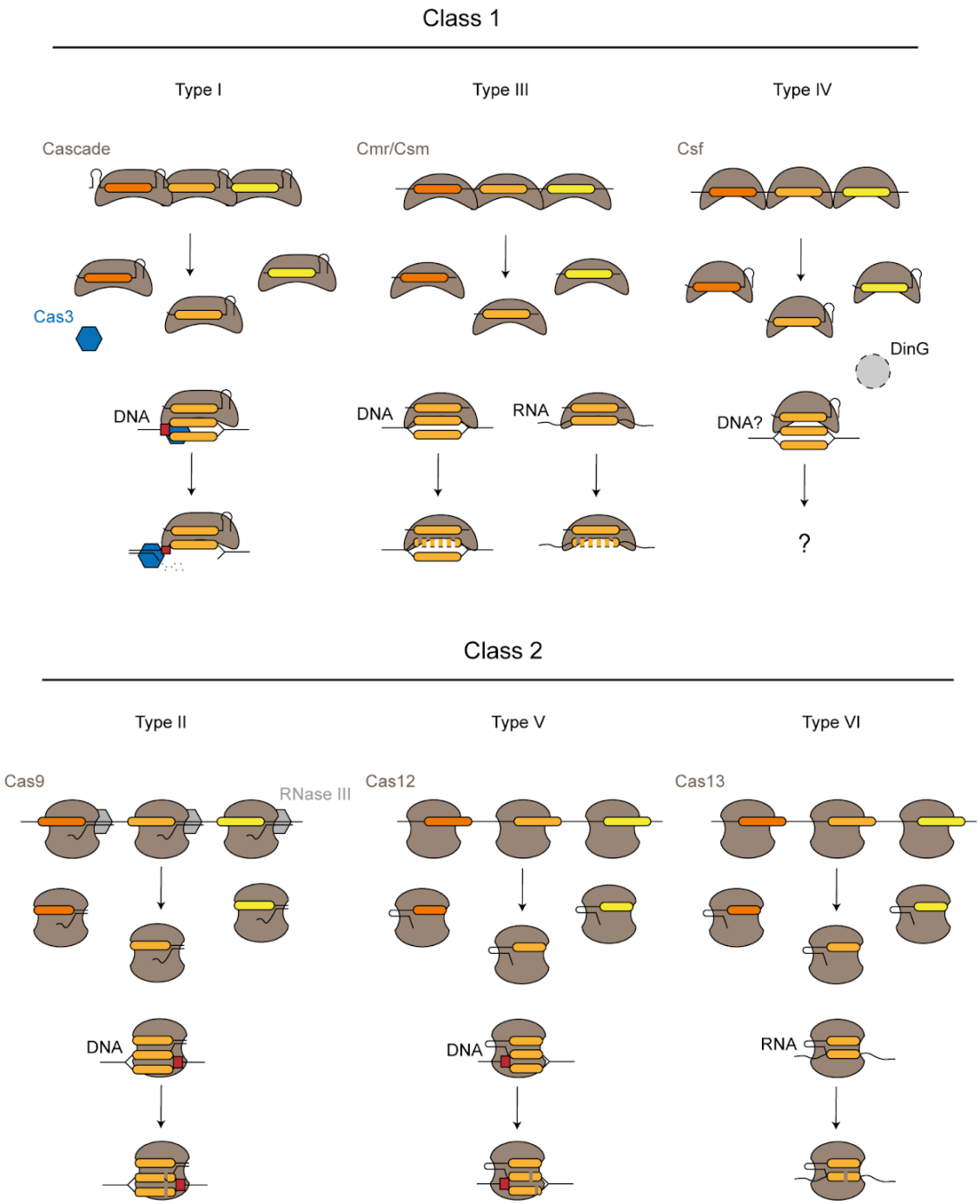


Figure 2: High diversity of CRISPR-Cas systems. The two classes of CRISPR-Cas systems are depicted consisting of six different types. The modes of action for the different CRISPR-Cas types are shown in a simplified version.

Class 1 systems contain the most abundant CRISPR-Cas system in nature - type I - representing about 50% of all CRISPR-Cas systems (15, 26). Type I systems canonically bind dsDNA by Cascade (CRISPR-associated complex for antiviral defense) that consist of multiple Cas proteins bound to a mature crRNA (27) (**Figures 1** and **2**). Upon recognition and binding of a DNA target, Cascade recruits the nuclease Cas3 (28), which leads to DNA degradation (29). Type III systems recognize RNA (30, 31) (**Figure 2**). Target RNA lacking complementarity with the repeat portion of the crRNA is bound by an RNP complex including Cas10-Csm or Cas10-Cmr protein complexes (24, 32). Subsequently, sequence-specific RNase activity and ssDNase activity is stimulated (32). Furthermore, cyclic oligoadenylates are produced by Cas10 that activate non-specific RNA cleavage by accessory proteins (33–35). The least studied CRISPR-Cas system is type IV, which is the last member of class 1 systems. The RNP complex in type IV systems is formed by Csf proteins and the crRNA and can lead to plasmid clearance (36). Type IV-A systems utilize the endonuclease DinG for their activity, while other subtypes lack this accessory protein (15, 36). Nevertheless, the exact mechanism of action for type IV systems is still not fully understood.

Class 2 systems contain the most studied CRISPR-Cas type - type II. Type II systems rely on Cas9 for interference activity but also require *trans*-activating crRNA (tracrRNA) to target dsDNA (37) (**Figure 2**). Binding of Cas9 to a target region next to a PAM sequence activates DNA cleavage activity of the effector protein (37). Type V systems only rely on tracrRNA or scoutRNA (short-complementarity untranslated RNA) in some subtypes (38, 39). Generally, Cas12 is guided to its dsDNA target by the mature crRNA followed by cleaved of the double-stranded target DNA (40) and non-specific ssDNA cleavage (41). Type VI systems represent the last member of class 2 CRISPR-Cas systems and are the only members of both classes systems that exclusively target RNA. Recognition of a target RNA sometimes requires a PFS (protospacer flanking site) or lack of extended complementarity between crRNA and target RNA (25, 42). Cas13, the effector protein of type VI binds its RNA target and subsequent RNA-cleavage activity gets activated. Thereby target-specific RNA but also non-specific RNA-cleavage can occur (42–45).

Mostly, CRISPR-Cas systems ensure the survival of their host by destruction of foreign nucleic acids that clears the intruder (37, 46). However, as some systems have the ability to non-specifically cleave DNA and/or RNA, these systems can also cleave or degrade host DNA/RNA (35, 43, 47, 48). Such collateral activity results in cell death or dormancy and turns the CRISPR-Cas system into an Abi system.

Anti-CRISPR proteins can counteract CRISPR-Cas immunity

As phages and bacteria constantly evolve in the battle for survival, phages also developed diverse mechanisms to circumvent bacterial attacks (49, 50). Here, we concentrate on phage mechanisms to escape from targeting by CRISPR-Cas systems. One means to escape from attack by CRISPR-Cas systems are mutations within the phage genome. Thereby mutating the spacer-complementary region or the adjacent PAM can avert CRISPR-Cas induced targeting (51). However, mutations in the phage genome might come with a fitness cost as expression of important genes could be impaired. Anti-CRISPR proteins (Acrs) represent a method to inactivate CRISPR-Cas systems without the need for genome mutations.

To-date, Acrs inhibiting all types of CRISPR-Cas systems besides type IV are known, although not for all subtypes (52–64). Acrs inhibiting type I-G CRISPR-Cas systems are still missing and only one inhibiting protein each is found for the type I-A, I-B and I-D systems (52, 53, 55). The amount of type II Acrs differs a lot between subtypes, with many Acrs discovered against type II-A, none for type II-B and few for type II-C (58, 59, 65–76). Only two type III Acrs are known so far (60, 61), although AcrIII-1 might be active against multiple type III systems as it leads to degradation of the signaling molecule cA_4 that plays an important role in many type III Acrs (60). For type V systems solely Acrs inhibiting type V-A (54, 62) were discovered. Finally, in type VI systems few Acrs were discovered inhibiting type VI-A and -B, although many of them are under debate (53, 63, 64, 77). Nevertheless, this list will most likely change in the near future as the search for new Acrs is still ongoing.

As these various Acrs inhibit a diverse set of CRISPR-Cas systems, the mode of action of these proteins is also highly variable. The identified mechanisms can be grouped in four general categories (78) with AcrIII-1 being the single member of one group. The remaining three groups are: prevention of CRISPR-Cas complex assembly, inhibition of effector binding, and suppression of effector cleavage. Mostly, Acrs achieve the former named outcomes by direct interaction with Cas proteins and only few Acrs were identified so far that use enzymatic strategies to achieve their goal (60, 73, 79–81). Nevertheless, the mechanism of action of many Acrs is still unknown.

The search for new Acrs could therefore not only discover enzymes inhibiting additional subtypes but also expand the already known mechanistic set used by Acrs. The first Acrs were identified in 2013 as type I-F counterplayers in *Pseudomonas* spp phages (57) followed by Acrs against the type I-E system discovered in the same phage group (56). But as most Acrs do not share sequence similarity, the search for additional Acrs was challenging. A putative transcriptional regulator containing a helix-turn-helix motif that was found to be encoded immediately upstream of the few identified Acrs eased the search for new Acrs (57). A “guilt-

by-association" method was employed to search for these helix-turn-helix containing proteins, termed anti-CRISPR associated (Aca) proteins, and novel Acrs were discovered encoded adjacent to *aca* genes (54, 59, 76, 82). To-date, 13 different Aca proteins are known (54, 57, 59, 75, 82–85) and their function as regulators controlling Acr expression was assigned (86–88).

Identification of new Acrs did not solely rely on Aca proteins and other means were established that for instance utilized phage infections or self-targeting spacers. In the first example, bacteria are infected by various phages and the efficiency of the bacterial encoded CRISPR-Cas system in defending phage infection was assessed (55, 65, 66). Potential Acr-bearing phages can be identified by successful bacterial infection. The individual genes encoded in the phage genome can then be tested separately for their function in blocking CRISPR-Cas immunity. The last example mentioned here interrogates bacteria that encode spacers with complementarity to their own genome. As cleavage of the bacterial genome is expected to lead to autoimmunity, the presence of self-targeting CRISPR-Cas systems can indicate the existence of Acrs that protect the bacterium from lethal self-targeting (58, 62). By using these and other methods (89), the discovery of new Acrs is just a matter of time.

Self-targeting CRISPR-Cas systems can harbor alternative functions

Self-targeting CRISPR-Cas systems are not always a burden to their host as they can unleash non-canonical functions of CRISPR-Cas systems (90, 91). One example is the primarily DNA-binding type I-F CRISPR-Cas system of *Pseudomonas aeruginosa.* Using a partial complementary crRNA, the bacterium converts the type I-F system into an RNA-targeting system that leads to degradation of lasR mRNA (92). As a result, *P. aeruginosa* faces a decreased immune reaction during host invasion (**Figure 3**). The same CRISPR-Cas system also harbors a second alternative function that is not necessarily beneficial for the bacterium. A partially complementary region in the genome of *P. aeruginosa* is targeted by the type I-F system, leading to nicking by Cas3 and induction of the SOS pathway. As a consequence, swarming motility is impaired and cell death of biofilm forming cells while sparing planktonic cells is induced (93–95) (**Figure 3**). Other CRISPR-Cas systems use self-targeting spacers for gene regulation (96–98). *Francisella novicida* utilizes Cas9 and scaRNAs (small CRISPR/Cas-associated RNAs) to regulate virulence-attenuating genes (96). A type I-B system becomes "addictive" by silencing the toxin CreT with the CRISPR RNA-resembling antitoxin (CreA) RNA that guides Cascade to a partial complementarity target in the *creT* promoter region (98) (**Figure 3**). These examples exemplify that self-targeting spacers can be beneficial to the host in various ways.

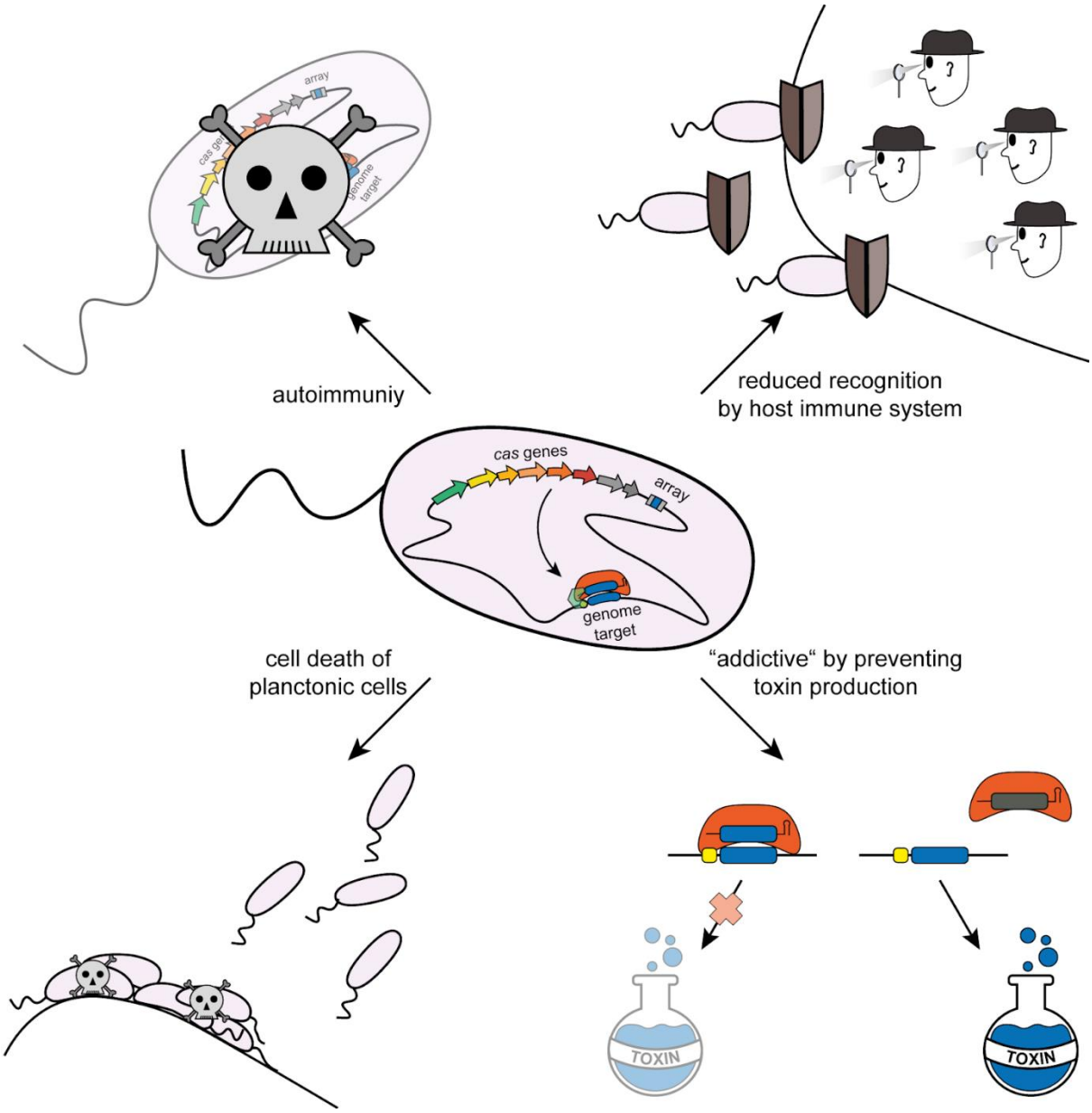


Figure 3: Self-targeting CRISPR-Cas systems can lead to several outcomes. Self-targeting CRISPR-Cas systems can lead to cell death by autoimmunity (upper left) or reduced recognition by the host immune system during bacterial infection (upper right). Self-targeting was also shown to induce cell death of biofilm forming cells while sparing planktonic cells (lower left). Finally, a CRISPR-Cas system can become addictive to its host by Cascade-induced gene regulation preventing toxin production (lower right).

Also, mobile genetic elements can use crRNAs complementary to the bacteria’s genome for their own purpose. Prominent examples that were recently discovered are CRISPR-Cas associated transposons (CASTs) (99–106). The Tn7-like transposons utilize CRISPR-Cas systems for RNA-guided transposition (103, 104). CASTs miss the mobilization gene *tnsE* and often *tnsD* required for transformation (101, 106). Instead, they harbor a CRISPR-Cas system lacking an acquisition module and a nuclease. A crRNA complementary to a bacterial genome or a mobile genetic element can guide insertion of the CAST within a region downstream of the target site (103, 104). So far, type I-F, type I-B and type V-K CRISPR-Cas systems are

shown to cooperate with Tn7-like transposons (102–104). A bioinformatic search for additional CASTs also identified type I-C and type IV systems associated with Tn7-like transposons and non-Tn7-CASTs associated with type I-E or inactivated type V systems, yet their functionality still needs to be explored (100). In summary, self-targeting CRISPR-Cas systems can be used by bacteria or mobile genetic elements to execute functions beyond adaptive immunity.

Applications of type I CRISPR-Cas systems

CRISPR-Cas systems are not only known for their biological functions but are also widely utilized as biotechnological tools. Class 2 systems are intensively used in CRISPR-Cas based applications. The preference of class 2 systems is mostly due to them utilizing one single protein for all functions and not consisting of a multi-protein complex like class 1 systems (15). Cas9 is the most established CRISPR-Cas system for biotechnological tools, nevertheless, Cas12 is now also heavily utilized and Cas13 gains more attention due to its RNA-targeting properties (107). Applications of class 2 CRISPR-Cas systems range from genome engineering in bacteria (108) over diagnostic tools (109) to even disease treatment in clinical trials (110).

Nevertheless, technologies involving class 1 systems, mostly focusing on type I, are emerging (111). As type I systems are the most common CRISPR-Cas system in nature (15, 26), harnessing type I systems within their natural host seems attractive. Especially as exogenous expression of Cas9 or Cas12a can lead to cytotoxicity and difficulties during plasmid delivery can be faced (112, 113). By utilizing the endogenous type I CRISPR-Cas systems, there is no need to transform a plasmid encoding for the effector. Solely a minimal CRISPR array targeting the genomic location selected for editing and a repair template for homology directed repair that includes the desired edits has to be transformed (114) (**Figure 4A**). If an endogenous system efficiently targets its own genome without a repair template, bacteria are usually prone to cell death (115). Therefore, lethal chromosomal targeting can be utilized to reprogram endogenous CRISPR-Cas systems into specific antimicrobials (116, 117) (**Figure 4B**). Furthermore, by mutation or deletion of the nuclease Cas3, endogenous type I CRISPR-Cas systems can be turned into gene regulators (118–120) (**Figure 4A**).

Unique properties of type I systems further promote their use in biotechnology. For instance, genome-editing with type I systems is less prone to off-target effects than type II systems. This is partly due to Cascade scanning DNA for complementary targets before recruitment of the nuclease Cas3. Thereby, an additional surveillance level is introduced upstream of nuclease cleavage that is not present in type II systems (121). Furthermore, Cas3 possesses the unique function of generating long range deletions in bacteria and in human cells that are challenging with Cas9 (121–124) (**Figure 4C**). Additionally, the multi-protein

nature of type I CRISPR effectors bears unique possibilities. As different tasks during CRISPR-Cas immunity are performed by separate proteins (e.g. nuclease: Cas3 (29), crRNA processing: Cas5 or Cas6 (125–127), PAM recognition: Cas8 (128, 129), backbone of Cascade: Cas7 (130)), Cas proteins can be used individually or easily left out to achieve diverse functions. One example is Cas3, as its removal turns type I systems, as mentioned previously, into gene regulators (118–120). A second example is Cas6. The RNA-binding properties of the small Cas6 protein can be reprogrammed to function in a highly sensitive and specific RNA-tracking platform (131). A third example utilizes the fact that Cascade complexes are assembled with a defined stoichiometry ($Cas8_1$-$Cse2_2$-$Cas7_6$-$Cas5_1$-$Cas6_1$ for *E. coli* type I-E Cascade) (27) (**Figure 1**). Thereby, each Cas7 interacts with 6 nts of the spacer and each Cse2 interacts with two Cas7 (128). By extending the spacer length by 6 nts or 12 nts, an additional Cas7 or two Cas7 and one Cse2, respectively, are included in Cascade formation (132). The enlarged Cascade can enhance gene silencing at some targets, potentially enabling fine-tuning of Cascade-binding. Summarizing, even though type I systems are not extensively used as biotechnological tools yet, some unique properties are already utilized. Nonetheless, due to the high diversity in type I CRISPR-Cas systems and their high prevalence in prokaryotes, additional functions likely await their discovery and use in new technologies.

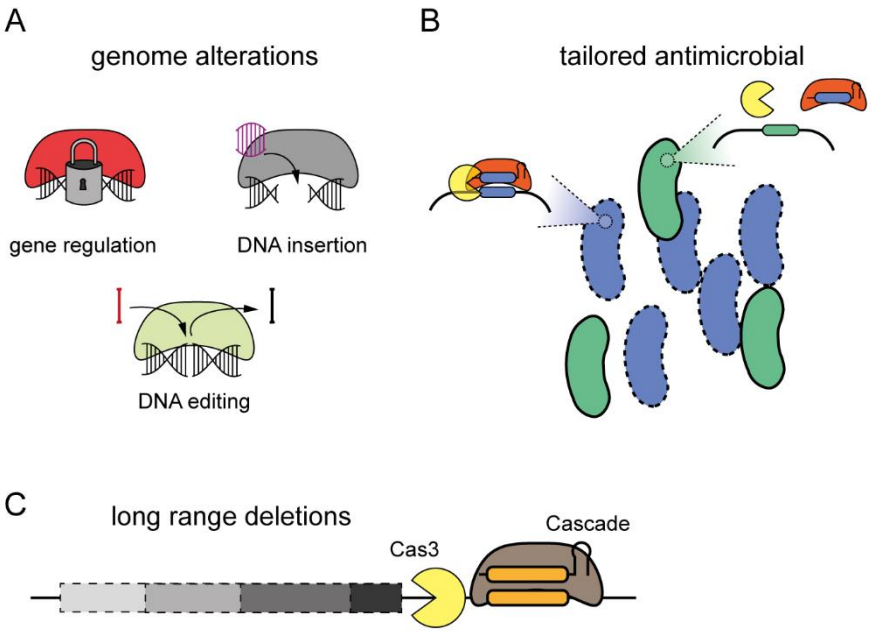


Figure 4: Type I CRISPR-Cas systems utilized as technological tools. (**A**) Endogenous type I systems can be used to perform changes in the genome or to regulate gene expression. (**B**) Type I systems can also be turned into tailored antimicrobials where they kill specific bacteria while sparing others. (**C**) Cas3 has a unique function in introducing long range genomic deletions.

Overcoming challenges in studying type I CRISPR-Cas systems by the use of cell-free transcription-translation systems

To further uncover new functions and potential new applications of CRISPR-Cas systems, there is a need to advance investigations of type I systems to unravel the high diversity of these systems and their unique properties. Unfortunately, the fact that type I systems are utilizing multiple proteins to execute their functions did not only lead to unique biotechnological tools but also hampered the investigation and fundamental understanding of many type I systems. Four Cas proteins in type I-C systems are the minimal number of proteins that are involved in type I interference (15). Thereby, Cas8, Cas5, and Cas7 form the Cascade and Cas3 is recruited as the DNA-degrading nuclease (126, 133). Type I-E systems however consist of five Cas proteins forming Cascade and including the nuclease, six Cas proteins are required for efficient DNA degradation (27). If type I systems are studied *in vitro*, all required Cas proteins need to be overexpressed and extracted either separately or as an already formed Cascade complex additionally extracting Cas3 (21). Optimizing plasmid extraction can be time-consuming and expensive. Cell-based experiments encompass different challenges. First of all, expression of unknown proteins can be toxic (105). Secondly, several Cas proteins need to be expressed at once. Therefore, multiple Cas proteins have to be encoded on one plasmid to minimize the total number of plasmids that need to be transformed. The cloning process for these huge plasmids can be demanding.

To overcome all of these challenges, cell-free systems come in handy. Transcription-translation (TXTL) systems contain an intact transcription and translation machinery, therefore, providing circular or linear DNA is sufficient for RNA and protein expression (134). For investigation of CRISPR-Cas systems, the TXTL system based on *E. coli* cell lysate was established (135). This system can be generated from *E. coli* strains with the desired genotype or can be purchased ready-to-use (136). It can also be handled in small volumes of a few microliters, which facilitates multiplexed experiments (137). As TXTL can be supplemented with polymerases like T7 RNA-polymerase that are not native to *E. coli*, genes can be cloned under promoters inactive within the cloning strain, circumventing potential protein toxicity (134). Furthermore, multiple plasmids can be added removing the need to minimize the number of plasmids used. So far, TXTL was mostly used to interrogate single effector CRISPR-Cas systems (62, 137–140), although it holds great potential to facilitate the future work in exploring the more complicated but very diverse multi-protein effector CRISPR-Cas systems and accelerate their use in technologies (105, 135).

In this work, TXTL was used in Chapter 2 and Chapter 3 to develop new techniques to overcome challenges in investigating type I CRISPR-Cas systems (105). These newly

established approaches were used to uncover properties of the recently discovered CASTs in Chapter 3 and study self-targeting CRISPR-Cas systems in Chapter 3 and Chapter 4. The results of this work lay the foundation for uncovering novel alternative functions of CRISPR-Cas systems that are reviewed in Chapter 1, and their use as new technologies.

Chapter 1: CRISPR-Cas systems and the paradox of self-targeting spacers

The content of this chapter was previously published and reproduced with permission from Frontiers Media SA.

Wimmer, F., & Beisel, C. L. (**2020**). CRISPR-Cas systems and the paradox of self-targeting spacers. *Frontiers in Microbiology*, 3078.

Author contributions:

FW and CB conceived and wrote the manuscript.

Abstract

CRISPR-Cas immune systems in bacteria and archaea record prior infections as spacers within each system's CRISPR arrays. Spacers are normally derived from invasive genetic material and direct the immune system to complementary targets as part of future infections. However, not all spacers appear to be derived from foreign genetic material and instead can originate from the host genome. Their presence poses a paradox, as self-targeting spacers would be expected to induce an autoimmune response and cell death. In this review, we discuss the known frequency of self-targeting spacers in natural CRISPR-Cas systems, how these spacers can be incorporated into CRISPR arrays, and how the host can evade lethal attack. We also discuss how self-targeting spacers can become the basis for alternative functions performed by CRISPR-Cas systems that extend beyond adaptive immunity. Overall, the acquisition of genome-targeting spacers poses a substantial risk but can aid in the host's evolution and potentially lead to or support new functionalities.

Introduction

CRISPR-Cas systems represent highly diverse adaptive immune systems found in many bacteria and most archaea (Barrangou et al., 2007; Sorek et al., 2013; Koonin et al., 2017). These systems consist of two general parts: Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) arrays and CRISPR-associated (Cas) proteins. CRISPR arrays represent the immunological memory of prior infections encoded within individual spacers separated by conserved repeats. Cas proteins carry out the adaptive immune

functions. The Cas proteins are highly diverse, resulting in CRISPR-Cas systems currently being grouped into two classes, six types, and over 30 subtypes (Makarova et al., 2015; Koonin et al., 2017; Koonin and Makarova, 2019).

While the specific proteins and biomolecular mechanisms vary, all systems act through three general steps as part of adaptive immunity. The first step, acquisition, incorporates pieces of invading nucleic acids, called protospacers, as new spacers within the CRISPR array. The protospacers are often selected based on the presence of a flanking protospacer adjacent motif (PAM) (Yosef et al., 2013; Wang et al., 2015). Acquisition requires the universal Cas proteins Cas1 and Cas2 (Yosef et al., 2012; Nuñez et al., 2014), although other accessory factors such as Cas4 (Kieper et al., 2018), Csa1 (Liu T. et al., 2017), Csn2 (Heler et al., 2015; Wei et al., 2015) and reverse transcriptase (RT) (Kojima and Kanehisa, 2008; Simon and Zimmerly, 2008; Silas et al., 2016) can also be involved. In type II CRISPR-Cas systems, the effector nuclease Cas9 can also play an essential role in the acquisition of new spacers (Heler et al., 2015; Wei et al., 2015). The acquired spacers serve as DNA records of prior infections that are passed to the host's progeny.

The second and third steps involve the biogenesis of CRISPR RNAs (crRNAs) from the CRISPR arrays followed by crRNA-directed immune defense. As part of crRNA biogenesis, the CRISPR array is, for most cases, transcribed into a long precursor CRISPR RNA (pre-crRNA) and processed into mature crRNAs by Cas proteins. In some cases, processing involves accessory factors such as RNase III (Carte et al., 2008; Deltcheva et al., 2011; Behler et al., 2018; Lee et al., 2018, 2019). The crRNA then forms a complex with Cas effector proteins to target foreign nucleic acids. Class 2 CRISPR-Cas systems rely on only one protein to bind and cleave their targets, with type II systems and some type V systems also requiring a *trans*-activating crRNA (tracrRNA) for effector complex formation (Deltcheva et al., 2011; Shmakov et al., 2015; Zetsche et al., 2015). Class I systems in contrast rely on multiple proteins that form a multi-subunit effector complex (Brouns et al., 2008; Hale et al., 2009). The resulting ribonucleoprotein complex then surveils the host's cytoplasm for DNA and/or RNA sequences that are complementary to the spacer and flanked either by a PAM or a sequence lacking complementarity to the corresponding portion of the crRNA repeat (Mojica et al., 2005; Marraffini and Sontheimer, 2010; Leenay and Beisel, 2017; Meeske and Marraffini, 2018).

One commonality across CRISPR-Cas systems is their reliance on the array-encoded spacers to direct CRISPR-based immunity. To-date, only 1–19% of identified spacers have been matched to potential protospacer sites, where most of the assigned spacers appear to be derived from the genome of bacteriophages (herein called phages), archaeal viruses (herein called viruses), plasmids or other organisms (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005; Marraffini and Sontheimer, 2008; Brodt et al., 2011; Bikard et al., 2012;

Shmakov et al., 2017). However, many of the assigned spacers match sequences within the host genome, what are generally called self-targeting spacers.

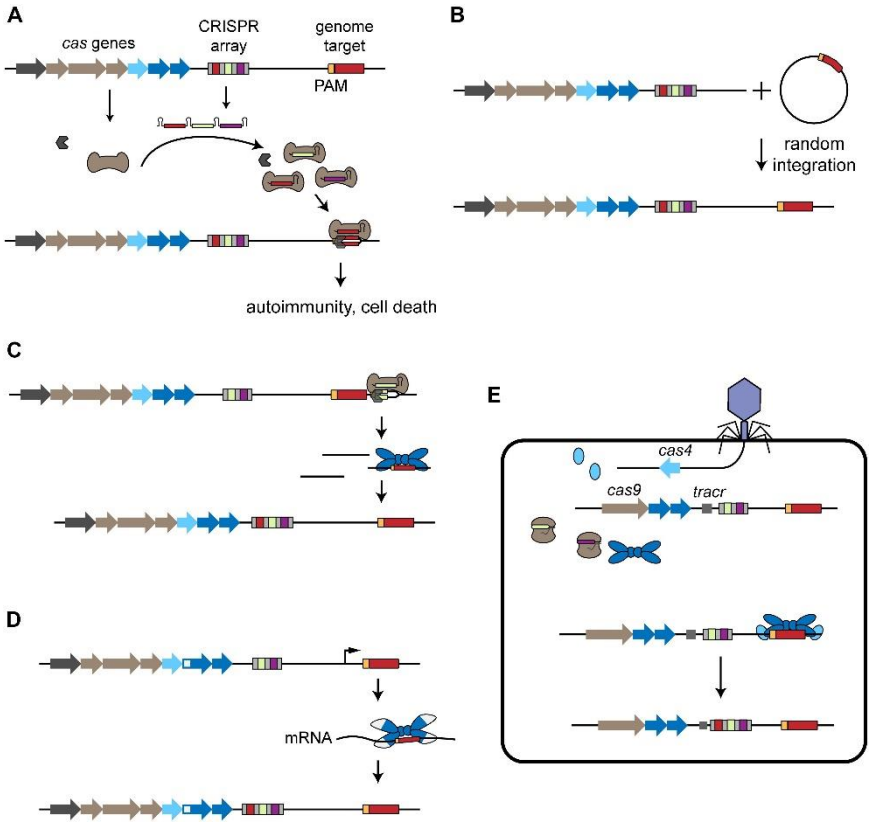


Figure 1: Acquisition of self-targeting spacers. (**A**) Overview of self-targeting by CRISPR-Cas systems. The CRISPR array is transcribed and processed into individual crRNAs that form a ribonucleoprotein complex with the Cas effector proteins (brown). One of the crRNAs encodes a self-targeting spacer (red) that directs binding to the complementary protospacer sequence (red) flanked by a PAM (orange) located on the genome, leading to autoimmunity and cell death. (**B**) Mobile genetic elements harboring a CRISPR-Cas target sequence can be incorporated into the host chromosome, leading to self-targeting. (**C**) Primed acquisition. The CRISPR effector complex recognizes a target, potentially generating cleaved products. These products can then be incorporated into the CRISPR array by the acquisition complex (blue), leading to acquisition of self-targeting spacers. (**D**) Spacer acquisition from RNA. RT-Cas1 forms a complex with Cas2 (white and blue) and leads to incorporation of self-targeting spacers derived from the host's RNA. (**E**) Virally driven acquisition of self-targeting spacers. The phage injects its genome into the host cell and the encoded *cas4* is expressed. In cooperation with the host's endogenous acquisition complex, the phage-derived Cas4 leads to the incorporation of genome-derived spacers into the host's CRISPR array.

Self-targeting spacers are unexpected due to an observed preference toward acquiring foreign genetic material (Levy et al., 2015) and heavy cytotoxicity to the host because self-targeting of the host's chromosome would lead to cell death (Stern et al., 2010; Jiang et al., 2013; Vercoe et al., 2013; Gomaa et al., 2014; **Figure 1A**). Here, we review the presence and consequences of self-targeting spacers. We address the known distribution of self-targeting spacers in sequenced CRISPR-Cas systems. We then discuss different mechanisms of acquisition that could generate self-targeting spacers and how these organisms can survive despite the potential for chromosomal targeting and autoimmunity. Finally, we report some of the beneficial functions that have been associated with the self-targeting spacers that can imbue CRISPR-Cas systems with functionalities that extend beyond adaptive immunity. This content greatly expands on an earlier mini-review on the consequences of chromosomal

targeting (Heussler and O'Toole, 2016) and incorporates recently reported examples of self-targeting reflecting alternative functions of these prevalent adaptive immune systems.

Natural occurrence of self-targeting spacers

Multiple studies have explored the source of spacers in diverse CRISPR-Cas systems, with recurring observations of self-targeting spacers. In the first broad study of matching protospacers, 88 of the analyzed 4,500 spacers were similar to known sequences, and 35% of these spacers matched chromosomal DNA not directly related to foreign genetic elements (Mojica et al., 2005). Separately, a study from 2008 found that 7% of spacers in different CRISPR-Cas systems from *Streptococcus thermophilus* matched chromosomal sequences (Horvath et al., 2008). One year later, the same group analyzed CRISPR-Cas systems from a more diverse set of lactic acid bacteria, reporting that 23 of the 104 spacers matched the chromosome (Horvath et al., 2009). Shortly thereafter, one study analyzed the CRISPR arrays of the 330 prokaryotes containing CRISPR-Cas systems included in the CRISPRdb database (Grissa et al., 2007) in 2010, with self-targeting spacers comprising 0.4% of all spacers (including the vast majority of spacers with no assignable protospacers) and appearing in 18% of the included prokaryotic genomes (Stern et al., 2010).

The number of sequenced organisms has increased over time, allowing more recent studies to more deeply and widely interrogate spacer origins. For instance, one study in 2017 screened 50,000 completely or partially assembled genomes, while another study in 2018 used the online tool CRISPRminer to evaluate more than 60,000 organisms harboring a CRISPR array (Shmakov et al., 2017; Zhang et al., 2018). Shmakov et al. assigned protospacer locations to 7% of the detected 363,460 unique spacers, with ~6% of these spacers matching prokaryotic genomes and 16% of these genome-matching spacers being potentially unrelated to (pro-)viral sequences (Shmakov et al., 2017). The study with CRISPRminer reported 22,110 self-targeting events in publications (Stern et al., 2010; Rauch et al., 2017; Watters et al., 2018) and could predict 6,260 additional putative self-targeting spacers in 4,136 organisms, implying that ~7% of the genomes within their database should harbor at least one self-targeting spacer (Zhang et al., 2018).

The natural acquisition of self-targeting spacers has also been observed as part of adaptive evolution studies between phages and their prokaryotic host. Two key studies relied on a strain of the bacterium *S. thermophilus* harboring two type II-A CRISPR-Cas systems (Paez-Espino et al., 2013, 2015). In these studies, only 0.01 – 0.04% of the observed new spacers matched the genome. These frequencies are lower than those reported in the large-scale bioinformatics studies, although this discrepancy can be attributed in part to the selective pressure exerted by the actively infecting phages.

Incorporation of self-targeting spacers

Given the frequency of self-targeting spacers and their potential for autoimmunity, we next discuss the circumstances under which a self-targeting spacer can be acquired. In particular, we consider five general scenarios that have been reported: naïve acquisition from DNA, protospacers within a transferred mobile genetic element (MGE), primed adaptation, naïve acquisition from RNA, and phage/virus-triggered acquisition from host DNA. For many of these scenarios, we address the extent to which acquisition differentiates between chromosomal and foreign genetic material and the known associated mechanisms. We finally must note that our understanding of CRISPR-based acquisition is still developing, and other mechanisms within the diversity of CRISPR-Cas systems likely await discovery.

Naïve acquisition

Naïve acquisition leads to the incorporation of new spacers without any influence from the existing pool of spacers. Cas1 and Cas2 are required while Cas4, Csn2 or Cas9 may be additionally needed depending on the system sub-type (Yosef et al., 2012; Nuñez et al., 2014; Heler et al., 2015; Wei et al., 2015; Kieper et al., 2018). It was known for many years that protospacers were commonly flanked by PAMs to allow targeting by the effector proteins to differentiate between self and non-self targets (Deveau et al., 2008; Horvath et al., 2008). However, it remained unclear how the acquisition machinery differentiated between invader and chromosomal DNA. In one of the first studies to systematically interrogate spacer acquisition, Levy and coworkers sequenced over 38 million newly acquired spacers following plasmid-based expression of Cas1 and Cas2 in an *Escherichia coli* strain harboring a CRISPR array but lacking endogenous *cas* genes. They found that spacers were preferentially acquired from replication forks, presumably due to stalling during replication and degradation by RecBCD. This preference resulted in 100-fold to 1,000-fold enrichment of spacers derived from a resident plasmid compared to the chromosome. The high-copy number plasmids present most of the replication forks in a replicating cell, partly explaining the preference toward high copy plasmids (Levy et al., 2015).

Another critical factor identified by Levy and coworkers was the presence of Chi sites. These sequence motifs interact with and prevent DNA degradation by RecBCD (Smith, 2012) at the sites of double-stranded DNA breaks that often occur at stalled replication forks (Kuzminov, 2001; Michel et al., 2001). Due to the fact that Chi sites occur approximately every 5 kb in the *E. coli* genome (El Karoui et al., 1999), these Chi sites were hypothesized to mark the host DNA as “self” and prevent acquisition of spacers from the host’s genome. The plasmids contained fewer Chi sites, likely further contributing to preferential acquisition from this DNA. Linear viral DNA would also offer a preferred substrate for RecBCD, resulting in DNA

fragments that can be used to generate new spacers (Levy et al., 2015). Phages are known to encode RecBCD inhibitors and some also encode a large number of Chi sites in their genome (Friedman and Hays, 1986; Murphy and Lewis, 1993; Bobay et al., 2013), thus potentially countering acquisition by CRISPR-Cas systems.

Liu et al. observed a different element influencing naïve acquisition in *Sulfolobus islandicus*, an archaeon that encodes one type I-A CRISPR-Cas system and two type III-B CRISPR-Cas systems. Following overexpression of Csa3a that drives expression of the type I-A acquisition genes, *S. islandicus* integrated spacers from the *csa3a* expression plasmid as well as from its own genome with a high bias toward the plasmid (Liu et al., 2015). Interestingly, for deletion mutants lacking RNA processing or nuclease activity, <28% of spacers were derived from the plasmid (Liu T. et al., 2017). While this fraction was far less than the >90% in a previous study (Liu et al., 2015), it still reflected preferential acquisition from plasmids when taking into account the relative length of the plasmid and chromosomal DNA (Liu T. et al., 2017). The stronger preference for plasmid DNA in the presence of an active CRISPR-Cas system may be explained in part by the cytotoxicity of genome targeting by the active but not impaired system upon self-targeting.

Spacer acquisition in type II CRISPR-Cas systems also appears to differ for active versus impaired CRISPR-Cas systems. Wei et al. (2015) looked at acquisition requirements in a type II-A CRISPR-Cas system by expressing the different CRISPR-Cas components on plasmids and monitoring spacer acquisition. They found that acquisition required the presence of Cas9, in contrast to spacer acquisition by Cas1 and Cas2 in type I CRISPR-Cas systems. Interestingly, the authors found that the cleavage activity of Cas9 contributed to an observed preference for acquisition from plasmid DNA. Specifically, by using a mutated Cas9 that disrupts its cleavage activity (dCas9), the authors shifted the fraction of plasmid-derived spacers from 68% to 4%, representing a loss of preference given the matching ratio of plasmid DNA to genomic DNA (Wei et al., 2015). In total, naïve acquisition by different types of CRISPR-Cas systems can lead to the incorporation of self-targeting spacers, although foreign genetic material is the predominant source of spacers. It would be interesting to investigate if the above reported phenomena can also be observed in different organisms or other CRISPR-Cas systems that rely on additional Cas proteins for acquiring new spacers.

Protospacer within transferred mobile genetic elements

Many self-targeting spacers identified in nature bear homology to MGEs such as transposons or prophages/proviruses that have been incorporated into the genome. These spacers could have been acquired prior to the incorporation of the MGE as a preventative measure, or afterward to induce cell death and prevent further spread of the MGE (**Figure 1B**). All evidence

of this mechanism comes from bioinformatic experiments. Looking at self-targeting spacers with 100% complementary to their predicted protospacer region, Stern et al. (2010) found an approximately equal distribution of protospacers from mobile elements encoded in the chromosome and non-mobile elements (47% vs. 53%). In comparison, Shmakov et al. (2017) assigned 83% of the self-targeting spacers to (pro-)phage sequences. The difference might arise from the greater abundance of sequenced MGEs over time (Geer et al., 2010; Akhter et al., 2012). Nevertheless, these frequencies leave ample spacers derived from non-mobile elements.

Primed adaptation

A different potential means of incorporating self-targeting spacers is through primed adaptation (or primed acquisition). Acquisition of spacers under primed adaptation requires target recognition with pre-existing spacers that are partially or fully complementary to the foreign DNA. Recognition leads to the acquisition of multiple spacers from sites in close proximity to the existing protospacer (Datsenko et al., 2012; Swarts et al., 2012; Richter et al., 2014; Jackson et al., 2019; **Figure 1C**).

Bioinformatic evidence indicates that primed adaptation is widespread in type I and type II CRISPR-Cas systems (Nicholson et al., 2019). Primed adaptation by type I systems involves degradation of the target site by Cas3 and incorporation of the degradation products as new spacers by Cas1 and Cas2 (Künne et al., 2016). Primed adaptation by type II systems is not well understood, although Nicholson et al. (2019) proposed two possible pathways: one that involves a main role of Cas9, and another involving host-specific processes such as DNA repair producing pre-spacers at the sites of target cleavage.

Regardless of the exact mechanism, primed adaptation is expected to preferentially incorporate foreign genetic material due to the pre-existence of more spacers derived from non-chromosomal elements. However, primed acquisition of host DNA could occur upon targeting MGE that were incorporated into a bacterial or archaeal genome (Nicholson et al., 2019). Primed acquisition outside of the borders of the MGE could also be triggered, leading to incorporation of non-mobile self DNA from the chromosome. Finally, spacers that evolved to target foreign DNA might prime with similar sequences in chromosomal DNA (Staals et al., 2016), where prior work showed that priming can occur even with 13 mutations in the target site relative to the pre-existing spacer (Fineran et al., 2014).

Naïve acquisition of RNA-derived spacers

One unique mode of acquisition is through relatively rare Cas proteins that recognize RNA rather than DNA. These proteins include a RT often translationally fused to Cas1 or to a fusion

between Cas1 and the Cas6 protein responsible for crRNA biogenesis. This unique RNA-acquiring machinery is predominantly associated with type III CRISPR-Cas systems but is also found with type I-E and type VI-A systems (Kojima and Kanehisa, 2008; Simon and Zimmerly, 2008; Toro and Nisa- Martínez, 2014; Silas et al., 2017; Toro et al., 2019a,b). For the few examples that have been studied, these RTs reverse-transcribe an acquired RNA into DNA to produce a substrate for acquisition (Silas et al., 2016). If the RNA-derived spacers are derived from host RNA, the associated type III CRISPR-Cas systems can now target the host and lead to autoimmunity (**Figure 1D**). Interestingly, self-targeting spacers have been found in three strains encoding a RT as part of their type III CRISPR-Cas systems (Silas et al., 2017; Zhang et al., 2018).

Other systems solely encode RT and Cas1 and lack all other Cas proteins, holding the potential to acquire self-targeting spacers without inducing autoimmunity. As one example, *Rivularia* sp. PCC 7116 encodes Cas1, Cas2, and RT in a distinct genomic island compared to the other CRISPR-Cas systems present in that bacterium. The CRISPR array associated with Cas1, Cas2 and RT harbors a spacer matching a hypothetical gene encoded on the bacterial chromosome (Silas et al., 2017; Kersey et al., 2018; Zhang et al., 2018). The lack of effector proteins suggests that these systems are used for alternative functions rather than immunity, although this has not been investigated to-date.

The unique sourcing of spacers from RNA raises questions about how the acquisition machinery selected some RNA sequences over others. Silas and coworkers sequenced the spacer content in an open-air culture of *Arthrospira platensis*, which encodes a RT-Cas1 fusion as part of its type III-B CRISPR-Cas system (Silas et al., 2017). Most of the associated protospacers could not be identified, and the few that could be identified traced to DNA viruses. Schmidt and coworkers were able to gain more extensive insights by monitoring spacer acquisition in *E. coli* following plasmid-based expression of the type III *Fusicatenibacter saccharivorans* RT-Cas1 and Cas2. While spacers were derived from RNAs encoded in the chromosome and plasmid, there was a strong preference for A/T-rich sequences at the ends of highly expressed genes. Interestingly, there was no obvious preference for a flanking motif or for plasmid-encoded RNAs (Schmidt et al., 2018). Further studies are needed to fully understand preferences exhibited by type III CRISPR-Cas systems for RNA acquisition.

Acquisition of self-targeting spacers triggered by foreign invaders

There is also evidence that phages can encode Cas proteins that drive endogenous CRISPR-Cas systems to preferentially acquire self-targeting spacers. The first direct evidence comes from studying the origin of spacers encoded within the CRISPR array of *Campylobacter jejuni* PT14 harboring a minimal type II-C CRISPR-Cas system (Hooton and Connerton, 2014). While

the spacers do not share 100% sequence identity with any known sequences, some of the spacers partially matched chromosomal sequences in the PT14 genome. Tracking spacer content in a co-culture of PT14 cells and CP8/CP30A phage revealed that all newly acquired spacers were derived from the host's chromosome and not the phage (Hooton and Connerton, 2014). The phage encoded a copy of the *cas4* gene involved in protospacer maturation as part of many CRISPR-Cas systems (Zhang et al., 2012; Lemak et al., 2013; Kieper et al., 2018; Lee et al., 2018), while the endogenous type II-C CRISPR-Cas system normally lacks this gene. The authors therefore attributed the unexpected self-targeting acquisition events to the phage encoded Cas4 (Hooton and Connerton, 2014; **Figure 1E**).

Further evidence that viral Cas4 can impact host acquisition was found in *S. islandicus*. Zhang and coworkers evaluated the impact of the viral *cas4* gene found in a *Sulfolobus* spindle-shaped virus by transforming a plasmid encoding the viral *cas4* into *S. islandicus*. Cells harboring the plasmid exhibited less frequent spacer acquisition, although the frequency of spacers acquired from the plasmid or chromosome did not change. Furthermore, overexpression of host Cas4 from a plasmid also led to reduced spacer acquisition. These findings suggest that overproduction of Cas4 can in some cases disable spacer acquisition. One explanation is that the viral encoded Cas4 serves as an anti-CRISPR protein (Acr) by preventing spacer acquisition and in turn enabling escape from CRISPR-Cas targeting (Zhang et al., 2019). While more work is needed to elucidate the underlying role of the virally encoded Cas4, these examples and the many other instances of virally encoded Cas4 (Krupovic et al., 2015; Hudaiberdiev et al., 2017) suggest the intriguing possibility that phages and viruses could be actively directing the acquisition of spacers.

Surviving self-targeting by CRISPR-Cas systems

Unrelated to how prokaryotes incorporate spacers that target their own genome, cells must overcome self-targeting by their own CRISPR-Cas system to survive. CRISPR-based interference against the host's own genome is expected to lead to lethal autoimmunity due to the nuclease cutting within or close to their target site. Repair mechanisms in prokaryotes are often not efficient enough to fix CRISPR-Cas induced DNA damage, and DNA breaks often result in cell death (Stern et al., 2010; Jiang et al., 2013; Vercoe et al., 2013; Gomaa et al., 2014). Nevertheless, many different examples exist in which a self-targeting spacer can be tolerated. Below we describe each known mechanism.

Active DNA repair

CRISPR-based targeting would be expected to induce irreparable damage, lest the cells repair invading genetic material and allow an infection to persist. Accordingly, many studies have

reported that chromosomal targeting by CRISPR nucleases is cytotoxic in different bacteria and archaea (Stern et al., 2010; Jiang et al., 2013; Vercoe et al., 2013; Bikard et al., 2014; Citorik et al., 2014; Gomaa et al., 2014; Li Y. et al., 2016). That said, there exist examples in which intrinsic DNA repair mechanisms such as homology-directed repair (HDR), non-homologous end joining (NHEJ), and alternative end-joining (A-EJ) mechanisms allow cell survival (Chayot et al., 2010; Tong et al., 2015; Cui and Bikard, 2016; Stachler et al., 2017; **Figure 2A**).

One potential mechanism is HDR through an additional copy of the chromosome. Cui and Bikard first observed this phenomenon when evaluating the consequences of targeting the *E. coli* chromosome with heterologously expressed Cas9 (Cui and Bikard, 2016). They found that targeting different sites within non-essential genes resulted in RecA-mediated HDR. Targeting did induce the SOS DNA-damage response, although the cells maintained their viability. In a separate example, Stachler and coworkers reported that the archaeon *Haloferax volcanii* could tolerate chromosomal targeting through its endogenous type I-B CRISPR-Cas system (Stachler et al., 2017). However, the tolerance could be attributed in part to the endogenous CRISPR array providing most of the crRNAs in the effector complexes. Deleting the Cas6 processing protein and expressing a mature self-targeting crRNA resulted in a fitness defect that was strengthened by expressing the crRNA at higher levels. The extent of self-targeting in the presence of the endogenous CRISPR array therefore was sufficiently weak to allow repair through HDR and the roughly 20 copies of the *H. volcanii* genome (Zerulla et al., 2014; Stachler et al., 2017). In both of these examples, there would likely be some selective pressure to disrupt self-targeting given the need for continuous repair.

Non-homologous end joining and alternative end-joining offer distinct repair mechanisms that permanently alter the target site, preventing further attack by CRISPR-Cas systems. NHEJ does not utilize a repair template and instead repairs double-stranded DNA breaks (DSBs) by adding insertions or deletions (indels) to the site of the DSB. Some prokaryotes possess relatively unsophisticated NHEJ machinery compared to eukaryotes, typically comprised of the complexes Ku and LigD (Aravind and Koonin, 2001; Weller et al., 2002; Gong et al., 2005; Bowater and Doherty, 2006; Shuman and Glickman, 2007; Tong et al., 2015). Some bacteria such as *E. coli* lacking Ku and LigD can utilize phage ligases to mediate NHEJ-like repair of CRISPR-Cas induced DSBs (Su et al., 2019). While NHEJ efficiently repairs DNA cleaved by some CRISPR nucleases in eukaryotic cells, some CRISPR-Cas induced DNA damage in prokaryotes is still highly cytotoxic when NHEJ is active (Xu et al., 2015; Bernheim et al., 2017). A-EJ is a repair mechanism that relies on microhomology-mediated end joining and largely leads to deletions. DSBs induce extensive end-resection that is mostly dependent on RecBCD, while Ligase-A repairs the break by joining micro-homologous regions of 1 to 9 bps (Chayot et

al., 2010; **Figure 2A**). Prior work suggested that A-EJ led to large deletions following genomic attack by a type I-F CRISPR-Cas system in *Pectobacterium atrosepticum* (Vercoe et al., 2013). Whether repair occurs through NHEJ or A-EJ, the resulting genome would be less susceptible (or even completely unsusceptible) to follow-up attack through the self-targeting spacer.

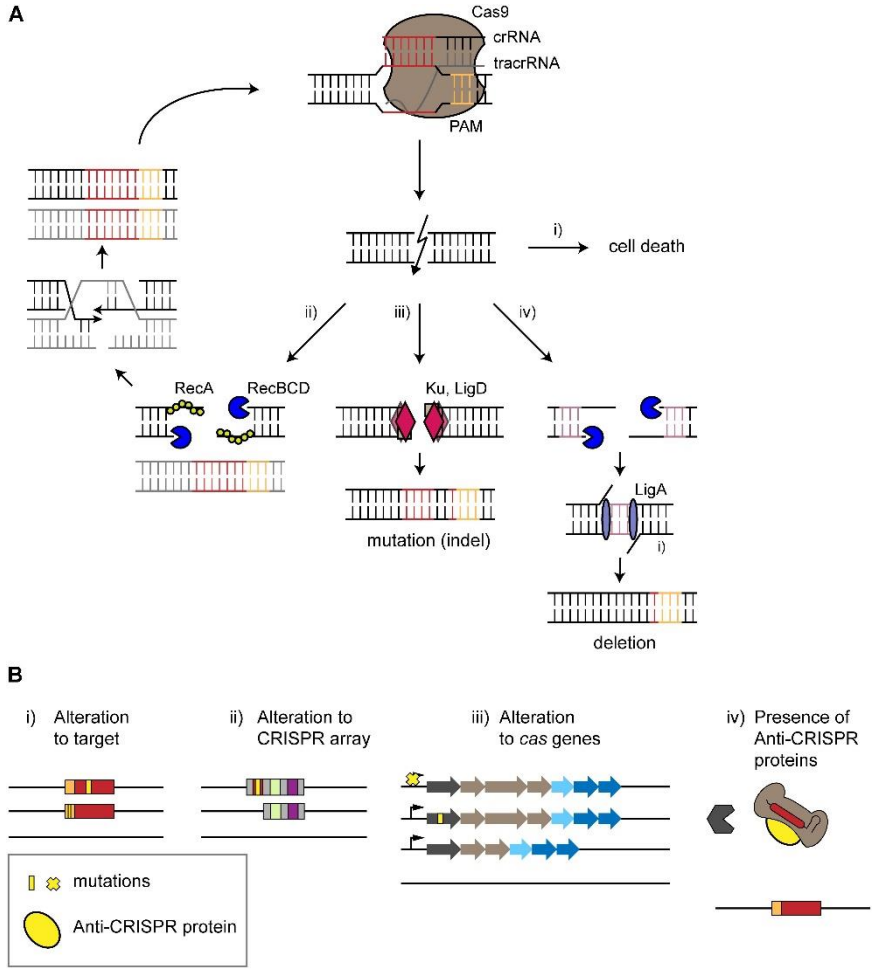


Figure 2: Surviving self-targeting. (**A**) Survival of CRISPR-Cas targeting by intrinsic repair mechanisms. The type II effector protein Cas9 is used as an example. The CRISPR effector complex binds to its target in the genome (red) next to a PAM (orange), leading to a double-stranded break (DSB) causing different outcomes. (**i**) Cell death occurs if the break is not repaired. (**ii**) Homology-directed repair (HDR) restores the target site in the presence of an intact copy of the chromosome. DNA ends of the DSB undergo trimming in a Rec-dependent manner. HDR leads to the restoration of a chromosome that undergoes further attack by the CRISPR-Cas system. (**iii**) Non-homologous end joining leads to the formation of an insertion or deletion (indel). End joining is mediated by the repair proteins Ku and LigD. (**iv**) Alternative end-joining leads to deletions. DNA ends are trimmed by RecBCD until micro-homologous regions (purple) are reached. These regions are then ligated by LigA, resulting in deletions. (**B**) Escape from autoimmunity through mutations, deletions, or active inhibition. (**i,ii**) mutations (yellow) or deletions within the protospacer, PAM or CRISPR array disrupts self-targeting. (**iii**) Mutation of the *cas* operon, inhibition of Cas expression or deletion of a *cas* gene or the entire locus can also prevent self-targeting. (**iv**) Anti-CRISPR proteins encoded within an integrated prophage can block CRISPR-Cas interference through different mechanisms, such as binding the Cas effector protein to prevent PAM recognition.

Mutations disrupting CRISPR-based targeting

Mutations can prevent efficient CRISPR targeting in multiple ways. One way is mutation of the target site such as through NHEJ or A-EJ, impacting spacer complementarity or PAM recognition (**Figure 2B**). Two studies evaluating self-targeting through type II-A systems in *S.*

thermophilus reported not only mutations of the targeted *lacZ* gene but also deletion of the gene (Selle et al., 2015; Cañez et al., 2019). In one of the studies assessing self-targeting in the *S. thermophilus* strain LMD-9, targeting resulted in loss of ~1.2 kb that included the *lacZ* gene. These deletions appeared to arise via genomic island excision via recombination between two flanking insertion-sequence elements (Selle et al., 2015) that occur quite frequently in *S. thermophilus* (Bolotin et al., 2004). In contrast, another study reported an ~40-kb deletion upon targeting *lacZ* in the *S. thermophilus* strain DGCC7710 that shares 99.2% sequence homology to LMD-9. No insertion-sequence elements could be detected within 50 kb flanking the *lacZ* gene, potentially explaining why the same escape mechanism observed in LMD-9 did not take place in DGCC7710. Recombination here might have happened between two regions encoding two *galE* genes sharing 86% nucleotide identity located 3 kb upstream and 30 kb downstream of *lacZ* (Cañez et al., 2019). One interesting possibility is that these large deletions existed in a small fraction of the cell population, where CRISPR-based targeting allowed this sub-population to survive (Selle et al., 2015; Cañez et al., 2019). The different outcomes of self-targeting in *S. thermophilus* LMD-9 and DGCC7710 highlight the different escape mechanisms that can occur even between strains of the same species.

Escape from lethal self-targeting can not only occur via target mutation but also via mutations or deletions within the CRISPR array or the *cas* genes (**Figure 2B**). In the same study noted above (Cañez et al., 2019), the authors also investigated the escape mechanism by targeting *lacZ* with the endogenous type I-E CRISPR-Cas system in *S. thermophilus* DGCC7710. Surprisingly, no deletions in the target site could be observed, and escape mutants consistently harbored defective plasmids missing the targeting spacer and one repeat likely caused by recombination between repeats that eliminated the self-targeting spacer (Cañez et al., 2019; **Figure 2B**). Thus, escape mechanisms can differ not only between strains but also between CRISPR-Cas systems. Loss of the plasmid-encoded spacer also occurred as the principal mode of escape when targeting *E. coli*'s genome with its endogenous type I-E system (Gomaa et al., 2014). Separately, as an example of disrupting *cas* genes, *Lactobacillus acidophilus* NCFM appears to have deleted its entire *cas* gene cassette to avoid lethal self-targeting by six genome-targeting spacers encoded in the CRISPR array (Stern et al., 2010; Kersey et al., 2018; Zhang et al., 2018). Furthermore, mutations in the *cas* genes were reported when targeting the staphylococcal cassette chromosome mec (SSCmec) in *Staphylococcus aureus* through the endogenous III-A CRISPR-Cas system (Guan et al., 2017) or when targeting different sites through the I-F CRISPR-Cas system native to *P. atrosepticum* (Dy et al., 2013). Between disruption of the spacer or *cas* genes, explicit loss of an endogenous self-targeting spacer has been less reported in natural systems. However, this can be explained

by bioinformatic searches having difficulties detecting loss of self-targeting spacers in genome databases given that the rest of the CRISPR array may still be intact.

Self-targeting by a CRISPR-Cas system also does not need to drive only one mode of escape. For instance, in the example of self-targeting through the type III-A CRISPR-Cas system in *S. aureus*, the authors reported different mutations or deletions in the escape mutants (Guan et al., 2017). Large deletions that included the target site occurred in ~90% of the escape mutants, while spacer mutations or loss-of-function mutations in *cas* genes were also detected. Separately, in the example of self-targeting through the type I-F CRISPR-Cas system in *P. atrosepticum*, the bacterium harbors one naturally occurring self-targeting spacer that is not cytotoxic due to a mutation in the target's PAM (Dy et al., 2013). Transformation of plasmids harboring other self-targeting spacers further led to different sized deletions of regions containing the protospacer or removal of the *cas* operon. The frequency of one escape mode over another likely depends on different factors such as the frequency of background mutation and recombination, the types of mutations that can form, and the fitness defect that they introduce.

Partially complementary spacers directing target binding but not cleavage

Mutations to the target site or the spacer can result in partial complementarity between spacers and their protospacers. For some systems, partial complementarity eliminates target cleavage but can preserve target binding. Comparison between off-target binding by dCas9 and off-target cleavage by Cas9 demonstrated extensive off-target binding but not cleavage (Wu et al., 2014). Another study also showed that partial target complementarity could allow an active Cas9 to bind but not cleave DNA, resulting in transcriptional silencing (Bikard et al., 2013). Wu and coworkers proposed a model which would explain the higher specificity of Cas9 by taking binding at the seed region into consideration. They hypothesized that PAM recognition by the Cas9:crRNA complex leads to DNA melting and enables base pairing between the spacer and the complementary seed region. As long as complementarity exists through the seed region, partial base pairing can allow target binding without cleavage (Wu et al., 2014). As a result, organisms could harbor spacers with partial complementarity to their own genome that would still drive target recognition but not autoimmunity.

RNA targeting

While we have focused on CRISPR-Cas systems that explicitly target DNA, the type III and VI systems naturally target RNA as part of immune defense (Hale et al., 2009, 2012; Abudayyeh et al., 2016), with distinct implications for self-targeting. Type III CRISPR-Cas systems are capable of targeting DNA and RNA. The system's Csm or Cmr effector complex is guided to

RNA targets complementary to the crRNA, triggering the sequence-specific RNase activity of Csm3 or Cmr4, respectively (Benda et al., 2014; Goldberg et al., 2014; Tamulaitis et al., 2014; Samai et al., 2015). Lack of complementarity between the 5′crRNA handle and the target RNA activates single-stranded DNase activity by Cas10 (Jung et al., 2015; Kazlauskiene et al., 2016), although there is evidence of a 3′ RNA PAM motif that suggests diverging criteria for target selection across type III systems (Elmore et al., 2016). Furthermore, target recognition by the type III effector complex triggers Cas10 to produce cyclic adenylates. These molecules in turn activate the CRISPR accessory protein Csm6/Csx1, leading to non-specific RNA degradation to assist in viral defense (Kazlauskiene et al., 2017; Niewoehner et al., 2017; Rouillon et al., 2018).

In contrast to type III CRISPR-Cas systems, type VI systems represent the only systems known to-date that exclusively target RNA (Abudayyeh et al., 2016). Cas13, the type VI effector protein, recognizes complementary RNA sequences as long as the repeat-portion of the crRNA cannot extensively base pair with the target (Meeske and Marraffini, 2018). Upon target recognition, Cas13 undergoes a conformational change that activates the effector's ribonuclease domain, resulting in non-specific cleavage of the proximal portions of the target RNA (Liu et al., 2017a,b). The effector domain remains highly active even after cleavage of local RNAs, leading to the extensive degradation of cellular RNAs. The degradation can be sufficiently extensive to shut down the host's growth, resulting in a reversible dormancy state (Abudayyeh et al., 2016; Meeske et al., 2019). The activity of type VI CRISPR-Cas targeting also had a more severe effect on the fitness of *E. coli* during high production of target RNA, potentially allowing the cell to survive self-targeting by Cas13 if the target RNA is not highly expressed (Abudayyeh et al., 2016) and sparing the cells from self-targeting induced dormancy.

Another consequence of RNA-based (self-)targeting is type III systems and Cas13 ignoring transcriptionally silent targets. Activation of type III and VI systems only upon RNA recognition would be particularly important for temperate phages and viruses whose lytic genes are repressed during lysogeny (Johnson et al., 1981). Therefore, if a spacer directs the Csm/Cmr effector complex or Cas13 to an RNA necessary for the lytic cycle, then only the lysogens entering the lytic cycle will be targeted. Tolerance of a prophage has been shown for the type III-A system in *Staphylococcus epidermidis* that actively targets its own prophages only upon transition into the lytic cycle (Goldberg et al., 2014). By only targeting phages and viruses in the lytic cycle, cells are able to maintain any potentially positive functions that might arise from a prophage/provirus encoded in their genome and prevent cell death during the invader's lytic phase.

Genome-encoded anti-CRISPR proteins

Escape from targeting is not limited to genetically disrupting the CRISPR-Cas system or its target; another means involves inhibiting CRISPR-Cas activity in *trans* by Acrs. These proteins allow phages/viruses to thwart immunity by CRISPR-Cas systems (Pawluk et al., 2018). So far, Acrs have been identified that inhibit different subtypes of type I, II, III, and V CRISPR-Cas systems (Bondy-Denomy et al., 2013; Pawluk et al., 2014, 2016a,b; Hynes et al., 2017; Rauch et al., 2017; He et al., 2018; Marino et al., 2018; Watters et al., 2018; Bhoobalan-Chitty et al., 2019), and Acrs against type IV and VI systems likely await discovery. The Acrs identified to-date have exhibited remarkable diversity in their sequence and in their mechanism of action, such as blocking DNA binding, preventing effector complex formation, sequestering the nuclease into dimers, blocking nuclease activity or preventing nuclease recruitment (Bondy-Denomy et al., 2015; Pawluk et al., 2018; Thavalingam et al., 2019).

Acrs allow phages and viruses to not only escape attack by CRISPR-Cas systems but also protect a lysogenized phage/provirus (not to mention the host chromosome) from an endogenous CRISPR-Cas system encoding a viral-targeting or chromosomal-targeting spacer (**Figure 2B**). Therefore, a genome encoding both a CRISPR-Cas system and a self-targeting spacer could potentially also encode an Acr. Rauch et al. (2017) hypothesized that self-targeting spacers would indicate the presence of an inhibiting Acr, which led them to identify four Acrs encoded in prophage regions of *Listeria monocytogenes* that inhibit Cas9. Separately, Watters et al. (2018) and Marino et al. (2018) used a similar approach to identify Acrs in *Moraxella bovoculi* active against type I and type V CRISPR-Cas systems. Given the success in identifying Acrs in prokaryotes harboring self-targeting CRISPR-Cas systems, this mechanism could principally explain the natural appearance of self-targeting spacers.

Self-targeting spacers underlying alternative functions of CRISPR-Cas systems

We have described how self-targeting spacers can be acquired and how cells can avoid the cytotoxic impact of self-targeting. In some of these cases, self-targeting could reflect an alternative function of the CRISPR-Cas system. Here, we describe different examples in which self-targeting has impacted the host or in which a mechanism has been reported that could impact host behavior, potentially foreshadowing an alternative function. These examples can be divided into four categories: genome evolution, RNA degradation, transcriptional repression, and foreign invaders co-opting self-targeting CRISPR-Cas systems. While conserved examples of CRISPR-Cas systems performing alternative functions have not been

described, there has been a steady increase in anecdotal examples that suggest that CRISPR-Cas systems can stray from adaptive immunity, with varying benefits to the host.

Genome evolution

One reported outcome of acquiring self-targeting spacers is genome evolution by forcing the host to mutate in order to escape autoimmunity. While this mechanism still reflects active DNA targeting through the standard steps of CRISPR-based immunity and thus may not represent a “true” alternative function, we still consider this an alternative function because of the large-scale change in genomic content that can confer benefits to the host. Specifically, chromosomal targeting can lead to mutations or small deletions in the target gene. These deletions can also be much larger and encompass many surrounding non-targeted genes. While any loss of an essential gene would be lethal, these larger deletions could also provide a fitness advantage by generating new phenotypes or reducing the overall size of the genome, and remodeling of pathogenicity islands could cause a change in bacterial virulence (Vercoe et al., 2013; Westra et al., 2014). Besides triggering active mutations, self-targeting by CRISPR-Cas systems can also select for a small sub-population already lacking the target (Dy et al., 2013; Selle et al., 2015).

Self-targeting by CRISPR-Cas systems can further lead to bacterial or archaeal evolution by disrupting an important gene and forcing the organism to adapt to this change. One important example comes from the bacterium *Pelobacter carbinolicus*. Unlike other members of the *Geobacteraceae* family, *P. carbinolicus* cannot reduce Fe(III) as part of its metabolism (Richter et al., 2007). This phenotype is potentially caused by an existing spacer within the endogenous type I-E CRISPR-Cas system that is complementary to a region within the histidyl-tRNA synthetase gene *hisS*. A lack of histidyl-tRNA synthetase would lead to reduced translation of proteins with multiple closely spaced histidines. The *hisS*-targeting spacer is located opposite of the end of the CRISPR array where new spacers are added, suggesting that the uptake of this spacer did not occur recently. Supporting the active targeting of *hisS*, transforming the self-targeting spacer and the *hisS* gene from *P. carbinolicus* into a genetically tractable strain of the related species *Geobacter sulfurreducens* resulted in few transformants, and these transformants grew poorly. *P. carbinolicus* has also lost or mutated multiple genes with high histidine content that are still present in closely related species, potentially also explaining the loss of Fe(III)-respiration (Aklujkar and Lovley, 2010). It would be interesting to see how the endogenous I-E system is impacting HisS expression without driving lethal autoimmunity, where we expect the mechanism to fall under one of the categories below.

CRISPR-Cas induced mRNA cleavage

Not all CRISPR-Cas systems solely target DNA, wherein RNA targeting could modulate gene expression without inducing cytotoxicity. To-date, type III, type VI, some type I, and some type II CRISPR-Cas systems have been shown to target RNA (Hale et al., 2009, 2012; O'Connell et al., 2014; Samai et al., 2015; Abudayyeh et al., 2016; Li R. et al., 2016; Dugar et al., 2018; Rousseau et al., 2018; Strutt et al., 2018). In the event that RNA but not DNA is targeted, self-targeting spacers would not necessarily result in autoimmunity but instead could degrade mRNA and lead to changes in gene expression.

The type III-B CRISPR-Cas system in *Myxococcus xanthus* is a potential example that degrades mRNA, although this mechanism remains to be fully established (Wallace et al., 2014). As part of the study, the authors performed a transposon screen in a Δ*pilA* strain lacking the type IV pilus required for exopolysaccharide production. They isolated a mutant with a transposon inserted into the CRISPR3 array, which coincided with restored exopolysaccharide production and impaired fruiting body development. Wallace et al. (2014) proposed a mechanism in which the transposon enhanced pre-crRNA processing, leading to crRNA-dependent regulation of exopolysaccharide production and fruiting body development. Other possibilities are that the repertoire of crRNAs includes a portion of the transposon, altering the targeting potential of the array. Given more recent reports of type III-B systems targeting transcriptionally active DNA (Peng et al., 2015; Estrella et al., 2016), other mechanisms may be at work in *M. xanthus* harboring the transposon insertion.

Another alternative function via self-targeting that appears to involve mRNA degradation allows the pathogen *Pseudomonas aeruginosa* to evade immune detection (**Figure 3**). The type I-F system in *P. aeruginosa* strain UCBPP-PA14 encodes one spacer within its CRISPR1 array that bears partial complementarity to the chromosomally encoded *lasR* gene. LasR is a bacterial quorum sensing regulator whose regulon includes virulence-associated factors presumably detected through Toll-like receptor 4 in mammals. The self-targeting spacer did not lead to any detectable cleavage of the chromosomal DNA but instead appeared to cleave the lasR mRNA. Downregulation of this receptor in turn led to a reduced pro-inflammatory response. The suspected target within the lasR mRNA spans 12 nts, with one internal mismatch and base pairs with the 3′ end of the spacer. Mutational analysis further revealed that disrupting a 5′-GGN-3′ sequence immediately upstream of the lasR target as well as the following 8 base pairs blocked mRNA target degradation (Li R. et al., 2016).

As a brief follow-up to this study, Müller-Esparza and Randau searched for other potentially targeted mRNAs within the *P. aeruginosa* UCBPP-PA14 strain based on potential target sites that include the upstream 5′-GGN-3′ sequence followed by nine complementary nts. They

could identify 189 putative targeted mRNAs, suggesting that additional requirements such as mRNA secondary structure are needed for mRNA targeting. Therefore, further studies are necessary to clarify the requirements for mRNA degradation by the type I-F CRISPR-Cas system in this strain of *P. aeruginosa* and the many other organisms encoding these systems (Müller-Esparza and Randau, 2017).

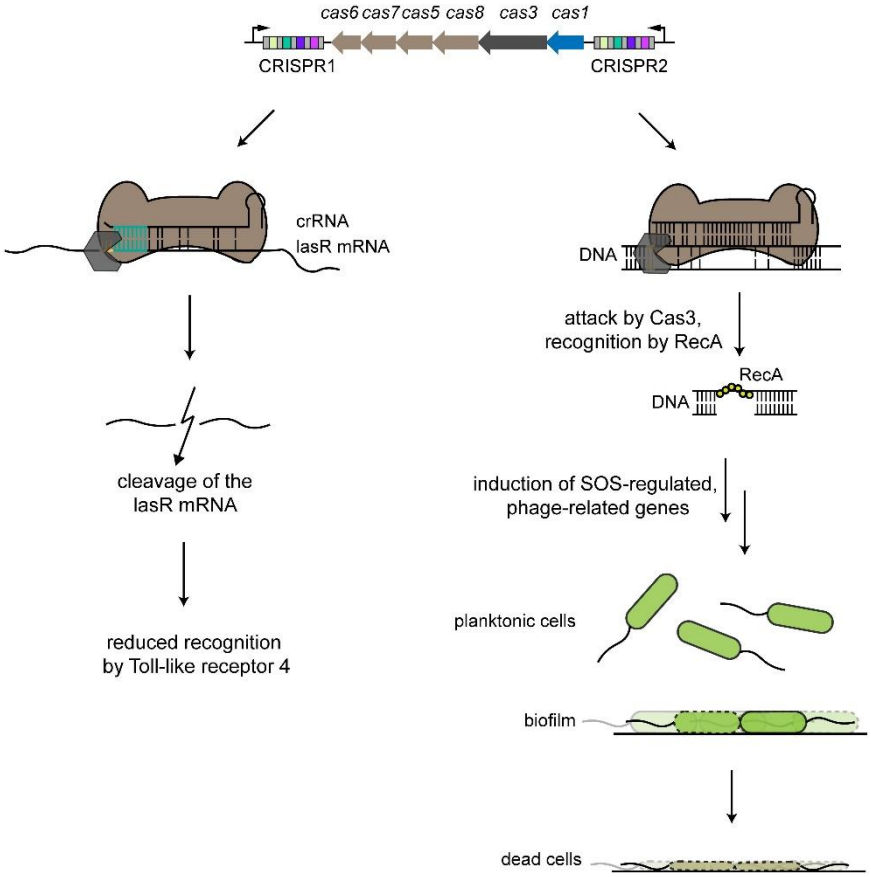


Figure 3: Examples of alternative CRISPR-Cas functions. The type I-F CRISPR-Cas system in *P. aeruginosa* harbors two CRISPR arrays that account for two different alternative functions. The left side shows partial binding between a crRNA derived from the CRISPR1 array and the lasR mRNA, with an indispensable interaction region of 8 nts (turquoise). The Cas effector complex (brown) binds to the target region with an adjacent recognition motif (orange), with some involvement of the Cas3 nuclease. lasR mRNA is then degraded, leading to reduced host recognition by Toll-like receptor 4 during an infection. The right side shows partial binding of a crRNA derived from the CRISPR2 array to a prophage region. Binding by the Cas effector complex recruits Cas3, resulting in nicking of one strand of the target DNA. Recognition by RecA triggers intrinsic processes that lead to induction of SOS-regulated, phage-related genes that lead to cell death of cells specifically forming a biofilm, while planktonic cells are unaffected.

Cas9 is traditionally seen as a DNA-targeting nuclease, yet emerging examples have revealed that some Cas9s can also target RNA (O'Connell et al., 2014; Rousseau et al., 2018; Strutt et al., 2018). Original studies of the Cas9 from *Streptococcus pyogenes* suggested that the effector protein could differentiate between RNA and DNA (Gasiunas et al., 2012), wherein RNA targeting could only be achieved by hybridizing RNA with a PAM-presenting oligonucleotide (PAMmer) (O'Connell et al., 2014; Nelles et al., 2016). Later, it was shown that some Cas9 proteins can cleave RNA even in the absence of a PAMmer. Specifically, the Cas9 from the type II-C system in *Neisseria meningitidis* was shown to cleave RNA *in vitro*, while

Cas9 from the type II-A system in *S. aureus* and the type II-C system in *C. jejuni* were shown to cleave RNA *in vitro* and *in vivo* (Dugar et al., 2018; Rousseau et al., 2018; Strutt et al., 2018). In all of these cases, RNA targeting did not require a flanking recognition motif. In the example from *C. jejuni*, the naturally occurring spacers were shown to bind and, in some cases, drive Cas9-mediated cleavage of endogenous RNAs. These spacers only exhibited partial complementarity to their targets, and the associated DNA sequences were not flanked by recognized PAMs, preventing genome cleavage. Dugar and coworkers did not explicitly identify a phenotype associated with RNA targeting by the endogenous Cas9 (Dugar et al., 2018), although Strutt et al. (2018) demonstrated that the Cas9 from *S. aureus* could inhibit gene expression through programmable RNA targeting in *E. coli* without leading to cell death. The above mentioned examples show that some DNA targeting systems can also target RNA, with the potential for these same systems to modulate gene expression by RNA degradation in their native hosts.

CRISPR-Cas induced DNA damage response

The type I-F CRISPR-Cas system in *P. aeruginosa* UCBPP-PA14 performs a distinct alternative function that induces the SOS response, preventing biofilm formation and impairing swarming motility (Zegans et al., 2009; Cady and O'Toole, 2011). A key factor was the presence of a partial match between a spacer within the CRISPR2 array and a sequence present within the lysogenized phage DSM3 (**Figure 3**). The authors showed that the observed phenotype was dependent not on the presence of the lysogenized phage but rather solely on the target sequence. The presence of the CRISPR-Cas system and the PAM-flanked protospacer led surface-attached cells to undergo cell death, explaining the lack of biofilm formation. The proposed mechanism-of-action involved the recruitment of Cas3 upon binding of the Cascade-crRNA complex to the region of partial complementarity, which recruited RecA and activated the SOS response upon nicking of one DNA strand. Activated RecA also triggered a pathway that led to accumulation of phage-related genes that induced cell death upon surface attachment (Heussler et al., 2015). The ensuing questions are whether this same phenomenon can be found in other biofilm-forming bacteria and whether partial genome targeting can induce other phenotypes.

Transcriptional regulation

Beyond RNA targeting, CRISPR-Cas systems have the potential to regulate transcription through partial spacer complementarity or due to the presence of an inactivated nuclease (Sampson et al., 2013, 2019; Ratner et al., 2019). Partial complementarity resulted in regulation of transcription in *Francisella novicida* by so-called scaRNAs (small CRISPR/Cas-

associated RNAs). ScaRNAs were encoded close to the CRISPR array associated with the type II CRISPR-Cas system in *F. novicida*. Strictly speaking, the scaRNA-based mechanism is not dependent on a self-targeting spacer but rather on the scaRNA acting as a crRNA. Originally it was hypothesized that the scaRNA targets RNA (Sampson et al., 2013), but later it was shown that the scaRNA hybridizes with the tracrRNA and directs Cas9 to the partially complementary 5′ UTR of its endogenous DNA targets. DNA binding of the target results in transcriptional repression (Ratner et al., 2019; Sampson et al., 2019). In the case of *F. novicida*, targeting with the scaRNA-tracrRNA-Cas9 complex resulted in transcriptional repression of four genes contributing to its virulence by facilitating evasion from immune detection. DNA cleavage by Cas9 is prevented through only partial complementarity of the scaRNA to the target site (Ratner et al., 2019).

Aside from transcriptional repression by DNA binding near promoter regions, another means to regulate transcription is through disruption of the Cas nuclease's active site. This phenomenon can occur in type I systems that lack the effector protein Cas3 but have an intact Cascade complex (Luo et al., 2015). It is also possible to disrupt the nucleolytic activity of a Cas effector protein by mutating the active site. For example, alanine substitutions in the HNH and RuvC domains in the single effectors Cas9 or Cas12a result in a catalytically dead protein that can bind a target but not cleave it (Bikard et al., 2013; Qi et al., 2013; Leenay et al., 2016). While mutations that solely inactivate cleavage are much less likely than deleterious mutations to the nuclease, either means would result in CRISPR machinery that tightly binds DNA, thereby blocking transcription. Natural examples of catalytically dead CRISPR-Cas systems acting as gene regulators have not been reported, although the ease in disrupting *cas3* in the highly prevalent type I systems would suggest that nature has regularly sampled this alternative function. Screening for CRISPR-Cas systems harboring inactive nucleases and self-targeting spacers or spacers with partial complementarity to the genome might lead to the discovery of further CRISPR-based gene regulatory systems.

Invaders co-opting CRISPR-Cas self-targeting

There is also evidence of foreign invaders co-opting CRISPR-Cas systems to either promote the spread of MGEs or weaken the host's adaptive immunity through self-targeting spacers. Recent publications described CRISPR-Cas systems associated with Tn7-like transposons that led to spacer-directed insertion of the transposon (Peters et al., 2017; Klompe et al., 2019; Strecker et al., 2019). The transposon portion of the system generally consists of *tnsB*, *tsnC*, and *tniQ* (a *tnsD* homolog), yet it lacks *tnsD* and *tnsE* normally responsible for recognition of the attachment site (Waddell and Craig, 1988, 1989). Instead, the CRISPR-Cas portion of the system, which lacks nuclease activity and the acquisition machinery, directs transposon

insertion up to ~80 nts downstream of the target site. Because the target site is preserved, an integrated CRISPR transposon would inherently encode a self-targeting spacer (Klompe et al., 2019; Strecker et al., 2019). Nevertheless, the self-targeting spacer appears to be no longer functional due to the lack of multiple transposon insertions at the same target site (Strecker et al., 2019). Acquiring different spacers targeting within the bacteria's genome would allow the transposon to insert itself elsewhere in the genome, although it is not known how new spacers can be acquired due to the lack of acquisition machinery.

Beyond transposons, phages and viruses also represent types of mobile genetic elements that have co-opted CRISPR-Cas systems for their own purposes. It is reported that some phages or viruses harbor at least parts of CRISPR-Cas systems (Seed et al., 2013; Hooton and Connerton, 2014; Krupovic et al., 2015; Levasseur et al., 2016; Hudaiberdiev et al., 2017; Naser et al., 2017; Dou et al., 2018). One noteworthy example comes from the lysogenic CP8/CP30A phage in *C. jejuni* described earlier. This phage encodes a *cas4*-like gene that is responsible for spacer acquisition within the type II-C CRISPR-Cas system targeting the host's genome. The authors hypothesized that these self-targeting spacers might provide a benefit for the phage infecting *C. jejuni* and assist in phage-mediated escape from CRISPR attack (Hooton and Connerton, 2014). Phages and viruses could escape from the host immune system by forcing the organism to use its endogenous CRISPR-Cas system for autoimmunity rather than for attacking viral invaders. Furthermore, the organism might mutate or delete its CRISPR-Cas system to prevent cell death and with this also lose the ability to target invading phages or viruses. In total, these examples show that the host and its invaders can utilize CRISPR-Cas systems and their encoded self-targeting spacers for different purposes.

Conclusion and future perspectives

Self-targeting spacers occur surprisingly often in nature, albeit less frequently than spacers matching sequences from known phages, viruses or plasmids. The apparent paradox between the presence of these spacers and their presumed autoimmunity can be resolved in two general ways. These spacers could represent less frequent but important biological "accidents" that compel cells to reduce or eliminate the impact of self-targeting. Alternatively, the cells could be actively using these self-targeting spacers for other purposes that extend beyond adaptive immunity. Both have been reported in the literature, with only a few examples of the latter. However, alternative functions through self-targeting spacers represent an underexplored area of research in CRISPR biology that could yield exciting new insights and tools. Below, we describe multiple opportunities for future research to uncover further instances of alternative functions, advance our understanding of CRISPR biology and evolution, and expand the available toolbox of CRISPR technologies.

One potential focus of future work is on CRISPR-Cas systems encoding multiple self-targeting spacers or on organisms encoding multiple CRISPR-Cas systems. A few examples of bacteria and archaea encoding self-targeting spacers have been reported (Stern et al., 2010) but never explored experimentally. While these examples were categorized as non-effective targeting due to the lack of an apparent PAM, mutated adjacent repeats, extended base pairing with the repeat or lack of some *cas* genes, these sequences could lead to some level of targeting. For instance, CRISPR nucleases are increasingly known to recognize non-canonical PAM sequences (Leenay and Beisel, 2017), and the absence of some *cas* genes could still allow some functions. The accumulation of multiple self-targeting spacers would also suggest a positive selective pressure. One exception could be the disruption of all but Cas1 and Cas2, possibly resulting in acquisition without negative selection against self-targeting spacers. The occurrence of prokaryotes with multiple CRISPR-Cas systems suggests the possibility that some systems could fulfill the canonical CRISPR function as an adaptive immune system and the others might perform alternative functions.

Another potential focus of future work is identifying spacers exhibiting partial complementarity to the host's genome. As described above, many CRISPR-Cas systems can still bind but not cleave partially complementary targets, resulting in transcriptional repression. Partial complementarity would also allow RNA targeting by some effector proteins, potentially allowing post-transcriptional regulation of endogenous genes. Standard searches for protospacers readily exclude partially matching sequences, owing in part to the difficulty in eliminating false positives. However, regardless of the source of these spacers, partial complementarity with the genome could drive alternative functions. More work is needed to understand what types of mismatches allow different CRISPR-Cas systems to bind but not cleave their targets. This information could then be fed into search algorithms tasked with identifying targets as potential sources of CRISPR-Cas systems moonlighting as gene regulators.

Anti-CRISPR proteins could also provide a potential source for alternative functions. As described above, one strategy to find new anti-CRISPR proteins is to identify organisms with self-targeting spacers (Rauch et al., 2017; Watters et al., 2018). However, the search could be reversed: identifying organisms that harbor both Acrs and CRISPR-Cas systems as potential candidates for identifying systems exhibiting alternative functions. For instance, an encoded Acr that blocks cleavage but not binding activity of the nuclease could convert the immune system into a transcriptional regulator (Pawluk et al., 2018). Discovering new Acrs still remains a major challenge, although further discoveries will enable the search for Acrs tied to alternative functions.

Beyond the discovery of novel instances of functions extending beyond adaptive immunity, interrogating how CRISPR-Cas systems exhibit alternative functions and cope with self-targeting continues to open new biotechnological applications. For instance, the recently discovered CRISPR transposons encoding genome-targeting spacers can serve as powerful tools to insert genes (Klompe et al., 2019; Strecker et al., 2019). Genome-targeting spacers have also been used with classical CRISPR-Cas systems to generate large deletions, representing important capabilities for genome engineering and minimization (Jiang et al., 2013; Oh and van Pijkeren, 2014). As there exist other means by which cells can escape autoimmunity, steps may be necessary to ensure target deletion is the predominant mode of escape. Beyond genome editing, self-targeting with endogenous CRISPR-Cas systems can be part of programmable gene regulation. The endogenous system can be rendered cleavage-deficient while preserving DNA binding activity (Luo et al., 2015). Efforts to interrogate escape from self-targeting have also revealed that gene regulation can be achieved without altering the endogenous system, such as by employing Acrs that inhibit cleavage activity but not DNA binding or by expressing partially complementary spacers. Finally, insights into self-targeting lend to employing endogenous CRISPR-Cas systems as programmable antimicrobials. If the endogenous system is fully active, self-targeting spacers can be used to kill specific bacteria (Bikard et al., 2014; Citorik et al., 2014; Gomaa et al., 2014). If the endogenous system is inhibited by an Acr, relieving expression or activity of these Acrs could unleash lethal autoimmunity, particularly if the endogenous system acquired self-targeting spacers. Further efforts to discover and elucidate new alternative functions could inspire the next generation of CRISPR technologies, emphasizing the need to further investigate the role of self-targeting CRISPR-Cas systems.

References

Abudayyeh, O. O., Gootenberg, J. S., Konermann, S., Joung, J., Slaymaker, I. M., Cox, D. B. T., et al. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science 353:aaf5573. doi: 10.1126/science.aaf5573

Akhter, S., Aziz, R. K., and Edwards, R. A. (2012). PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. Nucleic Acids Res. 40:e126. doi: 10.1093/nar/gks406

Aklujkar, M., and Lovley, D. R. (2010). Interference with histidyl-tRNA synthetase by a CRISPR spacer sequence as a factor in the evolution of *Pelobacter carbinolicus*. BMC Evol. Biol. 10:230. doi: 10.1186/1471-2148-10-230

Aravind, L., and Koonin, E. V. (2001). Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. Genome Res. 11, 1365–1374. doi: 10.1101/gr.181001

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709–1712. doi: 10.1126/science.1138140

Behler, J., Sharma, K., Reimann, V., Wilde, A., Urlaub, H., and Hess, W. R. (2018). The host-encoded RNase E endonuclease as the crRNA maturation enzyme in a CRISPR-Cas subtype III-Bv system. Nat. Microbiol. 3, 367–377. doi: 10.1038/s41564-017-0103-5

Benda, C., Ebert, J., Scheltema, R. A., Schiller, H. B., Baumgärtner, M., Bonneau, F., et al. (2014). Structural model of a CRISPR RNA-silencing complex reveals the RNA-target cleavage activity in Cmr4. Mol. Cell 56, 43–54. doi: 10.1016/j.molcel.2014.09.002

Bernheim, A., Calvo-Villamañán, A., Basier, C., Cui, L., Rocha, E. P. C., Touchon, M., et al. (2017). Inhibition of NHEJ repair by type II-A CRISPR-Cas systems in bacteria. Nat. Commun. 8:2094. doi: 10.1038/s41467-017-02350-1

Bhoobalan-Chitty, Y., Johansen, T. B., Di Cianni, N., and Peng, X. (2019). Inhibition of type III CRISPR-Cas immunity by an archaeal virus-encoded anti-CRISPR protein. Cell 179:448-458.e11. doi: 10.1016/j.cell.2019.09.003

Bikard, D., Euler, C. W., Jiang, W., Nussenzweig, P. M., Goldberg, G. W., Duportet, X., et al. (2014). Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. Nat. Biotechnol. 32, 1146–1150. doi: 10.1038/nbt.3043

Bikard, D., Hatoum-Aslan, A., Mucida, D., and Marraffini, L. A. (2012). CRISPR interference can prevent natural transformation and virulence acquisition during *in vivo* bacterial infection. Cell Host Microbe 12, 177–186. doi: 10.1016/j.chom.2012.06.003

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L. A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res. 41, 7429–7437. doi: 10.1093/nar/gkt520

Bobay, L.-M., Touchon, M., and Rocha, E. P. C. (2013). Manipulating or superseding host recombination functions: a dilemma that shapes phage evolvability. PLoS Genetics 9:e1003825. doi: 10.1371/journal.pgen.1003825

Bolotin, A., Quinquis, B., Renault, P., Sorokin, A., Ehrlich, S. D., Kulakauskas, S., et al. (2004). Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. Nat. Biotechnol. 22, 1554–1558. doi: 10.1038/nbt1034

Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151, 2551–2561. doi: 10.1099/mic.0.28048-0

Bondy-Denomy, J., Garcia, B., Strum, S., Du, M., Rollins, M. F., Hidalgo-Reyes, Y., et al. (2015). Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins. Nature 526, 136–139. doi: 10.1038/nature15254

Bondy-Denomy, J., Pawluk, A., Maxwell, K. L., and Davidson, A. R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature 493, 429–432. doi: 10.1038/nature11723

Bowater, R., and Doherty, A. J. (2006). Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. PLoS Genet. 2:e8. doi: 10.1371/journal.pgen.0020008

Brodt, A., Lurie-Weinberger, M. N., and Gophna, U. (2011). CRISPR loci reveal networks of gene exchange in archaea. Biol. Direct 6:65. doi: 10.1186/1745-6150-6-65

Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., et al. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321, 960–964. doi: 10.1126/science.1159689

Cady, K. C., and O'Toole, G. A. (2011). Non-identity-mediated CRISPR-bacteriophage interaction mediated via the Csy and Cas3 proteins. J. Bacteriol. 193, 3433–3445. doi: 10.1128/JB.01411-10

Cañez, C., Selle, K., Goh, Y. J., and Barrangou, R. (2019). Outcomes and characterization of chromosomal self-targeting by native CRISPR-Cas systems in *Streptococcus thermophilus*. FEMS Microbiol. Lett. 366:fnz105. doi: 10.1093/femsle/fnz105

Carte, J., Wang, R., Li, H., Terns, R. M., and Terns, M. P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. Genes Dev. 22, 3489–3496. doi: 10.1101/gad.1742908

Chayot, R., Montagne, B., Mazel, D., and Ricchetti, M. (2010). An end-joining repair mechanism in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. 107, 2141–2146. doi: 10.1073/pnas.0906355107

Citorik, R. J., Mimee, M., and Lu, T. K. (2014). Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. Nat. Biotechnol. 32, 1141–1145. doi: 10.1038/nbt.3011

Cui, L., and Bikard, D. (2016). Consequences of Cas9 cleavage in the chromosome of *Escherichia coli*. Nucleic Acids Res. 44, 4243–4251. doi: 10.1093/nar/gkw223

Datsenko, K. A., Pougach, K., Tikhonov, A., Wanner, B. L., Severinov, K., and Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat. Commun. 3:945. doi:10.1038/ncomms1937

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., et al. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602–607. doi: 10.1038/nature09886

Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Fremaux, C., Boyaval, P., et al. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J. Bacteriol. 190, 1390–1400. doi: 10.1128/jb.01412-07

Dou, C., Yu, M., Gu, Y., Wang, J., Yin, K., Nie, C., et al. (2018). Structural and mechanistic analyses reveal a unique Cas4-like protein in the mimivirus virophage resistance element system. iScience 3, 1–10. doi: 10.1016/j.isci.2018.04.001

Dugar, G., Leenay, R. T., Eisenbart, S. K., Bischler, T., Aul, B. U., Beisel, C. L., et al. (2018). CRISPR RNA-dependent binding and cleavage of endogenous RNAs by the *Campylobacter jejuni* Cas9. Mol. Cell 69:893-905.e7. doi: 10.1016/j.molcel.2018.01.032

Dy, R. L., Pitman, A. R., and Fineran, P. C. (2013). Chromosomal targeting by CRISPR-Cas systems can contribute to genome plasticity in bacteria. Mob. Genet. Elements 3:e26831. doi: 10.4161/mge.26831

El Karoui, M., Biaudet, V., Schbath, S., and Gruss, A. (1999). Characteristics of Chi distribution on different bacterial genomes. Res. Microbiol. 150, 579–587. doi: 10.1016/s0923-2508(99)00132-1

Elmore, J. R., Sheppard, N. F., Ramia, N., Deighan, T., Li, H., Terns, R. M., et al. (2016). Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR–Cas system. Genes Dev. 30, 447–459. doi: 10.1101/gad.272153.115

Estrella, M. A., Kuo, F.-T., and Bailey, S. (2016). RNA-activated DNA cleavage by the Type III-B CRISPR–Cas effector complex. Genes Dev. 30, 460–470. doi: 10.1101/gad.273722.115

Fineran, P. C., Gerritzen, M. J. H., Suárez-Diez, M., Künne, T., Boekhorst, J., van Hijum, S. A. F. T., et al. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. Proc. Natl. Acad. Sci. U.S.A. 111, E1629–E1638. doi: 10.1073/pnas.1400071111

Friedman, S. A., and Hays, J. B. (1986). Selective inhibition of *Escherichia coli* RecBC activities by plasmid-encoded GamS function of phage lambda. Gene 43, 255–263. doi: 10.1016/0378-1119(86)90214-3

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. U.S.A. 109, E2579–E2586.

Geer, L. Y.,Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., et al. (2010). The NCBI BioSystems database. Nucleic Acids Res. 38, D492–D496. doi: 10.1093/nar/gkp858

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. doi: 10.1038/nature13637

Gomaa, A. A., Klumpe, H. E., Luo, M. L., Selle, K., Barrangou, R., and Beisel, C. L. (2014). Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. mBio 5:e928-13. doi: 10.1128/mBio.00928-13

Gong, C., Bongiorno, P., Martins, A., Stephanou, N. C., Zhu, H., Shuman, S., et al. (2005). Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. Nat. Struct. Mol. Biol.12, 304–312. doi: 10.1038/nsmb915

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics 8:172. doi: 10.1186/1471-2105-8-172

Guan, J., Wang, W., and Sun, B. (2017). Chromosomal targeting by the type III-A CRISPR-Cas system can reshape genomes in *Staphylococcus aureus*. mSphere 2:e403-17. doi: 10.1128/msphere.00403-17

Hale, C. R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., et al. (2012). Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. Mol. Cell 45, 292–302. doi: 10.1016/j.molcel.2011.10.023

Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., et al. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell 139, 945–956. doi: 10.1016/j.cell.2009.07.040

He, F., Bhoobalan-Chitty, Y., Van, L. B., Kjeldsen, A. L., Dedola, M., Makarova, K. S., et al. (2018). Anti-CRISPR proteins encoded by archaeal lytic viruses inhibit subtype I-D immunity. Nat.Microbiol. 3, 461–469. doi: 10.1038/s41564-018-0120-z

Heler, R., Samai, P., Modell, J. W., Weiner, C., Goldberg, G. W., Bikard, D., et al. (2015). Cas9 specifies functional viral targets during CRISPR–Cas adaptation. Nature 519, 199–202. doi: 10.1038/nature14245

Heussler, G. E., Cady, K. C., Koeppen, K., Bhuju, S., Stanton, B. A., and O'Toole, G. A. (2015). Clustered regularly interspaced short palindromic repeat-dependent, biofilm-specific death of *Pseudomonas aeruginosa* mediated by increased expression of phage-related genes. mBio 6:e129-15. doi: 10.1128/mBio.00129-15

Heussler, G. E., and O'Toole, G. A. (2016). Friendly fire: biological functions and consequences of chromosomal targeting by CRISPR-Cas systems. J. Bacteriol. 198, 1481–1486. doi: 10.1128/JB.00086-16

Hooton, S. P. T., and Connerton, I. F. (2014). *Campylobacter jejuni* acquire new host-derived CRISPR spacers when in association with bacteriophages harboring a CRISPR-like Cas4 protein. Front. Microbiol. 5:744. doi: 10.3389/fmicb.2014.00744

Horvath, P., Coûté-Monvoisin, A.-C., Romero, D. A., Boyaval, P., Fremaux, C., and Barrangou, R. (2009). Comparative analysis of CRISPR loci in lactic acid bacteria genomes. Int. J. Food Microbiol. 131, 62–70. doi: 10.1016/j.ijfoodmicro.2008.05.030

Horvath, P., Romero, D. A., Coûté-Monvoisin, A.-C., Richards, M., Deveau, H., Moineau, S., et al. (2008). Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. J. Bacteriol. 190, 1401–1412. doi: 10.1128/jb.01415-07

Hudaiberdiev, S., Shmakov, S., Wolf, Y. I., Terns, M. P., Makarova, K. S., and Koonin, E. V. (2017). Phylogenomics of Cas4 family nucleases. BMC Evol. Biol. 17:232. doi: 10.1186/s12862-017-1081-1

Hynes, A. P., Rousseau, G. M., Lemay, M.-L., Horvath, P., Romero, D. A., Fremaux, C., et al. (2017). An anti-CRISPR from a virulent streptococcal phage inhibits *Streptococcus pyogenes* Cas9. Nat.Microbiol. 2, 1374–1380. doi: 10.1038/s41564-017-0004-7

Jackson, S. A., Birkholz, N., Malone, L. M., and Fineran, P. C. (2019). Imprecise spacer acquisition generates CRISPR-Cas immune diversity through primed adaptation. Cell Host Microbe 25:250-260.e4. doi: 10.1016/j.chom.2018.12.014

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. 31, 233–239. doi: 10.1038/nbt.2508

Johnson, A. D., Poteete, A. R., Lauer, G., Sauer, R. T., Ackers, G. K., and Ptashne, M. (1981). l Repressor and cro—components of an efficient molecular switch. Nature 294, 217–223. doi: 10.1038/294217a0

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. doi: 10.1016/j.str.2015.01.021

Kazlauskiene, M., Kostiuk, G., Venclovas, ˇ C, Tamulaitis, G., and Siksnys, V. (2017). A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. Science 357, 605–609. doi: 10.1126/science.aao0100

Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, ˇ C, and Siksnys, V. (2016). Spatiotemporal control of type III-A CRISPR-Cas immunity: coupling DNA degradation with the target RNA recognition. Mol. Cell 62, 295–306. doi: 10.1016/j.molcel.2016.03.024

Kersey, P. J., Allen, J. E., Allot, A., Barba, M., Boddu, S., Bolt, B. J., et al. (2018). Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. Nucleic Acids Res. 46, D802–D808. doi: 10.1093/nar/gkx1011

Kieper, S. N., Almendros, C., Behler, J., McKenzie, R. E., Nobrega, F. L., Haagsma, A. C., et al. (2018). Cas4 facilitates PAM-compatible spacer selection during CRISPR adaptation. Cell Rep. 22, 3377–3384. doi: 10.1016/j.celrep.2018.02.103

Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S., and Sternberg, S. H. (2019). Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. Nature 571, 219–225. doi: 10.1038/s41586-019-1323-z

Kojima, K. K., and Kanehisa, M. (2008). Systematic survey for novel types of prokaryotic retroelements based on gene neighborhood and protein architecture. Mol. Biol. Evol. 25, 1395–1404. doi: 10.1093/molbev/msn081

Koonin, E. V., and Makarova, K. S. (2019). Origins and evolution of CRISPR-Cas systems. Philos. Trans. R. Soc. Lond. B Biol. Sci. 374:20180087. doi: 10.1098/rstb.2018.0087

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

Krupovic, M., Cvirkaite-Krupovic, V., Prangishvili, D., and Koonin, E. V. (2015). Evolution of an archaeal virus nucleocapsid protein from the CRISPR-associated Cas4 nuclease. Biol. Direct. 10:65. doi: 10.1186/s13062-015-0093-2

Künne, T., Kieper, S. N., Bannenberg, J. W., Vogel, A. I. M., Miellet, W. R., Klein, M., et al. (2016). Cas3-derived target DNA degradation fragments fuel primed CRISPR adaptation. Mol. Cell 63, 852–864. doi: 10.1016/j.molcel.2016.07.011

Kuzminov, A. (2001). Single-strand interruptions in replicating chromosomes cause double-strand breaks. Proc. Natl. Acad. Sci. U.S.A. 98, 8241–8246. doi: 10.1073/pnas.131009198

Lee, H., Dhingra, Y., and Sashital, D. G. (2019). The Cas4-Cas1-Cas2 complex mediates precise prespacer processing during CRISPR adaptation. eLife 8:e44248. doi: 10.7554/eLife.44248

Lee, H., Zhou, Y., Taylor, D. W., and Sashital, D. G. (2018). Cas4-dependent prespacer processing ensures high-fidelity programming of CRISPR arrays. Mol. Cell 70:48-59.e5. doi: 10.1016/j.molcel.2018.03.003

Leenay, R. T., and Beisel, C. L. (2017). Deciphering, communicating, and engineering the CRISPR PAM. J. Mol. Biol. 429, 177–191. doi: 10.1016/j.jmb.2016.11.024

Leenay, R. T.,Maksimchuk, K. R., Slotkowski, R. A., Agrawal, R. N., Gomaa, A. A., Briner, A. E., et al. (2016). Identifying and visualizing functional PAM diversity across CRISPR-Cas Systems. Mol. Cell 62, 137–147. doi: 10.1016/j.molcel.2016.02.031

Lemak, S., Beloglazova, N., Nocek, B., Skarina, T., Flick, R., Brown, G., et al. (2013). Toroidal structure and DNA cleavage by the CRISPR-associated [4Fe-4S] cluster containing Cas4 nuclease SSO0001 from *Sulfolobus solfataricus*. J. Am. Chem. Soc. 135, 17476–17487. doi: 10.1021/ja408729b

Levasseur, A., Bekliz, M., Chabrière, E., Pontarotti, P., La Scola, B., and Raoult, D. (2016). MIMIVIRE is a defence system in Mimivirus that confers resistance to virophage. Nature 531, 249–252. doi: 10.1038/nature17146

Levy, A., Goren, M. G., Yosef, I., Auster, O., Manor, M., Amitai, G., et al. (2015). CRISPR adaptation biases explain preference for acquisition of foreign DNA. Nature 520, 505–510. doi: 10.1038/nature14302

Li, R., Fang, L., Tan, S., Yu, M., Li, X., He, S., et al. (2016). Type I CRISPR-Cas targets endogenous genes and regulates virulence to evade mammalian host immunity. Cell Res. 26, 1273–1287. doi: 10.1038/cr.2016.135

Li, Y., Pan, S., Zhang, Y., Ren, M., Feng, M., Peng, N., et al. (2016). Harnessing Type I and Type III CRISPR-Cas systems for genome editing. Nucleic Acids Res. 44:e34. doi: 10.1093/nar/gkv1044

Liu, L., Li, X., Ma, J., Li, Z., You, L., Wang, J., et al. (2017a). The molecular architecture for RNA-guided RNA cleavage by Cas13a. Cell 170:714-726.e10. doi: 10.1016/j.cell.2017.06.050

Liu, L., Li, X., Wang, J., Wang, M., Chen, P., Yin, M., et al. (2017b). Two distant catalytic sites are responsible for C2c2 RNase activities. Cell 168:121-134.e10. doi: 10.1016/j.cell.2016.12.031

Liu, T., Liu, Z., Ye, Q., Pan, S., Wang, X., Li, Y., et al. (2017). Coupling transcriptional activation of CRISPR-Cas system and DNA repair genes by Csa3a in *Sulfolobus islandicus*. Nucleic Acids Res. 45, 8978–8992. doi: 10.1093/nar/gkx612

Liu, T., Li, Y., Wang, X., Ye, Q., Li, H., Liang, Y., et al. (2015). Transcriptional regulator-mediated activation of adaptation genes triggers CRISPR *de novo* spacer acquisition. Nucleic Acids Res. 43, 1044–1055. doi: 10.1093/nar/gku1383

Luo, M. L., Mullis, A. S., Leenay, R. T., and Beisel, C. L. (2015). Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. Nucleic Acids Res. 43, 674–681. doi: 10.1093/nar/gku971

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. Nat. Rev. Microbiol. 13, 722–736. doi: 10.1038/nrmicro3569

Marino, N. D., Zhang, J. Y., Borges, A. L., Sousa, A. A., Leon, L. M., Rauch, B. J., et al. (2018). Discovery of widespread type I and type V CRISPR-Cas inhibitors. Science 362, 240–242. doi: 10.1126/science.aau5174

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

Marraffini, L. A., and Sontheimer, E. J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature 463, 568–571. doi: 10.1038/nature08703

Meeske, A. J., and Marraffini, L. A. (2018). RNA guide complementarity prevents self-targeting in type VI CRISPR systems. Mol. Cell 71:791-801.e3. doi: 10.1016/j.molcel.2018.07.013

Meeske, A. J., Nakandakari-Higa, S., and Marraffini, L. A. (2019). Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. Nature 570, 241–245. doi: 10.1038/s41586-019-1257-5

Michel, B., Flores, M. J., Viguera, E., Grompone, G., Seigneur, M., and Bidnenko, V. (2001). Rescue of arrested replication forks by homologous recombination. Proc. Natl. Acad. Sci. U.S.A. 98, 8181–8188. doi: 10.1073/pnas.111008798

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. 60, 174–182. doi: 10.1007/s00239-004-0046-3

Müller-Esparza, H., and Randau, L. (2017). Commentary: type I CRISPR-Cas targets endogenous genes and regulates virulence to evade mammalian host immunity. Front. Microbiol. 8:319.

Murphy, K. C., and Lewis, L. J. (1993). Properties of *Escherichia coli* expressing bacteriophage P22 Abc (anti-RecBCD) proteins, including inhibition of Chi activity. J. Bacteriol. 175, 1756–1766. doi: 10.1128/jb.175.6.1756-1766.1993

Naser, I. B., Hoque, M. M., Nahid, M. A., Tareq, T. M., Rocky, M. K., and Faruque, S. M. (2017). Analysis of the CRISPR-Cas system in bacteriophages active on epidemic strains of *Vibrio cholerae* in Bangladesh. Sci. Rep. 7:14880. doi: 10.1038/s41598-017-14839-2

Nelles, D. A., Fang, M. Y., O'Connell, M. R., Xu, J. L., Markmiller, S. J., Doudna, J. A., et al. (2016). Programmable RNA tracking in live cells with CRISPR/Cas9. Cell 165, 488–496. doi: 10.1016/j.cell.2016.02.054

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 Biol. 16, 566–576. doi: 10.1080/15476286.2018.1509662

Niewoehner, O., Garcia-Doval, C., Rostøl, J. T., Berk, C., Schwede, F., Bigler, L., et al. (2017). Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. Nature 548, 543–548. doi: 10.1038/nature23467

Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., and Doudna, J. A. (2014). Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. Nat. Struct. Mol. Biol. 21, 528–534. doi: 10.1038/nsmb.2820

O'Connell, M. R., Oakes, B. L., Sternberg, S. H., East-Seletsky, A., Kaplan, M., and Doudna, J. A. (2014). Programmable RNA recognition and cleavage by CRISPR/Cas9. Nature 516, 263–266. doi: 10.1038/nature13769

Oh, J.-H., and van Pijkeren, J.-P. (2014). CRISPR–Cas9-assisted recombineering in *Lactobacillus reuteri*. Nucleic Acids Res. 42:e131-31. doi: 10.1093/nar/gku623

Paez-Espino, D., Morovic, W., Sun, C. L., Thomas, B. C., Ueda, K.-I., Stahl, B., et al. (2013). Strong bias in the bacterial CRISPR elements that confer immunity to phage. Nat. Commun. 4:1430. doi: 10.1038/ncomms2440

Paez-Espino, D., Sharon, I., Morovic, W., Stahl, B., Thomas, B. C., Barrangou, R., et al. (2015). CRISPR immunity drives rapid phage genome evolution in *Streptococcus thermophilus*. mBio 6:230. doi: 10.1128/mBio.00262-15

Pawluk, A., Amrani, N., Zhang, Y., Garcia, B., Hidalgo-Reyes, Y., Lee, J., et al. (2016a). Naturally occurring off-switches for CRISPR-Cas9. Cell 167:1829-1838.e9. doi: 10.1016/j.cell.2016.11.017

Pawluk, A., Staals, R. H. J., Taylor, C., Watson, B. N. J., Saha, S., Fineran, P. C., et al. (2016b). Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. Nat. Microbiol. 1:16085. doi: 10.1038/nmicrobiol.2016.85

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

Pawluk, A., Davidson, A. R., and Maxwell, K. L. (2018). Anti-CRISPR: discovery, mechanism and function. Nat. Rev. Microbiol. 16, 12–17. doi: 10.1038/nrmicro.2017.120

Peng, W., Feng, M., Feng, X., Liang, Y. X., and She, Q. (2015). An archaeal CRISPR type III-B system exhibiting distinctive RNA targeting features and mediating dual RNA and DNA interference. Nucleic Acids Res. 43, 406–417. doi: 10.1093/nar/gku1302

Peters, J. E., Makarova, K. S., Shmakov, S., and Koonin, E. V. (2017). Recruitment of CRISPR-Cas systems by Tn7-like transposons. Proc. Natl. Acad. Sci. U.S.A. 114, E7358–E7366. doi: 10.1073/pnas.1709035114

Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology 151, 653–663. doi: 10.1099/mic.0.27437-0

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., et al. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152, 1173–1183. doi: 10.1016/j.cell.2013.02.022

Ratner, H. K., Escalera-Maurer, A., Le Rhun, A., Jaggavarapu, S., Wozniak, J. E., Crispell, E. K., et al. (2019). Catalytically active Cas9 mediates transcriptional interference to facilitate bacterial virulence. Mol. Cell 75:498-510.e5. doi: 10.1016/j.molcel.2019.05.029

Rauch, B. J., Silvis, M. R., Hultquist, J. F., Waters, C. S., McGregor, M. J., Krogan, N. J., et al. (2017). Inhibition of CRISPR-Cas9 with bacteriophage proteins. Cell 168:150-158.e10. doi: 10.1016/j.cell.2016.12.009

Richter, C., Dy, R. L., McKenzie, R. E., Watson, B. N. J., Taylor, C., Chang, J. T., et al. (2014). Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. Nucleic Acids Res. 42, 8516–8526. doi: 10.1093/nar/gku527

Richter, H., Lanthier, M., Nevin, K. P., and Lovley, D. R. (2007). Lack of electricity production by *Pelobacter carbinolicus* indicates that the capacity for Fe(III) oxide reduction does not necessarily confer electron transfer ability to fuel cell anodes. Appl. Environ. Microbiol. 73, 5347–5353. doi: 10.1128/aem.00804-07

Rouillon, C., Athukoralage, J. S., Graham, S., Grüschow, S., and White, M. F. (2018). Control of cyclic oligoadenylate synthesis in a type III CRISPR system. eLife 7:e36734.

Rousseau, B. A., Hou, Z., Gramelspacher, M. J., and Zhang, Y. (2018). Programmable RNA cleavage and recognition by a natural CRISPR-Cas9 system from *Neisseria meningitidis*. Mol. Cell 69:906-914.e4. doi: 10.1016/j.molcel.2018.01.025

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. doi: 10.1016/j.cell.2015.04.027

Sampson, T. R., Saroj, S. D., Llewellyn, A. C., Tzeng, Y.-L., and Weiss, D. S. (2013). A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. Nature 497, 254–257. doi: 10.1038/nature12048

Sampson, T. R., Saroj, S. D., Llewellyn, A. C., Tzeng, Y.-L., and Weiss, D. S. (2019). Author Correction: a CRISPR/Cas system mediates bacterial innate immune evasion and virulence. Nature 570, E30–E31. doi: 10.1038/s41586-019-1253-9

Schmidt, F., Cherepkova, M. Y., and Platt, R. J. (2018). Transcriptional recording by CRISPR spacer acquisition from RNA. Nature 562, 380–385. doi: 10.1038/s41586-018-0569-1

Seed, K. D., Lazinski, D. W., Calderwood, S. B., and Camilli, A. (2013). A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. Nature 494, 489–491. doi: 10.1038/nature11927

Selle, K., Klaenhammer, T. R., and Barrangou, R. (2015). CRISPR-based screening of genomic island excision events in bacteria. Proc. Natl. Acad. Sci. U.S.A. 112, 8076–8081. doi: 10.1073/pnas.1508525112

Shmakov, S., Abudayyeh, O. O., Makarova, K. S., Wolf, Y. I., Gootenberg, J. S., Semenova, E., et al. (2015). Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. Mol. Cell 60, 385–397. doi: 10.1016/j.molcel.2015.10.008

Shmakov, S. A., Sitnik, V., Makarova, K. S., Wolf, Y. I., Severinov, K. V., and Koonin, E. V. (2017). The CRISPR spacer space is dominated by sequences from species-specific mobilomes. mBio 8:e1397-17. doi: 10.1128/mBio.01397-17

Shuman, S., and Glickman, M. S. (2007). Bacterial DNA repair by non-homologous end joining. Nat. Rev. Microbiol. 5, 852–861.

Silas, S., Makarova, K. S., Shmakov, S., Páez-Espino, D., Mohr, G., Liu, Y., et al. (2017). On the origin of reverse transcriptase-using CRISPR-Cas systems and their hyperdiverse, enigmatic spacer repertoires. mBio 8:e897-17. doi: 10.1128/mBio.00897-17

Silas, S., Mohr, G., Sidote, D. J., Markham, L. M., Sanchez-Amat, A., Bhaya, D., et al. (2016). Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. Science 351:aad4234. doi: 10.1126/science.aad4234

Simon, D. M., and Zimmerly, S. (2008). A diversity of uncharacterized reverse transcriptases in bacteria. Nucleic Acids Res. 36, 7219–7229. doi: 10.1093/nar/gkn867

Smith, G. R. (2012). How RecBCD enzyme and Chi promote DNA break repair and recombination: a molecular biologist's view. Microbiol. Mol. Biol. Rev. 76, 217–228. doi: 10.1128/MMBR.05026-11

Sorek, R., Lawrence, C. M., and Wiedenheft, B. (2013). CRISPR-mediated adaptive immune systems in bacteria and archaea. Annu. Rev. Biochem. 82, 237–266. doi: 10.1146/annurev-biochem-072911-172315

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

Stachler, A.-E., Turgeman-Grott, I., Shtifman-Segal, E., Allers, T., Marchfelder, A., and Gophna, U. (2017). High tolerance to self-targeting of the genome by the endogenous CRISPR-Cas system in an archaeon. Nucleic Acids Res. 45, 5208–5216. doi: 10.1093/nar/gkx150

Stern, A., Keren, L., Wurtzel, O., Amitai, G., and Sorek, R. (2010). Self-targeting by CRISPR: gene regulation or autoimmunity? Trends Genetics 26, 335–340. doi: 10.1016/j.tig.2010.05.008

Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J. L., Makarova, K. S., Koonin, E. V., et al. (2019). RNA-guided DNA insertion with CRISPR-associated transposases. Science 365, 48–53. doi: 10.1126/science.aax9181

Strutt, S. C., Torrez, R. M., Kaya, E., Negrete, O. A., and Doudna, J. A. (2018). RNA-dependent RNA targeting by CRISPR-Cas9. eLife 7:e32724. doi: 10.7554/eLife.32724

Su, T., Liu, F., Chang, Y., Guo, Q., Wang, J., Wang, Q., et al. (2019). The phage T4 DNA ligase mediates bacterial chromosome DSBs repair as single component non-homologous end joining. Synth Syst. Biotechnol. 4, 107–112. doi: 10.1016/j.synbio.2019.04.001

Swarts, D. C., Mosterd, C., van Passel, M. W. J., and Brouns, S. J. J. (2012). CRISPR interference directs strand specific spacer acquisition. PLoS One 7:e35888. doi: 10.1371/journal.pone.0035888

Tamulaitis, G., Kazlauskiene, M., Manakova, E., Venclovas, ˇ C, Nwokeoji, A. O., Dickman, M. J., et al. (2014). Programmable RNA shredding by the type III-A CRISPR-Cas system of *Streptococcus thermophilus*. Mol. Cell 56, 506–517. doi: 10.1016/j.molcel.2014.09.027

Thavalingam, A., Cheng, Z., Garcia, B., Huang, X., Shah, M., Sun, W., et al. (2019). Inhibition of CRISPR-Cas9 ribonucleoprotein complex assembly by anti-CRISPR AcrIIC2. Nat. Commun. 10:2806. doi: 10.1038/s41467-019-10577-3

Tong, Y., Charusanti, P., Zhang, L., Weber, T., and Lee, S. Y. (2015). CRISPR-Cas9 based engineering of actinomycetal genomes. ACS Synth. Biol. 4, 1020–1029. doi: 10.1021/acssynbio.5b00038

Toro, N., Martínez-Abarca, F., Mestre, M. R., and González-Delgado, A. (2019a). Multiple origins of reverse transcriptases linked to CRISPR-Cas systems. RNA Biol. 16, 1486–1493. doi: 10.1080/15476286.2019.1639310

Toro, N., Mestre, M. R., Martínez-Abarca, F., and González-Delgado, A. (2019b). Recruitment of reverse transcriptase-Cas1 fusion proteins by Type VI-A CRISPR-Cas systems. Front.Microbiol. 10:2160. doi: 10.3389/fmicb.2019.02160

Toro, N., and Nisa-Martínez, R. (2014). Comprehensive phylogenetic analysis of bacterial reverse transcriptases. PLoS One 9:e114083. doi: 10.1371/journal.pone.0114083

Vercoe, R. B., Chang, J. T., Dy, R. L., Taylor, C., Gristwood, T., Clulow, J. S., et al. (2013). Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. PLoS Genet.9:e1003454. doi: 10.1371/journal.pgen.1003454

Waddell, C. S., and Craig, N. L. (1988). Tn7 transposition: two transposition pathways directed by five Tn7-encoded genes. Genes Dev. 2, 137–149. doi: 10.1101/gad.2.2.137

Waddell, C. S., and Craig, N. L. (1989). Tn7 transposition: recognition of the attTn7 target sequence. Proc. Natl. Acad. Sci. U.S.A. 86, 3958–3962. doi: 10.1073/pnas.86.11.3958

Wallace, R. A., Black, W. P., Yang, X., and Yang, Z. (2014). A CRISPR with roles in *Myxococcus xanthus* development and exopolysaccharide production. J. Bacteriol. 196, 4036–4043. doi: 10.1128/JB.02035-14

Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M., et al. (2015). Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems. Cell 163, 840–853. doi: 10.1016/j.cell.2015.10.008

Watters, K. E., Fellmann, C., Bai, H. B., Ren, S. M., and Doudna, J. A. (2018). Systematic discovery of natural CRISPR-Cas12a inhibitors. Science 362, 236–239. doi: 10.1126/science.aau5138

Wei, Y., Terns, R. M., and Terns, M. P. (2015). Cas9 function and host genome sampling in Type II-A CRISPR–Cas adaptation. Genes Dev. 29, 356–361. doi:10.1101/gad.257550.114

Weller, G. R., Kysela, B., Roy, R., Tonkin, L. M., Scanlan, E., Della, M., et al. (2002). Identification of a DNA nonhomologous end-joining complex in bacteria. Science 297, 1686–1689. doi: 10.1126/science.1074584

Westra, E. R., Buckling, A., and Fineran, P. C. (2014). CRISPR–Cas systems: beyond adaptive immunity. Nat. Rev. Microbiol. 12, 317–326. doi: 10.1038/nrmicro3241

Wu, X., Scott, D. A., Kriz, A. J., Chiu, A. C., Hsu, P. D., Dadon, D. B., et al. (2014). Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nat. Biotechnol. 32, 670–676. doi: 10.1038/nbt.2889

Xu, T., Li, Y., Shi, Z., Hemme, C. L., Li, Y., Zhu, Y., et al. (2015). Efficient genome editing in *Clostridium cellulolyticum* via CRISPR-Cas9 Nickase. Appl. Environ. Microbiol. 81, 4423–4431. doi: 10.1128/AEM.00873-15

Yosef, I., Goren, M. G., and Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. Nucleic Acids Res. 40, 5569–5576. doi: 10.1093/nar/gks216

Yosef, I., Shitrit, D., Goren, M. G., Burstein, D., Pupko, T., and Qimron, U. (2013). DNA motifs determining the efficiency of adaptation into the *Escherichia coli* CRISPR array. Proc. Natl. Acad. Sci. U.S.A. 110, 14396–14401. doi: 10.1073/pnas.1300108110

Zegans, M. E., Wagner, J. C., Cady, K. C., Murphy, D. M., Hammond, J. H., and O'Toole, G. A. (2009). Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviors of *Pseudomonas aeruginosa*. J. Bacteriol. 191, 210–219. doi: 10.1128/JB.00797-08

Zerulla, K., Chimileski, S., Näther, D., Gophna, U., Thane Papke, R., and Soppa, J. (2014). DNA as a phosphate storage polymer and the alternative advantages of polyploidy for growth or survival. PLoS One 9:e94819. doi: 10.1371/journal.pone.0094819

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163, 759–771. doi: 10.1016/j.cell.2015.09.038

Zhang, F., Zhao, S., Ren, C., Zhu, Y., Zhou, H., Lai, Y., et al. (2018). CRISPRminer is a knowledge base for exploring CRISPR-Cas systems in microbe and phage interactions. Commun. Biol. 1:180. doi: 10.1038/s42003-018-0184-6

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. doi: 10.1371/journal.pone.0047232

Zhang, Z., Pan, S., Liu, T., Li, Y., and Peng, N. (2019). Cas4 nucleases can effect specific integration of CRISPR spacers. J. Bacteriol. 201:e747-18. doi: 10.1128/JB.00747-18

Chapter 2: A TXTL-Based Assay to Rapidly Identify PAMs for CRISPR-Cas Systems with Multi-Protein Effector Complexes

The content of this chapter was previously published and reproduced with permission from Springer Nature AG & Co. KGaA.

Wimmer, F., Englert, F., & Beisel, C. L. (**2022**). A TXTL-Based Assay to Rapidly Identify PAMs for CRISPR-Cas Systems with Multi-Protein Effector Complexes. In *Cell-Free Gene Expression* (pp. 391-411). Humana, New York, NY**.**

Abstract

Type I CRISPR-Cas systems represent the most common and diverse type of these prokaryotic defense systems and are being harnessed for a growing set of applications. As these systems rely on multi-protein effector complexes, their characterization remains challenging. Here, we report a rapid and straightforward method to characterize these systems in a cell-free transcription-translation (TXTL) system. A ribonucleoprotein complex is produced and binds to its target next to a recognized PAM, thereby preventing the targeted sequence from being cleaved by a restriction enzyme. Selection for uncleaved targeted plasmids leads to an enrichment of recognized sequences within a PAM library. This assay will aid the exploration of CRISPR-Cas diversity and evolution and help contribute new systems for CRISPR technologies and applications.

Introduction

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas (CRISPR-associated) systems are adaptive immune systems present in many bacteria and most archaea [1, 2]. Adaptive immunity is conducted in three general steps: acquisition, expression, and immunity. In the first step (acquisition), foreign DNA or RNA is recognized, and short fragments, called protospacers, are integrated into the CRISPR array as spacers separated by conserved repeats [3]. The selected protospacers are normally flanked by a protospacer adjacent motif (PAM) unique to each system and used to differentiate the invader-associated sequence from the spacer integrated into the CRISPR array [4, 5]. In the second step (expression), CRISPR arrays are transcribed and processed into mature CRISPR RNAs (crRNAs) [6]. These crRNAs then form a complex with the Cas effector nuclease. In the third

step (immunity), the resulting ribonucleoprotein complex screens DNA or RNA present in the cell for sequences complementary to the spacer and flanked by the PAM [7]. Upon recognition, the complex cleaves the recognized target through different mechanisms, completing the CRISPR-Cas system's task as an adaptive immune system [8].

Despite the three general steps of adaptive immunity, CRISPR-Cas systems are highly diverse. To-date, two classes, six types, and more than 30 subtypes have been classified. class I and II systems are divided based on the presence of a multi-protein effector complex or a single-effector protein, respectively [9]. Type I systems are part of class I and represent the majority of all CRISPR-Cas systems, yet remain understudied compared to class II systems. This is in part due to type I systems involving four to eight proteins in the effector complex at different stoichiometries (**Figure 1A**). Almost all of these proteins form a complex with the crRNA called Cascade (CRISPR-associated complex for antiviral defense) and screen invading DNA for its target. Upon target recognition, Cas3 is recruited to nick the non-target strand and processively degrade this strand in the 3′-to-5′ direction [10, 11]. While harnessing this complex as a technology is far more challenging than harnessing a class II single-effector protein (e.g., Cas9 from type II systems), applications with type I CRISPR-Cas systems are now emerging. In particular, type I systems have been employed for genome editing and gene regulation in bacteria, archaea, and eukaryotes or as tailored-spectrum antimicrobials against bacterial pathogens [12–28]. To further explore the extensive diversity of CRISPR biology and advance the existing suite of CRISPR technologies, there is an opportunity to accelerate the characterization of type I systems.

Cell-free transcription-translation (TXTL) systems offer a convenient way to rapidly characterize CRISPR-Cas systems. TXTL systems are typically based on an *E. coli* cell lysate with the transcriptional and translational machinery being retained. With TXTL, DNA can be added to produce RNAs and proteins in minutes to hours, allowing for subsequent biochemical assays without protein purification or cell culturing [29]. So far, TXTL-based methods have been established to characterize the expression and immunity steps of CRISPR [30–33]. CRISPR arrays are efficiently transcribed and processed in TXTL, and processing can be visualized via northern blotting or RNA-Seq analyses [30, 31]. The targeting activity of Cas nucleases and crRNAs can also be assessed with TXTL by directing the CRISPR effector complex to target a fluorescence reporter plasmid, inhibiting fluorescence production [32, 33]. Targeting activity can also be used to screen for anti-CRISPR proteins (Acrs) that inhibit different steps of CRISPR immunity [33–37]. Within these various characterization approaches, arguably the most important is determining the PAM. A number of *in vivo* and *in vitro* PAM determination methods have been reported [38], although they have typically been limited to single-effector nucleases. Here, we report a binding-based PAM assay in TXTL that

is well suited for multi-subunit CRISPR effectors [46]. This assay differs from our previously performed TXTL-PAM assay [33, 39], as it relies on target binding rather than cleavage. Furthermore, the assay enriches rather than depletes recognized PAM sequences, reducing the required sequencing depth to identify even weakly recognized PAMs. Overall, the reported PAM assay enables a fast and easy method to rapidly determine PAM requirements of CRISPR-Cas systems with multi-protein effector complexes.

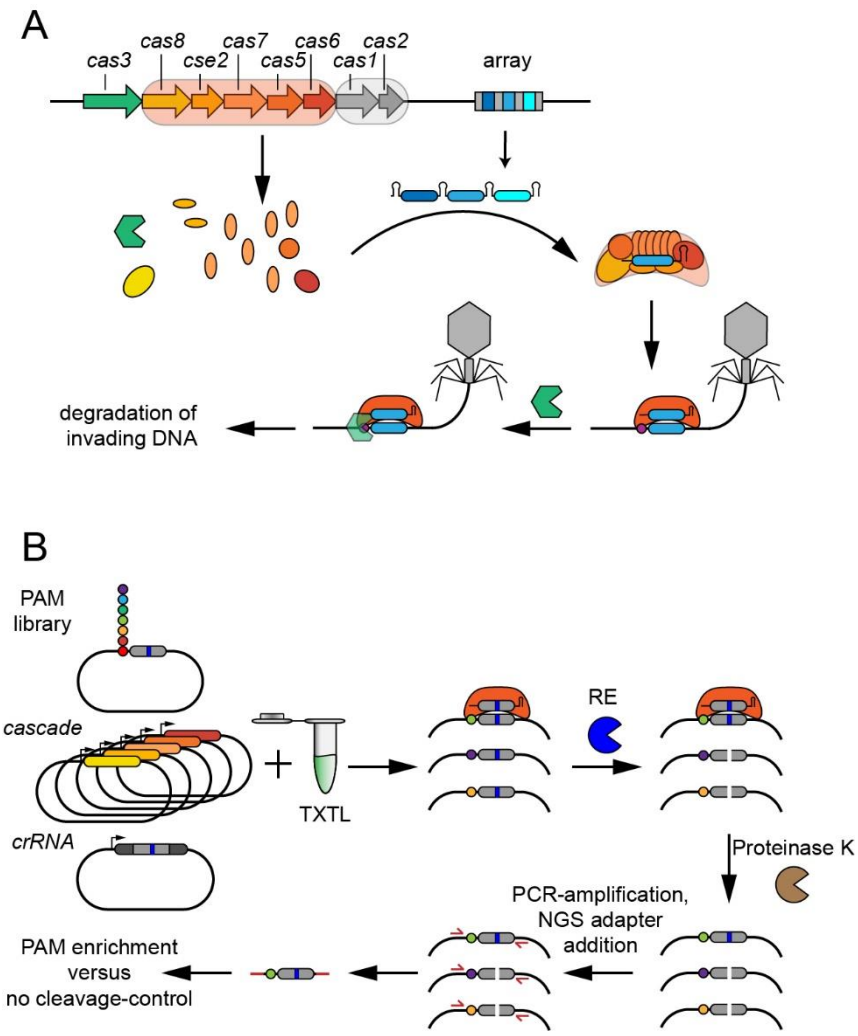


Figure 1: (**A**) Overview of DNA targeting with type I CRISPR-Cas systems. The type I-E CRISPR-Cas system from *E. coli* is used as a representative example. The *cas* genes encoding the proteins that form the Cascade complex are encircled in orange. The genes responsible for spacer acquisition are encircled in gray. As part of adaptive immunity, the system expresses the Cas proteins and transcribes the CRISPR array. The transcribed array is then processed into individual crRNAs that form a multi-protein ribonucleoprotein complex called Cascade. This complex screens DNA for protospacers comprising complementary sequences to the spacer (blue) and flanked by a PAM (purple). After Cascade binds a protospacer, it recruits Cas3 (green) to degrade the DNA. When the target is located in an invader such as a phage, recognition leads to DNA degradation and clearance of the invader. (**B**) Overview of the TXTL-based PAM determination assay. As part of the assay, three sets of plasmids are added to the TXTL mix: a plasmid encoding a PAM library (pGFP-PacI-5N) next to a targeted sequence (gray) that contains a restriction enzyme (RE) recognition site (blue), plasmids encoding for Cascade (pEcCas8, pEcCse2, pEcCas7, pEcCas5, pEcCas6) and a plasmid encoding a crRNA (pEc-crRNA1). The ribonucleoprotein complex (orange) is produced and binds to targets flanked by a recognized PAM. A subsequent digestion step with the RE results in cleavage of target sequences not bound by Cascade. Proteinase K (brown) is then used to remove all proteins. Adapters for NGS are added to the undigested PAM-containing plasmids by PCR. NGS sequencing results in the enrichment of recognized PAMs in the digested samples in comparison to samples without RE digestion

Materials

Reagents and Kits

1. 3 M sodium acetate, pH 5.2: 3 M sodium acetate, adjust pH with CH3COOH.
2. 70% ethanol (EtOH): prepare EtOH solution in dH2O by measuring EtOH and dH2O separately before combining.
3. Arbor Biosciences myTXTL Sigma 70 Master Mix kit.
4. LB-medium: 1% tryptone, 0.5% yeast extract, 86.6 mM NaCl, autoclave solution.
5. NGS library purification kit (e.g., AMPure beads).
6. Nuclease appropriate for NGS library preparation.
7. PCR purification kit.
8. Plasmid Midiprep kit.
9. Proteinase K (20 mg/µL).
10. Restriction enzyme (here PacI).
11. SOC medium: SOB medium, 10 mM MgCl2, 20 mM glucose, sterilize by passing solution through a 0.2 µm filter.
12. SOB medium: 2% tryptone, 0.5% yeast extract, 8.6 mM NaCl, 2.5 mM KCl, adjust pH to 7.0 with 5 N NaOH, autoclave solution.

Equipment

1. 96-well V bottom plate.
2. Cover mat for 96-well plate.

Methods

This binding-based PAM assay in TXTL results in an enrichment of positive PAMs (**Figure. 1B**). The CRISPR-Cas multi-protein complex is expressed, and the crRNA is transcribed. An effector complex is formed and binds at its target region flanked by recognized PAMs within a PAM library. Thus, the targeted plasmid is protected from digestion of a restriction enzyme (RE) (here PacI) that has its recognition site within the crRNA complementary region. The TXTL reaction is then digested with the RE, and a proteinase K digestion is performed. The remaining DNA is extracted by ethanol precipitation and sent for next-generation sequencing (NGS). We also provide a protocol for a quality check with Sanger sequencing or quantitative PCR (qPCR) before sending the samples for NGS. Finally, we provide an example by

characterizing the PAM requirements for the well-known *E. coli* Type I-E CRISPR-Cas system. All plasmids and primers that are used are stated in **Tables 1** and **2**, respectively.

Table 1: List of plasmids used

Name	Internal number	How to obtain	Benchling Link
pEc-crRNA1	CBS-1272	Addgene # 170088	https://benchling.com/s/seq-TnglgDNlecSHLmvamOql
pEc-crRNA2	CBS-2206	Addgene # 170089	https://benchling.com/s/seq-zsn7pzxiagaIAfG4kjO6
pEc-crRNAnt	CBS-212	[33]	https://benchling.com/s/seq-DFDGZdbiIESw3ElXz2iy
pEcCas5	CBS-189	Addgene # 170090	https://benchling.com/s/seq-RVJdQ9UfNPxyvc1nvD6l
pEcCas6	CBS-186	Addgene # 170091	https://benchling.com/s/seq-jbB5Es4jNHNtyKVr9Ctj
pEcCas7	CBS-194	Addgene # 170092	https://benchling.com/s/seq-YC8qRHgsEStzLtMu51iU
pEcCas8	CBS-196	Addgene # 170093	https://benchling.com/s/seq-FCnUAtortgKbxfUOAtK9
pEcCse2	CBS-184	Addgene # 170094	https://benchling.com/s/seq-eEYwschE6o4eaH5yKHw5
pGFP-ATAAC	CBS-2816	Addgene # 170095	https://benchling.com/s/seq-KiDRmcAwnw7WybwoZuK5
pGFP-CAAAG	CBS-2188	Addgene # 170096	https://benchling.com/s/seq-2s53R7nPYemkAgsq2EQg
pGFP-CAATG	CBS-2190	Addgene # 170097	https://benchling.com/s/seq-rnN1p3zVbrooInMHWVoV
pGFP-GTAAT	CBS-2762	Addgene # 170098	https://benchling.com/s/seq-f63QZs2FKTHAxS2EonHz
pGFP-GTATT	CBS-2754	Addgene # 170099	https://benchling.com/s/seq-VpH65rIVdKZQ8hkPj6f8
pGFP-PacI	CBS-332	Addgene # 170100	https://benchling.com/s/seq-TWxPKgiuHeuxelvn0WzG
pGFP-PacI-5N	CBS-1851	constructed based on pGFP-PacI	https://benchling.com/s/seq-6r4nLvHPddju0vBhANN5
pT7RNAP	CBS-344	Addgene # 170101	https://benchling.com/s/seq-PDEMzFwKlCSs9iRDTFTJ

Table 2: List of primers used

Name	Sequence	Purpose	Source
pr-01	5′-AATTCTGGCGAATCCTTTAATTAACTGAC-3′	PAM library introduction	this study
pr-02	5′-NNNNNAGACGAAAGGGCCTCGTGATAC-3′	PAM library introduction	this study
pr-03	5′-GGCGACACGGAAATGTTGAAT-3′	Sanger sequencing primer	this study
pr-04	5′-GCTGCAACCATTATCACCGC-3′	Sanger sequencing primer	this study
pr-05	5′-TATCACGAGGCCCTTTCGTC-3′	qPCR primer PAM library	this study
pr-06	5′-TCTGAATTGCAGCATCCGGT-3′	qPCR primer PAM library	this study
pr-07	5′-AACGTGGCGAGAAAGGAAGG-3′	qPCR primer pT7RNAP	this study
pr-08	5′-CGCTCGCGTATCGGTGATTC-3′	qPCR primer pT7RNAP	this study

Table 2 (continued)

Name	Sequence	Purpose	Source
pr-09	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCTTATCACGAGGCCCTTTCGT*C-3′	NGS library generation	this study
pr-10	5′-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCtCGTTTTCTGGCTGGTCAGTT*A-3′	NGS library generation	this study
pr-11	5′-AATGATACGGCGACCACCGAGATCTACACTTGGACTTACACTCTTTCCCTACACGAC*G-3′	NGS library generation	this study
pr-12	5′-CAAGCAGAAGACGGCATACGAGATGCGTTGGAGTGACTGGAGTTCAGACGTG*T-3′	NGS library generation	this study

PAM library construction

The PAM library should be at least one nucleotide longer than the expected PAM. We use five nucleotides (1024 combinations) in the example here, as the demonstration system from *E. coli* traditionally has a 3-nt PAM. For the PAM library, choose a plasmid with a unique RE recognition site within an untranscribed region. Our construct utilizes a unique restriction site recognized by PacI and additionally encodes for deGFP (pGFP-PacI). Make sure the RE can be heat inactivated. If introduction of a unique RE recognition site is necessary, this can be done by PCR mutagenesis of the chosen plasmid with primers including the recognition site at the 5′ end of either one or both primers followed by circularization and template removal such as with a Kinase-Ligase-DpnI (KLD) enzyme mix (see **Notes 1** and **2**).

The region targeted by the CRISPR-Cas system of interest should be chosen to span the RE recognition site. The PAM library (pGFP-PacI-5N) is then constructed adjacent to the targeted region (**Figure 2A**). For introduction of the PAM library, mutagenic primers are used (pr-01, pr-02) that amplify the whole plasmid at the desired site and include randomized nucleotides at their 5′ end. The resulting PCR product is then circularized, and the original DNA template is removed, such as with a Kinase-Ligase-DpnI (KLD) enzyme mix (see **Note 1**).

1. Transform library construct (pGFP-PacI-5N) in competent *E. coli*.
2. Recover transformed cells in 1 mL SOC medium without antibiotics for 1 h.
3. Add the recovered cell suspension in 50 mL LB medium with appropriate antibiotics and incubate overnight at an appropriate temperature.
4. Use a plasmid Midiprep kit to isolate the library plasmid from the 50 mL culture.
5. Re-clean and concentrate the plasmid DNA, if necessary, by a PCR purification kit.
6. Amplify the library plasmid (pGFP-PacI-5N) with primers spanning the PAM library (pr-03, pr-04) and send for Sanger sequencing to check for library quality (**Figure 2B**).

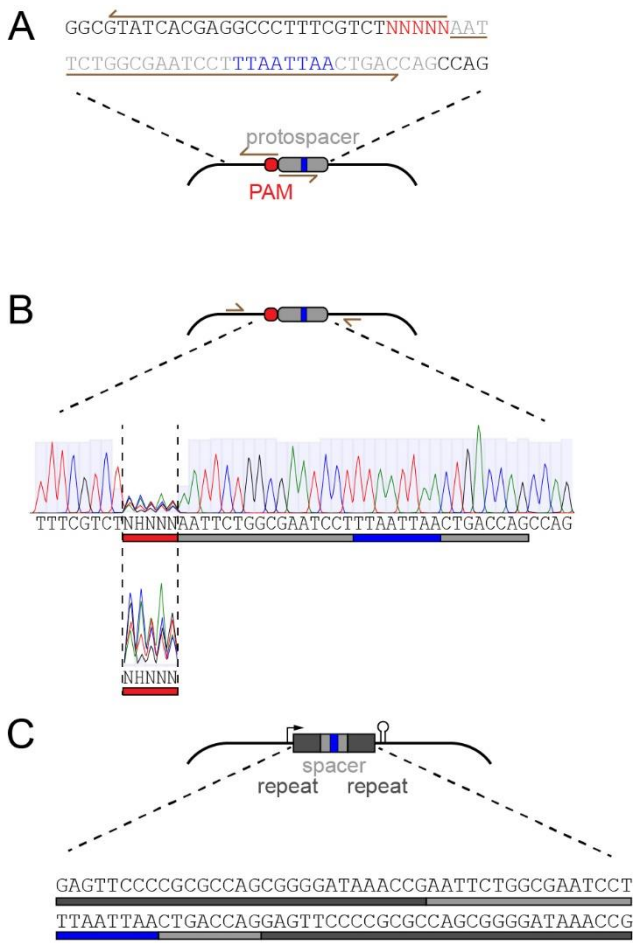


Figure 2: Construction of the PAM library and crRNA expression construct. (**A**) Overview of the PAM library containing plasmid (pGFP-PacI-5N). The PAM (red) consists of five random nucleotides and is located upstream of the protospacer (gray) that includes the RE recognition site (blue). Primers for library construction (pr-01, pr-02) are shown in brown. (**B**) Preliminary check of PAM library cloning by Sanger sequencing. To verify the cloned PAM library, a region containing the PAM library is amplified (primers are shown in brown) (pr-03, pr-04) and sent for Sanger sequencing. A representative sequencing trace is shown with nucleotide peaks. Below: a zoomed-in view of the library region. The nucleotide distribution does not need to be entirely equal, as enriched sequences are normalized to sequences in undigested samples as part of the NGS analysis. (**C**) The designed array sequence (pEc-crRNA1) from the I-E system in *E. coli* that is designed to target the PAM library-flanked target. The array is flanked by an upstream promoter and a downstream terminator. Repeats are shown in dark gray and the spacer is shown in light gray. The crRNA spacer spans the RE recognition site (blue).

CRISPR-Cas plasmid design and preparation

Cas proteins that form the Cascade can either be cloned on separate plasmids each (pEcCas8, pEcCse2, pEcCas7, pEcCas5, pEcCas6) or as an operon on one plasmid (see **Notes 3** and **4**). Use a protein expression vector with a strong promoter, an appropriate ribosomal binding site, and a terminator as the backbone (see **Note 5**). CRISPR arrays are cloned as a repeat-spacer-repeat (pEc-crRNA1) with a strong promoter and flanked by a terminator (**Figure 2C**). The spacer sequence herein is identical to the sequence downstream of the PAM library, covering the RE-recognition site. Transform all plasmids in competent *E. coli* and isolate the plasmids with a plasmid Midiprep kit followed by a PCR purification step.

PAM Assay

1. Prepare TXTL reaction on ice (**Table 3)** (see **Note 8**).
2. Carefully vortex the reactions and spin them down shortly.
3. Incubate at 29°C for 16 h (see **Notes 8** and **9**).
4. Dilute 3 µL of the TXTL reaction in 1.197 µL of nuclease-free water (see **Note 10**).
5. Prepare RE digestion reaction (**Table 4**), always prepare a control reaction with nuclease-free water instead of the RE.
6. Digest for 1 h at appropriate temperature (37°C for PacI).
7. Heat inactivate RE (20 min at 65°C for PacI).
8. Prepare proteinase K digestion (**Table 5**) and incubate for 1 h at 45°C.
9. Inactivate proteinase K for 5 min at 95°C.

Table 3: Components for the TXTL reaction of the PAM determination assay

Component	Volume (µL)	Initial concentration (nM)	Final concentration (nM)
TXTL	4.5	-	-
E. coli Cascade plasmids (pEcCas8, pEcCse2, pEcCas7, pEcCas5, pEcCas6) (see **Note 7**)	0.4	45	3
crRNA plasmid (pEc-crRNA1)	0.5	12	1
PAM library plasmid (pGFP-PacI-5N)	0.4	15	1
IPTG	0.06	50,000	500
T7RNAP (pT7RNAP)	0.1	12	0.2
Water	0.04	-	-

Table 4: Components for the RE digestion reaction

Component	Volume (µL)	Initial concentration	Final concentration
TXTL dilution	500	-	-
RE Buffer	56.1	10x	1x
RE (PacI)/ water	5	10 units/µL	0.09 units/µL

Table 5: Components for the Proteinase K digestion reaction

Component	Volume (µL)	Initial concentration (mg/mL)	Final concentration (mg/mL)
Digestion reaction	561.1	-	-
Proteinase K	1.4	20	0.05

DNA extraction

DNA is extracted by EtOH precipitation (see **Note 11**). All centrifugation steps are done at maximum speed and at 4°C.

1. Divide each sample in two equal parts of 280 µL so the following EtOH precipitation can be carried out in 1.5-mL tubes.
2. Add 3 M sodium acetate, pH 5.2 and ice-cold 100% EtOH to your samples (**Table 6**).
3. Mix by vortexing and spin down quickly.
4. Immediately store at -80°C in a pre-cooled rack for at least 20 min (see **Note 12**).
5. Spin your samples for 15 min (see **Note 13**).
6. Carefully remove liquid by decanting.
7. Add 200 µL of ice-cold 70% EtOH to your samples.
8. Spin for 10 min (see **Note 13**).
9. Repeat steps 6–8.
10. Remove liquid completely with a pipette, being careful to not touch the side of your tube with the DNA pellet.
11. Evaporate the remaining liquid by placing the tube at 50°C with an open lid.
12. Add 10 µL of nuclease-free water to the tube, be careful not to touch the pellet.
13. Incubate at 65°C for 10 min.
14. Vortex vigorously.
15. Combine your divided samples.

Table 6: Components for ethanol precipitation reaction

Component	Volume (µL)
Proteinase K digestion reaction	280
100% EtOH, ice-cold	700
3M sodium acetate, pH 5.2	28

Quality check

The final output of the PAM assay requires NGS. However, there are some methods to estimate if the assay was successful and that provide some indication of the expected PAMs. Specifically, Sanger sequencing of the PAM library containing region gives indication of the recognized PAMs if compared to sequencing results of samples lacking RE-digestion. Analysis of protection of RE-digestion by qPCR can provide an idea if the Cascade-crRNA complex was able to bind to its target.

Sanger sequencing

1. PCR-amplify a region containing the PAM library from the targeted plasmid in the digested and the undigested (control) sample (pr-03, pr-04).
2. Use a PCR purification kit to purify the amplicons.
3. Send the amplicons for Sanger sequencing using one or both primers used for amplification.
4. A noticeable difference in the sequencing results within the PAM library region between the digested and undigested samples indicates recognition of specific PAMs and thus enrichment of the recognized PAMs in the Sanger sequencing file (**Figure 3A**).

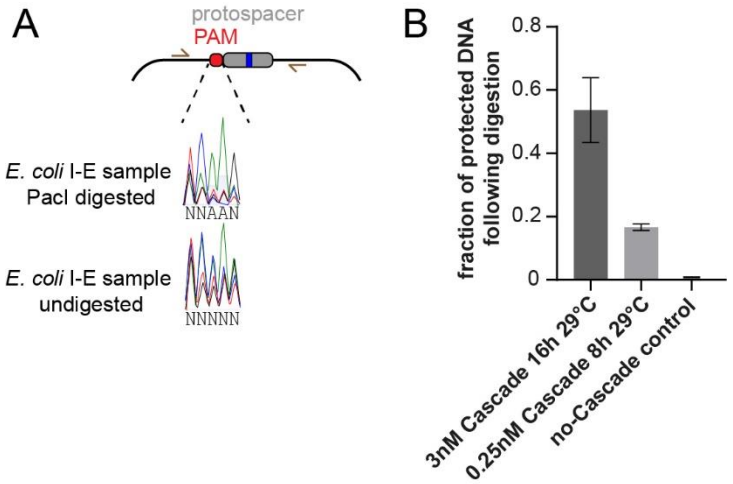


Figure 3: Preliminary assessment of a PAM assay readout. (**A**) Preliminary assessment by Sanger sequencing. An amplicon is generated from the targeted plasmid (pGFP-PacI-5N) that includes the PAM library region (red) and the targeted region (gray) with the RE recognition site (blue). Primers (pr-03, pr-04) are shown in brown. The nucleotide peaks from Sanger sequencing are shown for a sample that was digested with PacI and a sample that was not digested with PacI. The digested sample shows some increased peaks compared to the undigested sample, suggesting successful PAM recognition during the PAM assay. (**B**) Preliminary assessment with qPCR. The relative amount of uncleaved DNA in a digested sample is compared to the undigested control. The fraction of protected DNA is then determined based on $2^{-\Delta\Delta Ct}$ values from qPCR. An amplicon from the plasmid without the RE recognition site (pT7RNAP) was used as the reference sequence. Results are shown after applying the assay to the *E. coli* I-E Cascade following a TXTL reaction with 3 nM Cascade plasmid for 16 h or 0.25 nM Cascade plasmid for 8 h. Less protection, which can produce a stricter PAM profile, is achieved when using a lower Cascade plasmid concentration or a shorter incubation time (see **Note 8**). A control reaction without the Cascade or crRNA plasmids in the TXTL reaction shows negligible protection. Bars reflect the mean and error bars reflect the standard deviation from the duplicate reactions.

Quantitative PCR (qPCR)

1. Design two primer pairs. One amplifies a 100–250 bp long region containing the PAM library (pr-05, pr-06). The other amplifies a plasmid added to every single reaction but is not digested by your RE (pT7RNAP, pr-07, pr-08) and generates a similar amplicon size.
2. Use a qPCR kit without a reverse-transcription step to amplify the DNA obtained with the PAM assay with the primer pairs from step 1.

3. Determine the fold change between digested and undigested samples. Calculate $2^{-\Delta\Delta Ct}$ using the undigested sample as the control sample and the amplicon from the plasmid without RE recognition site (pT7RNAP) as the reference sequence. The fold change shows how much of your targeted plasmid is protected by the CRISPR-Cas systems and not prone to RE digestion. A fold change of 0.01 or lower indicates no binding of the CRISPR-Cas system (**Figure 3B**).

NGS Library Preparation

1. Amplify the EtOH precipitated DNA from Subheading "**DNA extraction**" with a nuclease appropriate for NGS library preparation adding Illumina sequencing primer binding sites on both ends of the amplicons containing the PAM library region (pr-09, pr10) (**Figure 4**).

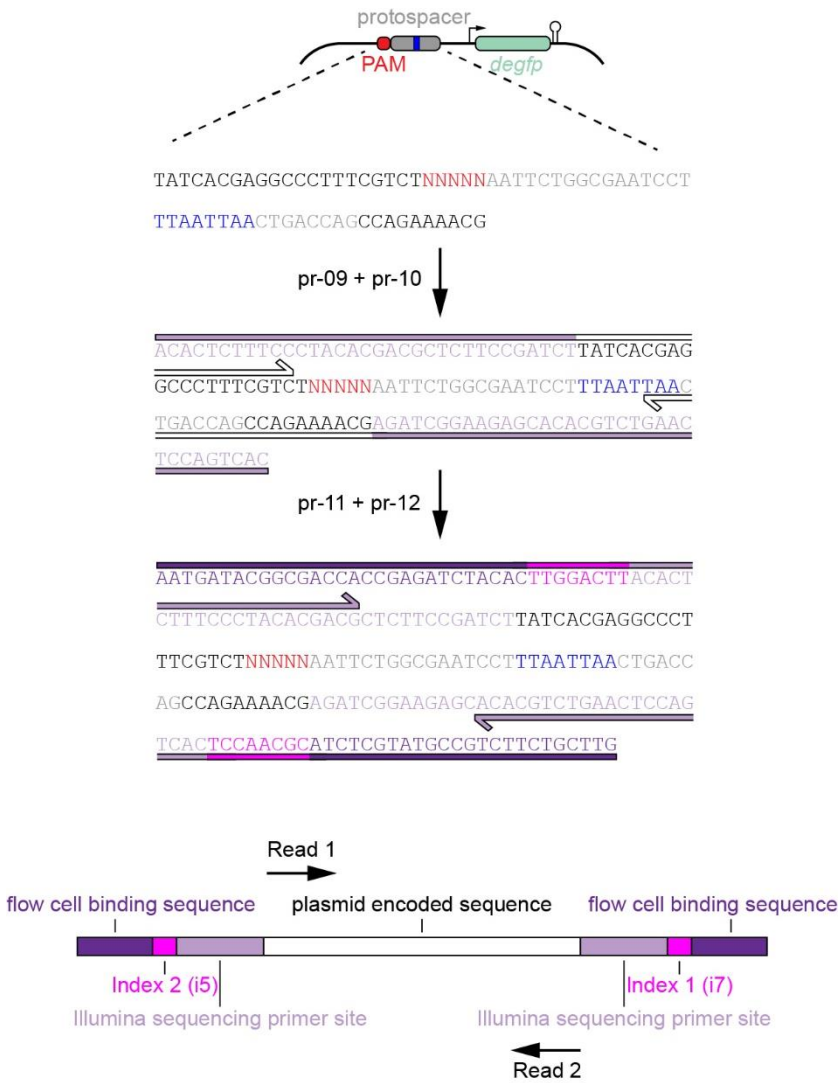


Figure 4: NGS library preparation. A region of the targeted plasmid is amplified that spans the PAM region (red) and the protospacer region (gray) including the RE recognition site (blue). Primers (pr-09, pr-10) are used that add the Illumina sequencing primer sites (light purple). Flow cell binding sites (dark purple) and the i7 and i5 indices are added in a second PCR (pr-11, pr-12). Directions of Read1 and Read2 generated by NGS are shown.

2. Check the amplicons on an agarose gel.
3. Purify the amplicon with an appropriate method, e.g., use AMPure beads.
4. Add the flow cell binding sequences and unique dual indices to both sides of the amplicon by amplifying the amplicon at the Illumina sequencing primer binding sites with a nuclease appropriate for NGS library preparation (pr-11, pr-12) (**Figure 4**).
5. Check the amplicons on an agarose gel.
6. Purify the amplicon with an appropriate method, e.g., use AMPure beads.
7. Sequence the amplicon with a NovaSeq 6000 Sequencing System with 50 bp paired-end reads.

NGS Data Analysis

Data analysis is done according to Leenay et al. [40]. A detailed protocol can be found there. The analysis starts with raw.fastq files. The following protocol is based on the sequence of **Figure 4** and can be adjusted according to the plasmid encoding the PAM library.

1. List the nucleotides from the randomized region with the following code:
   Read 1:
   grep '[TCAG][TCAG][TCAG][TCAG][TCAG]AATTCTGGCGAATCCTTTAATTAA' Sample1.fastq | cut -c 22-26 | sort | uniq -c | sort -nr | less > Sample1List.txt
   Read 2:
   grep ' TTAATTAAAGGATTCGCCAGAATT[TCAG][TCAG][TCAG][TCAG][TCAG]' Sample1.fastq | cut –c 43-47 | sort | uniq -c | sort -nr | less > Sample1List.txt
2. Import the .txt list into Microsoft Excel and sort it with the Sort and Filter tool.
3. Calculate PAM enrichment with the following formula:

$$\text{Enrichment} = \frac{\text{total reads non} - \text{digested sample}}{\text{total reads digested sample}} \times \frac{\text{reads digested sample}}{\text{reads non} - \text{digested sample}}$$

4. Use the calculated PAM enrichments and the PAM sequences to generate Krona Plots by adding them to the KronaExcelTemplate (available at https://github.com/marbl/Krona/wiki) and using Category 1 for the nucleotide adjacent to the protospacer (see **Note 14**) (**Figure 5**).
5. The generated Krona Plot can be viewed as a .html file with any web browser. The file can be downloaded and modified (**Figure 6**) (see **Note 15**).

Enrichment	Category 1	Category 2	Category 3	Other categories		5‘ to 3‘
1.28074176	A	A	A	A	A	AAAAA
2.29633284	C	A	A	A	A	AAAAC
2.66738671	G	A	A	A	A	AAAAG
1.19200502	T	A	A	A	A	AAAAT
0.25554045	A	C	A	A	A	AAACA
0.04757172	C	C	A	A	A	AAACC
0.30977032	G	C	A	A	A	AAACG
0.05010758	T	C	A	A	A	AAACT
2.05049297	A	G	A	A	A	AAAGA
0.85070642	C	G	A	A	A	AAAGC
2.72242137	G	G	A	A	A	AAAGG
0.67600828	T	G	A	A	A	AAAGT
0.66324258	A	T	A	A	A	AAATA
1.02993144	C	T	A	A	A	AAATC
2.55832849	G	T	A	A	A	AAATG
0.14667648	T	T	A	A	A	AAATT
0.05604735	A	A	C	A	A	AACAA
0.14719521	C	A	C	A	A	AACAC
1.48221946	G	A	C	A	A	AACAG

Figure 5: Populating the Krona Excel template to generate the PAM wheel. Positioning of the calculated PAM enrichments and the sorted PAM in the Excel template file are shown. The PAM sequences are depicted in their 5′ to 3′ direction. To achieve a 5′ to 3′ order from the outer circle to the inner circle in the PAM wheel, the nt in the -5 position is placed in Category 1, the nt in the -4 position in Category 2, and so on.

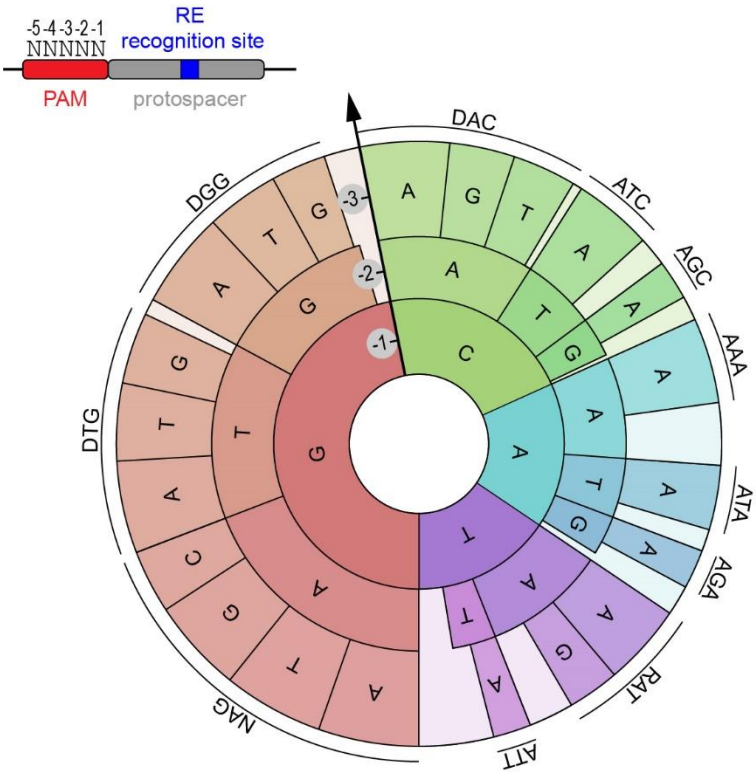


Figure 6: PAM wheel for the PAM determination assay conducted with Cascade from *E.coli*'s Type I-E system. The Krona Plot representing recognized PAMs generated from the example data with 3 nM Cascade and an incubation time of 16 h is shown. Thereby, only three nts are depicted as positions -4 and -5 did not show any nucleotide specificity. The section size of a PAM sequence is proportional to its enrichment during the PAM assay

Data Validation

TXTL can be also used to validate the PAMs that were found in the Krona Plot. A convenient method for this is to target the promoter region of a deGFP encoding plasmid (pGFP-PacI)

(**Figure 7A**). Clone each PAM-of-interest upstream of the -35 region of the promoter with the targeted sequence covering parts of the promoter (**Figure 7B**). Recognition of the PAM and thus binding of the Cascade-crRNA complex results in repression of deGFP expression, halting the build-up of fluorescence. Fold changes between a reaction containing a CRISPR array with a non-targeting spacer and a reaction containing a CRISPR array targeting the *degfp*-promoter region indicate the functionality of the chosen PAM.

1. Prepare TXTL reactions on ice (**Table 7**) (see **Note 16**). Always include a background control consisting of TXTL and water only.
2. Carefully vortex the reactions and spin them down shortly.
3. Incubate reactions at 29°C for 4 h (see **Note 9**).
4. Add reporter plasmid (here pGFP-CAAAG, pGFP-CAATG, pGFP-ATAAC, pGFP-GTAAT or pGFP-CTATT) to TXTL reaction (**Table 7**), add water to your background control instead.
5. Carefully vortex and briefly spin down your reactions.
6. Load at least technical duplicates of 5 µL each in a 96-well plate with V-shaped bottom.
7. Seal the loaded plate with a cover mat to prevent evaporation over time.
8. Measure deGFP fluorescence (Ex 485 nm, Em 528 nm) (see **Note 17**) in a plate reader pre-warmed to 29°C (see **Note 9**). Measure fluorescence every 3 min for up to 16 h.
9. Subtract the values of the background control from your samples for every measured timepoint.
10. Calculate fold-changes by dividing the fluorescence of the nontargeting control by the targeting control. High fold-changes represent highly recognized PAMs, low/no fold-changes represent low/no recognition (**Figure 7C**).

Table 7: Components for the TXTL reporter assay to validate identified PAMs

Component	Volume (µL)	Initial concentration (nM)	Final concentration (nM)
TXTL	9	-	-
E. coli Cascade plasmids (pEcCas8, pEcCse2, pEcCas7, pEcCas5, pEcCas6) (see **Note 7**)	0.5	12	0.5
crRNA plasmid (pEc-crRNA2 or pEc-crRNAnt)	1	12	1
Reporter plasmid (pGFP-CAAAG, pGFP-CAATG, pGFP-ATAAC, pGFP-GTAAT or pGFP-CTATT) (added later)	0.5	24	1
IPTG	0.12	50,000	500
T7RNAP (pT7RNAP)	0.2	12	0.2
Water	0.68	-	-

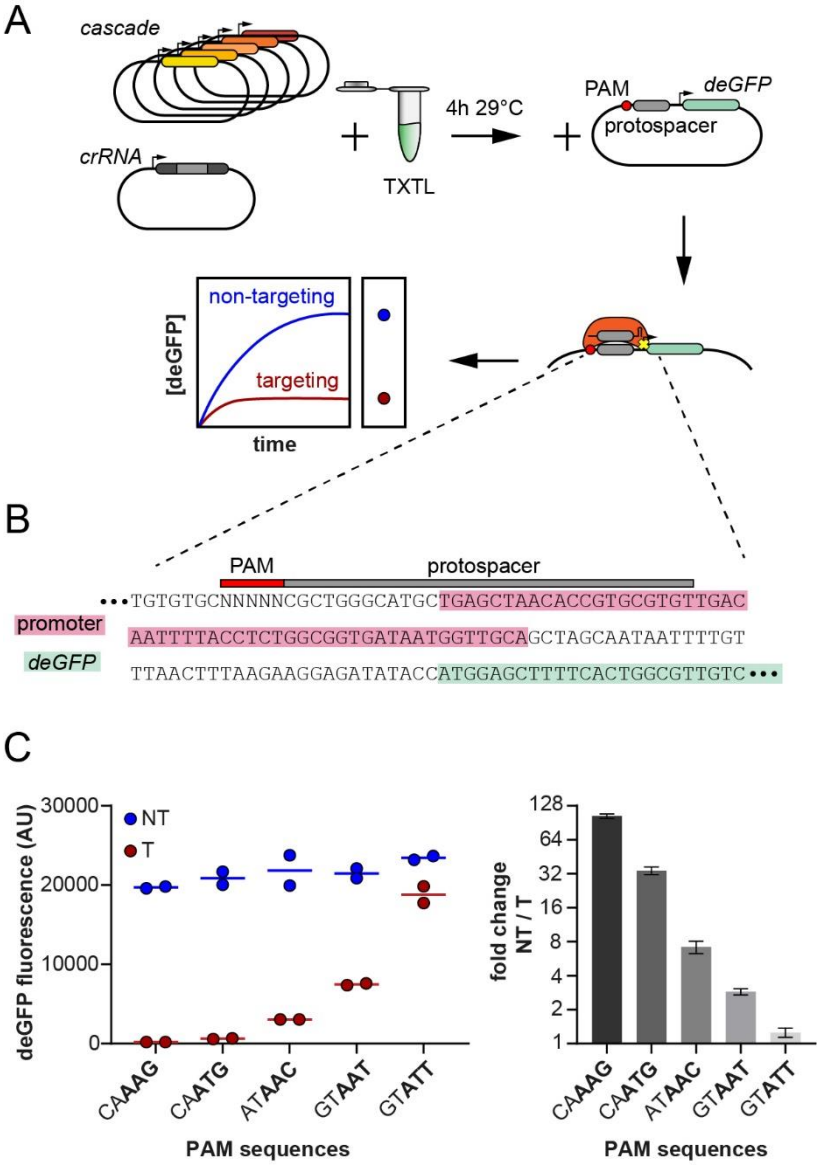


Figure 7: Validation of enriched PAM sequences using a TXTL-based reporter assay. (**A**) Overview of the reporter assay. Plasmids encode the Cascade proteins (pEcCas8, pEcCse2, pEcCas7, pEcCas5, pEcCas6) as well as a crRNA targeting the deGFP promoter (pEc-crRNA2) or a non-targeting crRNA (pEc-crRNAnt). These components are pre-expressed in TXTL at 29°C for 4 h to allow for ribonucleoprotein complex formation prior to expressing the reporter. Addition of a targeted deGFP reporter plasmid (pGFP-CAAAG, pGFP-CAATG, pGFP-ATAAC, pGFP-GTAAT or pGFP-CTATT) leads to binding of the protospacer by the ribonucleoprotein complex with the targeting crRNA, blocking transcription of the reporter. The rate of binding and the efficiency of transcriptional blocking impacts the accumulation of deGFP and the resulting fluorescence of the TXTL reaction. Under this setup, better-recognized PAMs result in less deGFP accumulation and fluorescence. (**B**) Sequence of the targeted plasmid. A region within the targeted plasmids containing the PAM (red), the protospacer (gray), the promoter (light red), and beginning of the deGFP coding region (light green) is shown. The PAM sequence is located upstream of the promoter, limiting interference with deGFP expression on its own and can be replaced with any sequence-of-interest. (**C**) deGFP expression levels and fold changes based on endpoint fluorescence. deGFP fluorescence levels of reactions with a targeting crRNA (T) or a non-targeting crRNA (NT) are shown on the left. The bigger the difference in fluorescence between NT and T, the better the PAM is recognized. Fold changes represent the ratio of deGFP fluorescence for the non-targeting reaction over the targeting reaction. Bars reflect the mean and error bars reflect the standard deviation from the duplicate reactions

Notes

1. Online tools can be used to design the mutagenic primers, e.g., NEBase Changer.
2. A deGFP-expressing plasmid can be used for introduction of the unique RE recognition site for an easy visual control of the functionality of the TXTL reaction and for use in later validation (here pGFP-PacI). A slightly green reaction mix after the incubation time indicates active protein expression (see **Note 17**).
3. Cloning all Cas proteins on separate plasmids allows for the addition of Cas protein encoding plasmids in a stoichiometric manner, while cloning Cascade proteins as an operon on one plasmid facilitates handling.
4. Plasmids can be exchanged with linear DNA if GamS is added to the TXTL reaction (final concentration 2 μM) or if χ sites are included in the linear construct to prevent RecBCD-induced DNA degradation [41, 42].
5. If a protein-of-interest is toxic, inducible promoters such as T7 promoter or IPTG-inducible promoters can be used.
6. We recommend preparing a MasterMix with TXTL, water and inducer and/or T7 RNA polymerase plasmid if necessary. We do not recommend including any other component (e.g., PAM library plasmid) into the MasterMix to ensure highest independence between replicates.
7. If Cas proteins are encoded on separate plasmids, prepare a MasterMix combining all Cas-encoding plasmids in one sample. Add every plasmid according to the stoichiometry of the encoded Cas protein. Concentration of all Cas-encoding plasmids should result in an initial concentration of 45 or 12 nM. If Cas proteins are encoded on one plasmid, prepare a stock concentration of 45 or 12 nM.
8. If less Cascade protein production is required, e.g., to only select for strong PAMs, add less plasmids encoding for Cas proteins and/or shorten the incubation time.
9. 29°C is commonly used for TXTL reactions and optimal for deGFP production, although the temperature can be varied between 25 and 42°C [43] and can impact the expression and activity of some Cas nucleases [33].
10. A 1:400 dilution of the TXTL reaction is optimized for the RE PacI. If other enzymes are used, different dilutions of TXTL can be tested for optimal results.
11. We recommend using EtOH precipitation over column-based purification due to the small amount of plasmid DNA remaining in the reaction and the lower DNA recovery of column-based purification compared to EtOH precipitation.
12. This step can proceed overnight.
13. Always place the tube in the centrifuge in the same orientation (e.g., lid pointing toward the center of the rotor) so you know where your DNA pellet is located.

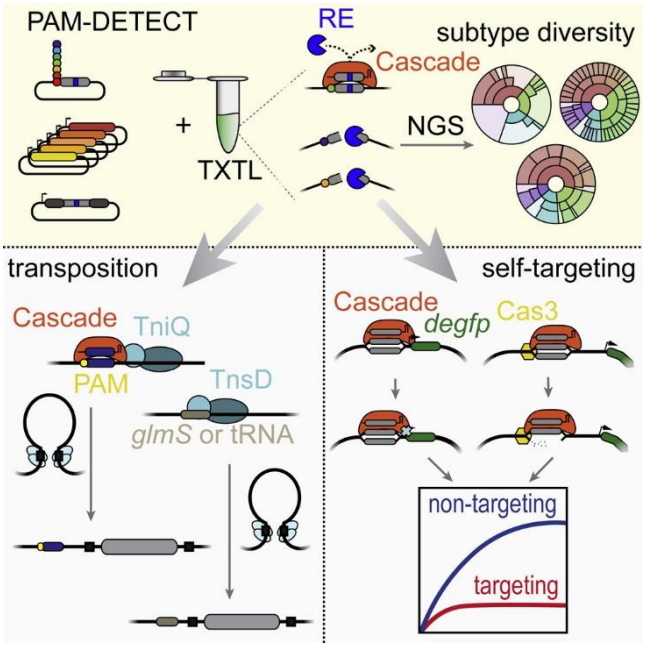
14. People with programming experience can also use the source code and run it on a local machine (see https://github.com/marbl/Krona/wiki for more information).
15. Other visualization methods besides Krona Plots can be used. Examples are sequence logos or motif plots [40, 44, 45].
16. Determination of the optimal Cascade plasmid concentration may be required if overall fluorescence is low.
17. deGFP can be exchanged with other fluorescence reporters.
18. We recommend preparing a MasterMix with TXTL, water, Cascade plasmids, crRNA plasmid and inducer and/or T7 RNA polymerase plasmid if necessary.

References

1. Barrangou R, Fremaux C, Deveau H et al (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science 315:1709–1712
2. Sorek R, Lawrence CM, Wiedenheft B (2013) CRISPR-mediated adaptive immune systems in bacteria and archaea. Annu Rev Biochem 82:237–266
3. Heler R, Marraffini LA, Bikard D (2014) Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. Mol Microbiol 93:1–9
4. Yosef I, Shitrit D, GorenMGet al (2013) DNA motifs determining the efficiency of adaptation into the *Escherichia coli* CRISPR array. Proc Natl Acad Sci U S A 110:14396–14401
5. Wang J, Li J, Zhao H et al (2015) Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems. Cell 163:840–853
6. Charpentier E, Richter H, van der Oost J, White MF (2015) Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. FEMS Microbiol Rev 39:428–441
7. Leenay RT, Beisel CL (2017) Deciphering, communicating, and engineering the CRISPR PAM. J Mol Biol 429:177–191
8. Marraffini LA, Sontheimer EJ (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet 11:181–190
9. Makarova KS, Wolf YI, Iranzo J et al (2020) Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol 18:67–83
10. Brouns SJJ, Jore MM, Lundgren M et al (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321:960–964
11. Westra ER, van Erp PBG, Künne T et al (2012) CRISPR immunity relies on the consecutive binding and degradation of negatively super-coiled invader DNA by Cascade and Cas3.Mol Cell 46:595–605
12. Csörgő B, León LM, Chau-Ly IJ et al (2020) A compact Cascade-Cas3 system for targeted genome engineering. Nat Methods 17:1183–1190
13. Chen Y, Liu J, Zhi S et al (2020) Repurposing type I–F CRISPR–Cas system as a transcriptional activation tool in human cells. Nat Commun 11:1–14
14. Dolan AE, Hou Z, Xiao Y et al (2019) Introducing a spectrum of long-range genomic deletions in human embryonic stem cells using type I CRISPR-Cas. Mol Cell 74:936–950.e5
15. Xu Z, Li M, Li Y et al (2019) Native CRISPR-Cas-mediated genome editing enables dissecting and sensitizing clinical multidrug-resistant *P. aeruginosa*. Cell Rep 29:1707–1717.e3
16. Luo ML, Mullis AS, Leenay RT, Beisel CL (2015) Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. Nucleic Acids Res 43:674–681
17. Hidalgo-Cantabrana C, Barrangou R (2020) Characterization and applications of Type I CRISPR-Cas systems. Biochem Soc Trans 48:15–23
18. Morisaka H, Yoshimi K, Okuzaki Y et al (2019) CRISPR-Cas3 induces broad and unidirectional genome editing in human cells. Nat Commun 10:5302
19. Cheng F, Gong L, Zhao D et al (2017) Harnessing the native type I-B CRISPR-Cas for genome editing in a polyploid archaeon. J Genet Genomics 44:541–548
20. Li Y, Pan S, Zhang Y et al (2016) Harnessing type I and type III CRISPR-Cas systems for genome editing. Nucleic Acids Res 44:e34

21. Cameron P, Coons MM, Klompe SE et al (2019) Harnessing type I CRISPR-Cas systems for genome engineering in human cells. Nat Biotechnol 37:1471–1477
22. Pickar-Oliver A, Black JB, Lewis MM et al (2019) Targeted transcriptional modulation with type I CRISPR-Cas systems in human cells. Nat Biotechnol 37:1493–1501
23. Pyne ME, Bruder MR, Moo-Young M et al (2016) Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *Clostridium*. Sci Rep 6:25666
24. Hidalgo-Cantabrana C, Goh YJ, Pan M et al (2019) Genome editing using the endogenous type I CRISPR-Cas system in *Lactobacillus crispatus*. Proc Natl Acad Sci U S A 116:15774–15783
25. Zheng Y, Han J, Wang B et al (2019) Characterization and repurposing of the endogenous type I-F CRISPR-Cas system of *Zymomonas mobilis* for genome engineering. Nucleic Acids Res 47:11461–11475
26. Rath D, Amlinger L, Hoekzema Met al (2015) Efficient programmable gene silencing by Cascade. Nucleic Acids Res 43:237–246
27. Gomaa AA, Klumpe HE, Luo ML et al (2014) Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. Mbio 5:e00928–e00913
28. Yosef I, Manor M, Kiro R, Qimron U (2015) Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria. Proc Natl Acad Sci USA 112:7267–7272
29. Silverman AD, Karim AS, Jewett MC (2020) Cell-free gene expression: an expanded repertoire of applications. Nat Rev Genet 21:151–170
30. Liao C, Ttofali F, Slotkowski RA et al (2019) Modular one-pot assembly of CRISPR arrays enables library generation and reveals factors influencing crRNA biogenesis. Nat Commun 10:2948
31. Liao C, Slotkowski RA, Achmedov T, Beisel CL (2019) The *Francisella novicida* Cas12a is sensitive to the structure downstream of the terminal repeat in CRISPR arrays. RNA Biol 16:404–412 32.
32. Marshall R, Beisel CL, Noireaux V (2020) Rapid testing of CRISPR nucleases and guide RNAs in *an E. coli* cell-free transcription-translation system. STAR Protocols 1:100003
33. Marshall R, Maxwell CS, Collins SP et al (2018) Rapid and scalable characterization of CRISPR technologies using an *E. coli* cell-free transcription-translation system. Mol Cell 69:146–157.e3
34. Bondy-Denomy J, Pawluk A, Maxwell KL, Davidson AR (2013) Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature 493:429–432
35. Davidson AR, Lu W-T, Stanley SY et al (2020) Anti-CRISPRs: protein inhibitors of CRISPR-Cas systems. Annu Rev Biochem 89:309–332
36. Wandera KG, Collins SP, Wimmer F et al (2020) An enhanced assay to characterize anti-CRISPR proteins using a cell-free transcription-translation system. Methods 172:42–50
37. Watters KE, Fellmann C, Bai HB et al (2018) Systematic discovery of natural CRISPR-Cas12a inhibitors. Science 362:236–239
38. Collias D, Beisel CL (2021) CRISPR technologies and the search for the PAM-free nuclease. Nat Commun 12:555
39. Maxwell CS, Jacobsen T, Marshall R et al (2018) A detailed cell-free transcription-translation-based assay to decipher CRISPR protospacer-adjacent motifs. Methods 143:48–57
40. Leenay RT, Maksimchuk KR, Slotkowski RA et al (2016) Identifying and visualizing functional PAM diversity across CRISPR-Cas systems. Mol Cell 62:137–147
41. Sitaraman K, Esposito D, Klarmann G et al (2004) A novel cell-free protein synthesis system. J Biotechnol 110:257–263
42. Marshall R, Maxwell CS, Collins SP et al (2017) Short DNA containing χ sites enhances DNA stability and gene expression in *E. coli* cell-free transcription-translation systems. Biotechnol Bioeng 114:2137–2141
43. Shin J, Noireaux V (2010) Efficient cell-free expression with the endogenous *E. coli* RNA polymerase and sigma factor 70. J Biol Eng 4:8
44. Schneider TD, Stephens RM (1990) Sequence logos: a new way to display consensus sequences. Nucleic Acids Res 18:6097–6100
45. Collias D, Leenay RT, Slotkowski RA et al (2020) A positive, growth-based PAM screen identifies noncanonical motifs recognized by the *S. pyogenes* Cas9. Sci Adv 6:eabb4054
46. Wimmer F, Mougiakos I, Englert F, Beisel CL (2021) Rapid cell-free characterization of multi-subunit CRISPR effectors and transposons bioRxiv 2021.10.18.464778

Chapter 3: Rapid cell-free characterization of multi-subunit CRISPR effectors and transposons



The content of this chapter was previously published and reproduced with permission from Cell Press.

Wimmer, F.*, Mougiakos, I.*, Englert, F., & Beisel, C. L. (**2022**). Rapid cell-free characterization of multi-subunit CRISPR effectors and transposons. *Molecular Cell*.

* These authors contributed equally

Author contributions:

Conceptualization, F.W., I.M., and C.L.B.; methodology, F.W., I.M., and C.L.B.; software, F.W. and I.M.; validation, F.W., I.M., and F.E.; investigation, F.W., I.M., and F.E.; writing – original draft, F.W., I.M., and C.L.B.; writing – review & editing, F.W., I.M., F.E., and C.L.B.; visualization, F.W., I.M., and C.L.B.; supervision, C.L.B.; funding acquisition, C.L.B.

Summary

CRISPR-Cas biology and technologies have been largely shaped to-date by the characterization and use of single-effector nucleases. By contrast, multi-subunit effectors dominate natural systems, represent emerging technologies, and were recently associated with RNA-guided DNA transposition. This disconnect stems from the challenge of working with multiple protein subunits *in vitro* and *in vivo*. Here, we apply cell-free transcription-translation (TXTL) systems to radically accelerate the characterization of multi-subunit CRISPR effectors and transposons. Numerous DNA constructs can be combined in one TXTL reaction, yielding defined biomolecular readouts in hours. Using TXTL, we mined phylogenetically diverse I-E effectors, interrogated extensively self-targeting I-C and I-F systems, and elucidated targeting rules for I-B and I-F CRISPR transposons using only DNA-binding components. We further recapitulated DNA transposition in TXTL, which helped reveal a distinct branch of I-B CRISPR transposons. These capabilities will facilitate the study and exploitation of the broad yet underexplored diversity of CRISPR-Cas systems and transposons.

Introduction

CRISPR-Cas systems endow prokaryotes with an adaptive defense against invading elements and possess effector nucleases that have become versatile biomolecular tools (Barrangou and Doudna, 2016; Pickar-Oliver and Gersbach, 2019). These systems are remarkably diverse, with two classes, six types, over 30 subtypes, and a few subtype variants defined to-date (Makarova et al., 2020). The two classes are distinguished based on whether the effector nuclease responsible for CRISPR RNA (crRNA)-directed immune defense comprises a multi-protein complex (class 1) or a single multi-domain protein (class 2). Within these classes, class 2 systems have been the most extensively explored. For example, comprehensive determination of target-flanking protospacer-adjacent motifs (PAMs) (Leenay and Beisel, 2017) has been conducted on more than 100 class 2 effectors spanning at least 15 subtypes (Collias and Beisel, 2021) but only on 10 class 1 effectors spanning 7 subtypes (**Table S1**). Despite this discrepancy, class 1 systems represent over 75% of all CRISPR-Cas systems found in nature, contain phylogenetically diverse proteins possessing unique mechanisms of action (Makarova et al., 2015), and are associated with emerging alternative functions (Li et al., 2021). The associated machinery has also been recently applied as tools in mammalian and plant cells, offering distinct means of achieving gene regulation, genome editing, and variable chromosomal deletions (Cameron et al., 2019; Chen et al., 2020; Dolan et al., 2019; Liu et al., 2018; Morisaka et al., 2019; Osakabe et al., 2020; Pickar-Oliver et al., 2019; Zheng et al., 2020). Finally, a subset of class 1 systems contain Tn7-like transposon genes and were shown to mediate crRNA-directed transposition. These CRISPR-associated transposons

(CASTs) have since been employed in bacteria for the efficient, programmable, and multiplexed insertion of donor DNA exceeding 10 kb (Chen et al., 2021; Klompe et al., 2019; Park et al., 2021; Petassi et al., 2020; Peters et al., 2017; Rybarski et al., 2021; Saito et al., 2021; Strecker et al., 2019; Vo et al., 2021). These examples highlight the potential of further exploring and harnessing class 1 CRISPR-Cas systems and CASTs.

The disconnect between the broad relevance of class 1 systems and the few well-characterized examples can be largely attributed to the challenge of working with multiple protein subunits. Cell-based assays are complicated by the need to encode and optimally express multiple subunits from a minimal number of constructs, whereas *in vitro* assays require intensive purification of multi-subunit complexes – tasks that are far simpler for single-effector nucleases. A promising alternative came with the advent of cell-free transcription-translation (TXTL) systems and their use for rapidly and scalably characterizing CRISPR-Cas systems (Garamella et al., 2016; Jiao et al., 2021; Liao et al., 2019a, 2019b; Marshall et al., 2018; Maxwell et al., 2018; Silverman et al., 2020; Watters et al., 2018). As part of a TXTL reaction, circular or linear DNA constructs are added to the TXTL mix, resulting in the transcription and translation of the encoded products in minutes to hours. In our prior work, we showed that TXTL could functionally express the type I effector complex Cascade (CRISPR-associated complex for antiviral defense) that yielded transcriptional repression of a reporter gene (Marshall et al., 2018). However, all other implementations of TXTL to-date have focused on single-effector nucleases (Khakimzhan et al., 2021; Liao et al., 2019a, 2019b; Wandera et al., 2020; Watters et al., 2018). Here, we leverage TXTL to rapidly characterize diverse type I systems and transposons, allowing ortholog mining, characterization of self-targeting systems, and harnessing of CASTs. The resulting capabilities should accelerate the exploration and exploitation of this broad yet understudied branch of CRISPR biology.

Design – PAM-DETECT: a TXTL-based enrichment assay for PAM determination

One of the defining features of DNA-targeting CRISPR-Cas systems is the PAM (Leenay and Beisel, 2017). This collection of sequences always flanks a crRNA target and allows the effector nuclease to discriminate between self (the associated spacer in the CRISPR array) and non-self (the targeted invader). The associated sequences can vary widely even between close homologs (Collias and Beisel, 2021). Given that prior comprehensive PAM determination assays applied to class 1 systems involved laborious *in vitro* or cell-based assays (**Table S1**), we devised a TXTL-based assay that could elucidate the complete PAM profile recognized by an effector complex but without the need for protein purification or cellular expression (**Figures 1A** and **1B**; **Methods S1**). The assay involves expressing the crRNA and the three to five Cas

proteins to form Cascade, which then binds target DNA. Even though Cascade binding normally recruits the endonuclease Cas3 (Hochstrasser et al., 2014; Huo et al., 2014; Westra et al., 2012; Xiao et al., 2017) to nick and processively degrade the non-target strand of DNA (Gong et al., 2014; Huo et al., 2014; Mulepati and Bailey, 2013; Westra et al., 2012; Xiao et al., 2017, 2018), Cascade strongly binds DNA even without Cas3 (Jackson et al., 2014; Jore et al., 2011; Mulepati et al., 2012; Westra et al., 2012). As part of the TXTL-based assay, Cascade binds target DNA flanked by a library of potential PAM sequences. A restriction enzyme is then introduced that cleaves a sequence within the DNA target. As a result, DNA containing a recognized PAM sequence is protected by the bound Cascade from cleavage, thereby enriching this sequence within the library. Next-generation sequencing (NGS) is then performed to quantify the relative frequency of each PAM sequence before and after restriction digestion. We call this assay PAM-DETECT (PAM DETermination with Enrichment-based Cell-free TXTL). From the addition of the DNA constructs to the isolation of library DNA for NGS, the entire process requires 13-23 h.

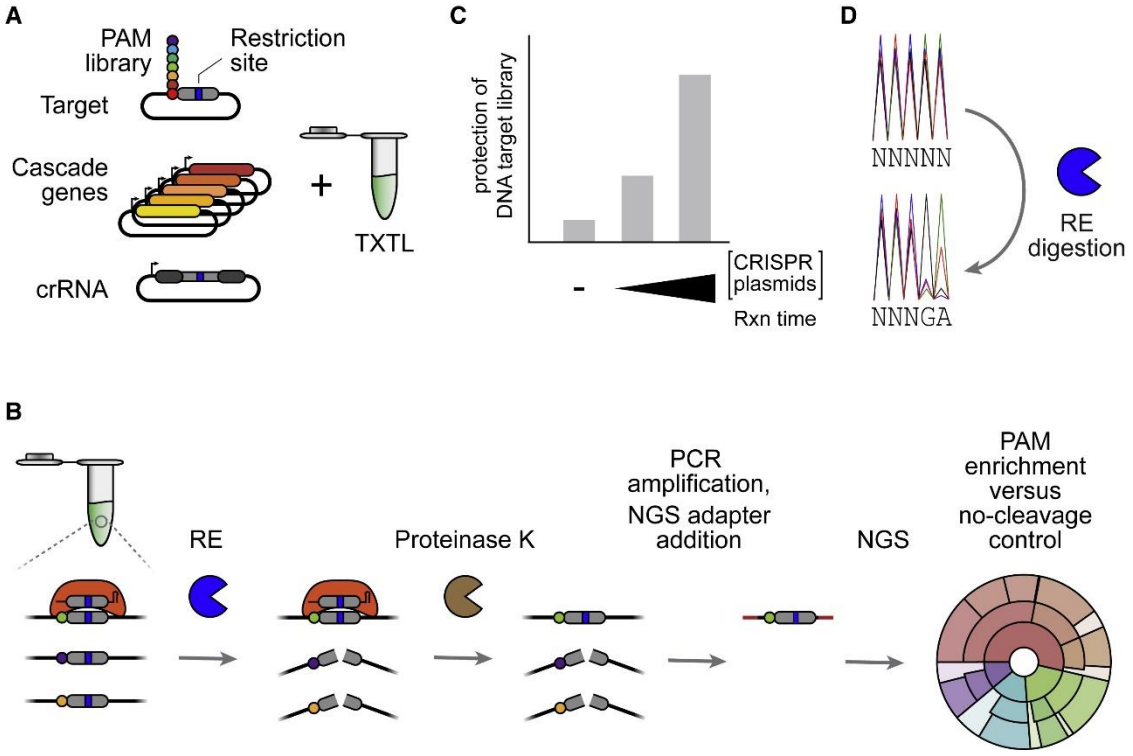


Figure 1: PAM-DETECT, a TXTL-based PAM determination assay for multi-protein CRISPR effectors. (**A**) DNA components added to a TXTL reaction to perform PAM-DTECT. The Cascade genes can be encoded on separate plasmid, as shown here, or as an operon. (**B**) Steps comprising PAM-DETECT. RE, restriction enzyme. (**C**) Determination of PAM enrichment by Sanger sequencing.

As part of PAM-DETECT, we devised two parallel checkpoints to assess the extent of library protection and PAM enrichment prior to submitting samples for NGS. For the first checkpoint **(Figure 1C**), quantitative PCR (qPCR) is applied with a digested and undigested library to measure the extent to which the library was protected by Cascade binding. Given that excess

effector can boost the prevalence of less-preferred PAM sequences (Karvelis et al., 2015), the qPCR results can indicate the stringency of the determined PAM sequences. Fortunately, the conditions of PAM-DETECT can be readily tuned by changing the concentration of the added DNA constructs and the time allowed for Cascade expression and DNA binding. For the second checkpoint (**Figure 1D**), the digested and undigested libraries are subjected to Sanger sequencing. Elevated peaks in the digested versus undigested sample reflect enrichment of those bases at that PAM position, providing a preliminary indication of the determined PAM.

Results

PAM-DETECT validated with the canonical type I-E CRISPR-Cas system from *Escherichia coli*

To evaluate PAM-DETECT, we began with Cascade encoded by the type I-E CRISPR-Cas system from *Escherichia coli* (*E. coli*) (**Figure 2A**). As part of its extensive characterization, the effector complex has been subjected to multiple comprehensive PAM determination assays (Caliando and Voigt, 2015; Fineran et al., 2014; Fu et al., 2017; Leenay et al., 2016; Musharova et al., 2019; Xue et al., 2015), establishing a complex landscape principally composed of the canonical PAM sequences AAG, AGG, ATG, and GAG (written 5′ to 3′) located on the non-target strand immediately upstream of the sequence matching the crRNA guide. We applied PAM-DETECT by encoding the five Cascade genes and a targeting single-spacer CRISPR array on six separate plasmids and combining these plasmids with a 5-base PAM target library harboring a PacI restriction site (**Figure 2A**). To explicitly evaluate the impact of excess effector complexes, we tested two different conditions: one with 0.25 nM of Cascade-encoding plasmids and 6-h reaction time for low Cascade expression/binding and another with 3 nM of Cascade-encoding plasmids and 16-h reaction time for high Cascade expression/binding. The qPCR check showed significant DNA protection compared with the control lacking Cascade, with ~2-fold more protection for the high-versus low-Cascade condition (**Figure 2B**). In parallel, the Sanger-sequencing checkpoint showed enrichment of an AAG motif compared with the undigested control, where the motif was more pronounced for the low-Cascade condition (**Figure 2C**). The checkpoints were in line with the protection of DNA sequences related to the known PAM, with heightened protection for the high-Cascade condition.

Proceeding to NGS, we visualized the results as a PAM wheel to capture both individual sequences and enrichment scores (Leenay et al., 2016; **Figure 2D**). The PAM wheel for the low-Cascade condition captured the four known canonical PAMs as well as other well-recognized PAM sequences (e.g., TAG and AAC). The PAM wheel for the high-Cascade condition included these PAM sequences as well as other PAM sequences that were less

enriched (e.g., AAA and AAT) or negligibly enriched (e.g., CAG and ATT) for the low-Cascade condition (**Figure 2D**). The differences in PAM profiles demonstrate how PAM-DETECT can be readily tuned by varying plasmid concentration and reaction time.

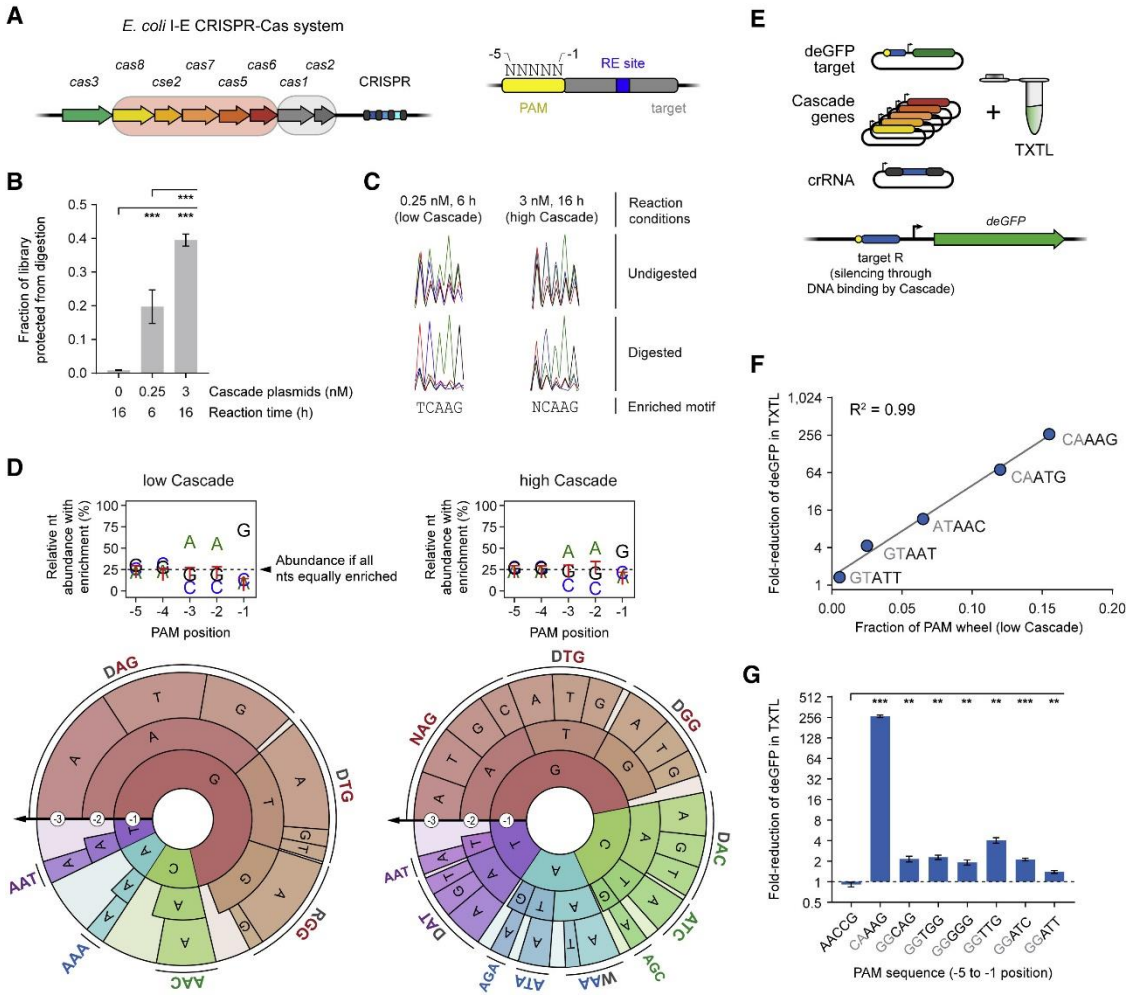


Figure 2: Validation pf PAM-DETECT with the I-E CRISPR-Cas system from *E. coli*. (**A**) The type I-E CRISPR-Cas systems from *E. coli*. The genes encoding the Cascade complex are in the light orange box, and the genes encoding the acquisition proteins are in the gray box. Right: 5N library of potential PAM sequences used with PAM-DETECT. (**B**) Extent of PAM library protection under conditions resulting in low or high levels of Cascade based on qPCR. Library protection compares the library with and without RE digestion. (**C**) Effect of low or high levels of Cascade based on Sanger sequencing. Over-representation of T and C at the -5 and -4 position, respectively, can be explained by the library generation, as TCAAG represented one of the most prevalent sequences in the library. (**D**) Nucleotide-enrichment plots and PAM wheels based on conducting PAM-DETECT with low or high levels of Cascade. Individual sequences comprising at least 2% of the PAM wheel are shown. Results represent the average of duplicate independent experiments. The size of the arc for an individual sequence corresponds to its relative enrichment within the library. (**E**) Overview of the TXTL-based PAM validation assay. PAM sequences are tested by Cascade binding target R flanked by the tested PAM. Because target R overlaps the promoter driving expression of deGFP, target binding blocks deGFP expression. Target R is distinct from the restriction site-containing target used with PAM-DETECT. Fold-reduction is calculated based on a non-targeting crRNA control. (**F**) Correlation between PAM enrichment from PAM-DETECT and gene repression in TXTL. Enrichment was based on the low-Cascade condition. Enrichment values represent the mean of duplicate PAM-DETECT assays, whereas fold-reduction values represent the mean of triplicate TXTL assays. Fold-reduction was calculated based on a non-targeting crRNA control. (**G**) TXTL validation of PAM sequences identified by PAM-DETECT but not previously by PAM-SCANR. CAAAG serves as a positive control. The AACCG self PAM matches the 3′ end of the repeat and is the reference for statistical analyses.

Error bars in (B) and (G) indicate the mean and standard deviation of triplicate independent experiments. ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, and ns: $p > 0.05$.

To validate the results, we applied TXTL to silence plasmid-based expression of deGFP, a truncated version of eGFP that is more efficiently translated in cell-free systems (Shin and Noireaux, 2010, 2012). By targeting a distinct target sequence overlapping the deGFP promoter (**Figure 2E**; **Table S2**), the PAM sequence could be altered without affecting the promoter sequence. For representative PAM sequences, the fold-repression of deGFP production versus a non-targeting control strongly correlated with the enrichment score of each sequence in PAM-DETECT for the low-Cascade condition (R^2 = 0.99) (**Figure 2F**). Applying the same assay to PAM sequences enriched under the high-Cascade condition but not detected with our previous PAM-SCANR method (Leenay et al., 2016), we measured modest but significant deGFP repression (**Figure 2G**). These validation experiments show that PAM-DETECT can produce comprehensive and quantitative PAM profiles.

Distinct PAM profiles pervade I-E CRISPR-Cas systems

Nuclease mining has been highly successful for identifying single-effector nucleases such as Cas9 with a wide spectrum of PAMs (Gasiunas et al., 2020; Zetsche et al., 2020) and thus could be highly valuable when applied to class 1 systems. Focusing again on the I-E subtype, we began by identifying diverse Cas8e proteins responsible for PAM recognition within Cascade (van Erp et al., 2015; Hayes et al., 2016) from mesophilic bacterial strains. We divided the identified set of 213 Cas8e proteins in groups according to the amino-acid sequence of the highly variable L1 loop within the N-terminal domain (**Table S3**) reported to stabilize Cas8e-PAM interactions (Tay et al., 2015; Xiao et al., 2017). We selected 11 representative I-E systems reflecting some of the most abundant L1 motifs (**Figures 3A** and **S1**). The resulting Cascade complexes could be readily characterized with PAM-DETECT in parallel despite involving 55 Cascade genes and 11 single-spacer arrays, each in separate plasmids. We selected the high-Cascade conditions (3-nM plasmids and 16-h reaction time) given the uncertainty about how well a given system would be functionally expressed in TXTL. All but one system yielded significant enrichment of the PAM library, compared with a non-digested control (**Figure S1A**), allowing us to determine a large number of PAM profiles.

PAM-DETECT revealed a broad range of recognized PAMs (**Figures 3A** and **S1B**). The PAM profile most distinct from that associated with the *E. coli* Cascade was recognized by Cascade from *Streptococcus thermophilus* DGCC 7710 (Sth). This profile comprised any sequence with an A or T at the -1 position as well as (S = G and C) and ATS, which included the few PAM sequences previously confirmed to bind purified Cascade *in vitro* (Sinkunas et al., 2013). Most remaining systems generally recognized AAG as a dominant PAM sequence, although there were notable deviations and additions. For example, one system from *Azotobacter chroococcum* NCIMB 8003 (Ac2) principally recognized AA, whereas another

system from *Paracoccus* sp. J4 (Ps) preferentially recognized AAC. Interestingly, Ac2 and a separate system from *Azotobacter chroococcum* NCIMB 8003 (Ac3) are present in the same bacterium, suggesting that their partially overlapping PAM profiles could confer redundancy in immune defense as reported for co-occurring type I and type III systems (Silas et al., 2017). The distinct PAM profiles that gave measurable activity in the deGFP silencing assay in TXTL confirmed the trends observed with the PAM wheels (**Figure 3B**). Given that type I-E systems represent one of the most abundant CRISPR-Cas subtypes in nature (Makarova et al., 2015), our initial characterization suggests that a far greater diversity of recognized PAM profiles likely exists.

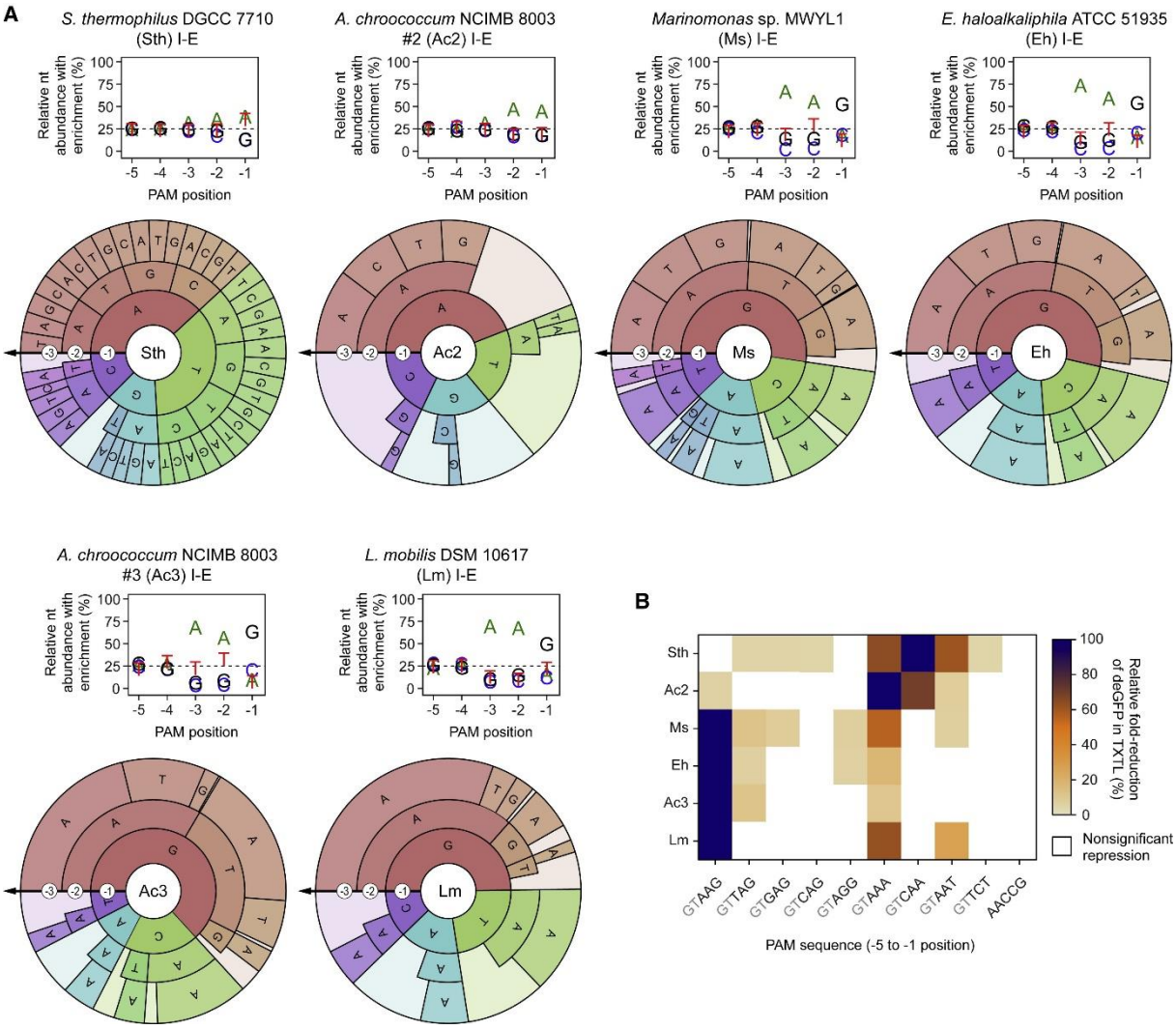


Figure 3: Harnessing the functional diversity of I-E CRISPR-Cas systems. (**A**) Nucleotide enrichment plots and PAM wheels for selected I-E systems subjected to PAM-DETECT. Ac1 (**Figure S1**), Ac2, and Ac3 are present in the same bacterium. Individual sequences comprising at least 2% of the PAM wheel are shown. Plots and PAM wheels are averages of duplicate independent experiments. (**B**) Comparison of PAM recognition between systems. Recognition was determined by assessing the repression of a deGFP reporter in TXTL. Values represent the mean of three TXTL experiments. Fold-reduction values that are not significantly different from that of the non-targeting crRNA control ($p > 0.05$) are shown as white squares. The PAM sequence showing the highest fold reduction for each system was set to 100%. AACCG matches the 3′ end of the repeat for most of the systems.

Extensive self-targeting I-C and I-F1 CRISPR-Cas systems in *Xanthomonas albilineans* are functionally encoded

Beyond nuclease mining, PAM-DETECT can be further applied to interrogate systems that deviate from traditional immune defense. Prominent examples are self-targeting CRISPR-Cas systems that encode crRNAs targeting chromosomal locations (Wimmer and Beisel, 2019). Although self-targeting is considered inherently incompatible with a functional CRISPR-Cas system (Gomaa et al., 2014; Stern et al., 2010; Vercoe et al., 2013), accumulating examples provide important counterpoints where the systems tolerate or even utilize self-targeting crRNAs (Li et al., 2021; Marino et al., 2018; Rauch et al., 2017; Watters et al., 2018; Yin et al., 2019). PAM-DETECT and TXTL therefore could accelerate the characterization of these unique systems.

We specifically focused on two extensively self-targeting CRISPR-Cas systems within the plant pathogen *Xanthomonas albilineans* CFBP7063. This bacterium encodes two CRISPR-Cas systems (I-C and I-F1), each harboring the full cohort of *cas* genes (**Figure 4A**). Furthermore, of the 64 spacers present across the six CRISPR arrays, 24 (38%) at least partially match sites in the chromosome or one plasmid (**Table S4**; **Figure S2A**) with a common set of flanking PAMs (**Figure 4B**). The ensuing questions are whether they could lead to autoimmunity through their self-targeting spacers.

We first performed PAM-DETECT using Cascade from both CRISPR-Cas systems to assess whether both are functionally encoded and what PAM profiles they recognize (**Figure 4C**). Either Cascade protected a small portion of the DNA library (~2% for I-C, ~6% for I-F1) from restriction digestion (**Figure S2B**), indicating functional expression of all Cascade subunits. PAM-DETECT further revealed PAM profiles that overlapped – but were not identical to – the I-C and I-F1 systems with even a moderately mapped PAM profile (Almendros et al., 2012; Leenay et al., 2016; Rao et al., 2017; Rollins et al., 2015; Tuminauskaite et al., 2020; Zheng et al., 2019). In particular, the I-C system from *X. albilineans* recognizes TTC followed by TTT and CTC, whereas the characterized I-C system from *Bacillus halodurans* recognizes TTC followed by CTC and then TCC (Leenay et al., 2016) and the I-C system from *Legionella pneumophila* recognizes TTC followed by TTT and CTT (Rao et al., 2017). Separately, the I-F1 system from *X. albilineans* recognizes CC as the strongest PAM similar to other I-F systems (Almendros et al., 2012; Rollins et al., 2015; Tuminauskaite et al., 2020; Zheng et al., 2019), although the *X. albilineans* system also can recognize a G and T but not an A at the -2 position and, in this case, could tolerate a CC PAM shifted upstream by one nucleotide. The recognized PAMs of both I-C and I-F1 systems further overlapped with the PAM sequences flanking the self-targets for 87% of the I-C self-targets (TTC, TTT, and CTC) and all I-F1 self-targets (CC

and CCT) (**Figures 4B** and **4C**). Testing individual PAMs in TXTL using gene repression with Cascade confirmed that the I-C system could recognize not only TTC but also TTT and CTC (**Figure 4D**). Similarly, the I-F1 system could recognize the CC PAM associated with almost all self-targets. PAM-DETECT thus can be implemented beyond I-E systems, and it indicated that the interrogated I-C and I-F1 systems in *X. albilineans* are capable of binding the vast majority of self-targeting sites in the genome.

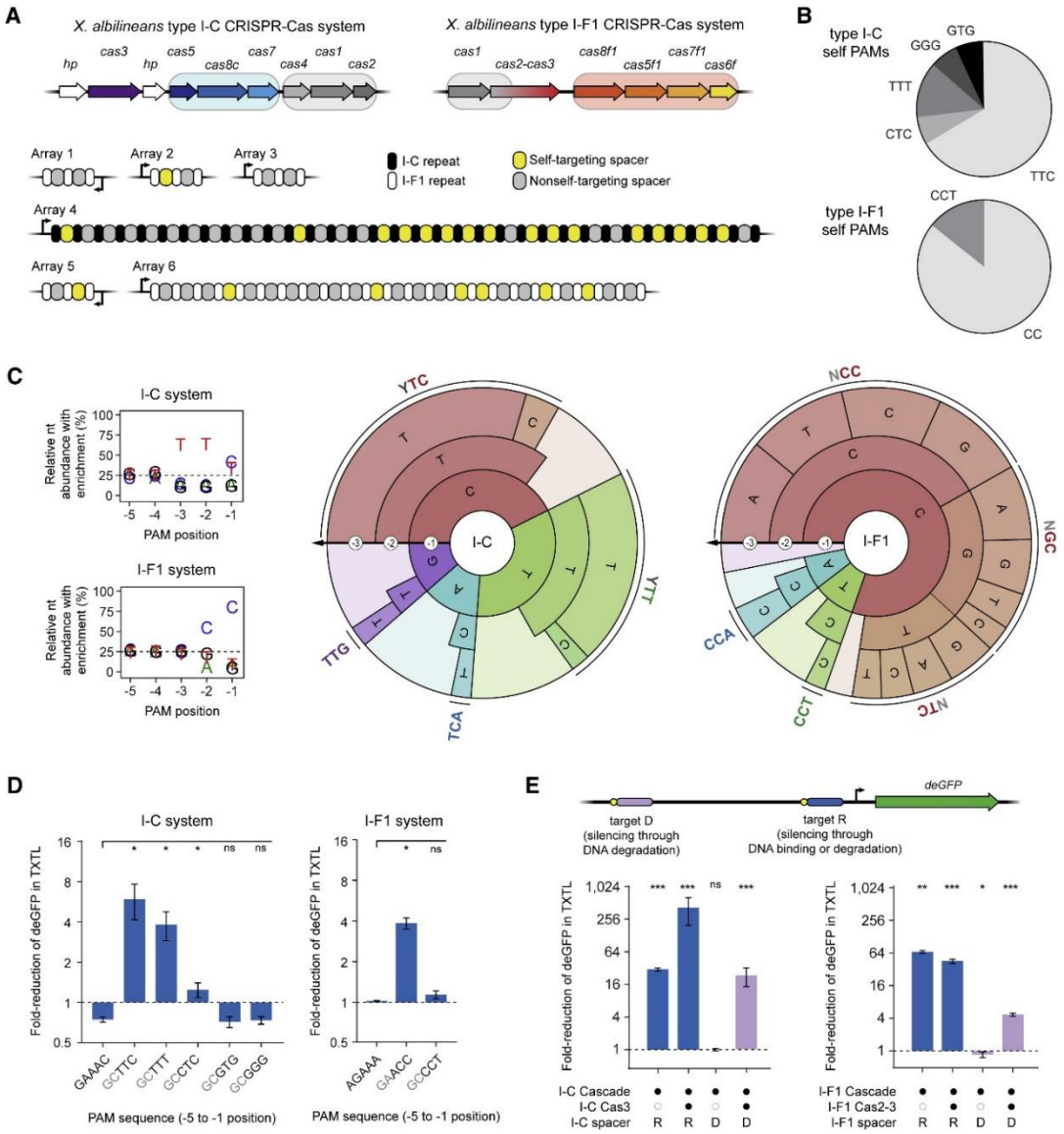


Figure 4: Interrogating extensive self-targeting for two Type I CRISPR-Cas systems in *X. albilineans*. (**A**) Overview of the I-C and I-F1 CRISPR-Cas systems and self-targeting spacers. The genes encoding the Cascade complex are in the light blue box (I-C) or the light orange box (I-F1), and the genes encoding the acquisition proteins are in the gray box. (**B**) Distribution of PAMs associated with the self-targets. See **Figure S2** for the self-target location and **Table S4** for the self-target sequences. (**C**) Nucleotide-enrichment plots and PAM wheels based on conducting PAM-DETECT. Individual sequences comprising at least 2% of the PAM wheel are shown. Plots and PAM wheels are averages of duplicate independent experiments. (**D**) Validation of PAMs associated with self-targets in TXTL. See **Figure 2E** for details. The self PAMs GAAAC (I-C) and AGAAA (I-F1) are references for statistical analyses. (**E**) Assessing DNA binding by Cascade and DNA degradation by Cas3 in TXTL. See **Figure 2E** for details about target R. Targeting far upstream of the promoter (target D) can reduce deGFP levels only through degradation of the plasmid. The non-targeting crRNA control is the reference for statistical analyses.
Errors bars in (D) and (E) indicate the mean and standard deviation of triplicate independent experiments. ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, and ns: $p > 0.05$.

If the Cas3 endonuclease for either system is functionally encoded and expressed, then recognition of these self-targeting sites should prove lethal to this bacterium. We therefore reconfigured the TXTL assay to evaluate the extent to which the I-C or I-F1 Cas3 could elicit DNA degradation (**Figure 4E**). The DNA target was placed in the backbone of the deGFP reporter ~200 bps upstream of the deGFP promoter (target D) flanked by a TTC (I-C) or CC (I-F1) PAM. Under this setup, loss of deGFP fluorescence would occur only if the backbone is nicked or cleaved, leading to DNA degradation by RecBCD (Marshall et al., 2018). For both CRISPR-Cas systems, this target site location resulted in targeted deGFP silencing following expression of Cascade and Cas3 but not Cascade alone (**Figure 4E**). The extent of deGFP silencing was less than that when targeting the deGFP promoter (target R), which can be explained by silencing through target R requiring Cascade binding versus silencing through target D requiring Cascade binding, Cas3 cleavage, and RecBCD degradation. Similar extents of deGFP silencing through Cas3 were observed when testing two native spacer:self-target pairs for each system (**Figure S2C**). We conclude that Cas3 is functionally encoded and would lead to lethal self-targeting unless Cascade or Cas3 is fully silenced in this bacterium, or another mechanism is in place to inhibit Cascade and/or Cas3 activity.

The I-F CRISPR transposon from *Vibrio cholera* recognizes an extremely flexible PAM profile

The demonstrated applicability of PAM-DETECT for diverse type I CRISPR-Cas systems created a unique opportunity: applying the same assay to CASTs. Of the three known CAST types (I-B, I-F, and V-K), two (I-B and I-F) rely on Cascade for DNA target recognition (Klompe et al., 2019; Saito et al., 2021). Recognition then leads to integration of the transposon DNA at a defined distance downstream of the target. Characterization of these systems to-date has relied on encoding a crRNA, all CRISPR and transposon components, and donor DNA flanked by the transposon ends in bacteria to achieve targeted transposition. However, the reliance of I-B and I-F CASTs on Cascade offers an opportunity to express only these CAST components as part of PAM-DETECT to elucidate rules for DNA target recognition.

We began with the I-F CAST from *Vibrio cholerae* (*V. cholerae*) (VcCAST) that exhibited robust DNA integration in *E. coli* and has been used for multiple applications in bacteria (Klompe et al., 2019; Vo et al., 2021; **Figure 5A**). Prior screening of individual potential PAM sequences via transposition in *E. coli* established a general preference for a C at the -2 position (Klompe et al., 2019), although a comprehensive PAM remained to be determined. We applied PAM-DETECT by expressing the three Cascade genes along with *tniQ* responsible for recruiting the other three transposon proteins (TnsA, TnsB, and TnsC), as the role of TniQ in DNA target recognition remained to be established (Klompe et al., 2019; Petassi et al., 2020;

Vo et al., 2021). High-Cascade conditions (3-nM plasmids and 16-h reaction time) protected 57% of the DNA, leading us to also perform PAM-DETECT with low-Cascade conditions (0.25-nM plasmids and 6-h reaction time) that exhibited 25% DNA protection (**Figure S3A**). The resulting PAM profile was remarkably flexible, with a preference for a C and bias against an A at the -2 position (**Figures 5B** and **S3B**). We further noticed that recognition of a G or T at the -2 position could be enhanced with a C at the -1 position or and A at the -3 position. Separately, an A at the -2 position could be rescued with a C at the -3 position (**Figures 5B** and **S3B**). Recognition was maintained even in the absence of *tniQ* (**Figures 5C** and **S3C**). The results from PAM-DETECT therefore suggest that Cascade from the I-F VcCAST recognizes a remarkably flexible PAM profile with preferences extending beyond a simple consensus sequence.

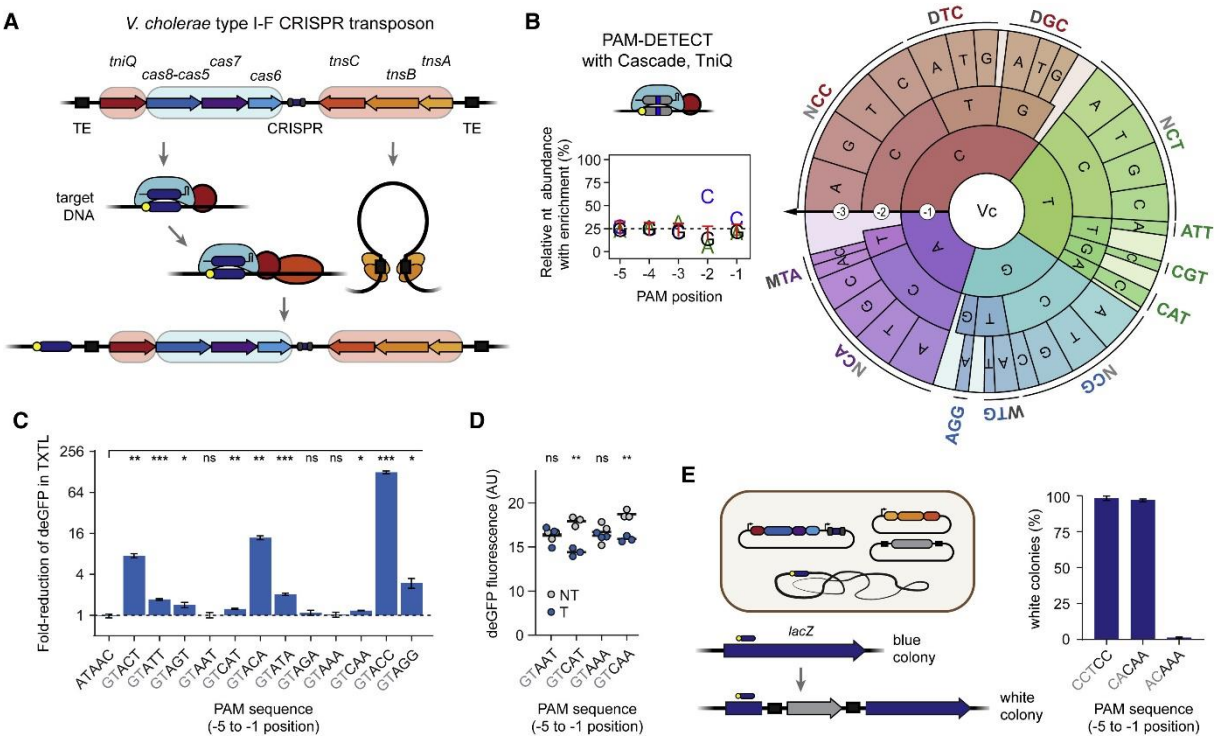


Figure 5: Interrogating the PAM profile of VcCAST. (**A**) Overview of VcCAST and its mechanism of transposition. (**B**) Nucleotide-enrichment plot and PAM wheel based on conducting PAM-DETECT with Cascade and TniQ. Individual sequences comprising at least 1% of the PAM wheel are shown. The plot and PAM wheel are averages of duplicate independent experiments. (**C**) Validation of PAMs in TXTL. See **Figure 2E** for details. The ATAAC self PAM is the reference for statistical analyses. (**D**) Individual measurements of endpoint deGFP levels in TXTL. Triplicate values are shown for selected PAMs with a targeting (T) or non-targeting (NT) crRNA. See (C) for details. (**E**) Validation of PAM recognition for DNA transposition in *E. coli*. Donor DNA is inserted within the *lacZ* gene, preventing the formation of blue colonies on IPTG and X-gal. The targets for the CAA and AAA PAMs are shifted by one nucleotide. See **Figure S3**.
Error bars in (C–E) indicate the mean and standard deviation of triplicate independent experiments. ***p < 0.001, **p < 0.01, *p < 0.05, and ns: p > 0.05.

To evaluate the PAM profile output by PAM-DETECT, we first employed our TXTL-based deGFP silencing assay (**Figure 5C**). Cascade most strongly recognized PAM sequences with C at the -2 position, with the greatest performance for CC. Deviating from CC reduced but did not eliminate measurable silencing as long as A was not present at the -2 and -3 positions.

Interestingly, whereas AAA and AAT yielded no measurable deGFP silencing, replacing A with C at the -3 position restored measurable silencing, albeit with low activity (**Figure 5D**). To assess how these small but measurable differences impact DNA transposition, we employed the previously described transposition system *in E. coli* (Klompe et al., 2019) conducted at 30°C for higher integration efficiency (Vo et al., 2021). Using this experimental setup, we found that CAA but not AAA yielded robust DNA transposition despite the targets being separated by only one base (**Figures 5E**, **S3D**, and **S3E**). Furthermore, the measured transposition efficiency was similar for CAA and CC. Therefore, even low levels of gene silencing with Cascade in TXTL can translate into efficient transposition in *E. coli*.

The I-B2 CRISPR transposon from *Rippkaea orientalis* recognizes a less flexible PAM profile

We next turned to I-B CASTs. Two examples of I-B CASTs were experimentally characterized recently, revealing that a second TniQ (renamed TnsD) drives DNA transposition at conserved sites flanking tRNAs or *glmS* independently of Cascade or a crRNA (Saito et al., 2021). Type I-B CASTs were further split into two subtypes (I-B1 and I-B2) based on TnsA and TnsB proteins being fused or separate, the general genetic organization of the CAST locus, and crRNA-independent insertion flanking tRNAs or *glmS*.

While exploring examples within the I-B CASTs, we noticed a further division within the I-B2 subtype typified by *tnsD* flanking the Cascade genes rather than the other transposon genes (**Figure 6A**). This organization more closely parallels that of I-B1 CASTs (Saito et al., 2021) but still possesses the *tnsAB* fusion and the presence of tRNAs flanking the CASTs indicative of I-B2 CASTs. The division of the I-B2 CASTs in two clades, denoted hereafter as I-B2.1 and I-B2.2, was further supported by the higher shared similarity of the TnsAB, TnsC, TnsD and TniQ proteins from systems that belong to each clade (**Figures 6A** and **S4A**). The Cascade protein sequences were similar across all I-B CASTs We chose the I-B2.2 CAST from *Rippkaea orientalis* (*R. orientalis*) (RoCAST) as a representative example to characterize.

PAM-DETECT yielded a PAM profile for the RoCAST Cascade dominated by ATG (**Figures 6B**, **S4B**, and **S4C**), matching the PAM recognized by the one previously characterized I-B2.1 CAST from *Peltigera membranacea cyanobiont* 210A (PmcCAST) (Saito et al., 2021). This match was expected given the high similarity (65%-81%) between the protein components forming PmcCAST and RoCAST Cascade. However, single-base perturbations to ATG (e.g., GTG) could be recognized by the RoCAST even under low-Cascade conditions. The TXTL-based deGFP silencing assay confirmed recognition of ATG as well as the single-base perturbations (**Figure 6C**). We further showed that PAM-DETECT can be applied to the

previously characterized I-B1 CRISPR transposon from *Anabaena variabilis* ATCC 29413 (AvCAST) (Saito et al., 2021; **Figures S5A** and **S5B**).

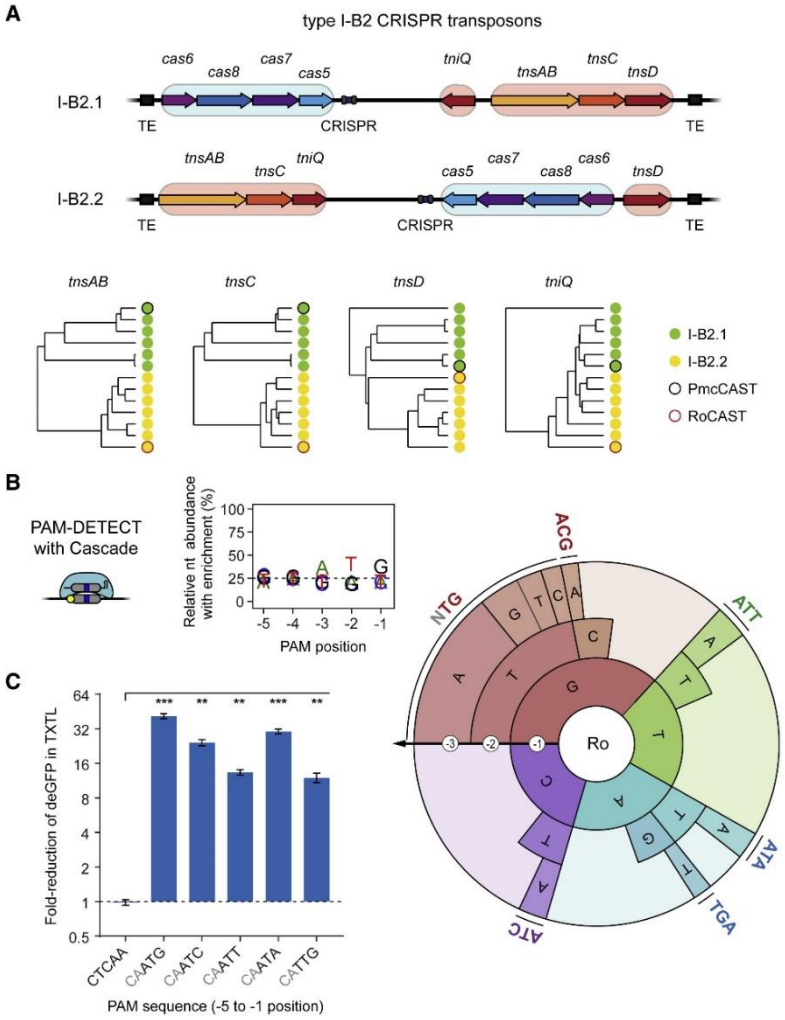


Figure 6: Interrogating PAM requirements of the *Rippkaea orientalis* I-B2.2 CRISPR transposon. (**A**) Overview of I-B2.1 and I-B2.2 CRISPR transposons. The two are divided based in the gene organization within each transposon. Phylogenetic trees are shown for the transposon genes. The PmcCAST from the I-B2.1 branch was previously characterized (Saito et al., 2021). (**B**) Nucleotide-enrichment plot and PAM wheel based on conducting PAM-DETECT with Cascade from RoCAST. Individual sequences comprising at least 2% of the PAM wheel are shown. The plot and PAM wheel are averages of duplicate independent experiments. (**C**) Validation of PAMs in TXTL. See **Figure 2E** for details. The CTCAA self PAM is the reference for statistical analyses.
Error bars in (C) indicate the mean and standard deviation of triplicate independent experiments. ***p < 0.001, **p < 0.01, *p < 0.05, and ns: p > 0.05.

DNA transposition by CRISPR transposons can be recapitulated in TXTL

An ensuing question is how insights into PAM recognition translate into DNA transposition. As *in vitro* or cell-based assays are slow and laborious, we instead sought to recapitulate transposition in TXTL (**Figure 7A**). We began with the VcCAST. Combining DNA constructs encoding a targeting single-spacer array, three Cascade genes, four transposon genes, donor DNA flanked by the transposon ends, and a target construct resulted in measurable DNA transposition in both orientations by PCR (**Figure S6A**), even though the transposition

efficiency in TXTL was too low to be effectively quantified by qPCR (**Table S5**). Sanger sequencing of the PCR products revealed the core transposon ends as well as the distance between the target site and insertion site that aligned with prior work (**Figure S6A**). We were also able to reconstitute transposition in TXTL for the I-B1 AvCAST (**Figure S5C**). TXTL thus can be used to recapitulate DNA transposition by CASTs, allowing elucidation of the transposon ends and insertion sites.

DNA transposition by CRISPR transposons can be recapitulated in TXTL

An ensuing question is how insights into PAM recognition translate into DNA transposition. As *in vitro* or cell-based assays are slow and laborious, we instead sought to recapitulate transposition in TXTL (**Figure 7A**). We began with the VcCAST. Combining DNA constructs encoding a targeting single-spacer array, three Cascade genes, four transposon genes, donor DNA flanked by the transposon ends, and a target construct resulted in measurable DNA transposition in both orientations by PCR (**Figure S6A**), even though the transposition efficiency in TXTL was too low to be effectively quantified by qPCR (**Table S5**). Sanger sequencing of the PCR products revealed the core transposon ends as well as the distance between the target site and insertion site that aligned with prior work (**Figure S6A**). We were also able to reconstitute transposition in TXTL for the I-B1 AvCAST (**Figure S5C**). TXTL thus can be used to recapitulate DNA transposition by CASTs, allowing elucidation of the transposon ends and insertion sites.

DNA transposition in TXTL with the *Rippkaea orientalis* CAST establishes a distinct branch within I-B2 CRISPR transposons

We next evaluated DNA transposition in TXTL with the I-B2.2 RoCAST (**Figure 7B**). Because the ends of this transposon were unclear, we constructed a donor DNA construct flanked by two 250-bp sequences predicted to contain the right and left RoCAST ends. We combined the donor DNA and target DNA flanked by an ATG PAM with constructs encoding the I-B2.2 Cascade genes, transposase genes, and a single-spacer CRISPR array. The TXTL reactions resulted in measurable crRNA-directed transposition in both orientations by PCR (**Figure 7C**). Sanger sequencing of the PCR products revealed the core transposon ends along with five bases that are duplicated as part of transposition (**Figure 7B**), similar to other CASTs (Klompe et al., 2019).

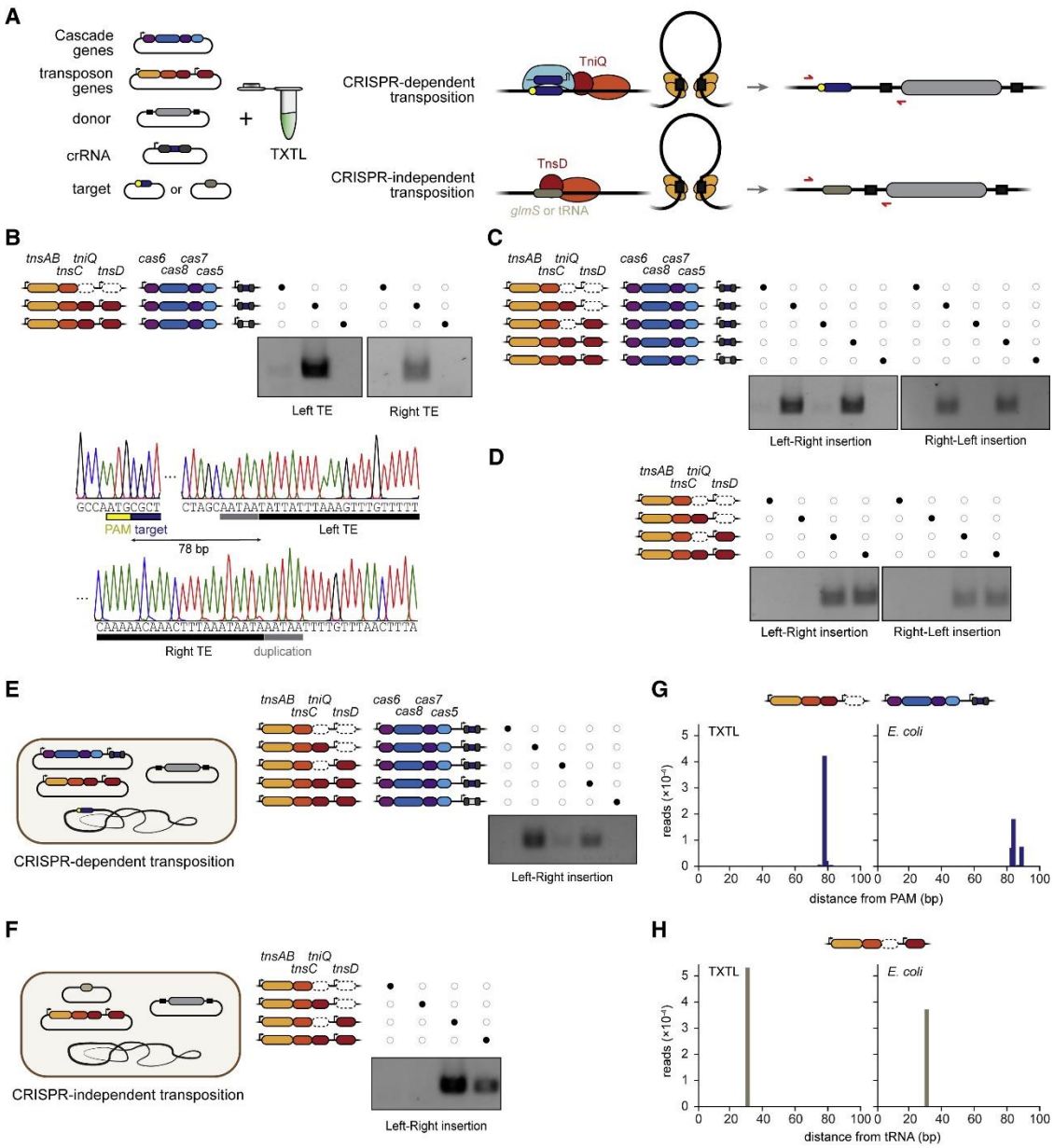


Figure 7: Investigating transposition of RoCAST in TXTL and in *E. coli*. (**A**) Overview of the TXTL-based transposition assay. (**B**) CRISPR-dependent transposition and determination of transposon ends and insertion distance using the TXTL-based transposition assay with RoCAST. PCR products are specific to the left-right orientation. (**C**) CRISPR-dependent transposition in TXTL. PCR products span the crRNA target site and the beginning of the cargo for both orientations of transposon insertion. (**D**) CRISPR-independent transposition in TXTL. PCR products span the end of the *tRNA-Leu* gene and the beginning of the cargo for both orientations of transposon insertion. (**E**) CRISPR-dependent transposition in *E. coli*. PCR products span the crRNA target site and the beginning of the cargo (left-right orientation). (**F**) CRISPR-independent transposition in *E. coli*. PCR products span the TnsD target site and the beginning of the cargo (left-right orientation). (**G** and **H**) Assessment of insertion distances for CRISPR-dependent transposition in TXTL and in *E. coli*. The constructs lacking *tnsD* (G) or *tniQ* (H) were used. Transposition was determined by NGS of the PCR product spanning the crRNA target site and the beginning of the cargo (left-right orientation) (G) or spanning the end of the *tRNA-Leu* gene and the beginning of the cargo (left-right orientation) (H).
All gel images are representative of at least duplicate independent experiments.

Recent work revealed that I-B CASTs possess two distinct modes of transposition: CRISPR-dependent transposition through TniQ and DNA targeting by Cascade and CRISPR-independent transposition through TnsD (Saito et al., 2021). For CRISPR-dependent transposition, TXTL reactions yielded a more pronounced PCR product when including rather

than excluding TniQ, although modest but detectable crRNA-dependent transposition was detected even in the absence of TniQ and TnsD (**Figures 7B**, **7C** and **S6B**). By contrast, TniQ was reported to be required for crRNA-dependent transposition by the I-B1 AvCAST (**Figure S5C**) and the I-B2.1 PmcCAST (Saito et al., 2021). For CRISPR-independent transposition, we swapped the crRNA target for the *tRNA-Leu* gene naturally flanking RoCAST in the *R. orientalis* genome. CRISPR-independent transposition was detected in both orientations (**Figure 7D**). Transposition required TnsAB, TnsC, and TnsD, whereas removing TnsD or replacing it with TniQ eliminated transposition.

We finally asked how the properties of RoCAST observed in TXTL translate *in vivo* (**Figures 7E** and **7F**). For CRISPR-dependent transposition, we targeted the *lacZ* gene in the *E. coli* genome at a site flanked by an ATG PAM. Over-expressing Cascade proved to be cytotoxic, reflecting challenges to characterizing CASTs *in vivo*, although the cytotoxicity could be relieved with minimal induction of Cascade expression. In line with the TXTL results, CRISPR-dependent transposition was measurable by PCR in *E. coli* strains expressing the Cascade, TnsAB, TnsC and TniQ proteins, albeit only for the left-to-right insertion orientation (**Figure 7E**). Removing TnsD boosted this mode of transposition (**Figure 7E**). Somewhat paralleling the TXTL results, less efficient transposition was measurable by PCR in the absence of TniQ but not both TniQ and TnsD (**Figures 7E** and **S6C**). For CRISPR-independent transposition, we targeted a vector carrying the terminal region of the *tRNA-Leu* gene from the *R. orientalis* genome. Matching the TXTL results, TnsAB, TnsC, and TnsD proteins were necessary for transposition (**Figure 7F**).

To compare the insertion distances between the target and the inserted donor DNA in TXTL and in *E. coli*, the PCR products were subjected to NGS. For CRISPR-dependent transposition, transposition in TXTL consistently occurred 78 bps downstream of the PAM, while transposition in *E. coli* principally occurred within a window of 83-89 bps downstream of the PAM (**Figure 7G**). The difference may be attributed to the use of different target sites and insertion contexts as was previously reported for the I-B1 AvCAST (Saito et al., 2021). For CRISPR-independent transposition, transposition in TXTL and in *E. coli* both occurred 31 bps downstream of the *tRNA-Leu* gene (**Figure 7H**). The insertion distances for both modes of transposition are comparable to the insertion windows identified for the other characterized I-B2 system (Saito et al., 2021). Overall, these findings demonstrate that insights from TXTL-based transposition translate into *in vivo* settings.

Discussion

PAM-DETECT offers important advantages over current cell-based and *in vitro*-based methods that should accelerate characterization of class 1 CRISPR-Cas systems and

transposons. As one advantage, PAM-DETECT can be completed in under 1 day when starting from purified DNA constructs and ending with amplicons for NGS. By contrast, cell-based methods require DNA transformation, culturing, growth, and in some cases screening methods that require access to specialized instruments (e.g., fluorescence-activated cell sorting for PAM-SCANR; Leenay et al., 2016) that can require days to weeks. *In vitro* assays can also require more time and extensive optimization irrelevant to TXTL, such as combining the constructs into a small set of compatible plasmids with appropriate expression, purifying components, and tackling issues of toxicity. As a second advantage, the ability to conduct reactions in a few microliters allows PAM-DETECT to be readily scaled, facilitating the parallel interrogation of multiple systems under different reaction conditions. Given these advantages, TXTL-based characterization of class 1 systems could become a standard means to explore these abundant and diverse systems.

We further leveraged TXTL to accelerate the validation and extension of our results from PAM-DETECT. We frequently employed a deGFP repression assay in which target binding by Cascade blocks deGFP expression. One potential limitation is that binding may not correspond to DNA degradation through Cas3, as was reported to some degree for DNA binding and degradation by the I-E system (Xue et al., 2015). However, as part of characterizing the self-targeting CRISPR-Cas systems in *X. albilineans*, we showed that the repression assay could be readily modified to specifically assess DNA degradation by Cas3. Finally, we showed that DNA transposition by CASTs could be fully recapitulated in TXTL. We were able to recapitulate CRISPR-dependent and CRISPR-independent transposition by I-B and I-F CASTs, suggesting that transposition could be recapitulated for V-K CASTs in TXTL as well (Saito et al., 2021; Strecker et al., 2019). With these additional assays in place, TXTL can be applied well beyond PAM determination.

One major application we pursued was mining the natural diversity of I-E CRISPR-Cas systems. Using PAM-DETECT, we evaluated 11 different systems representing diverse sequences within the variable L1 loop of the Cas8e protein. The identified PAMs deviated from those associated with *E. coli*'s I-E system, suggesting that a far broader range of PAMs could be revealed by further interrogating the diversity of these systems. Whether the diversity parallels that observed for Cas9 nucleases remains to be seen and could reflect the distinct forces that shaped the evolution of each system type (Gasiunas et al., 2020). A similar approach could be particularly powerful for mining I-C and I-Fv Cascade complexes that require the fewest number of canonical Cas proteins (Hochstrasser et al., 2016; Pausch et al., 2017). Complexes could be mined exhibiting not only unique PAM preferences but also smaller proteins, altered temperature ranges, or enhanced binding and cleavage activities. Given the proliferation of engineered single effectors with altered PAM recognition (Collias and Beisel,

2021), TXTL could be applied to characterize any similarly engineered variants of type I systems.

Beyond mining orthologs within a CRISPR-Cas subtype, PAM-DETECT offered a powerful means to interrogate CRISPR-Cas systems with potentially unique properties. We specifically focused on a I-C system and a I-F1 system present in *X. albilineans* that encode a large repertoire of self-targeting spacers. Although genetic deactivation of the CRISPR machinery is thought to be a common means of resolving otherwise lethal self-targeting (Stern et al., 2010), we showed that Cascade and Cas3 were functionally encoded and could recognize PAMs flanking the vast majority of self-targets. These findings instead suggest that the expression or activity of the CRISPR machinery is inhibited, preventing lethal self-targeting. We speculate that anti-CRISPR proteins are responsible for the lack of autoimmunity, as VirSorter (Roux et al., 2015) predicts three prophage regions in the genome of *X. albilineans* (**Figure S2A**) and anti-CRISPR proteins (Acrs) are known to often be located in such regions (Davidson et al., 2020). Future work could interrogate what prevents not only these systems from lethal self-targeting not only in *X. albilineans* but also CRISPR-Cas systems with self-targeting spacers in many other organisms. This work could reveal additional classes of Acrs as well as instances of CRISPR-Cas systems performing functions extending beyond adaptive immunity.

As a final example, we applied TXTL to characterize a distinct branch of I-B2 CASTs. When exploring I-B2 CASTs, we noticed a clear division in the genetic organization of these CASTs that paralleled phylogenetic trees for the transposon genes. We found that CRISPR-dependent transposition could occur in TXTL in the absence of TniQ for one branch (I-B2.2), contrasting with the essential role of TniQ described for the other branch (I-B2.1) and subtype (I-B1) (Saito et al., 2021). The type V-K CAST from *Scytonema hofmanni* (ShCAST) was similarly shown to transpose *in vitro* in the absence of TniQ (Strecker et al., 2019), whereas a recent structural study showed that ShTniQ takes part in the formation of the ShCAST transposition complex but is not required for the complex's catalytic function (Querques et al., 2021). Regardless of its biological relevance, TniQ-independent transposition likely reflects distinct biomolecular mechanisms and interactions for this branch of I-B CASTs that further support some division in categorization. As only a small number of CASTs have been characterized to-date, further exploring these unique mobile genetic elements could reveal additional properties and provide CASTs for further technological development and application.

Limitations of the study

Although PAM-DETECT offers numerous advantages over existing PAM determination assays, it comes with some limitations. First, PAM-DETECT is best suited to systems from

mesophilic organisms because the activity of our TXTL system is restricted to 25°C-42°C (Sun et al., 2013), although the DNA-binding and restriction steps could be conducted at elevated temperatures. Second, no simple and rapid means exist to quantify protein production and complex formation in TXTL, which would help differentiate between poor expression and poor binding activity. Third, PAM-DETECT is less suited to probe PAM dependencies for imperfect targets, where bound Cascade would be more likely to dissociate and allow cleavage by the restriction enzyme. Finally, when applying TXTL to characterize DNA transposition by CASTs, we identified some discrepancies between our TXTL results and our *in vivo* results, including the transposition efficiency and the necessity of TniQ for RoCAST.

Star Methods

Key Resource Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
For bacterial strains see Table S6	N/A	N/A
Chemicals, peptides, and recombinant proteins		
PacI, recombinant	New England Biolabs	Cat#R0547S
CutSmart® Buffer	New England Biolabs	Cat#B7204S
Proteinase K (illustra™ Bacteria genomicPrep Mini-Spin-Kit)	Cytiva	Cat#28-9042-58
AMPure XP	Beckman Coulter	Cat#A63881
Isopropyl-ß-D-thiogalactopyranosid (IPTG)	Carl Roth	Cat#2316.4; CAS: 367-93-1
Critical commercial assays		
myTXTL Sigma 70 Master Mix Kit	Arbor Bioscience	Cat#507096
KAPA HiFi HotStart ReadyMix PCR Kit	KAPA Biosystems	Cat# KK2600
SsoAdvanced Universal SYBR Green Supermix	Bio-Rad Laboratories	Cat#1725271
Q5 Hot Start High-Fidelity 2X Master Mix	New England Biolabs	Cat#M0494L
illustra™ Bacteria genomicPrep Mini-Spin-Kit	Cytiva	Cat#28-9042-58
Deposited data		
Raw and analyzed data	This study	GEO: GSE179614
Raw Gel Images	This study	Mendeley Data: https://doi.org/10.17632/99g7j7rz7r.1

Key Resource Table (continued)

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides		
For oligonucleotides see Table S6	N/A	N/A
Recombinant DNA		
For plasmids see Table S6	N/A	N/A
Software and algorithms		
PAM wheel script	(Ondov et al., 2011)	https://github.com/marbl/Krona/wiki
PAM analysis (R)	(Marshall et al., 2018)	https://bitbucket.org/csmaxwell/ crispr-txtl-pam-counting-script/ src/master/example-analysis/
PROMALS3D	(Pei et al., 2008)	http://prodata.swmed. edu/promals3d/
cd-hit	(Huang et al., 2010)	https://github.com/ weizhongli/cdhit/wiki
mcl algorithm	**(Enright et al., 2002)**	**https://github.com/micans/mcl**
Gismo	(Neuwald and Liu, 2004)	http://gismo.igs.umaryland.edu/
blast+	(Altschul et al., 1990)	http://blast.ncbi.nlm.nih.gov// blast.ncbi.nlm.nih.gov/Blast.cgi
JalView	(Waterhouse et al., 2009)	http://www.jalview.org/ getdown/release/
T-Coffee	(Di Tommaso et al., 2011)	http://tcoffee.crg.cat/
Hmmer	(Eddy, 2009)	http://hmmer.org/
VirSorter v1.0.3	(Roux et al., 2015)	https://github.com/simroux/VirSorter
Other		
Detailed protocol for PAM-DETECT	this study	Methods S1

Method details

Plasmid construction

Standard cloning methods Gibson Assembly, Site Directed Mutagenesis (SDM) and Golden Gate were used to clone plasmids used in TXTL experiments. pPAM_library containing a PAM library with five randomized nucleotides was generated by SDM on p70a-deGFP_PacI with primers FW531 and FW532 (**Table S6**). Single-spacer CRISPR arrays were generated either with Golden Gate adding spacer sequences in a plasmid containing two repeat sequences interspaced by two BaeI or BbsI restriction sites or by SDM on pEc_gRNA1, pEc_gRNA2 or pEc_gRNAnt to change the repeat sequences to match the tested CRISPR systems. Plasmids harboring different PAM sequences for PAM validation assays were generated by SDM on p70a-deGFP_PacI. To generate plasmids encoding *X. albilineans* type I-C and type I-F1 Cas proteins, genomic DNA isolated from *Xanthomonas albilineans* CFBP7063 was PCR amplified using Q5 Hot Start High-Fidelity 2X Master Mix (NEB) and cloned into pET28a using Gibson

Assembly. All other plasmids were generated with Gibson Assembly or SDM (**Table S6**). All constructed plasmids were verified with Sanger sequencing.

For the VcCAST *in vivo* transposition experiments we cloned into the previously described pSL0284 vector (Klompe et al., 2019) two spacers targeting the *lacZ* gene of the *E. coli* BL21 (DE3) genome, yielding the pQCas_CAA and pQCas_AAA vectors. The protospacer targeted by the former vector has a 5′ CAA PAM, whereas the protospacer targeted by the latter vector has a 5′ AAA PAM.

For the RoCAST *in vivo* transposition experiments, genes encoding the *Rippkaea orientalis tnsAB, tnsC, tnsD and tniQ* were synthesized (Twist Bioscience) and cloned in the pET24a vector in various combinations, resulting in the construction of the pRoTnsABC, pRoTnsABCD, pRoTnsABCQ, pRoTnsABCDQ vectors (**Table S6**). The *Rippkaea orientalis* Cascade operon (*cas6, cas8, cas7, cas5*) was synthesized (Twist Bioscience) and cloned into the pCDFDuet-1 vector together with a *gfp* gene flanked by two BsaI restriction sites and the corresponding CRISPR direct repeats. Into the resulting pRoCascade_gfp vector we cloned a spacer targeting the *lacZ* gene of the *E. coli* BL21 (DE3) genome and a non-targeting control spacer, constructing the pRoCascade_T (targeting) and pRoCascade_NT (non-targeting) vectors, respectively (**Table S6**). DNA fragments encoding the right and left RoCAST ends were synthesized (IDT) and cloned into the pUC19 vector flanking a *cmr* gene, yielding pRoDonor (**Table S6**). A 105-bp long DNA fragment from the *Rippkaea orientalis* genome, encoding the region which is located right upstream of the left end of RoCAST and includes the last 74 bp of the *tRNA-Leu* gene, was synthesized (IDT) and cloned into the pCDFDuet-1 vector, resulting in the construction of the pRoTarget vector (**Table S6**).

PAM-DETECT

Methods S1 contains a protocol for performing PAM-DETECT. A plasmid with five randomized nucleotides flanking a target site covering a PacI restriction enzyme recognition site was constructed as described before. If Cas proteins required for Cascade formation were encoded on separate plasmids, a MasterMix with the required Cas protein encoding plasmids in their stoichiometric amount was prepared beforehand. Thereby, a stoichiometry of $Cas8e_1$-$Cse2_2$-$Cas7_6$-$Cas5_1$-$Cas6_1$ was used for all Type I-E systems. A 6 µL TXTL reaction was assembled consisting of 3-nM (high-Cascade) or 0.25-nM (low-Cascade) of the Cascade-encoding plasmid or the Cascade MasterMix, 4.5 µL myTXTL Sigma 70 Master Mix, 0.2 nM pET28a_T7RNAP, 0.5 mM IPTG, 1 nM gRNA-encoding plasmid and 1 nM pPAM_library. A negative control containing all components from the reaction besides the Cascade plasmids and the gRNA-expressing plasmid was included. PAM-DETECT assays assessing either the type I-C or the type I-F1 system in *X. albilineans* were lacking IPTG in their reactions. TXTL

reactions were incubated at 29°C for 6 h or 16 h. The samples were diluted 1:400 in nuclease-free H2O. 500 µL were digested at 37°C with PacI (NEB) at 0.09 units/µL in 1x CutSmart Buffer (NEB) for 1 h and 500 µL were used as a “non-digested” control by adding nuclease-free H2O instead of PacI. After inactivation of PacI at 65°C for 20 min, 0.05 mg/mL proteinase K (Cytiva) was added and incubated at 45°C for 1 h. After inactivation of Proteinase K at 95°C for 5 min, remaining plasmids were extracted via standard EtOH precipitation. Illumina adapters with unique dual indices were added by two amplification steps with KAPA HiFi HotStart Library Amplification Kit (KAPA Biosystems) and purified by AMPure XP (Beckman Coulter) after every PCR reaction. The first PCR reaction adds the Illumina sequencing primer sites with primers that can be found in **Table S6** using 15 µL of the EtOH-purified samples in a 50 µL reaction and 19 cycles. The second PCR adds the unique dual indices and the flow cell binding sequence using 1 ng purified amplicons generated with the first PCR using 18 cycles. The samples were submitted for next-generation sequencing with 50 bp paired-end reads with 1.25 or 2.0 million reads per sample on an Illumina NovaSeq 6000 sequencer. PAM wheels were generated according to Leenay et al. (2016) and Ondov et al. (2011). Nucleotide enrichment plot generation was adapted to the PAM analysis script from Marshall et al. (2018) by changing the script to visualize the probability of a given nucleotide at a given position. We started by normalizing the read counts of every PAM with the total number of reads. Next, we calculated the fold change for every PAM by determining the ratio of digested sample reads over undigested sample reads. The ratios for a given nucleotide at a given position were added up and divided by the sum of the ratios of all nucleotides at that given position and multiplied by 100. 25% represents no enrichment/depletion. All PAM-DETECT assays were done in duplicates and PAM wheel and nucleotide enrichment plots show averages. The generated NGS data have been deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) and are accessible through GEO Series accession number GSE179614 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE179614).

qPCR analysis of PAM-DETECT

To assess the remaining amount of PAM-library containing plasmid after conducting PAM-DETECT, qPCR was performed using SsoAdvanced Universal SYBR Green Supermix (Biorad) in 10 µL reactions. The reactions were quantified using a QuantStudio Real-Time PCR System (Thermo Fisher) with an annealing temperature of 68°C according to manufacturers’ instructions. All samples were prepared by using the liquid handling machine Echo525 (Beckman Coulter).

deGFP repression assays in TXTL

To assess activity of CRISPR-Cas systems, deGFP-repression assays in 3 µL TXTL reactions were conducted, measuring deGFP-expression over time in a 96-well V-bottom plate with BioTek Synergy H1 plate reader (BioTek) at 485/528 nm excitation/emission (Shin and Noireaux, 2012). All TXTL samples were either prepared by hand or by using the liquid handling machine Echo525 (Beckman Coulter).

3 µL TXTL reactions for PAM validation assays were prepared containing Cascade plasmid concentrations according to **Table S2**. If Cas proteins required for Cascade formation were encoded on separate plasmids, a MasterMix with the required Cas protein encoding plasmids in their stoichiometric amount was prepared beforehand. Thereby, a stoichiometry of $Cas5_1$-$Cas8_1$-$Cas7_7$ was used for *X. albilineans* type I-C, $Cas8f1_1$-$Cas5f1_1$-$Cas7f1_6$-$Cas6f_1$ was used for *X. albilineans* type I-F1 and $Cas8e_1$-$Cse2_2$-$Cas7_6$-$Cas5_1$-$Cas6_1$ was used for all type I-E systems. Other components included in the TXTL reactions were 2.25 µL myTXTL Sigma 70 Master Mix, 0.2 nM p70a_T7RNAP, 0.5 mM IPTG and 1 nM gRNA-encoding plasmid. After a 4 h pre-incubation at 29°C or 37°C that allowed the ribonucleoprotein complex of Cascade and crRNA to form, 1 nM reporter plasmid (pGFP_XXXXX) with various PAM sequences in close proximity to the promoter driving deGFP expression was added to the reaction to ensure Cascade-binding would lead to deGFP inhibition. The reactions were incubated for additional 16 h at 29°C or 37°C while measuring deGFP expression. The gRNAs were constructed to target a protospacer within the *degfp* promoter located adjacent to the various PAM sequences.

To test the cleavage and/or binding ability of the type I-C and the type I-F1 systems in *X. albilineans*, 3 µL TXTL assays were conducted containing Cascade-encoding plasmids in the stoichiometry as mentioned before. To test binding ability, 2.25 µL myTXTL Sigma 70 Master Mix, 0.2 nM p70a_T7RNAP, 0.5 mM IPTG, 1 nM gRNA1-, gRNA2-, gRNA6-, or gRNAnt-encoding plasmid and 1 nM or 0.25 nM Cascade MasterMix was added to a TXTL reaction for the type I-C and type I-F1 system, respectively. To test cleavage ability, 2.25 µL myTXTL Sigma 70 Master Mix, 0.2 nM p70a_T7RNAP, 0.5 mM IPTG, 1 nM gRNA1-, gRNA2, gRNA4, gRNA5, or gRNAnt-encoding plasmid, 1 nM Cascade MasterMix and 0.5 nM or 0.25 nM pXalb_IC_Cas3 or pXalb_IF_Cas2-3 was added to a TXTL reaction for the type I-C and type I-F1 system, respectively. After a 4 h pre-expression at 29°C, 1 nM p70a_deGFP reporter plasmid, p70a_deGFP_ICST1, p70a_deGFP_ICST2, p70a_deGFP_IF1ST1, or p70a_deGFP_IF1ST2 was added to the reactions and incubated for additional 16 h at 29°C while measuring deGFP-fluorescence. gRNA1 is designed to target a protospacer within the promoter driving deGFP expression adjacent to a type I-C TTC or a type I-F1 CC PAM to ensure Cascade-binding would lead to deGFP-inhibition. gRNA2, gRNA4 and gRNA5 were

designed to target a protospacer adjacent to a type I-C TTC or a type I-F1 CC PAM upstream of the promoter to ensure cleavage of the targeted plasmid would result in deGFP-inhibition whereas binding-only would result in deGFP-production. gRNAnt represents a non-targeting control.

Prophage prediction

Prophage regions in the genome of *X. albilineans* CFBP7063 were predicted using VirSorter v1.0.3 (Roux et al., 2015). Prophage sequences with category 5 and 6 were found and are shown in **Figure S2A**.

Transposition in TXTL

To assess crRNA-dependent transposition of the *Vibrio cholerae* Tn6677 I-F CAST in TXTL, 5 µL TXTL reactions containing 3.75 µL myTXTL Sigma 70 Master Mix, 0.2 nM p70a_T7RNAP, 0.5 mM IPTG, 1 nM of the previously described donor plasmid (pSL0527), 2 nM of the previously described TnsABC-plasmid (pSL0283) (Klompe et al., 2019), 1 nM p70a_deGFP and 1 nM pVch_IF_CasQ_gRNA3 or pVch_IF_CasQ_gRNAnt were prepared. The reactions were incubated at 29°C for 16 h. Transposition events were detected in a 1:400 dilution of the TXTL reaction by PCR amplification using Q5 Hot Start High-Fidelity 2X Master Mix (NEB) and combinations of donor DNA and genome specific primers. Transposition was verified by Sanger sequencing (**Table S6**).

crRNA-independent transposition of RoCAST in TXTL was performed in 3 µL TXTL reactions consisting of 2.25 µL myTXTL Sigma 70 Master Mix, 0.2 nM p70a_T7RNAP, 0.5 mM IPTG, 1 nM pRoTarget, 1 nM pRoDonor and 1 nM pRoTnsABC, pRoTnsABCD, pRoTnsABCQ or pRoTnsABCDQ. The reactions were incubated at 29°C for 16 h. Transposition events were detected in a 1:100 dilution of the TXTL reaction by PCR amplification using Q5 Hot Start High-Fidelity 2X Master Mix (NEB) and combinations of donor DNA and genome specific primers (**Table S6**). Transposition was verified by Sanger sequencing.

qPCR analysis of VcCAST transposition efficiency in TXTL

Pairs of VcCAST cargo and target plasmid-specific primers were designed to amplify 106-151 bp long fragments, resulting from VcCAST induced transposition in either orientation in TXTL reactions. The qPCR reactions were performed using SsoAdvanced Universal SYBR Green Supermix (Biorad) in 10 µL reactions. The reactions were quantified using a QuantStudio Real-Time PCR System (Thermo Fisher) with an annealing temperature of 60°C according to manufacturers' instructions.

We cloned the VcCAST cargo from the donor plasmid (pSL0527) into a position 50 bp downstream of the protospacer in the pGFP_GTACC target plasmid. The cargo was cloned in

both orientations (left to right and right to left transposon end) resulting in the construction of the pGFP_GTACC_LR and pGFP_GTACC_RL plasmids that mimic the two products of successful VcCAST-based transposition in TXTL. We then performed control TXTL transposition reactions, as previously described, altering the ratios of pGFP_GTACC to pGFP_GTACC_LR or pGFP_GTACC_RL plasmids in each reaction simulating variable transposition efficiencies. We tested the qPCR primer pairs with each of the control TXTL reactions and we detected transposition products in either orientation and at efficiencies as low as 0.5%.

We performed TXTL transposition reactions, as previously described, using either a targeting or a non-targeting gRNA expressing plasmid and three target plasmids with distinct PAM sequences (ACC, CAA, and AAA respectively) as their defining difference (**Table S5**). Samples from each TXTL reaction were analyzed by qPCR. The transposition efficiency for each reaction and for each primer pair/orientation was calculated as $2^{-\Delta\Delta Ct}$, where ΔΔCt was the difference between the ΔCt of an experimental TXTL reaction and a control TXTL reaction that contained either the pGFP_GTACC_LR or the pGFP_GTACC_RL as the only target plasmid (**Table S5**).

Transposition *in vivo*

For the crRNA-dependent transposition *in vivo* using the I-F CAST from *Vibrio cholerae* Tn6677, we employed the previously described transposition system (Klompe et al., 2019). We electroporated 30 ng of the pSL0283 vector with 30 ng of the pSL0527 vector and 30 ng of either the pQCas_CAA or pQCas_AAA vector into *E. coli* BL21(DE3) electrocompetent cells. We plated a fraction of each electroporation mixture on 100 mg/mL ampicillin, 50 mg/mL spectinomycin, 50 mg/mL kanamycin, 0.1 mM IPTG and 100 µg/mL X-gal containing LB-agar plates. The plates were incubated for 24 h at 30°C and the formed colonies were subjected to blue/white screening. Transposition events were identified by colony PCR using Q5 Hot Start High-Fidelity 2X Master Mix (NEB) and genome specific primers (**Table S6**).

For the crRNA-dependent transposition *in vivo* using RoCAST, we electroporated 30 ng of either pRoCascade_T or pRoCascade_NT vector with 30 ng of pRoDonor and 30 ng of either pRoTnsABC, pRoTnsABCD, pRoTnsABCQ or pRoTnsABCDQ vector into *E. coli* BL21(DE3) electrocompetent cells. We plated a fraction of each electroporation mixture on 100 mg/mL ampicillin, 50 mg/mL spectinomycin, and 50 mg/mL kanamycin containing LB-agar plates. The plates were incubated for 20 h at 37°C and the formed colonies were scraped and resuspended in LB liquid medium. A fraction of each cell suspension was re-plated on LB-agar plates supplemented with 100 mg/mL ampicillin, 50 mg/mL spectinomycin, 50 mg/mL kanamycin and 0.01 mM IPTG for induction of the expression of the Cascade and transposase proteins. The

plates were incubated 20 h at 37°C and all the formed colonies were scraped and resuspended in LB liquid medium. A fraction of each cell suspension was subjected to gDNA isolation using the illustra Bacteria genomicPrep Mini Spin Kit (Cytiva). Transposition events were identified by PCR using Q5 Hot Start High-Fidelity 2X Master Mix (NEB) and combinations of donor DNA and genome specific primers (**Table S6**).

For the crRNA-independent *in vivo* transposition using RoCAST, we electroporated 30 ng of the pRoTarget with 30 ng of pRoDonor and 30 ng of either the pRoTnsABC, pRoTnsABCD, pRoTnsABCQ or pRoTnsABCDQ vector into *E. coli* BL21(DE3) electrocompetent cells. We plated a fraction of each electroporation mixture on 100 mg/mL ampicillin, 50 mg/mL spectinomycin, and 50 mg/mL kanamycin containing LB-agar plates. The plates were incubated for 20 h at 37°C and the formed colonies were scraped and resuspended in LB liquid medium. A fraction of each cell suspension was re-plated on LB-agar plates supplemented with 100 mg/mL ampicillin, 50 mg/mL spectinomycin, 50 mg/mL kanamycin and 0.01 mM IPTG for induction of the expression of the transposase proteins. The plates were incubated 20 h at 37°C and all the formed colonies were scraped and resuspended in LB liquid medium. A fraction of each cell suspension was subjected to gDNA isolation using the illustra Bacteria genomicPrep Mini Spin Kit (Cytiva). Transposition events were identified by PCR using Q5 Hot Start High-Fidelity 2X Master Mix (NEB) and combinations of donor DNA and pRoTarget specific primers (**Table S6**).

Assessing transposition insertion point

To assess the exact insertion point of *Rippkaea orientalis* I-B2.2 CAST, *in vivo* and *in vitro,* transposition assays were conducted as previously described and the transposition products were PCR amplified and sent for next-generation sequencing. Illumina adapters with unique dual indices were added by two amplification steps with KAPA HiFi HotStart Library Amplification Kit (KAPA Biosystems) and each amplicon was purified by AMPure XP (Beckman Coulter). The first PCR reaction adds the Illumina sequencing primer sites with primers that can be found in **Table S6**, the second PCR adds the unique dual indices and the flow cell binding sequences. 2 µL of 1:100 dilutions were used in a 50 µL PCR reaction to amplify TXTL reactions using either 19 or 30 cycles. 50 ng of genomic DNA were used in a 50 µL PCR reaction to amplify *in vivo* transposition with either 19 or 30 cycles. 1 ng of purified TXTL or *in vivo*-amplicon were subjected to the second PCR using 18 cycles.

Library-pools consisting of six samples were submitted for next-generation sequencing with 300 bp paired-end reads with 0.15 million reads on an Illumina MiSeq machine.

The generated NGS data have been deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) and are accessible through GEO Series accession number GSE179614 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE179614).

Quantification and statistical analysis

deGFP repression assays in TXTL

The fluorescence background was subtracted from the endpoint deGFP values with TXTL samples consisting of only myTXTL Sigma 70 Master Mix and nuclease-free water. The resulting endpoint deGFP values were either depicted as averages of a targeting gRNA and a non-targeting gRNA or fold change-repression was calculated by the ratio of non-targeting over the targeting deGFP values. Significance was calculated with Welch's t-test. $P > 0.05$ is shown as ns, $P < 0.05$ is shown as *, $P < 0.01$ is shown as ** and $P < 0.001$ is shown as ***. Within the PAM validation assays represented as fold changes, significance was calculated between the fold change of a given PAM and the fold change of a PAM that corresponds to the 3′ end of the repeat of the tested CRISPR system. The fold changes of the PAM validation in **Figure 3B** are depicted in a heat map. Thereby a difference between a non-targeting sample and a targeting sample with a specific PAM resulting in $P > 0.05$ is shown in white and excluded from further analysis. For all other samples within the heat map, the fold changes were calculated as mentioned above and presented relative to the highest fold change within one system. Significance within the deGFP repression assays testing binding and cleavage ability of the type I-C and the type I-F1 system in *X. albilineans* was calculated with the targeting and non-targeting sample for each condition. For the endpoint measurements in **Figure 5C**, significance was calculated between a non-targeting sample and a targeting sample targeting the same PAM.

qPCR analysis for PAM-DETECT

Cq values were used to measure target amounts. To calculate the relative abundance of the PAM library containing plasmid in the digested sample to the non-digested sample, the relative plasmid amount was normalized to a control amplifying the pET28a-T7RNAP that has no PacI recognition site using the $2^{-\Delta\Delta Ct}$ method. Significance to the control sample lacking a CRISPR-Cas system was calculated with Welch's t-test. $P > 0.05$ is shown as ns, $P < 0.05$ is shown as *, $P < 0.01$ is shown as ** and $P < 0.001$ is shown as ***.

Assessing transposition insertion point

~15 nts long sequences 5′ of the transposon terminal left end were extracted, counted and sorted. The sequences were mapped to the targeted plasmid or the targeted genome tolerating

2 nts mismatches and the distance between the insertion point and the PAM upstream of the protospacer or the end of the *tRNA-Leu* gene was noted. To only depict reliable insertion points, we present insertion points with more than 20 reads. The insertion points are shown as bar graphs.

The processed NGS data have been deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) and are accessible through GEO Series accession number GSE179614 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE179614).

In silico selection of representative type I-E CRISPR-Cas systems for PAM-DETECT

HMM profiles for the Cas5e, Cas6e, Cas7e and Cas8e proteins were developed upon aligning the members of the corresponding protein families (Cas5e: pfam09704, TIGR1868, TIGR02593; Cas6e: pfam08798, TIGR01907; Cas7e: pfam09344, TIGR01869; Cas8e: pfam 09481, TIGR02547). A new HMM profile was generated for the less conserved Cse2 protein upon aligning sequences with known 3D structure using PROMALS3D server (Pei et al., 2008) followed by a series of iterative alignment/model building steps to include additional sequences and increase sequence diversity. For the aligning processes of all five proteins, sequences were dereplicated at 90% identity using cd-hit (Huang et al., 2010) (with options -c 0.90 -g 1 -aS 0.9). The dereplicated sequences were compared against each other using blastp from blast+ v2.6.0 (Altschul et al., 1990) with e-value 10e-05 and defaults for the rest of parameters. Hits were filtered to retain those at >=60% pairwise identity, and were next clustered using the mcl algorithm (Enright et al., 2002) with inflation parameter of 2.0. Clusters with >=10 members were aligned using Gismo (Neuwald and Liu, 2004) with default parameters, and consensus sequences were extracted from the alignments. These consensus sequences, as well as singletons and sequences from smaller clusters were aligned using Gismo (Neuwald and Liu, 2004). Alignments were manually curated to remove shorter sequences that did not have one or more of the active site positions and HMM profiles were generated using hmmbuild (Eddy, 2009). Hmmsearch (Eddy, 2009) using the generated HMM profiles against all public genomes (isolates, SAGs, and MAGs), and all public metagenomes resulted in hits which were subsequently aligned against the generated HMM profiles. After selecting gene arrays that have all five complete or nearly complete genes, we identified 6,964 arrays in public genomes and 5,000 arrays in public metagenomes. Aligned sequences for all proteins from the same array were concatenated, and the resulting sequences were dereplicated with cd-hit (Huang et al., 2010) at 90% identity, aligned over at least 90% of the shorter sequences. This resulted in 2,851 clusters, 1,799 from metagenomes and 1,052 from genomes. Whereas the alignment of the Cas8e proteins from these clusters showed high variability, the predicted L1 helix regions of the Cas8e, which have been shown to directly interact with the PAM (Xiao et al., 2017),

presented higher conservation. We generated a list with the L1 signatures from the de-replicated cluster set and we subsequently manually filtered out systems that do not belong to known cultured mesophilic bacteria (**Table S3**). From the resulting list we selected I-E CRISPR/Cas systems with a variety of L1 motifs for experimental validation with PAM-DETECT.

Comparative analysis of I-B CAST transposases

We searched previous literature (Peters et al., 2017; Saito et al., 2021) for *in silico* identified I-B2 CASTs, which contain a fused *tnsAB* gene and are easily distinguished from I-B1 CASTs, which contain separate *tnsA* and *tnsB* genes. We observed that one clade of the I-B2 CASTs encompasses systems with *tnsAB-tnsC-tnsD* operons while having the *tniQ* gene separated, whereas the other clade encompasses systems with *tnsAB-tnsC-tniQ* operons and the *tnsD* gene separated. We denoted the systems in the former clade as I-B2.1 CASTs and in the latter clade as I-B2.2 CASTs. We focused on the I-B2.2 CAST clade, that has no *in vitro* or *in vivo* characterized members, and we discarded from further analysis the systems that lacked at least one of the CRISPR-Cas or transposition genes (*tnsAB, tnsC, tnsD, tniQ, cas5, cas6, cas7, cas8*). We performed BlastP search (Altschul et al., 1990) using the TnsAB, TnsC, TnsD, TniQ proteins of each selected I-B2.2 system as queries, aiming to identify additional I-B2.2 CAST candidates. Our analysis yielded in total seven I-B2.2 systems and we selected six previously described I-B2.1 systems for phylogenetic analysis (Saito et al., 2021). The alignment of I-B2.1 and I-B2.2 transposition proteins was performed using T-Coffee (Di Tommaso et al., 2011), the phylogenetic trees were built using average distance and the BLOSUM62 matrix and they were visualized with JalView (Waterhouse et al., 2009).

***In silico* analysis of RoCAST**

We predicted the CRISPR array of RoCAST by uploading the *Rippkaea orientalis* genomic region between the *Rocas5* and *RotniQ* to CRISPRFinder (Grissa et al., 2007). The RoCAST ends were determined manually on Benchling by searching for repeat sequences of 20 nucleotides, with maximum 5 mismatched nucleotides, within the *R. orientalis* genomic regions 1 kb upstream of the *RotnsAB* and 1 kb downstream of the *RotnsD*. We identified two types of repeat sequences present in both regions in opposite orientations and a candidate duplication region. Notably, we identified five repeat sequences in the predicted left end region, with one of the repeat sequences located downstream of the predicted duplication site, hence outside of the predicted RoCAST limits. The TXTL transposition demonstrated that this repeat is not part of the RoCAST transposon.

References

Almendros, C., Guzmán, N.M., Díez-Villaseñor, C., García-Martínez, J., and Mojica, F.J.M. (2012). Target motifs affecting natural immunity by a constitutive CRISPR-Cas system in *Escherichia coli*. PLoS One *7*, e50797.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Barrangou, R., and Doudna, J.A. (2016). Applications of CRISPR technologies in research and beyond. Nature Biotechnology *34*, 933–941.

Caliando, B.J., and Voigt, C.A. (2015). Targeted DNA degradation using a CRISPR device stably carried in the host genome. Nat. Commun. *6*, 1–10.

Cameron, P., Coons, M.M., Klompe, S.E., Lied, A.M., Smith, S.C., Vidal, B., Donohoue, P.D., Rotstein, T., Kohrs, B.W., Nyer, D.B., et al. (2019). Harnessing type I CRISPR–Cas systems for genome engineering in human cells. Nat. Biotechnol. *37*, 1471–1477.

Chen, W., Ren, Z.-H., Tang, N., Chai, G., Zhang, H., Zhang, Y., Ma, J., Wu, Z., Shen, X., Huang, X., et al. (2021). Targeted genetic screening in bacteria with a Cas12k-guided transposase. Cell Rep. *36*, 109635.

Chen, Y., Liu, J., Zhi, S., Zheng, Q., Ma, W., Huang, J., Liu, Y., Liu, D., Liang, P., and Songyang, Z. (2020). Repurposing type I–F CRISPR–Cas system as a transcriptional activation tool in human cells. Nat. Commun. *11*, 1–14.

Collias, D., and Beisel, C.L. (2021). CRISPR technologies and the search for the PAM-free nuclease. Nat. Commun. *12*, 1–12.

Davidson, A.R., Lu, W.-T., Stanley, S.Y., Wang, J., Mejdani, M., Trost, C.N., Hicks, B.T., Lee, J., and Sontheimer, E.J. (2020). Anti-CRISPRs: protein inhibitors of CRISPR-Cas systems. Annu. Rev. Biochem. *89*, 309–332.

Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.-M., Taly, J.-F., and Notredame, C. (2011). T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res. *39*, W13–W17.

Dolan, A.E., Hou, Z., Xiao, Y., Gramelspacher, M.J., Heo, J., Howden, S.E., Freddolino, P.L., Ke, A., and Zhang, Y. (2019). Introducing a Spectrum of Long-Range Genomic Deletions in Human Embryonic Stem Cells Using Type I CRISPR-Cas. Mol. Cell *74*, 936–950.e5.

Eddy, S.R. (2009). A new generation of homology search tools based on probabilistic inference. Genome Inform. *23*, 205–211.

Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. *30*, 207–210.

Elmore, J., Deighan, T., and Westpheling, J. (2015). DNA targeting by the type IG and type IA CRISPR–Cas systems of *Pyrococcus furiosus*. Nucleic Acids *43*, 10353–10363.

Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. *30*, 1575–1584.

van Erp, P.B.G., Jackson, R.N., Carter, J., Golden, S.M., Bailey, S., and Wiedenheft, B. (2015). Mechanism of CRISPR-RNA guided recognition of DNA targets in Escherichia coli. Nucleic Acids Res. *43*, 8381–8391.

Fineran, P.C., Gerritzen, M.J.H., Suárez-Diez, M., Künne, T., Boekhorst, J., van Hijum, S.A.F.T., Staals, R.H.J., and Brouns, S.J.J. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. Proc. Natl. Acad. Sci. U. S. A. *111*, E1629–E1638.

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

Fu, B.X.H., Wainberg, M., Kundaje, A., and Fire, A.Z. (2017). High-throughput characterization of Cascade type I-E CRISPR guide efficacy reveals unexpected PAM diversity and target sequence preferences. Genetics *206*, 1727–1738.

Garamella, J., Marshall, R., Rustad, M., and Noireaux, V. (2016). The All *E. coli* TX-TL Toolbox 2.0: A platform for cell-free synthetic biology. ACS Synth. Biol. *5*, 344–355.

Gasiunas, G., Young, J.K., Karvelis, T., Kazlauskas, D., Urbaitis, T., Jasnauskaite, M., Grusyte, M.M., Paulraj, S., Wang, P.-H., Hou, Z., et al. (2020). A catalogue of biochemically diverse CRISPR-Cas9 orthologs. Nat. Commun. *11*, 5512.

Gomaa, A.A., Klumpe, H.E., Luo, M.L., Selle, K., Barrangou, R., and Beisel, C.L. (2014). Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. mBio *5*, e00928–13.

Gong, B., Shin, M., Sun, J., Jung, C.-H., Bolt, E.L., van der Oost, J., and Kim, J.-S. (2014). Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. Proc. Natl. Acad. Sci. U. S. A. *111*, 16359–16364.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. *35*, W52–W57.

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

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

Hochstrasser, M.L., Taylor, D.W., Kornfeld, J.E., Nogales, E., and Doudna, J.A. (2016). DNA targeting by a minimal CRISPR RNA-guided Cascade. Mol. Cell *63*, 840–851.

Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics *26*, 680–682.

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

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

Jiao, C., Sharma, S., Dugar, G., Peeck, N.L., Bischler, T., Wimmer, F., Yu, Y., Barquist, L., Schoen, C., Kurzai, O., et al. (2021). Noncanonical crRNAs derived from host transcripts enable multiplexable RNA detection by Cas9. Science *372*, 941–948.

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

Jung, C., Hawkins, J.A., Jones, S.K., Jr, Xiao, Y., Rybarski, J.R., Dillard, K.E., Hussmann, J., Saifuddin, F.A., Savran, C.A., Ellington, A.D., et al. (2017). Massively parallel biophysical analysis of CRISPR-Cas complexes on next generation sequencing chips. Cell *170*, 35–47.e13.

Karvelis, T., Gasiunas, G., Young, J., Bigelyte, G., Silanskas, A., Cigan, M., and Siksnys, V. (2015). Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. Genome Biol. *16*, 1–13.

Khakimzhan, A., Garenne, D., Tickman, B., Fontana, J., Carothers, J., and Noireaux, V. (2021). Complex dependence of CRISPR-Cas9 binding strength on guide RNA spacer lengths. Phys. Biol.

Klompe, S.E., Vo, P.L.H., Halpin-Healy, T.S., and Sternberg, S.H. (2019). Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. Nature *571*, 219–225.

Leenay, R.T., and Beisel, C.L. (2017). Deciphering, communicating, and engineering the CRISPR PAM. J. Mol. Biol. *429*, 177–191.

Leenay, R.T., Maksimchuk, K.R., Slotkowski, R.A., Agrawal, R.N., Gomaa, A.A., Briner, A.E., Barrangou, R., and Beisel, C.L. (2016). Identifying and visualizing functional PAM diversity across CRISPR-Cas systems. Mol. Cell *62*, 137–147.

Li, M., Gong, L., Cheng, F., Yu, H., Zhao, D., Wang, R., Wang, T., Zhang, S., Zhou, J., Shmakov, S.A., et al. (2021). Toxin-antitoxin RNA pairs safeguard CRISPR-Cas systems. Science *372*, eabe5601.

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

Liao, C., Slotkowski, R.A., Achmedov, T., and Beisel, C.L. (2019a). The *Francisella novicida* Cas12a is sensitive to the structure downstream of the terminal repeat in CRISPR arrays. RNA Biol. *16*, 404–412.

Liao, C., Ttofali, F., Slotkowski, R.A., Denny, S.R., Cecil, T.D., Leenay, R.T., Keung, A.J., and Beisel, C.L. (2019b). Modular one-pot assembly of CRISPR arrays enables library generation and reveals factors influencing crRNA biogenesis. Nat. Commun. *10*, 2948.

Liu, T., Pan, S., Li, Y., Peng, N., and She, Q. (2018). Type III CRISPR-Cas system: introduction and its application for genetic manipulations. Curr. Issues Mol. Biol. *26*, 1–14.

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

Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., et al. (2019). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat. Rev. Microbiol. *18*, 67–83.

Marino, N.D., Zhang, J.Y., Borges, A.L., Sousa, A.A., Leon, L.M., Rauch, B.J., Walton, R.T., Berry, J.D., Joung, J.K., Kleinstiver, B.P., et al. (2018). Discovery of widespread type I and type V CRISPR-Cas inhibitors. Science *362*, 240–242.

Marshall, R., Maxwell, C.S., Collins, S.P., Jacobsen, T., Luo, M.L., Begemann, M.B., Gray, B.N., January, E., Singer, A., He, Y., et al. (2018). Rapid and scalable characterization of CRISPR technologies using an *E. coli* cell-free transcription-translation system. Mol. Cell *69*, 146–157.e3.

Maxwell, C.S., Jacobsen, T., Marshall, R., Noireaux, V., and Beisel, C.L. (2018). A detailed cell-free transcription-translation-based assay to decipher CRISPR protospacer-adjacent motifs. Methods *143*, 48–57.

Morisaka, H., Yoshimi, K., Okuzaki, Y., Gee, P., Kunihiro, Y., Sonpho, E., Xu, H., Sasakawa, N., Naito, Y., Nakada, S., et al. (2019). CRISPR-Cas3 induces broad and unidirectional genome editing in human cells. Nat. Commun. *10*, 5302.

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

Mulepati, S., Orr, A., and Bailey, S. (2012). Crystal structure of the largest subunit of a bacterial RNA-guided immune complex and its role in DNA target binding. J. Biol. Chem. *287*, 22445–22449.

Musharova, O., Sitnik, V., Vlot, M., Savitskaya, E., Datsenko, K.A., Krivoy, A., Fedorov, I., Semenova, E., Brouns, S.J.J., and Severinov, K. (2019). Systematic analysis of Type I-E *Escherichia coli* CRISPR-Cas PAM sequences ability to promote interference and primed adaptation. Mol. Microbiol. *111*, 1558–1570.

Neuwald, A.F., and Liu, J.S. (2004). Gapped alignment of protein sequence motifs through Monte Carlo optimization of a hidden Markov model. BMC Bioinformatics *5*.

Ondov, B.D., Bergman, N.H., and Phillippy, A.M. (2011). Interactive metagenomic visualization in a Web browser. BMC Bioinformatics *12*, 385.

Osakabe, K., Wada, N., Miyaji, T., Murakami, E., Marui, K., Ueta, R., Hashimoto, R., Abe-Hara, C., Kong, B., Yano, K., et al. (2020). Genome editing in plants using CRISPR type I-D nuclease. Communications Biology *3*, 1–10.

Park, J.-U., Tsai, A.W.-L., Mehrotra, E., Petassi, M.T., Hsieh, S.-C., Ke, A., Peters, J.E., and Kellogg, E.H. (2021). Structural basis for target site selection in RNA-guided DNA transposition systems. Science *373*, 768–774.

Pausch, P., Müller-Esparza, H., Gleditzsch, D., Altegoer, F., Randau, L., and Bange, G. (2017). Structural variation of Type I-F CRISPR RNA guided DNA surveillance. Mol. Cell *67*, 622–632.e4.

Pei, J., Kim, B.-H., and Grishin, N.V. (2008). PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res. *36*, 2295–2300.

Petassi, M.T., Hsieh, S.-C., and Peters, J.E. (2020). Guide RNA categorization enables target site choice in Tn7-CRISPR-Cas transposons. Cell *183*, 1757–1771.e18.

Peters, J.E., Makarova, K.S., Shmakov, S., and Koonin, E.V. (2017). Recruitment of CRISPR-Cas systems by Tn7-like transposons. Proc. Natl. Acad. Sci. U. S. A. *114*, E7358–E7366.

Pickar-Oliver, A., and Gersbach, C.A. (2019). The next generation of CRISPR–Cas technologies and applications. Nat. Rev. Mol. Cell Biol. *20*, 490–507.

Pickar-Oliver, A., Black, J.B., Lewis, M.M., Mutchnick, K.J., Klann, T.S., Gilcrest, K.A., Sitton, M.J., Nelson, C.E., Barrera, A., Bartelt, L.C., et al. (2019). Targeted transcriptional modulation with type I CRISPR–Cas systems in human cells. Nat. Biotechnol. *37*, 1493–1501.

Querques, I., Schmitz, M., Oberli, S., Chanez, C., and Jinek, M. (2021). Target site selection and remodelling by type V CRISPR-transposon systems. Nature *599*, 497–502.

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

Rauch, B.J., Silvis, M.R., Hultquist, J.F., Waters, C.S., McGregor, M.J., Krogan, N.J., and Bondy-Denomy, J. (2017). Inhibition of CRISPR-Cas9 with bacteriophage proteins. Cell *168*, 150–158.e10.

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

Roux, S., Enault, F., Hurwitz, B.L., and Sullivan, M.B. (2015). VirSorter: mining viral signal from microbial genomic data. PeerJ *3*, e985.

Rybarski, J.R., Hu, K., Hill, A.M., Wilke, C.O., and Finkelstein, I.J. (2021). Metagenomic Discovery of CRISPR-Associated Transposons. bioRxiv.

Saito, M., Ladha, A., Strecker, J., Faure, G., Neumann, E., Altae-Tran, H., Macrae, R.K., and Zhang, F. (2021). Dual modes of CRISPR-associated transposon homing. Cell *184*, 2441–2453.e18.

Shin, J., and Noireaux, V. (2010). Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70. J. Biol. Eng. *4*, 8.

Shin, J., and Noireaux, V. (2012). An *E. coli* cell-free expression toolbox: application to synthetic gene circuits and artificial cells. ACS Synth. Biol. *1*, 29–41.

Silas, S., Lucas-Elio, P., Jackson, S.A., Aroca-Crevillén, A., Hansen, L.L., Fineran, P.C., Fire, A.Z., and Sánchez-Amat, A. (2017). Type III CRISPR-Cas systems can provide redundancy to counteract viral escape from type I systems. Elife *6*, e27601.

Silverman, A.D., Karim, A.S., and Jewett, M.C. (2020). Cell-free gene expression: an expanded repertoire of applications. Nat. Rev. Genet. *21*, 151–170.

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

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

Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J.L., Makarova, K.S., Koonin, E.V., and Zhang, F. (2019). RNA-guided DNA insertion with CRISPR-associated transposases. Science *365*, 48–53.

Sun, Z.Z., Hayes, C.A., Shin, J., Caschera, F., Murray, R.M., and Noireaux, V. (2013). Protocols for implementing an *Escherichia coli* based TX-TL cell-free expression system for synthetic biology. J. Vis. Exp. e50762.

Tay, M., Liu, S., and Yuan, Y.A. (2015). Crystal structure of *Thermobifida fusca* Cse1 reveals target DNA binding site. Protein Sci. *24*, 236–245.

Tuminauskaite, D., Norkunaite, D., Fiodorovaite, M., Tumas, S., Songailiene, I., Tamulaitiene, G., and Sinkunas, T. (2020). DNA interference is controlled by R-loop length in a type I-F1 CRISPR-Cas system. BMC Biol. *18*, 65.

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

Vo, P.L.H., Ronda, C., Klompe, S.E., Chen, E.E., Acree, C., Wang, H.H., and Sternberg, S.H. (2021). CRISPR RNA-guided integrases for high-efficiency, multiplexed bacterial genome engineering. Nat. Biotechnol. *39*, 480–489.

Wandera, K.G., Collins, S.P., Wimmer, F., Marshall, R., Noireaux, V., and Beisel, C.L. (2020). An enhanced assay to characterize anti-CRISPR proteins using a cell-free transcription-translation system. Methods *172*, 42–50.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.

Watters, K.E., Fellmann, C., Bai, H.B., Ren, S.M., and Doudna, J.A. (2018). Systematic discovery of natural CRISPR-Cas12a inhibitors. Science *362*, 236–239.

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

Wimmer, F., and Beisel, C.L. (2019). CRISPR-Cas systems and the paradox of self-targeting spacers. Front. Microbiol. *10*, 3078.

Xiao, Y., Luo, M., Hayes, R.P., Kim, J., Ng, S., Ding, F., Liao, M., and Ke, A. (2017). Structure basis for directional R-loop formation and substrate handover mechanisms in type I CRISPR-Cas system. Cell *170*, 48–60.e11.

Xiao, Y., Luo, M., Dolan, A.E., Liao, M., and Ke, A. (2018). Structure basis for RNA-guided DNA degradation by Cascade and Cas3. Science *361*, eaat0839.

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

Yin, Y., Yang, B., and Entwistle, S. (2019). Bioinformatics identification of anti-CRISPR loci by using homology, guilt-by-association, and CRISPR self-targeting spacer approaches. mSystems *4*, e00455–19.

Zetsche, B., Abudayyeh, O.O., Gootenberg, J.S., Scott, D.A., and Zhang, F. (2020). A survey of genome editing activity for 16 Cas12a orthologs. Keio J. Med. *69*, 59–65.

Zheng, Y., Han, J., Wang, B., Hu, X., Li, R., Shen, W., Ma, X., Ma, L., Yi, L., Yang, S., et al. (2019). Characterization and repurposing of the endogenous Type I-F CRISPR–Cas system of *Zymomonas mobilis* for genome engineering. Nucleic Acids Research *47*, 11461–11475.

Zheng, Y., Li, J., Wang, B., Han, J., Hao, Y., Wang, S., Ma, X., Yang, S., Ma, L., Yi, L., et al. (2020). Endogenous Type I CRISPR-Cas: from foreign DNA defense to prokaryotic engineering. Front Bioeng Biotechnol *8*, 62.

Supplementary Information

Supplementary figures

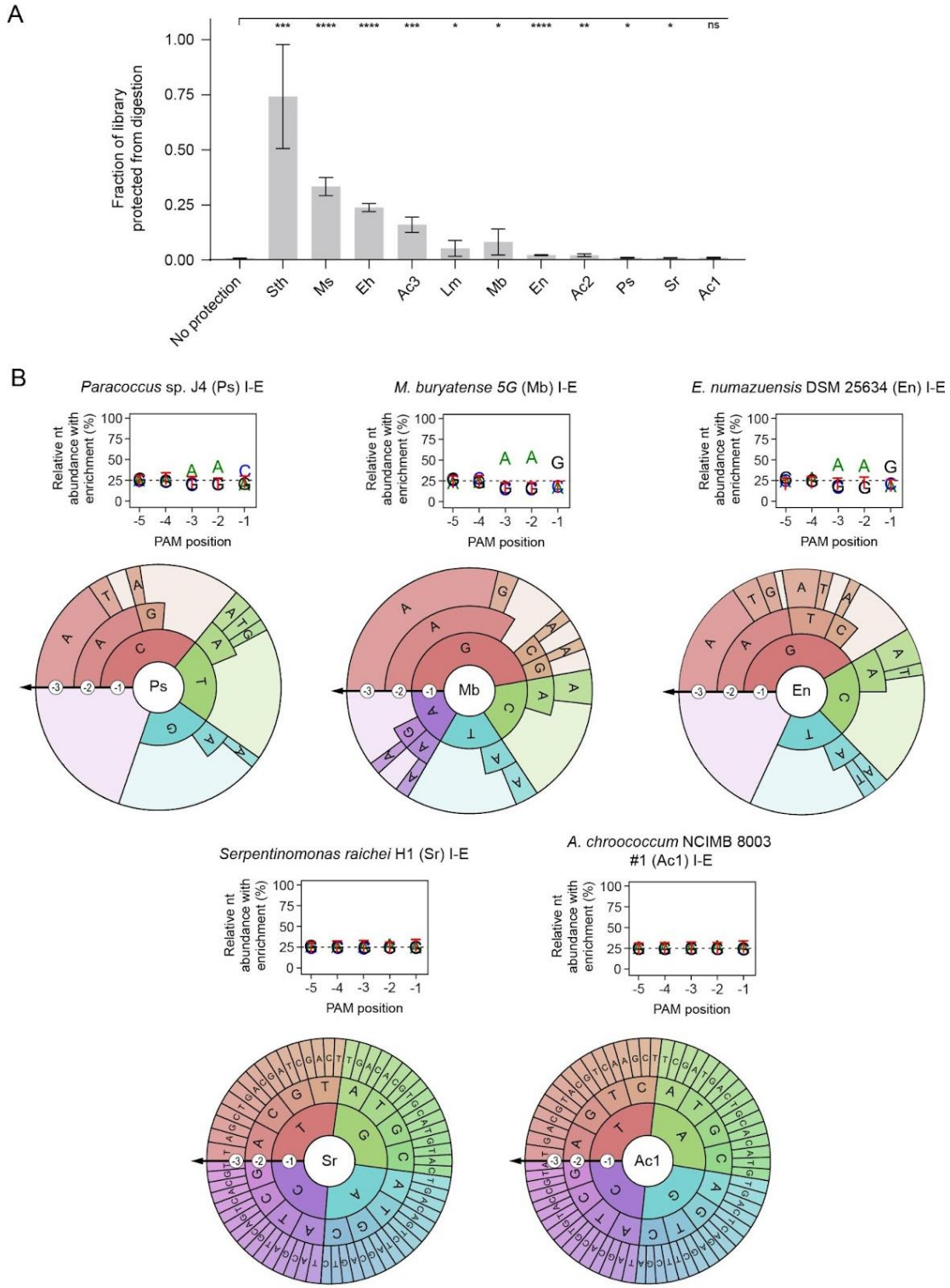


Figure S1: Additional type I-E CRISPR-Cas systems subjected to PAM-DETECT. Related to **Figure 3**. (**A**) Extent of PAM library protection for the mined I-E CRISPR-Cas systems. Error bars indicate the mean and standard deviation of triplicate independent experiments. ***: $p < 0.001$. **: $p < 0.01$. *: $p < 0.05$. ns: $p > 0.05$. (**B**) Nucleotide-enrichment plots and PAM wheels based on conducting PAM-DETECT with Cascade from additional I-E CRISPR-Cas systems. Individual sequences comprising at least 2% of the PAM wheel are shown. Results represent the average of duplicate independent experiments.

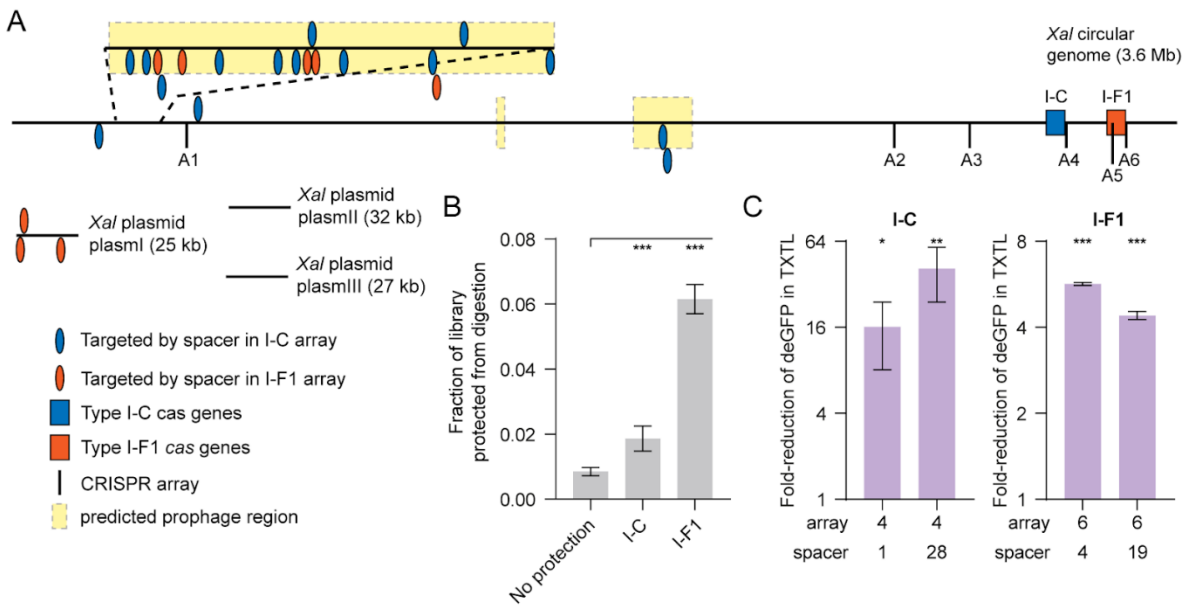


Figure S2: Genomic target sites, predicted prophage regions, library protection and cleavage of genomic target sites for the I-C and I-F1 CRISPR-Cas systems in *Xanthomonas albilineans*. Related to **Figure 4**. (**A**) Genomic architecture of the two CRISPR-Cas systems, their self-targets, and location of predicted prophage regions. The numbering of the arrays corresponds to those in **Figure 4A**. Placement of the ovals indicates whether the self-target corresponding to the spacer sequence is located on the top or bottom strand of the chromosome or plasmid. Prophage regions were predicted with VirSorter (Roux et al., 2015). (**B**) Extent of PAM library protection for both CRISPR-Cas systems. (**C**) Functional targeting in TXTL using native spacer:self-target pairs from *X. albilineans*. The origin of the native spacers is indicated below the graphs. Fold-change was calculated based on a non-targeting crRNA control. The non-targeting crRNA control is the reference for statistical analyses.
Error bars in B and C indicate the mean and standard deviation of triplicate independent experiments. ***: $p < 0.001$. **: $p < 0.01$. *: $p < 0.05$. ns: $p > 0.05$.

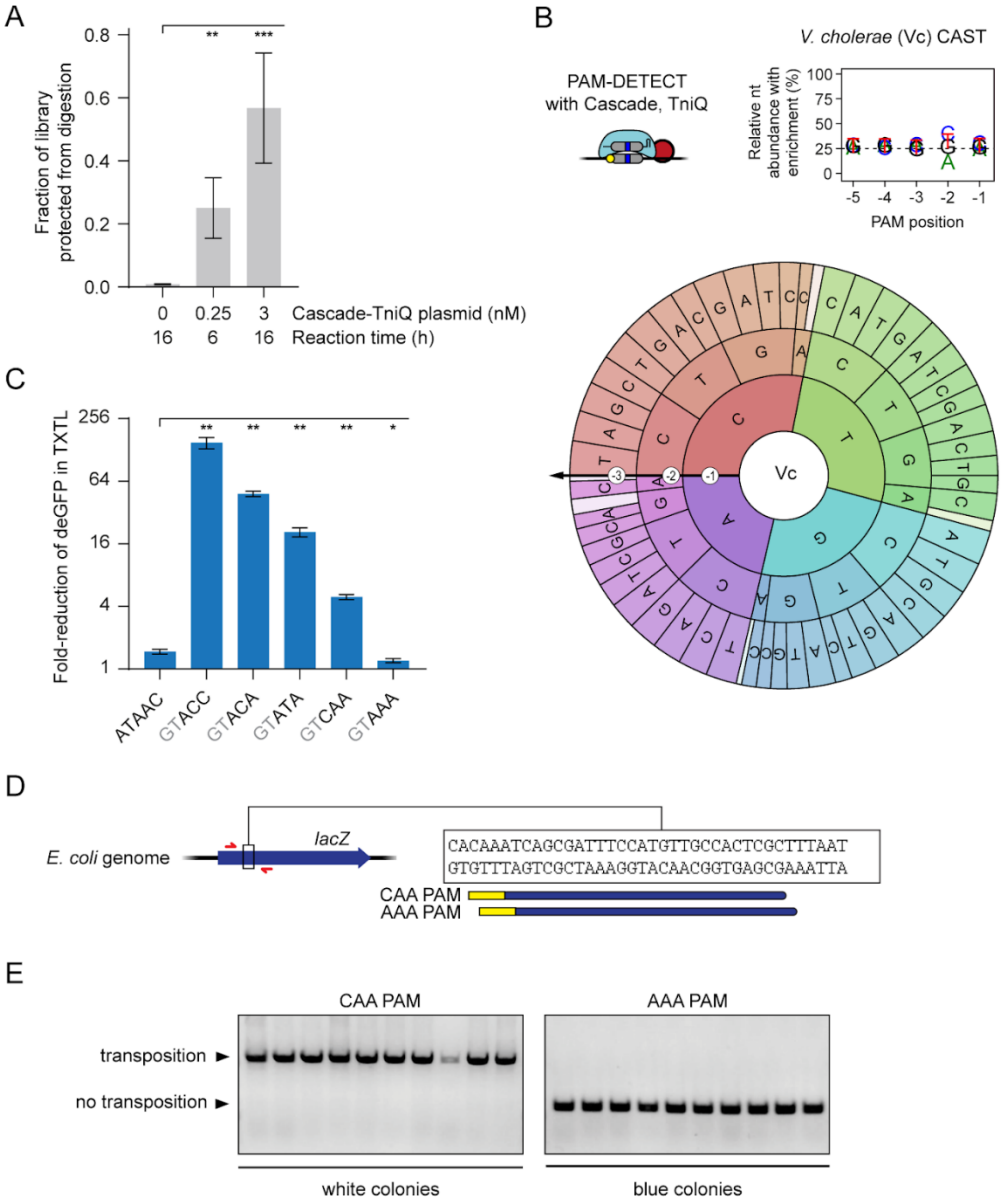


Figure S3: Further characterization of the *Vibrio cholerae* I-F CRISPR transposon. Related to **Figure 5**. (**A**) Extent of PAM library protection with Cascade from the VcCAST under different conditions. The low-Cascade conditions (0.25-nM plasmid, 6-h reaction time) was the basis of the PAM-DETECT output reported in **Figure 5B**. (**B**) Nucleotide-enrichment plot and PAM wheel based on conducting PAM-DETECT with Cascade and TniQ under high-Cascade conditions (3-nM plasmid, 16-h reaction time). Individual sequences comprising at least 1% of the PAM wheel are shown. The plot and PAM wheel are averages of duplicate independent experiments. (**C**) Dispensability of TniQ for Cascade binding in TXTL-based PAM validation assay. The TXTL-based deGFP repression assay performed in **Figure 5C** was conducted using Cascade alone. ATAAC matches the 3′ end of the repeat and therefore serves as a negative control. The ATAAC self PAM is the reference for statistical analyses. (**D**) Target locations within the *lacZ* gene associated with a CAA or AAA PAM. The locations are shifted by a single base. See **Figure 5E** for the extent of transposition for either site based on formation of blue or white colonies. (**E**) Colony PCR of representative colonies associated with the CAA or AAA PAM targets. Only white colonies were picked for CAA and only blue colonies were picked for AAA in part because colonies of the opposite color were rarely observed (see **Figure 5E**). The markers indicate the expected band sizes based on successful transposition or no transposition.
Error bars in A and B indicate the mean and standard deviation of triplicate independent experiments. ***: $p < 0.001$. **: $p < 0.01$. *: $p < 0.05$. ns: $p > 0.05$.

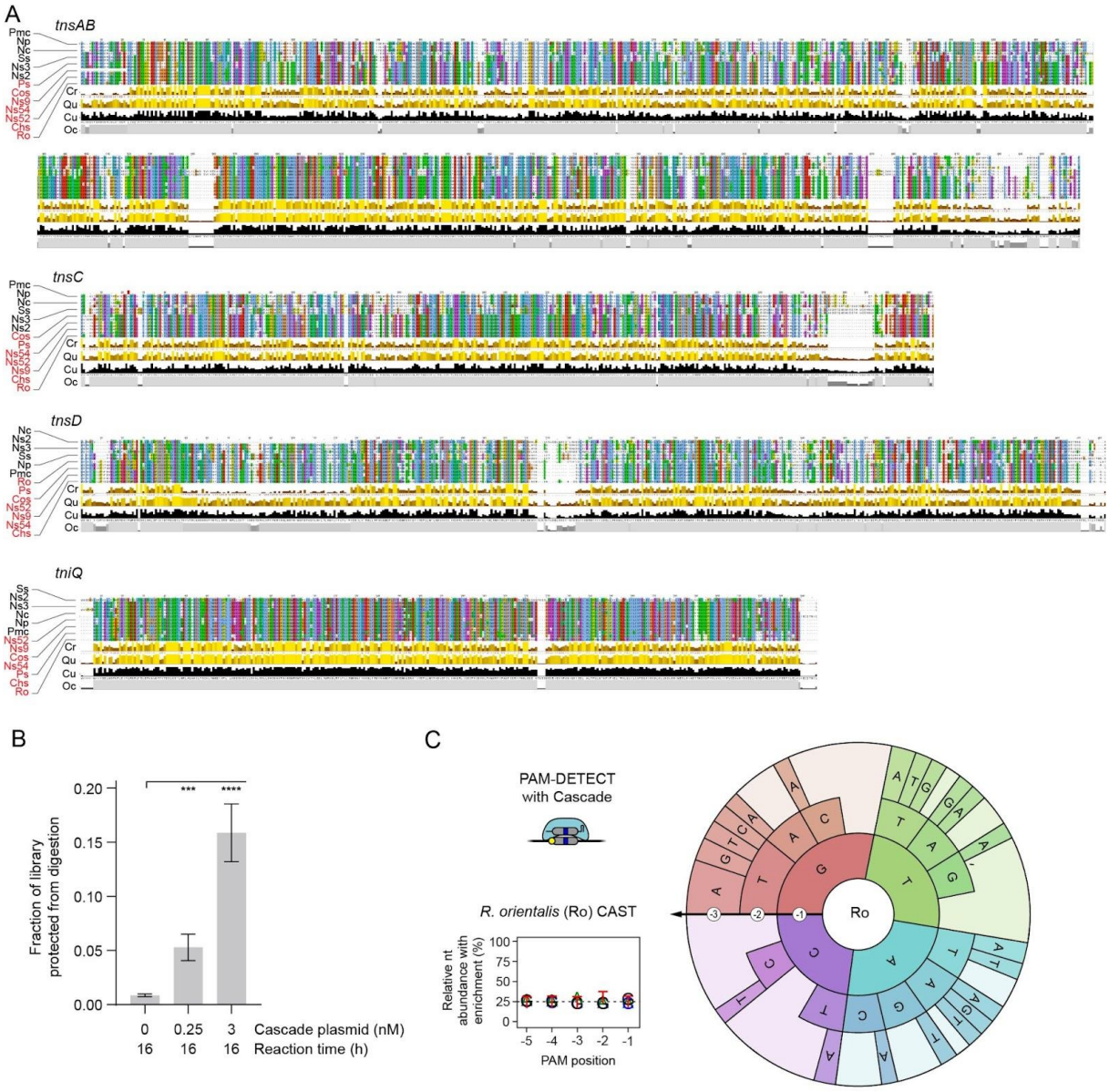


Figure S4: Interrogating the *Rippkaea orientalis* I-B2.2 CRISPR transposon. Related to **Figure 6**. (**A**) Sequence alignments of the TnsAB, TnsC, TnsD, and TniQ proteins from six previously reported I-B2.1 CAST systems and the seven I-B2.2 CAST systems identified in this study. The alignments were built with T-Coffee (Di Tommaso et al., 2011), visualized with Jalview and ordered according to the trees presented in **Figure 6A**. The conservation (Cr), quality (Qu), consensus (Cu), and occupancy (Oc) histograms from Jalview are presented below each alignment. The alignments are colored using the ClustalX color palette. Pmc: *Peltigera membranacea cyanobiont* 210A. Np: *Nostoc punctiforme* NIES-2108. Nc: *Nostoc carneum* NIES-207. Ss: *Stanieria* sp. NIES-3757. Ns3: *Nostoc* sp. NIES-3756. Ns2: *Nostoc* sp. NIES-2111. Ps: *Planktothrix serta*. Cos: *Coleofasciculus* sp. FACHB-SPT9. Ns9: *Nostoc* sp. NMS9. Ns54: *Nostoc* sp. CENA543. Ns52: *Nostoc* sp. C052. Chs: *Chroococcus* sp. FPU101. Ro: *Rippkaea orientalis*. Names in black and in red are associated with the I-B2.1 and I-B2.2 branch, respectively. (**B**) Extent of PAM library protection with Cascade from the *R. orientalis* CAST (RoCAST) under different conditions. The low-Cascade condition (0.25-nM plasmid, 16-h reaction time) was the basis of the PAM-DETECT output reported in **Figure 6B**. Error bars indicate the mean and standard deviation of triplicate independent experiments. ***: $p < 0.001$. **: $p < 0.01$. *: $p < 0.05$. ns: $p > 0.05$. (**C**) Nucleotide-enrichment plot and PAM wheel based on conducting PAM-DETECT with Cascade under high-Cascade conditions (3-nM Cascade plasmid, 16-h reaction time). Individual sequences comprising at least 2% of the PAM wheel are shown. The plot and PAM wheel are averages of duplicate independent experiments. See **Figure 6B** for the PAM-DETECT output with Cascade under low-Cascade conditions (0.25-nM Cascade plasmid, 16-h reaction time).

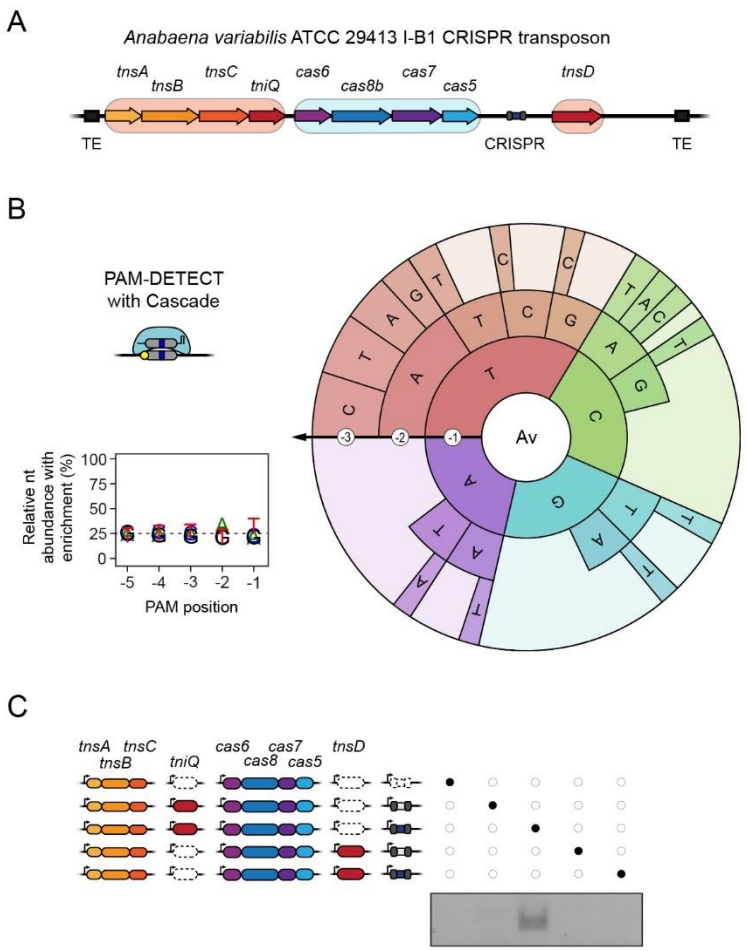


Figure S5: TXTL-based PAM determination and DNA transposition for the *Anabaena variabilis* ATCC 29413 I-B1 CRISPR transposon. Related to **Figure 6** and **7**. (**A**) The I-B1 CRISPR transposon from *A. variabilis* ATCC 29413 (AvCAST). (**B**) Nucleotide-enrichment plot and PAM wheel based on conducting PAM-DETECT with high-Cascade conditions (3-nM plasmid, 16-h reaction time). Individual sequences comprising at least 2% of the PAM wheel are shown. The plot and PAM wheel are averages of duplicate independent experiments. The predominant PAM (AT) matches that identified in recent work in *E. coli* (Saito et al., 2021). (**C**) CRISPR-dependent transposition with the AvCAST in TXTL. The crRNA target was flanked by an AT PAM. Transposition only occurred in the presence of TniQ.

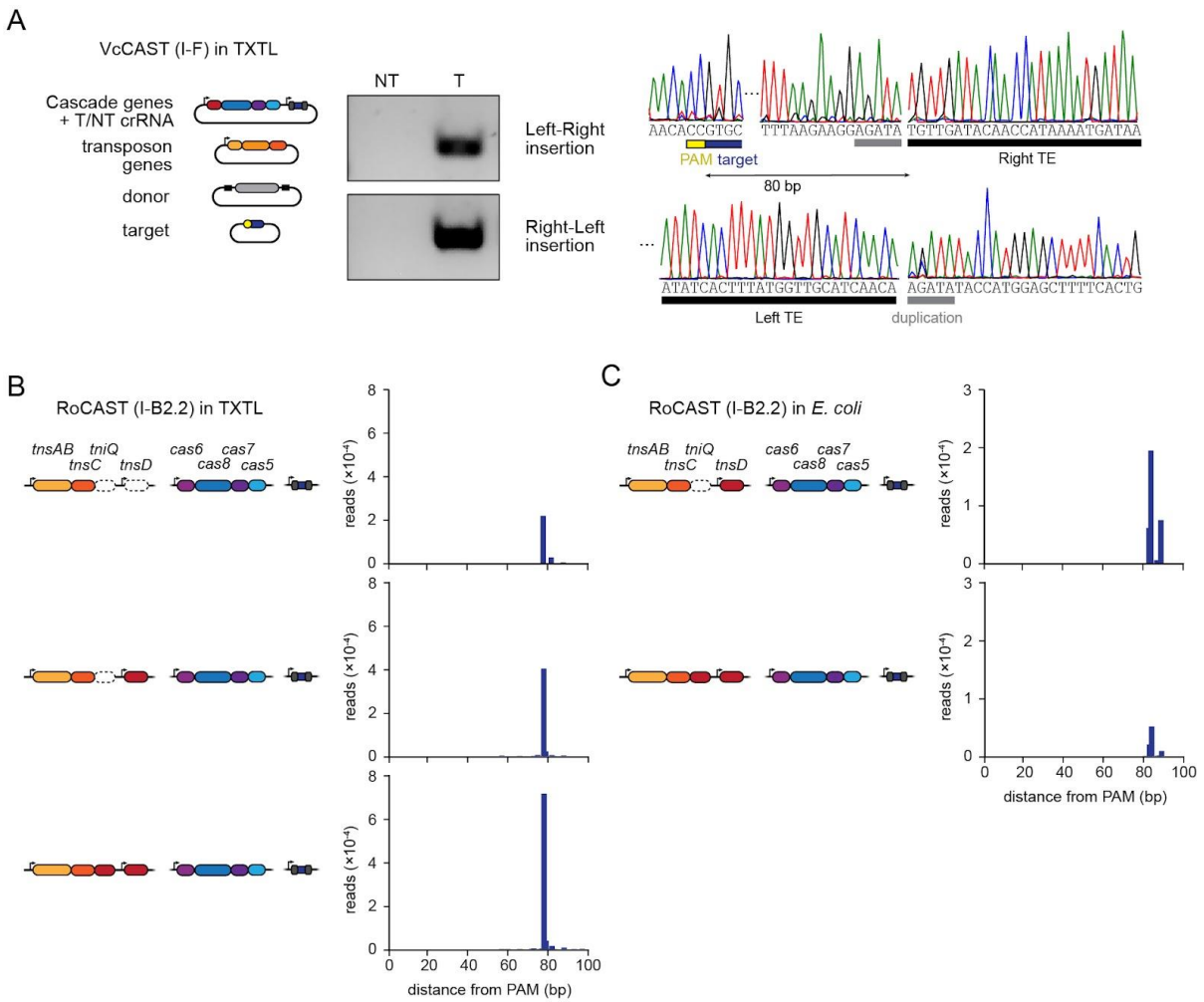


Figure S6: CRISPR-dependent transposition by the *V. cholerae* I-F CRISPR transposon and the *R. orientalis* I-B2.2 CRISPR transposon. Related to **Figure 7**. (**A**) CRISPR-dependent transposition with the *V. cholerae* CRISPR transposon (VcCAST) in TXTL. Top: transposition with a targeting (T) or non-targeting (NT) crRNA. Transposition was determined by amplifying across the junction of the target and inserted donor DNA. Bottom: Determination of transposon ends and insertion distance by Sanger sequencing. (**B**) CRISPR-dependent transposition with and without TniQ and TnsD with the *R. orientalis* I-B2.2 CRISPR transposon (RoCAST) in TXTL. Transposition was determined by next-generation sequencing of the PCR product spanning the crRNA target site and the beginning of the cargo (left-right orientation). (**C**) CRISPR-dependent transposition with and without TniQ with the *R. orientalis* I-B2.2 CRISPR transposon (RoCAST) in *E. coli*. Transposition was determined by next-generation sequencing of the PCR product spanning the crRNA target site and the beginning of the cargo (left-right orientation).

Supplementary tables

Table S1: Extent of PAM determination for type I CRISPR-Cas systems. Studies that tested at least 40 PAM sequences were included. Related to **Figure 1**.

System type	Organism	Number of tested PAMs	Method	Source
I-A	*P. furiosus*	64	Individual plasmids tested with plasmid interference assay	(Elmore et al., 2015)
I-B	*H. volcanii*	62	Individual plasmids tested with plasmid interference assay	(Fischer et al., 2012)
I-B	*H. hispanica*	64	Individual plasmids tested with plasmid interference assay	(Li et al., 2014)
I-C	*B. halodurans*	256	Library screening with PAMS-CANR	(Leenay et al., 2016)
I-C	*L. pneumophila*	64	Library screening with plasmid interference assay	(Rao et al., 2017)

Table S1 (continued)

System type	Organism	Number of tested PAMs	Method	Source
I-D	*M. aeruginosa*	256	Library screening with plasmid interference assay	(Osakabe et al., 2020)
I-E	*E. coli*	256	Library screening with PAMS-CANR	(Leenay et al., 2016)
I-E	*E. coli*	64	Library screening with plasmid loss assay	(Musharova et al., 2019)
I-E	*E. coli*	256	crRNA library targeting bacterial genome	(Fu et al., 2017)
I-E	*E. coli*	40	PAM-protospacer library screening with plasmid interference assay	(Fineran et al., 2014)
I-E	*E. coli*	64	Library screening with plasmid interference assay	(Xue et al., 2015)
I-E	*E. coli*	64	Library screening with phagemid transduction assay	(Caliando and Voigt, 2015)
I-E	*T. fusca*	4,096	Library screening with CHAMP	(Jung et al., 2017)
I-F	*Z. mobilis*	64	Individual plasmids tested with plasmid interference assay	(Zheng et al., 2019)
I-G	*P. furiosus*	64	Individual plasmids tested with plasmid interference assay	(Elmore et al., 2015)

Table S2: Cascade concentrations used in the TXTL-based PAM validation assays. If proteins required for Cascade formation were not encoded as an operon on one plasmid but encoded on separate plasmids, a MasterMix with all required plasmids in their stoichiometric amount was prepared beforehand (Cascade-MM). Related to **Figures 2**, **3**, **4**, **5** and **6**.

System	Cascade-MM	Operon	Final concentration	Reaction Temperature
E. coli type I-E	x		0.5 nM	29 °C
Azotobacter chroococcum (Ac2) type I-E	x		2 nM	29 °C
Azotobacter chroococcum (Ac3) type I-E	x		1 nM	29 °C
Leptothrix mobilis (Lm) type I-E	x		2 nM	29 °C
Ectothiorhodospira haloalkaliphila (Eh) type I-E	x		0.5 nM	29 °C
Marinomonas sp. (Ms) type I-E	x		0.5 nM	29 °C
Streptococcus thermophilus (St) type I-E	x		1 nM	37 °C
X. albilineans type I-C	x		0.5 nM	29 °C
X. albilineans type I-F1	x		0.25 nM	29 °C
V. cholerae I-F CAST		x	0.5 nM	29 °C
Rippkaea orientalis I-B CAST		x	0.25 nM	29 °C

Table S3: Cas8e proteins and the variable L1 loop from diverse I-E CRISPR-Cas systems from cultured mesophilic bacterial strains, related to Figure 3.

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
646098354	CaurA7_010100002214	CRISPR-associated protein, Cse1 family	2529292503	Derxia gummosa DSM 723	Yes	Mesophile	DFFTKR	D	FFT	K	22
637965378	Csal_0227	CRISPR-associated protein	645951870	Corynebacterium aurimucosum CN-1, ATCC 700975	Yes	Mesophile	DFFTMR	D	FFT	M	22
2516164566	DeslaDRAFT_0364	CRISPR-associated Cse1 family protein	2795385473	Actinorugispora endophytica DSM 46770	Yes	Mesophile	PFFTMR	P	FFT	M	22
637781729	Dde_0864	CRISPR-associated Cse1 family protein	2728369266	Allonocardiopsis opalescens DSM 45601	Yes	Mesophile	PFFTMR	P	FFT	M	22
2509058093	EcthaDRAFT_2723	CRISPR-associated protein, Cse1 family	2554235031	Nocardiopsis potens DSM 45234	Yes	Mesophile	PFFTMR	P	FFT	M	22
2524650249	C793_00642	CRISPR system Cascade subunit Cas8	2515154141	Nonomuraea coxensis DSM 45129	Yes	Mesophile	PFFTMR	P	FFT	M	22
2503822743	BI299_1241	CRISPR system Cascade subunit Cas8	2808606818	Streptomyces sp. Mg1	Yes	Mesophile	PFFTMR	P	FFT	M	22

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2512026936	TIIST44_03605	CRISPR-associated protein, Cse1 family	2528768220	Actinokineospora inagensis DSM 44258	Yes	Mesophile	PFFTNR	P	FFT	N	22
2677094550	Ga011593_1397	CRISPR-associated protein, Cse1 family	2524023137	Actinomyces gerencseriae DSM 6844	Yes	Mesophile	PFFTTR	P	FFT	T	22
643581720	Ddes_0923	CRISPR system Cascade subunit Cas8	2565956542	Actinomyces israelii DSM 43320	Yes	Mesophile	PFFTTR	P	FFT	T	22
2505969160	MRE50lv_1752	CRISPR system Cascade subunit Cas8	2548876973	Actinomyces massiliensis 4401292	Yes	Mesophile	PFFTTR	P	FFT	T	22
2515333929	GaB11_00156	CRISPR-associated protein, Cse1 family	2523231055	Actinomyces suimastitidis DSM 15538	Yes	Mesophile	PFFTTR	P	FFT	T	22
646475273	ROD_30461	CRISPR system Cascade subunit Cas8	2600255104	Actinomyces urogenitalis S6-C4	Yes	Mesophile	PFFTTR	P	FFT	T	22
2562267663	CSSP291_13490	CRISPR system Cascade subunit Cas8	2562617184	Actinomyces viscosus C505	Yes	Mesophile	PFFTTR	P	FFT	T	22

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2503698823	Dole_2984	CRISPR system Cascade subunit Cas8	2731639183	Compostimonas suwonensis DSM 25625	Yes	Mesophile	PFFTTR	P	FFT	T	22
2502436610	Dsarc_43440	CRISPR system Cascade subunit Cas8	2522572156	Granulicoccus phenolivorans DSM 17626	Yes	Mesophile	PFFTTR	P	FFT	T	22
642677895	Glov_2479	CRISPR system Cascade subunit Cas8	2524614761	Pseudoclavibacter soli DSM 23366	Yes	Mesophile	PFFTTR	P	FFT	T	22
2562615778	KPR_4123	CRISPR system Cascade subunit Cas8	2510461000	Saccharomonospora paurometabolica YIM 90007	Yes	Mesophile	PFFTTR	P	FFT	T	22
640805919	Mmwyl1_3547	CRISPR system Cascade subunit Cas8	2513237375	Actinomyces graevenitzii C83	Yes	Mesophile	QFFTTR	Q	FFT	T	22
2516127798	Metunv3DRAFT_0059	CRISPR-associated protein, Cse1 family	651324006	Actinomyces sp. oral taxon 448 F0400	Yes	Mesophile	QFFTTR	Q	FFT	T	22
2516958901	MetmiDRAFT_0040	CRISPR-associated protein, Cse1 family	647000206	Schaalia odontolytica F0309	Yes	Mesophile	QFFTTR	Q	FFT	T	22

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
649659694	NIDE1542	CRISPR-associated protein, Cse1 family	646564557	Nocardiopsis dassonvillei DSM 43111	Yes	Mesophile	RFFTMR	R	FFT	M	22
2508671946	FrCN3DRAFT_1636	CRISPR system Cascade subunit Cas8	2600255109	Arcanobacterium sp. S3PF19	Yes	Mesophile	AIFSPK	I	FSP	K	22
2579858212	FF36_06138	CRISPR system Cascade subunit Cas8	2563366755	Lactobacillus pasteurii CRBIP 24.76	Yes	Mesophile	AIFSPK	I	FSP	K	22
2566130684	ES1_14530	CRISPR system Cascade subunit Cas8	2558860180	Lactobacillus helveticus H9	Yes	Mesophile	SIFSPK	I	FSP	K	22
2515910635	B153DRAFT_05717	CRISPR-associated protein, Cse1 family	2671180689	Lactobacillus antri DSM 16041	Yes	Mesophile	DIFSPN	I	FSP	N	22
2561349138	HMPREF1503_0992	CRISPR-associated protein, Cse1 family	643886145	Anaerococcus lactolyticus ATCC 51172	Yes	Mesophile	ALFSPK	L	FSP	K	22
2766196881	Ga0137923_113674	CRISPR system Cascade subunit Cas8	2529292727	Anaerococcus prevotii ACS-065-V-Col13	Yes	Mesophile	ALFSPK	L	FSP	K	22

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2637682039	Ga0081847_11681	CRISPR system Cascade subunit Cas8	2562617091	Atopobium vaginae DSM 15829	Yes	Mesophile	ALFSPK	L	FSP	K	22
644130105	HMPREF0072_0861	CRISPR system Cascade subunit Cas8	2547132195	Kallipyga massiliensis ph2	Yes	Mesophile	ALFSPK	L	FSP	K	22
2529738225	HMPREF9290_0209	CRISPR system Cascade subunit Cas8	2513237395	Lactobacillus iners 7_1_47FAA	Yes	Mesophile	ALFSPK	L	FSP	K	22
2562961603	HMPREF0091_10521	CRISPR system Cascade subunit Cas8	2588253851	Mageeibacillus indolicus 0009-5 S7-24-11	Yes	Mesophile	ALFSPK	L	FSP	K	22
2547746337	NoneDRAFT_01582	cse1 family CRISPR-associated protein	648276710	Peptoniphilus duerdenii ATCC BAA-1640	Yes	Mesophile	ALFSPK	L	FSP	K	22
2514731304	HMPREF1027_00751	CRISPR system Cascade subunit Cas8	2547132135	Peptoniphilus senegalensis JC140	Yes	Mesophile	ALFSPK	L	FSP	K	22
2589002815	HMPREF1632_07170	CRISPR system Cascade subunit Cas8	2600255110	Peptostreptococcus sp. MV1	Yes	Mesophile	ALFSPK	L	FSP	K	22

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
648807133	HMPREF9225_1593	CRISPR system Cascade subunit Cas8	2558860327	Streptococcus mutans G123	Yes	Mesophile	ALFSPK	L	FSP	K	22
2547521535	PTSHGDRAFT_01752	CRISPR system Cascade subunit Cas8	2548876625	Streptococcus sobrinus TCI-352	Yes	Mesophile	ALFSPK	L	FSP	K	22
2600982446	Ga0060238_01794	CRISPR system Cascade subunit Cas8	2513237373	Streptococcus sp. oral taxon 058 F0407	Yes	Mesophile	ALFSPK	L	FSP	K	22
2559537887	SMU61_01467	putative CRISPR system CASCADE complex protein Cas8	648276711	Peptoniphilus sp. F0141	Yes	Mesophile	SLFSPK	L	FSP	K	22
2549198957	K33DRAFT_00764	CRISPR system Cascade subunit Cas8	2519899669	Peptostreptococcus anaerobius DSM 2949	Yes	Mesophile	SLFSPK	L	FSP	K	22
2514660375	HMPREF9184_00518	CRISPR system Cascade subunit Cas8	2558860334	Streptococcus mutans NFSM2	Yes	Mesophile	TLFSPR	L	FSP	R	22
648676147	HMPREF9131_1418	CRISPR system Cascade subunit Cas8	2534682084	Gardnerella vaginalis 0288E	Yes	Mesophile	AVFSPK	V	FSP	K	22

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2520565252	F823DRAFT_01904	CRISPR system Cascade subunit Cas8	2558860923	Lachnospiraceae bacterium MSX33	Yes	Mesophile	AYFSPK	Y	FSP	K	22
2559553035	SMU52_08806	CRISPR-associated protein	649989950	Lachnoanaerobaculum saburreum DSM 3986	Yes	Mesophile	AYFSPR	Y	FSP	R	22
2528202242	ThrDRAFT_00111	CRISPR-associated protein, Cse1 family	2527291627	Frankia casuarinae Thr	Yes	Mesophile	PLFSSR	L	FSS	R	22
2511519937	CDHC03_0039	CRISPR-associated protein, Cse1 family	643348538	Desulfovibrio desulfuricans ATCC 27774	Yes	Mesophile	THFDHE	H	FDH	E	16
644132132	HMPREF0294_0580	CRISPR-associated protein, Cse1 family	642979316	Desulfovibrio piger ATCC 29098	Yes	Mesophile	TLFDHA	L	FDH	A	16
2515367852	A3ECDRAFT_1587	CRISPR system Cascade subunit Cas8	2608642258	Geobacter pickeringii G13	Yes	Mesophile	TLFDHG	L	FDH	G	16
2536743910	HMPREF0737_01132	CRISPR system Cascade subunit Cas8	2516653075	Methylomicrobium buryatense 5G	Yes	Mesophile	TLFDHG	L	FDH	G	16

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2548979052	NoneDRAFT_00867	CRISPR-associated protein, Cse1 family	2502790015	Aminomonas paucivorans GLU-3, DSM 12260	Yes	Mesophile	VLFDHH	L	FDH	H	16
647128456	HMPREF0297_0689	CRISPR system Cascade subunit Cas8	2579778672	Serpentinomonas raichei H1	Yes	Mesophile	VLFDHI	L	FDH	I	16
2599121444	B842_11840	CRISPR-associated protein, Cse1 family	2639762859	Azotobacter chroococcum NCIMB 8003	Yes	Mesophile	VLFDHS	L	FDH	S	16
2515846639	B097DRAFT_01247	CRISPR system Cascade subunit Cas8	2728369519	Murinocardiopsis flavida DSM 45312	Yes	Mesophile	TLFDHT	L	FDH	T	16
650933874	CRES_2077	CRISPR-associated protein, Cse1 family	639633050	Pelobacter propionicus DSM 2379	Yes	Mesophile	VLFDHT	L	FDH	T	16
648802771	HMPREF0574_1669	CRISPR system Cascade subunit Cas8	2585428156	Desulfatibacillum alkenivorans DSM 16219	Yes	Mesophile	VLFDHV	L	FDH	V	16
2771599009	Ga0244577_101272	CRISPR-associated Cse1 family protein	2802428809	Leptothrix mobilis DSM 10617	Yes	Mesophile	VVFDHA	V	FDH	A	16

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2527181697	H567DRAFT_01731	CRISPR-associated protein, Cse1 family	2529293194	Pseudomonas stutzeri ZoBell 632, ATCC 14405	Yes	Mesophile	VVFDHA	V	FDH	A	16
2504981494	ParJ4_00016520	CRISPR-associated protein, Cse1 family	2526164741	Azohydromonas australica DSM 1124	Yes	Mesophile	VVFDHG	V	FDH	G	16
2534465703	BURK_035484	CRISPR-associated protein, Cse1 family	2517572136	Chitiniphilus shinanonensis DSM 23277	Yes	Mesophile	VVFDHH	V	FDH	H	16
2501025101	Dshi_3215	CRISPR system Cascade subunit Cas8	2513237199	Rubrivivax gelatinosus IL144	Yes	Mesophile	VVFDHM	V	FDH	M	16
2509034985	JonanDRAFT_0157	CRISPR-associated protein, Cse1 family	2582581266	Lampropedia hyalina DSM 16112	Yes	Mesophile	VVFDHS	V	FDH	S	16
2574431821	BR51DRAFT_02839	CRISPR-associated protein	646311913	Citrobacter rodentium ICC168	Yes	Mesophile	DHFIKR	H	FIK	R	15
2506860697	Lepil_0432	CRISPR system Cascade subunit Cas8	2561511151	Cronobacter sakazakii Sp291	Yes	Mesophile	DHFIKR	H	FIK	R	15

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2547424803	ENTHGDRAFT_02641	CRISPR-associated protein, Cse1 family	2503692001	Desulfococcus oleovorans Hxd3	Yes	Mesophile	DHFIKR	H	FIK	R	15
2513101709	EBL_c30680	CRISPR-associated protein, Cse1 family	2502422304	Desulfosarcina variabilis Montpellier	Yes	Mesophile	DHFIKR	H	FIK	R	15
2508721592	Thi970DRAFT_4956	CRISPR-associated protein, Cse1 family	642555130	Geobacter lovleyi SZ	Yes	Mesophile	DHFIKR	H	FIK	R	15
2753224586	Ga0154114_1119516	CRISPR system Cascade subunit Cas8	2561511244	Klebsiella pneumoniae rhinoscleromatis SB3432	Yes	Mesophile	DHFIKR	H	FIK	R	15
2652591118	Ga0080901_108923	CRISPR-associated protein, Cse1 family	640753033	Marinomonas sp. MWYL1	Yes	Mesophile	DHFIKR	H	FIK	R	15
2587210999	JCM10415DRAFT_02102	CRISPR system Cascade subunit Cas8	2515154210	Methyloversatilis universalis Fam500	Yes	Mesophile	DHFIKR	H	FIK	R	15
650242942	HMPREF9219_1120	CRISPR system Cascade subunit Cas8	2516653058	Methylovulum miyakonense HT12	Yes	Mesophile	DHFIKR	H	FIK	R	15

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2600979453	Ga0060234_00094	CRISPR-associated protein, Cse1 family	649633030	Nitrospira defluvii	Yes	Mesophile	DHFIKR	H	FIK	R	15
2563937305	BN53_08570	CRISPR-associated protein, Cse1 family	2565956522	Photorhabdus australis DSM 17609	Yes	Mesophile	DHFIKR	H	FIK	R	15
2559091643	LBH_1246	CRISPR-associated protein, Cse1 family	2518285522	Photorhabdus khanii NC19	Yes	Mesophile	DHFIKR	H	FIK	R	15
2673799539	Ga0106094_10163	CRISPR system Cascade subunit Cas8	2508501051	Thiocystis violascens 611, DSM 198	Yes	Mesophile	DLFIKQ	L	FIK	Q	15
2503575203	Corgl_0421	CRISPR-associated protein, Cse1 family	2599185147	Marinospirillum alkaliphilum DSM 21637	Yes	Mesophile	DLFIKR	L	FIK	Q	15
645972807	SSPB78_010100015740	CRISPR-associated protein, Cse1 family	651324072	Methylophaga aminisulfidivorans MP, KCTC 12909	Yes	Mesophile	DHFVKG	H	FVK	G	15
2532537093	SZN_19188	CRISPR-associated protein, Cse1 family	637000204	Pelobacter carbinolicus Bd1, GraBd1	Yes	Mesophile	DHFVKG	H	FVK	G	15

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2563896362	Draft03411	CRISPR-associated Cse1 family protein	2767802764	Alteromonadales bacterium BS08 (Bankia setacea isolate)	Yes	Mesophile	DHFVKR	H	FVK	G	15
2735483105	Ga0180969_3172	CRISPR-associated protein, Cse1 family	2519899679	Arhodomonas aquaeolei DSM 8974	Yes	Mesophile	DHFVKR	H	FVK	G	15
2795799541	Ga0310471_1048	CRISPR system Cascade subunit Cas8	2574179768	Azoarcus communis DSM 12120	Yes	Mesophile	DHFVKR	H	FVK	G	15
2729865294	Ga0181037_10923	CRISPR system Cascade subunit Cas8	2506783010	Leptonema illini 3055, DSM 21528	Yes	Mesophile	DHFVKR	H	FVK	G	15
2554398419	D459DRAFT_05204	CRISPR-associated protein, Cse1 family	2547132115	Metakosakonia massiliensis JC163	Yes	Mesophile	DHFVKR	H	FVK	G	15
2515777278	A3G7DRAFT_01590	CRISPR system Cascade subunit Cas8	2513020017	Shimwellia blattae DSM 4481	Yes	Mesophile	DHFVKR	H	FVK	G	15
2810858371	Ga0325148_1498	CRISPR-associated protein, Cse1 family	2508501048	Thiorhodovibrio sp. 970	Yes	Mesophile	THFVKG	H	FVK	G	15

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2529257774	H504DRAFT_06104	CRISPR system Cascade subunit Cas8	2751185772	Thauera selenatis AX, ATCC 55363	Yes	Mesophile	AHFVKR	H	FVK	R	15
2524213658	G448DRAFT_02739	CRISPR system Cascade subunit Cas8	2524614757	Aliagarivorans taiwanensis DSM 22990	Yes	Mesophile	DLFVKR	L	FVK	R	15
2566111715	O145DRAFT_00985	CRISPR-associated protein, Cse1 family	2524614513	Arenimonas composti TR7-09, DSM 18010	Yes	Mesophile	DLFVKR	L	FVK	R	15
2550597463	W5WDRAFT_02820	CRISPR-associated protein, Cse1 family	2522572179	Marinobacterium litorale DSM 23545	Yes	Mesophile	DLFVKR	L	FVK	R	15
2523431396	G438DRAFT_1973	CRISPR system Cascade subunit Cas8	2513237204	Pararhodospirillum photometricum DSM 122	Yes	Mesophile	DLFVKR	L	FVK	R	15
2600971568	Ga0060230_01895	CRISPR-associated protein, Cse1 family	2574179790	Endozoicomonas numazuensis DSM 25634	Yes	Mesophile	DLFVKT	L	FVK	R	15
2514134393	GLX_07820	CRISPR system Cascade subunit Cas8	2597490360	Corynebacterium humireducens NBRC 106098 Genome sequencing	Yes	Mesophile	FTMR		FTM	R	10

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2525711607	G462DRAFT_04124	CRISPR system Cascade subunit Cas8	2515154153	Corynebacterium pilosum DSM 20521	Yes	Mesophile	FTMR		FTM	R	10
2524670199	K311DRAFT_01352	CRISPR-associated protein	650716029	Corynebacterium resistens DSM 45100	Yes	Mesophile	FTMR		FTM	R	10
2523179699	G403DRAFT_02362	cse1 family CRISPR-associated protein	648276689	Mobiluncus curtisii ATCC 35241	Yes	Mesophile	FTMR		FTM	R	10
2514184385	RSPPHO_02014	CRISPR system Cascade subunit Cas8	2529293003	Actinobaculum massiliae ACS-171-V-Col2	Yes	Mesophile	DHFTMR	H	FTM	R	10
2574518771	numaz_04447	CRISPR system Cascade subunit Cas8	2513237262	Corynebacterium casei UCMA 3821	Yes	Mesophile	DLFTMR	L	FTM	R	10
640548325	Gura_0827	CRISPR system Cascade subunit Cas8	2513237174	Bifidobacterium asteroides ATCC 25910	Yes	Mesophile	FLFTMR	L	FTM	R	10
2684382492	Ga0111619_10498	CRISPR system Cascade subunit Cas8	2558860221	Corynebacterium vitaeruminis DSM 20294 Genome sequencing	Yes	Mesophile	EYFTMR	Y	FTM	R	10

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2735793740	Ga0183458_11476	CRISPR system Cascade subunit Cas8	2529292687	Corynebacterium urealyticum DSM 7111	Yes	Mesophile	YFTMR	Y	FTM	R	10
2514526072	HMPREF0731_3087	CRISPR-associated protein, Cse1 family	2597490209	Corynebacterium ureicelerivorans IMMIB RIV-2301	Yes	Mesophile	YFTMR	Y	FTM	R	10
2600875626	Ga0056720_02572	CRISPR-associated protein, Cse1 family	2516653079	Haloglycomyces albus DSM 45210	Yes	Mesophile	PLFTPF	L	FTP	F	10
2509358994	ParJ55DRAFT_00033330	CRISPR system Cascade subunit Cas8	2554235317	Lactobacillus fermentum F6	Yes	Mesophile	AVFTPR	V	FTP	R	10
2511496528	ClimR_02359	CRISPR system Cascade subunit Cas8	2503754026	Bifidobacterium longum infantis UCD299	Yes	Mesophile	KFFTTR	F	FTT	R	9
2641020404	Ga0069373_1780	CRISPR system Cascade subunit Cas8	2511231219	Cutibacterium acnes ATCC 11828	Yes	Mesophile	KFFTTR	F	FTT	R	9
2547802249	ETEEDRAFT_03273	CRISPR system Cascade subunit Cas8	2675903216	Propionibacterium cyclohexanicum DSM 16859	Yes	Mesophile	KFFTTR	F	FTT	R	9

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2556947344	Q370DRAFT_02355	CRISPR-associated protein, Cse1 family	2523533527	Actinomyces vaccimaxillae DSM 15804	Yes	Mesophile	PLFTTR	L	FTT	R	9
2526102338	G465DRAFT_01328	CRISPR-associated protein, Cse1 family	2506783060	Mycobacterium sp. JS623	Yes	Mesophile	KYFTTR	Y	FTT	R	9
642684646	Cphamn1_2159	CRISPR-associated protein, Cse1 family	637000198	Nocardia farcinica IFM 10152	Yes	Mesophile	KYFTTR	Y	FTT	R	9
642725403	Paes_1418	CRISPR-associated Cse1 family protein	2770939500	Nocardia neocaledoniensis DSM 44717	Yes	Mesophile	KYFTTR	Y	FTT	R	9
2555724984	A606_11705	CRISPR system Cascade subunit Cas8	2505679089	Cellulomonas fimi NRS 133, ATCC 484	Yes	Mesophile	PYFTTR	Y	FTT	R	9
2601009591	Ga0060241_02331	CRISPR-associated Cse1 family protein	2757320518	Saccharothrix australiensis DSM 43800	Yes	Mesophile	PYFTTR	Y	FTT	R	9
2597415109	LX16DRAFT_4360	CRISPR system Cascade subunit Cas8	2511231059	Corynebacterium diphtheriae HC03	Yes	Mesophile	FSMR		FSM	R	7

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2508736466	Thivi_0457	CRISPR-associated protein, Cse1 family	643886147	Corynebacterium glucuronolyticum ATCC 51867	Yes	Mesophile	FSMR		FSM	R	7
2599295717	Ga0002112_01367	CRISPR system Cascade subunit Cas8	2515154059	Corynebacterium ulceribovis DSM 45146	Yes	Mesophile	FSMR		FSM	R	7
650187958	HMPREF9436_01696	CRISPR-associated protein, Cse1 family	2541046979	Propionimicrobium lymphophilum ACS-093-V-SCH5	Yes	Mesophile	GLFSMR	L	FSM	R	7
650558791	FPR_15490	putative CRISPR system CASCADE complex protein Cas8	649989963	Lactobacillus iners LEAF 3008A-a	Yes	Mesophile	SIFSMK	I	FSM	K	7
2585132787	EJ14DRAFT_00920	CRISPR system Cascade subunit Cas8	2765235980	Streptomyces rubrolavendulae MJM4426	Yes	Mesophile	ALFSMR	L	FSM	R	7
2503957434	Spico_1126	CRISPR-associated protein, Cse1 family	2636415710	Bifidobacterium animalis lactis ATCC 27673	Yes	Mesophile	MLFSMR	L	FSM	R	7

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2514665395	HMPREF0045_01086	CRISPR system Cascade subunit Cas8	2531839682	Burkholderia sp. SJ98	Yes	Mesophile	HDLK		HDL	K	6
651677024	HMPREF9062_2116	CRISPR-associated protein, Cse1 family	2501004205	Dinoroseobacter shibae DFL-12, DSM 16493	Yes	Mesophile	HDLK		HDL	K	6
647126765	HMPREF0970_01439	CRISPR system Cascade subunit Cas8	2508501103	Jonquetella anthropi ADV 126, DSM 22815	Yes	Mesophile	HDLK		HDL	K	6
646837379	Ndas_1281	CRISPR system Cascade subunit Cas8	2651869728	Methylobacterium platani JCM 14648	Yes	Mesophile	HDLK		HDL	K	6
2637635308	Ga0069377_111220	CRISPR-associated protein, Cse1 family	643692051	Thauera aminoaromatica MZ1T	Yes	Mesophile	HDLK		HDL	K	6
2795707106	Ga0310538_106149	CRISPR system Cascade subunit Cas8	2547132082	Verminephrobacter aporrectodeae subsp. tuberculatae At4	Yes	Mesophile	HDLK		HDL	K	6
637126060	GSU1385	CRISPR-associated protein, Cse1 family	645058855	Streptomyces viridochromogenes DSM 40736	Yes	Mesophile	PFFSAR	P	FFS	A	5

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
647338324	DFW101DRAFT_1038	putative CRISPR-associated helicase Cas3 family protein	645951849	Streptomyces sp. SPB78	Yes	Mesophile	PFFSMR	P	FFS	M	5
2568553918	BR08DRAFT_02498	CRISPR-associated protein, Cse1 family	2531839181	Streptomyces zinciresistens K42	Yes	Mesophile	PFFSMR	P	FFS	M	5
651554917	HMPREF9439_02464	CRISPR system Cascade subunit Cas8	2563366745	Actinoalloteichus spitiensis RMV-1378	Yes	Mesophile	PFFSTR	P	FFS	T	5
2529306061	H566DRAFT_2644	CRISPR-associated Cse1 family protein	2734482175	Actinokineospora cianjurensis DSM 45657	Yes	Mesophile	PFFSTR	P	FFS	T	5
2566040115	BO26DRAFT_01316	CRISPR-associated protein, Cse1 family	637000075	Chromohalobacter salexigens 1H11, DSM 3043	Yes	Mesophile	DFFVKR	D	FFV	K	5
2518346120	PTE_03981	CRISPR-associated protein, Cse1 family	2516143004	Desulfonatronum lacustre Z-7951, DSM 10312	Yes	Mesophile	DFFVKR	D	FFV	K	5
643570220	Mpal_1607	CRISPR-associated protein, Cse1 family	637000095	Desulfovibrio alaskensis G20	Yes	Mesophile	DFFVKR	D	FFV	K	5

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2509742495	Desti_4554	CRISPR system Cascade subunit Cas8	2508501110	Ectothiorhodospira haloalkaliphila Imhoff 51/7, ATCC 51935	Yes	Mesophile	DFFVKR	D	FFV	K	5
2530838541	HMPREF9233_00304	CRISPR-associated protein, Cse1 family	2524614508	Halomonas jeotgali Hwa	Yes	Mesophile	DFFVKR	D	FFV	K	5
2523323381	G453DRAFT_00126	CRISPR-associated protein, Cse1 family	647000236	Desulfovibrio carbinoliphilus oakridgensis FW1012B	Yes	Mesophile	DFFIKR	D	FFI	K	3
2508699723	DesteDRAFT_0757	CRISPR system Cascade subunit Cas8	2568526008	Desulfovibrio gracilis DSM 16080	Yes	Mesophile	DFFIKR	D	FFI	K	3
2541323588	HMPREF9306_01427	CRISPR-associated protein, Cse1 family	651324084	Parasutterella excrementihominis YIT 11859	Yes	Mesophile	DFFIKR	D	FFI	K	3
2525313820	K318DRAFT_0497	CRISPR system Cascade subunit Cas8	2556921007	Microvirgula aerodenitrificans DSM 15089	Yes	Mesophile	NLFFNK	L	FFN	K	3
2517450211	S272_02933	CRISPR-associated protein, Cse1 family	2636415698	Serpentinomonas raichei A1	Yes	Mesophile	TFFNEA	T	FFN	E	3

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2533660215	ERJG_01842	CRISPR-associated Cse1 family protein	2795385449	Plasticicumulans lactativorans DSM 25287	Yes	Mesophile	TFFNPE	T	FFN	P	3
644957035	Caci_3909	CRISPR system Cascade subunit Cas8	2770939511	Bradymonas sediminis DSM 28820	Yes	Mesophile	HDIK		HDI	K	3
647545273	SCLAV_2741	CRISPR-associated protein, Cse1 family	2526164742	Desulfatiglans anilini DSM 4660	Yes	Mesophile	HDIK		HDI	K	3
640538636	Acry_3315	CRISPR-associated Cse1 family protein	2504756064	Paracoccus sp. J4	Yes	Mesophile	HDIK		HDI	K	3
2536530280	CGSMWGv0288E_05610	CRISPR-associated Cse1 family protein	2734482251	Humitalea rosea DSM 24525	Yes	Mesophile	DLLVHR	L	LVH	R	3
2555619118	LBFF_1329	CRISPR system Cascade subunit Cas8	2513237333	Roseomonas cervicalis ATCC 49957	Yes	Mesophile	DLLVHR	L	LVH	R	3
2521741006	H163DRAFT_04036	CRISPR-associated protein, Cse1 family	640427101	Acidiphilium cryptum JF-5	Yes	Mesophile	GPLVHP	P	LVH	P	3

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
650306251	HMPREF0381_0522	CRISPR-associated protein, Cse1 family	2639762859	Azotobacter chroococcum NCIMB 8003	Yes	Mesophile	VLTQSQ	L	TQS	Q	3
2559206578	B843_10185	CRISPR system Cascade subunit Cas8	2547132206	Edwardsiella tarda 080813	Yes	Mesophile	VLTQSQ	L	TQS	Q	3
2529594186	CU7111_1892	CRISPR system Cascade subunit Cas8	2556921006	Franconibacter pulveris DSM 19144	Yes	Mesophile	VLTQSQ	L	TQS	Q	3
2559151219	GbCGDNIH4_1693	CRISPR system Cascade subunit Cas8	2585427921	Lactobacillus paralimentarius TB 1	Yes	Mesophile	SIFAPK	I	FAP	K	2
647357455	HMPREF7215_1778	CRISPR system Cascade subunit Cas8	2534681768	Streptococcus ratti FA-1	Yes	Mesophile	SLFAPR	L	FAP	R	2
2652590425	Ga0080901_105814	CRISPR-associated protein, Cse1 family	2558860205	Granulibacter bethesdensis CGDNIH2	Yes	Mesophile	FDLK		FDL	K	2
643699249	Tmz1t_2229	CRISPR-associated protein, Cse1 family	2558860206	Granulibacter bethesdensis CGDNIH4	Yes	Mesophile	FDLK		FDL	K	2

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2547299932	VTUDRAFT_02456	CRISPR system Cascade subunit Cas8	2596583525	Desulforhopalus singaporensis DSM 12130	Yes	Mesophile	TLFDSA	L	FDS	A	2
643593574	A2cp1_0779	CRISPR-associated protein, Cse1 family	2526164518	Spirochaeta cellobiosiphila DSM 17781	Yes	Mesophile	VLFDSQ	L	FDS	Q	2
2502855286	Apau_0539	CRISPR-associated protein, Cse1 family	649989951	Faecalibacterium cf. prausnitzii KLE1255	Yes	Mesophile	RLFPLY	L	FPL	Y	2
2580493015	SRAHDRAFT_00756	CRISPR-associated protein, Cse1 family	650377940	Faecalibacterium prausnitzii SL3/3	Yes	Mesophile	RLFPLY	L	FPL	Y	2
639756392	Ppro_2342	CRISPR-associated protein, Cse1 family	2508501039	Frankia saprophytica CN3	Yes	Mesophile	PLFSAR	L	FSA	R	2
2588105914	EJ43DRAFT_04710	CRISPR-associated protein, Cse1 family	2579778521	Frankia torreyi CpI1-S	Yes	Mesophile	PLFSAR	L	FSA	R	2
2579857243	FF36_05168	CRISPR-associated protein, Cse1 family	2579778521	Frankia torreyi CpI1-S	Yes	Mesophile	PLFSSR	L	FSS	R	2

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2521849049	F591DRAFT_04162	CRISPR system Cascade subunit Cas8	2521172661	Paenibacillus ginsengihumi DSM 21568	Yes	Mesophile	RLFSSR	L	FSS	R	2
2517048649	HalalDRAFT_1689	CRISPR-associated protein, Cse1 family	2548876527	Brachybacterium squillarum M-6-3	Yes	Mesophile	FTLR		FTL	R	2
2523655867	G439DRAFT_0067	CRISPR-associated protein, Cse1 family	647000231	Corynebacterium jeikeium ATCC 43734	Yes	Mesophile	FTLR		FTL	R	2
2581724394	P304DRAFT_00951	CRISPR-associated protein, Cse1 family	2524614662	Brevibacterium album DSM 18261	Yes	Mesophile	AAFTQR	A	FTQ	R	2
2600354560	Ga0056080_0939	CRISPR system Cascade subunit Cas8	2517434006	Brevibacterium casei S18	Yes	Mesophile	QAFTQR	A	FTQ	R	2
2563314316	HMPREF0059_00198	CRISPR-associated protein, Cse1 family	2531839473	Escherichia coli M863	Yes	Mesophile	AFVNQP	A	FVN	Q	2
2731809968	Ga0181017_0211	CRISPR-associated protein, Cse1 family	637000120	Geobacter sulfurreducens PCA	Yes	Mesophile	CFVNEP	C	FVN	E	2

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2802460013	Ga0310535_2613	CRISPR-associated protein, Cse1 family	643348507	Anaeromyxobacter dehalogenans 2CP-1	Yes	Mesophile	HDVK		HDV	K	2
2531539294	PstZobell_18105	CRISPR-associated protein, Cse1 family	642555106	Anaeromyxobacter sp. K	Yes	Mesophile	HDVK		HDV	K	2
2509097257	SacglDRAFT_1985	CRISPR system Cascade subunit Cas8	2600255081	Haematobacter massiliensis CCUG 47968	Yes	Mesophile	TLMVRE	L	MVR	E	2
2506446541	FrEUN1f_2435	CRISPR system Cascade subunit Cas8	2509276015	Paracoccus sp. J55	Yes	Mesophile	DLMVRR	L	MVR	R	2
2507124138	Mycsm_07018	CRISPR-associated protein, Cse1 family	642555122	Chlorobium phaeobacteroides BS1	Yes	Mesophile	ILTQYQ	L	TQY	Q	2
637576995	nfa44270	CRISPR-associated protein, Cse1 family	642555149	Prosthecochloris aestuarii SK413, DSM 271	Yes	Mesophile	ILTQYQ	L	TQY	Q	2
2559148509	GbCGDNIH2_1693	CRISPR system Cascade subunit Cas8	2503538010	Coriobacterium glomerans PW2, DSM 20642	Yes	Mesophile	YLFAMK	L	FAM	K	1

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
642762231	AnaeK_0779	CRISPR-associated protein, Cse1 family	647000302	Pyramidobacter piscolens W5455	Yes	Mesophile	FDVK		FDV	K	1
648709640	DesfrDRAFT_2476	CRISPR system Cascade subunit Cas8	2502957028	Desulfatibacillum aliphaticivorans AK-01	Yes	Mesophile	QVFESQ	V	FES	Q	1
651614077	MAMP_02712	CRISPR-associated protein, Cse1 family	2515154197	Streptomyces scabrisporus DSM 41855	Yes	Mesophile	SWFGHV	W	FGH	V	1
637750548	Pcar_0957	CRISPR-associated protein, Cse1 family	2505679073	Methanocella arvoryzae MRE50	Yes	Mesophile	SHFHHG	H	FHH	G	1
2770888547	Ga0248310_110324	CRISPR system Cascade subunit Cas8	2596583683	Stackebrandtia albiflava DSM 45044	Yes	Mesophile	PLFIGR	L	FIG	R	1
2520590213	C516DRAFT_03486	CRISPR system Cascade subunit Cas8	2515154048	Candidatus Gilliamella apicola wkB11	Yes	Mesophile	DHFIKR	H	FIK	R	1
2535399218	SRA_06686	CRISPR system Cascade subunit Cas8	2600255117	Corynebacterium freneyi DNF00450	Yes	Mesophile	AMFIRR	M	FIR	R	1

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
643137403	DESPIG_00183	CRISPR-associated protein, Cse1 family	643348525	Methanosphaerula palustris E1-9c, DSM 19958	Yes	Mesophile	DHFLKR	H	FLK	R	1
2609281803	Ga0069007_111993	CRISPR-associated protein, Cse1 family	644736339	Catenulispora acidiphila ID139908, DSM 44928	Yes	Mesophile	PFLTMR	P	FLT	M	1
2517030209	METBUDRAFT_3498	CRISPR-associated protein, Cse1 family	2639762859	Azotobacter chroococcum NCIMB 8003	Yes	Mesophile	NLYFNK	Y	FNK		1
2641021862	Ga0069373_171542	CRISPR system Cascade subunit Cas8	2582581275	Alkalibacter saccharofermentans DSM 14828	Yes	Mesophile	RLFPQR	L	FPQ	R	1
2730702738	Ga0181034_12646	CRISPR-associated protein, Cse1 family	2503904012	Sphaerochaeta coccoides SPN1, DSM 17374	Yes	Mesophile	ILFQSQ	L	FQS	Q	1
2596708879	LX67DRAFT_03261	CRISPR system Cascade subunit Cas8	2565956546	Eubacterium siraeum V10Sc8a	Yes	Mesophile	RLFSDR	L	FSD	R	1
2526232065	K345DRAFT_02297	CRISPR type I-E/ECOLI-associated protein Cas8/Cse1	2515154167	Spirosoma panaciterrae DSM 21099	Yes	Mesophile	LLFSHD	L	FSH	D	1

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2556951212	Q352DRAFT_01755	CRISPR-associated protein, Cse1 family	2558860929	Olsenella uli MSTE5	Yes	Mesophile	RLFSLR	L	FSL	R	1
2514240275	BANAN_06875	CRISPR system Cascade subunit Cas8	2534682146	Rothia mucilaginosa M508	Yes	Mesophile	FSVR		FSV	R	1
2514274672	CCAS_06445	CRISPR system Cascade subunit Cas8	2513237236	Bifidobacterium animalis animalis ATCC 25527	Yes	Mesophile	RLFTIR	L	FTI	R	1
2514074346	BAST_00530	CRISPR system Cascade subunit Cas8	2509601019	Desulfomonile tiedjei DCB-1, DSM 6799	Yes	Mesophile	HFTKS	H	FTK	S	1
2730703684	Ga0181034_1458	CRISPR system Cascade subunit Cas8	2524614800	Jonesia quinghaiensis DSM 15701	Yes	Mesophile	QYFTVR	Y	FTV	R	1
645414198	SvirD4_010100032121	CRISPR system Cascade subunit Cas8	2513237191	Komagataeibacter medellinensis NBRC 3288	Yes	Mesophile	DLFVHR	L	FVH	R	1
2523097800	G531DRAFT_03367	CRISPR-associated protein, Cse1 family	2651869728	Methylobacterium platani JCM 14648	Yes	Mesophile	DHFVRR	H	FVR	R	1

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2525721595	G557DRAFT_0118	CRISPR-associated protein, Cse1 family	647533233	Streptomyces clavuligerus ATCC 27064	Yes	Mesophile	PFWSAR	P	FWS	A	1
2510465282	SacpaDRAFT_00042900	CRISPR-associated protein, Cse1 family	640427115	Geobacter uraniireducens Rf4	Yes	Mesophile	TLHDHG	L	HDH	G	1
2527172922	H537DRAFT_01000	CRISPR-associated protein, Cse1 family	648276636	Desulfovibrio fructosovorans JJ	Yes	Mesophile	HGEK		HGE	K	1
2517864160	C556DRAFT_00296	CRISPR system Cascade subunit Cas8	2523231030	Desulfovibrio putealis DSM 16056	Yes	Mesophile	HGIK		HGI	K	1
2514159863	RGE_06880	CRISPR system Cascade subunit Cas8	2508501043	Desulfovibrio termitidis HI1	Yes	Mesophile	HQIK		HQI	K	1
2585104745	EK00DRAFT_02472	CRISPR-associated protein, Cse1 family	2681813507	Insolitispirillum peregrinum integrum DSM 11589	Yes	Mesophile	VLLATQ	L	LAT	Q	1
2503010693	Dalk_4934	CRISPR system Cascade subunit Cas8	2521172637	Pannonibacter phragmitetus DSM 14782	Yes	Mesophile	DVLTHR	V	LTH	R	1

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2559129212	GbCGDNIH3_1380	CRISPR-associated protein, Cse1 family	2558860199	Granulibacter bethesdensis CGDNIH3	Yes	Mesophile	DVMVHR	V	MVH	R	1
641271477	Sare_1972	CRISPR system Cascade subunit Cas8	2541046977	Actinomyces europaeus ACS-120-V-Col10b	Yes	Mesophile	PYMVMR	Y	MVM	R	1
2516060777	A3ICDRAFT_08235	CRISPR system Cascade subunit Cas8	2579778963	Chrysiogenes arsenatis DSM 11915	Yes	Mesophile	TLNDHG	L	NDH	G	1
2641021869	Ga0069373_171549	CRISPR system Cascade subunit Cas8	2600254929	Marinospirillum celere DSM 18438	Yes	Mesophile	VLNQSQ	L	NQS	Q	1
2561332176	HMPREF1495_0701	CRISPR-associated protein, Cse1 family	2511231051	Chlorobaculum limnaeum RK-j-1	Yes	Mesophile	ILSQFQ	L	SQF	Q	1
2598725891	Ga0057669_02324	CRISPR-associated protein, Cse1 family	2524614869	Aquaspirillum serpens DSM 68	Yes	Mesophile	VLTQTQ	L	TQT	Q	1
2771547737	Ga0244503_10510	CRISPR-associated protein, Cse1 family	641228504	Salinispora arenicola CNS-205	Yes	Mesophile	VWFGHH	V	WFG	H	1

Table S3 (continued)

Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Cultured	Temperature Range	Cas8 active site	pre-motif AA	main motif	post-motif AA	hits
2506044905	Cfi_0001.00005110	CRISPR system Cascade subunit Cas8	2508501119	Saccharomonospora glauca K62, DSM 43769	Yes	Mesophile	VWHSHT	V	WHS	H	1
2758116095	Ga0197483_2742	CRISPR system Cascade subunit Cas8	2506381019	Frankia sp. EUN1f	Yes	Mesophile	VWLGHH	V	WLG	H	1
2525865037	G538DRAFT_0014	CRISPR-associated Cse1 family protein	2728369519	Murinocardiopsis flavida DSM 45312	Yes	Mesophile	MWLSHD	M	WLS	H	1
2541319828	HMPREF9238_01221	CRISPR system Cascade subunit Cas8	2554235357	Corynebacterium terpenotabidum Y-11 Genome sequencing	Yes	Mesophile	DLYTMR	L	YTM	R	1

Table S4: Self-targeting spacers and targets within *X. albilineans*, related to Figure 4.

position in the genome	genomic target sequence 5' -> 3' (matching spacer)	5' flanking sequence	associated spacer	associated spacer number	related CRISPR system	mis-matches
195944 - 195912	ATCGTGGCCAACGACGCCACG GTCAAGGGCGGCAC	GCGTG	ATCCTGGCTATTGTCGCAA TTGTCAAGGGCGGCGC	array 4 spacer24	I-C	9
223306 - 223274	CGCATTGACTCCCAGCGCGCAT ACGGCACTGCA	GATTC	CGCATTGACTCCCAGCGC GCATACGGCACTGCA	array 4 spacer1	I-C	0
224773 - 224737	ATGTACTACGTCCACGGGAGGG CCATCGCATGAGCCG	ACCTC	ATGTACTACGTACACGGGA GGGCCATCGCATGAGCCG	array 4 spacer30	I-C	1
225997 - 225966	TGGTGGACGCATCCGCGCAGC ACTTGCGCAGG	ACGCC	TGGTGGCCGCATCCGCGC AGCACTTGCGCAGG	array 6 spacer21	I-F1	1
226162 - 226125	GCATGTGCGAATGCACGACGGT CGGGGCCGAATGATGG	CATTC	GCATGTGCGAATGCACGA CGGTCGGGGCCGAATGAT GG	array4 spacer21	I-C	0
227986 - 227955	TGACACTTAGTTAAGCTTTTCAT GGATCACTC	CAACC	TGACACTTAGTTAAGCTTTT CATGGATCACTC	array6 spacer4	I-F1	0
232114 - 232080	CGACCGAGACCGACTCGCCTTC CAGTCCGCGCAGC	GATTT	CGACCGAGACCGACTCGC CTTCCAGCCCGCGCAGG	array4 spacer18	I-C	1
239454 - 239418	GGCACAAGCGTCCAGCCATCC GGCACACCCTGCAGGC	CGTTC	GGCACAAGCGTCCAGCCA TCCGGCACACCCTGCAGG C	array4 spacer32	I-C	0
241168 - 241135	GTGCTCCTCGATGACGGATCCG CAGTCGCTGCAG	TTTTC	GTGCTCCTCGATGACGGAT CCGCAGTCGCTGCAG	array4 spacer19	I-C	0
242415 - 242384	ATGTCCTGCGGCAGCAACGCG CCGATGGGGGC	CGACC	ATGTCCTGCGGCAGCAAC GCGCCGATGGGGGC	array6 spacer19	I-F1	0
242885 - 242918	TCCTAACCCTGTTGCGTCCGGA CGTGATCTGACA	AATTT	TCCTAACCCTGTTGCGTCC GGACGTGATCTGACG	array4 spacer16/20	I-C	1
242924 - 242893	CATTATTGTCAGATCACGTCCG GACGCAACAG	CTCCT	TATGATTGTCAGATCGCGT CCGGACGTAACAG	array5 spacer1	I-F1	4
245629 - 245593	AAGCGATCCGATGCGGTCGGC CATGCCGGCCTTGACT	GATTC	AAGCGATCCGATGCGGTC GGCCATGCCGGCCTTGAC G	array4 spacer17	I-C	1

Table S4 (continued)

position in the genome	genomic target sequence 5' -> 3' (matching spacer)	5' flanking sequence	associated spacer	associated spacer number	related CRISPR system	mis-matches
254149 - 254113	GCCAGTCAACGTGGAGAGATTCACGGCAGCGCGGCTT	CTTTC	GCCAGTCAACGTGGAGAGATTCACGGCAGCGCGGCTT	array4 spacer31	I-C	0
256054 - 256022	AATCGCCGCTGCGCCCGCACTGACCGATGTAC	TGTCC	AATCGCCGCAGCGCCCGCACTGACCGATGTAC	array2 spacer1	I-F1	1
260227 - 260261	GAGTTCAGCCGCGGCCCGCTGGTCTGGCGTTACAC	TGTTC	GAGTTCAGCCGCGGCCCGCTGGTCTGGCGTTACAG	array4 spacer25	I-C	1
271563 - 271528	AACGGTAGGAAGGCCGCTGAGGACTGGGCGCGCGCG	CGTTC	AACGGTAGGAAGGCCGCTGAGGACTGGGCGCGCGCG	array4 spacer28	I-C	0
373654 - 373688	CATTTCCTCCTATCCGCACCCGTGGCTGATGCCAG	CAGGG	GATTTCGTCCTCTCCGTAGCTGTATGGGCTGATGCCAG	array4 spacer29	I-C	9
1795082 - 1795047	CTTTGGAAATGCTCACCCATCCCATGCGCTATCTGC	ACTTC	CTTTGGAAATGCTCACCCACCCCATGCGCTATCTGG	array4 spacer23	I-C	2
1796072 - 1796038	GACCTCGCTGGCGTACTTATAAAGATTGATGGTCA	TGTTC	GACCTTGCTGGCGTACTCATAGAGATTGATGGTCG	array4 spacer12	I-C	4
83 - 52 (plasmid I)	TTCTGCGCCGCAATCACAATAGTTTGCATGAT	CGCCC	TTCTGCGCCGCAATCACAATAGTCTGCATGAT	array6 spacer16	I-F1	1
495 - 526 (plasmid I)	CATGCGCCAAGCGATCGAGGAAGGTGGGTTAA	GAACC	GATGCGCCAGGCCATCGAGGAAGGCGGGTTAA	array6 spacer15	I-F1	4
21805 - 21774 (plasmid I)	TCGCTGCGCCATAGATTCCGGCCGTCCACGTC	CCGCC	TCGCTGCGCCATAGATTCCGGCCGTCCACGTC	array6 spacer11	I-F1	0

Table S5: qPCR analysis of VcCAST transposition efficiency in TXTL, related to Figure 5.

Left to Right End transposition (left end)	
PAM	**2^(-ΔΔCt)**
GTACC	0.000106791
GTCAA	0.000166159
GTAAA	0.000141569
Left to Right End transposition (right end)	
PAM	**2^(-ΔΔCt)**
GTACC	0.000767149
GTCAA	0.000748387
GTAAA	0.000727097
Right to Left End transposition (right end)	
PAM	**2^(-ΔΔCt)**
GTACC	1.78568E-05
GTCAA	1.35544E-05
GTAAA	2.26264E-05
Right to Left End transposition (left end)	
PAM	**2^(-ΔΔCt)**
GTACC	2.25506E-05
GTCAA	3.08333E-05
GTAAA	2.73168E-05

Table S6: Lists of plasmids, primers, and strains used in this study, related to Figures 2, 3, 4, 5, 6, and 7 and STAR Methods.

PLASMID LIST					
Name	**Lab number**	**Description**	**Source**	**Addgene number**	**Link**
E. coli_I-E array_GFP_dropout	CBS-1268	Golden Gate GFP dropout vector to generate *E. coli* type I-E single arrays	this study	-	https://benchling.com/s/seq-0rwG4NfXvn550EFgLvF9
p70a_deGFP	CBS-338	encoding *degfp* with p70a promoter	commercially available from arbor bioscience	-	https://benchling.com/s/seq-luGf8hlOBTwWXilrHQ2q
p70a_deGFP_ICST1	CBS-4188	encoding IC self-target 1 upstream of *degfp*-promoter	this study	-	https://benchling.com/s/seq-CitRlnOFJYjnk4CE4Hvu
p70a_deGFP_ICST2	CBS-4189	encoding IC self-target 2 upstream of *degfp*-promoter	this study	-	https://benchling.com/s/seq-vh1UAwc6zeeekWJIKG5g
p70a_deGFP_IF1ST1	CBS-4190	encoding IF1 self-target 1 upstream of *degfp*-promoter	this study	-	https://benchling.com/s/seq-5IbDwTgrNDCzK4Rejf2X
p70a_deGFP_IF1ST2	CBS-4191	encoding IF1 self-target 2 upstream of *degfp*-promoter	this study	-	https://benchling.com/s/seq-mIYrTYOPg1AojPJ1iU1l
p70a_deGFP_PacI	CBS-332	Starting vector for pPAM_library	this study	170100	https://benchling.com/s/seq-I4jYts43tCNhqzsshRF5
p70a_T7RNAP	CBS-011	expressing T7 RNA-Polymerase	Garamella et al. 2016 (PMID: 26818434)	-	https://benchling.com/s/seq-C5XpSSJcu2SmYf7rjK7Z
pAc1_Cas5	CBS-1529	encoding *A. chroococcum* type I-E #1 *cas5*	this study*	-	https://benchling.com/s/seq-waIYSsvGAvlOSJTNdPev
pAc1_Cas6	CBS-1530	encoding *A. chroococcum* type I-E #1 *cas6*	this study*	-	https://benchling.com/s/seq-ASkZLMA2q0QmfEC7jmWF
pAc1_Cas7	CBS-1528	encoding *A. chroococcum* type I-E #1 *cas7*	this study*	-	https://benchling.com/s/seq-SwKnvN1W28aBEy7ITjXt
pAc1_Cas8	CBS-1526	encoding *A. chroococcum* type I-E #1 *cas8*	this study*	-	https://benchling.com/s/seq-iA9WQ2DN3qZMj7urKqbV

Table S6 (continued)

PLASMID LIST					
Name	**Lab number**	**Description**	**Source**	**Addgene number**	**Link**
pAc1_Cse2	CBS-1527	encoding *A. chroococcum* type I-E #1 *cse2*	this study*	-	https://benchling.com/s/seq-zWWZSpo2Kq75ZiEYVx7J
pAc2_Cas5	CBS-1534	encoding *A. chroococcum* type I-E #2 *cas5*	this study*	178737	https://benchling.com/s/seq-yV1Ufp8MJjfE5FngM5s5
pAc2_Cas6	CBS-1535	encoding *A. chroococcum* type I-E #2 *cas6*	this study*	178738	https://benchling.com/s/seq-mLvHziHQri4MfE8ULOny
pAc2_Cas7	CBS-1533	encoding *A. chroococcum* type I-E #2 *cas7*	this study*	178739	https://benchling.com/s/seq-GqcMhmFgSsbUCSYfFfmR
pAc2_Cas8	CBS-1531	encoding *A. chroococcum* type I-E #2 *cas8*	this study	178735	https://benchling.com/s/seq-xkjM2MrPuiAbZo6pwxsz
pAc2_Cse2	CBS-1532	encoding *A. chroococcum* type I-E #2 *cse2*	this study*	178736	https://benchling.com/s/seq-NqeHcYr0refZa9HVbabU
pAc2_gRNA1	CBS-2441	targeting next to PAM library	this study	178740	https://benchling.com/s/seq-5LQb079ptp2Bi0EVIpsW
pAc2_gRNAnt	CBS-3029	non-targeting gRNA	this study	-	https://benchling.com/s/seq-vq8soTShmA0AYrNhgaOH
pAc3_Cas5	CBS-1539	encoding *A. chroococcum* type I-E #3 *cas5*	this study*	178743	https://benchling.com/s/seq-U3efg1fPXXly0MH4g3rc
pAc3_Cas6	CBS-1540	encoding *A. chroococcum* type I-E #3 *cas6*	this study*	178744	https://benchling.com/s/seq-5Pw4CttK6lEGGbDXtUQU
pAc3_Cas7	CBS-1538	encoding *A. chroococcum* type I-E #3 *cas7*	this study*	178745	https://benchling.com/s/seq-dg26rU9hjsB7f5PButSd
pAc3_Cas8	CBS-1536	encoding *A. chroococcum* type I-E #3 *cas8*	this study*	178741	https://benchling.com/s/seq-XStQVbmNQzOcDoQFW9Kt
pAc3_Cse2	CBS-1537	encoding *A. chroococcum* type I-E #3 *cse2*	this study*	178742	https://benchling.com/s/seq-D1EXrvZtCvkRTnWmORlA
pAc3_gRNA1	CBS-2440	targeting next to PAM library, used for Ac3	this study	178746	https://benchling.com/s/seq-MZMeWGeydG5DHm4udG15
pAc3_gRNA2	CBS-3026	targeting protospacer in *gfp*-promoter, used for Ac3	this study	-	https://benchling.com/s/seq-o24LR2BFbSp2qiwGUKW0

Table S6 (continued)

PLASMID LIST					
Name	**Lab number**	**Description**	**Source**	**Addgene number**	**Link**
pAc3_gRNAnt	CBS-3028	non-targeting gRNA	this study	-	https://benchling.com/s/seq-qzioOKi4zHnbz8dRpRfX
pAnava_array_GFP_dropout	CBS-2491	Golden Gate GFP dropout vector to generate *A. variabilis* type I-B single arrays	this study	-	https://benchling.com/s/seq-o2CEkxrDymYYdjWekCk3
pAnava_Cas	CBS-3688	encoding *A. variabilis* type I-B CAST Cascade	this study**	-	https://benchling.com/s/seq-WlgEv6BgBLUwa8SXdpJD
pAnava_donor	CBS-3692	*A. variabilis* type I-B CAST left end, *A. variabilis* type I-B CAST right end	this study**	-	https://benchling.com/s/seq-PqgKq6cn4SsIjHUFZ4IK
pAnava_gRNA1	CBS-2540	targeting next to PAM library	this study	-	https://benchling.com/s/seq-OGsnEc7CQvfckmkK5cu1
pAnava_gRNA2	CBS-3311	targeting pGFP_GTAAT	this study	-	https://benchling.com/s/seq-TOOeCNYX4BqodBxDgHUV
pAnava_gRNAnt	CBS-3295	non-targeting gRNA	this study	-	https://benchling.com/s/seq-C5YHNDgyIzoo7XPA5V1Z
pAnava_TniQ	CBS-3690	encoding *A. variabilis* type I-B CAST *tniQ*	this study**	-	https://benchling.com/s/seq-Tv32jd8v4r5C4GY7oJC9
pAnava_TnsABC	CBS-3689	encoding *A. variabilis* type I-B CAST *tnsABC*	this study**	-	https://benchling.com/s/seq-7WpsFUpo9FnimhjmZMD9
pAnava_TnsD	CBS-3691	encoding *A. variabilis* type I-B CAST *tnsD*	this study**	-	https://benchling.com/s/seq-WIecPMYs2mUWBdHBupzI
pEc_Cas5	CBS-189	encoding *E. coli* type I-E *cas5*	this study*	170090	https://benchling.com/s/seq-65IrsL64NW5LLkogoKqn
pEc_Cas6	CBS-186	encoding *E. coli* type I-E *cas6*	this study*	170091	https://benchling.com/s/seq-16mAtf6O2oa0cHEcXRqf
pEc_Cas7	CBS-194	encoding *E. coli* type I-E *cas7*	this study*	170092	https://benchling.com/s/seq-IYo8oxft1A87j77g9OrG
pEc_Cas8	CBS-196	encoding *E. coli* type I-E *cas8*	this study*	170093	https://benchling.com/s/seq-9D5K9gDPRon56KQbH7uo

Table S6 (continued)

PLASMID LIST					
Name	**Lab number**	**Description**	**Source**	**Addgene number**	**Link**
pEc_Cse2	CBS-184	encoding *E. coli* type I-E *cse2*	this study*	170094	https://benchling.com/s/seq-XNZZg2gqmYlcd6sqpKzo
pEc_gRNA1	CBS-1272	targeting next to PAM library	this study	170088	https://benchling.com/s/seq-S2bC0HDstwufdMpoph8N
pEc_gRNA2	CBS-2206	targeting protospacer in *gfp*-promoter	this study	170089	https://benchling.com/s/seq-9VAV6h63G0leMsUPWYyD
pEc_gRNAnt	pCB709	non-targeting gRNA	Marshall et al. 2018 (PMID: 29304331)	-	https://benchling.com/s/seq-az94dUxjJhljqECXg2wW
pEh_Cas5	CBS-1594	encoding *E. haloalkaliphila* type I-E *cas5*	this study*	178755	https://benchling.com/s/seq-omIteC3ZL04M4BbkDy5D
pEh_Cas6	CBS-1595	encoding *E. haloalkaliphila* type I-E *cas6*	this study*	178756	https://benchling.com/s/seq-8B7xoqg0wdbmJCFyNvnG
pEh_Cas7	CBS-1593	encoding *E. haloalkaliphila* type I-E *cas7*	this study*	178757	https://benchling.com/s/seq-XikqAcNwPQ6Jn3jsejYE
pEh_Cas8	CBS-1591	encoding *E. haloalkaliphila* type I-E *cas8*	this study*	178753	https://benchling.com/s/seq-qBQqs2aoaGHQ3WRfn07C
pEh_Cse2	CBS-1592	encoding *E. haloalkaliphila* type I-E *cse2*	this study*	178754	https://benchling.com/s/seq-qu6L6GWwrnZuF4ceIILv
pEh_gRNA1	CBS-1959	targeting next to PAM library	this study	178758	https://benchling.com/s/seq-KyhX6EiKgLypXVqvNk7W
pEh_gRNA2	CBS-3025	targeting protospacer in *gfp*-promoter	this study	-	https://benchling.com/s/seq-RsZVBxcn452dBE0qPoJM
pEh_gRNAnt	CBS-1988	non-targeting gRNA	this study	-	https://benchling.com/s/seq-Q4h7t4giw4YP8AHVYhDj
pEn_Cas5	CBS-1549	encoding *E. numazuensis* type I-E *cas5*	this study*	-	https://benchling.com/s/seq-B2dTi4pCAkuTmbuK7YKa
pEn_Cas6	CBS-1550	encoding *E. numazuensis* type I-E *cas6*	this study*	-	https://benchling.com/s/seq-bwmnHz9BGew1LVGvI6Ns

Table S6 (continued)

PLASMID LIST					
Name	**Lab number**	**Description**	**Source**	**Addgene number**	**Link**
pEn_Cas8	CBS-1546	encoding *E. numazuensis* type I-E *cas8*	this study*	-	https://benchling.com/s/seq-P5RrYEcA8Pt7ccSGGEbN
pEn_Cse2	CBS-1547	encoding *E. numazuensis* type I-E *cse2*	this study*	-	https://benchling.com/s/seq-xNy79Ojzwpi0urGGdl6e
pEn_gRNA1	CBS-1852	targeting next to PAM library	this study	-	https://benchling.com/s/seq-nkZhOr92twojww7Sfd0m
pET28a_T7RNAP	CBS-344	expressing T7 RNA-Polymerase	this study	170101	https://benchling.com/s/seq-tYoRtnD5Nn5iSYvDGsHM
pGFP_AACCG	CBS-3341	AACCG PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-jl0z6o0z6lheJUVQeSAU
pGFP_AGAAA	CBS-2146	AGAAA PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-tDTRuMTbVChX2bYgkROz
pGFP_ATAAC	CBS-2816	ATAAC PAM close to *gfp*-promoter	this study	170095	https://benchling.com/s/seq-In9ePbqpnZPu0JHYrREn
pGFP_CAAAG	CBS-2188	CAAAG PAM close to *gfp*-promoter	this study	170096	https://benchling.com/s/seq-4kgZeiqPNuJtMN4VJkAP
pGFP_CAATA	CBS-3230	CAATA PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-tKAVnsx4DLeNbjm5CRLc
pGFP_CAATC	CBS-2194	CAATC PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-Q3baX4P7bQt1N8tezFLN
pGFP_CAATG	CBS-2190	CAATG PAM close to *gfp*-promoter	this study	170097	https://benchling.com/s/seq-M9ye6lFOwoUiok5QLPTq
pGFP_CAATT	CBS-3291	CAATT PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-qvYIUT3jkk75JSLElmhx
pGFP_CATTG	CBS-3228	CATTG PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-vvWQGmCnoQjv3NLlp9OS
pGFP_CTCAA	CBS-3229	CTCAA PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-TgF2hbrt14l0Sm6tmjOE
pGFP_GAAAC	CBS-2138	GAAAC PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-ISw1bavI2wtt85BeUc6S

Table S6 (continued)

PLASMID LIST					
Name	**Lab number**	**Description**	**Source**	**Addgene number**	**Link**
pGFP_GAACC	CBS-2139	GAACC PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-GpRvaHYioLYhzR7cF00V
pGFP_GCCTC	CBS-2133	GCCTC PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-NIvitQX8xZCTJJjL4ftX
pGFP_GCGGG	CBS-3232	GCGGG PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-bVWQlnCp9XWHLh7FXiAv
pGFP_GCGTG	CBS-3231	GCGTG PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-QuBccwYyEVhphkY1Xqpu
pGFP_GCTTC	CBS-2131	GCTTC PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-WYr0BbLVPddtrEXDdXfx
pGFP_GCTTT	CBS-2132	GCTTT PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-N96aJW2Icx1uR1of8rkC
pGFP_GGATC	CBS-2181	GGATC PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-k55b2z8Yfb6Id66TBznj
pGFP_GGATT	CBS-2182	GGATT PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-9MWXziqA7HaMzrn3tXNi
pGFP_GGCAG	CBS-2177	GGCAG PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-vURqX2UIKaJw8Q1juMGY
pGFP_GGCCT	CBS-2144	GGCCT PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-kLgbjg8g9OlvjfeU12mz
pGFP_GGGGG	CBS-2179	GGGGG PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-KcS8qeiNIoT7WD2MzimD
pGFP_GGTGG	CBS-2178	GGTGG PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-ZOioZH0wkoplSn3IGjLP
pGFP_GGTTG	CBS-2180	GGTTG PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-WpFKmrZjs6qsKGKJVtwd
pGFP_GTAAA	CBS-2765	GTAAA PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-OTpu5j0Itbr52FMZeMVI
pGFP_GTAAG	CBS-3336	GTAAG PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-I9FpWvLDUQtni3UxQzwg

Table S6 (continued)

PLASMID LIST					
Name	**Lab number**	**Description**	**Source**	**Addgene number**	**Link**
pGFP_GTAAT	CBS-2762	GTAAT PAM close to *gfp*-promoter	this study	170098	https://benchling.com/s/seq-yrVwQl3g4zZZcg72B9iy
pGFP_GTACA	CBS-2763	GTACA PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-YcJhXpP7Etsa1p9QE3Cy
pGFP_GTACC	CBS-3227	GTACC PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-8NVlKFmi7Tkt3elsKWCB
pGFP_GTACC_LR	-	VchCAST left end, *crm*, VchCAST right end cloned in pGFP_GTACC	this study	-	https://benchling.com/s/seq-uQoQh4U9JrbbVkCz3pKH
pGFP_GTACC_RL	-	VchCAST right end, *crm*, VchCAST left end cloned in pGFP_GTACC	this study	-	https://benchling.com/s/seq-yfWpmS8uZfvXdAmz6nGe
pGFP_GTACT	CBS-2761	GTACT PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-8eeObSUoehS4qhB7PiJe
pGFP_GTAGA	CBS-2756	GTAGA PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-wWnpiJEpfhS6Re5j3Rce
pGFP_GTAGG	CBS-3335	GTAGG PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-z1ICEHsLfRDlTUgtcqbE
pGFP_GTAGT	CBS-2755	GTAGT PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-9fjtFLYJcDzF3jDB5J3O
pGFP_GTATA	CBS-2764	GTATA PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-s5rL8znen8tQpevABZs1
pGFP_GTATT	CBS-2754	GTATT PAM close to *gfp*-promoter	this study	170099	https://benchling.com/s/seq-PRqm3HH0tRj19NmawUIi
pGFP_GTCAA	CBS-2758	GTCAA PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-KPTOPvAUpGq2HSUJuIzu
pGFP_GTCAG	CBS-3339	GTCAG PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-uidLVdAPXEIxdBYARDIW
pGFP_GTCAT	CBS-2757	GTCAT PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-Xv7oReCzngcJ99vSXhjA
pGFP_GTGAG	CBS-3338	GTGAG PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-tyUpoaHyaahWJ3DU28Ph

Table S6 (continued)

PLASMID LIST					
Name	**Lab number**	**Description**	**Source**	**Addgene number**	**Link**
pGFP_GTTAG	CBS-3337	GTTAG PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-vDQT38w9hgKhSDaBiiPW
pGFP_GTTCT	CBS-3340	GTTCT PAM close to *gfp*-promoter	this study	-	https://benchling.com/s/seq-EiufUXNJPpi31SLwwCuA
pLm_Cas5	CBS-1564	encoding *L. mobilis* type I-E *cas5*	this study*	178749	https://benchling.com/s/seq-81sgsCNlCSPcswirF7Jo
pLm_Cas6	CBS-1565	encoding *L. mobilis* type I-E *cas6*	this study*	178750	https://benchling.com/s/seq-MRwpIxWkblk1W3whef93
pLm_Cas7	CBS-1563	encoding *L. mobilis* type I-E *cas7*	this study*	178751	https://benchling.com/s/seq-FR9nUgxpGdTIanuiyUJJ
pLm_Cas8	CBS-1561	encoding *L. mobilis* type I-E *cas8*	this study*	178747	https://benchling.com/s/seq-YBHLYCGV6tTbQsPPr4Nu
pLm_Cse2	CBS-1562	encoding *L. mobilis* type I-E *cse2*	this study*	178748	https://benchling.com/s/seq-rPzYKdz8LhdAIydfZqCH
pLm_gRNA1	CBS-1957	targeting next to PAM library	this study	178752	https://benchling.com/s/seq-BxAQhcqkvZuaF5wzGPRq
pLm_gRNA2	CBS-3024	targeting protospacer in *gfp*-promoter	this study	-	https://benchling.com/s/seq-Cec28xSZLaiWYJN204UL
pLm_gRNAnt	CBS-1987	non-targeting gRNA	this study	-	https://benchling.com/s/seq-k6YfndaQt0WZ75xHx6cl
pMb_Cas5	CBS-1554	encoding *M. buryatense* type I-E *cas5*	this study*	-	https://benchling.com/s/seq-iKaDwVTaCCD5P7hvNqDr
pMb_Cas6	CBS-1555	encoding *M. buryatense* type I-E *cas6*	this study*	-	https://benchling.com/s/seq-UcTQWfRwb6AhXxphSScP
pMb_Cas7	CBS-1553	encoding *M. buryatense* type I-E *cas7*	this study*	-	https://benchling.com/s/seq-lnj7DEmpXXKmBLjUofO4
pMb_Cas8	CBS-1551	encoding *M. buryatense* type I-E *cas8*	this study*	-	https://benchling.com/s/seq-0cE68mZJZOe0cte1kzP9
pMb_Cse2	CBS-1552	encoding *M. buryatense* type I-E *cse2*	this study*	-	https://benchling.com/s/seq-UqDFpkZPQQML6zhqMSCp

Table S6 (continued)

PLASMID LIST					
Name	**Lab number**	**Description**	**Source**	**Addgene number**	**Link**
pMb_gRNA1	CBS-1955	targeting next to PAM library	this study	-	https://benchling.com/s/seq-Csp8sOaiVMv0zW2BudEE
pMs_Cas5	CBS-1584	encoding *Marinomonas* sp. type I-E *cas5*	this study*	178761	https://benchling.com/s/seq-ic0LxyAa9bBDQHeeK7iM
pMs_Cas6	CBS-1585	encoding *Marinomonas* sp. type I-E *cas6*	this study*	178762	https://benchling.com/s/seq-YOvzcmOTTXURI7F51Wff
pMs_Cas7	CBS-1583	encoding *Marinomonas* sp. type I-E *cas7*	this study*	178763	https://benchling.com/s/seq-IutRAOrFaVS9APO4aBKd
pMs_Cas8	CBS-1581	encoding *Marinomonas* sp. type I-E *cas8*	this study*	178759	https://benchling.com/s/seq-D5IUgOcweuDswRYCkEpn
pMs_Cse2	CBS-1582	encoding *Marinomonas* sp. type I-E *cse2*	this study*	178760	https://benchling.com/s/seq-lrxdesLRFsLl5ZI5qQ5h
pMs_gRNA1	CBS-1958	targeting next to PAM library	this study	178764	https://benchling.com/s/seq-n9U9BgpaBbVJHRd9R0QY
pMs_gRNA2	CBS-3015	targeting protospacer in *gfp*-promoter	this study	-	https://benchling.com/s/seq-UYXX3Vp0p2NoiXWq4Zdo
pMs_gRNAnt	CBS-1995	non-targeting gRNA	this study	-	https://benchling.com/s/seq-xkc6Vm6p5NQWiw6mLDJO
pPAM_library	CBS-1851	encoding a randomized PAM library with 5Ns	this study	-	https://benchling.com/s/seq-6z8hWsmd4dCSWJwycDfJ
pPs_Cas5	CBS-1514	encoding *Paracoccus* sp. type I-E *cas5*	this study*	-	https://benchling.com/s/seq-mJjNfojyuLcVMmA31xhT
pPs_Cas6	CBS-1515	encoding *Paracoccus* sp. type I-E *cas6*	this study*	-	https://benchling.com/s/seq-D5kx57klMRpChwYCTjwU
pPs_Cas7	CBS-1513	encoding *Paracoccus* sp. type I-E *cas7*	this study*	-	https://benchling.com/s/seq-m1roIm0L7eeElP99DEhP
pPs_Cas8	CBS-1511	encoding *Paracoccus* sp. type I-E *cas8*	this study*	-	https://benchling.com/s/seq-QPNPrM16LFm7yQfAGEgI
pPs_Cse2	CBS-1512	encoding *Paracoccus* sp. type I-E *cse2*	this study*	-	https://benchling.com/s/seq-voezCbebIsbubhYoqIuS

Table S6 (continued)

PLASMID LIST					
Name	**Lab number**	**Description**	**Source**	**Addgene number**	**Link**
pPs_gRNA1	CBS-1906	targeting next to PAM library	this study	-	https://benchling.com/s/seq-cW6E7js0YnL9rnQlOzxR
pQCas_AAA	CBS-3717	pCDFDuet-1*; VchtniQ, Vchcas8, Vchcas7, Vchcas6,* VchCAST CRISPR repeat, *ldhL* targeting spacer (protospacer with AAA PAM)	This study	-	https://benchling.com/s/seq-j6xOHmm2PKkEn1KOaLhN
pQCas_CAA	CBS-3713	pCDFDuet-1; *VchtniQ, Vchcas8, Vchcas7, Vchcas6,* VchCAST CRISPR repeat, *ldhL* targeting spacer (protospacer with CAA PAM)	This study	-	https://benchling.com/s/seq-nwzvFdfYQ36lr1puk2H1
pRo_array_GFP_dropout	CBS-2444	Golden Gate GFP dropout vector to generate RoCAST type I-B single arrays	this study	-	https://benchling.com/s/seq-xtp0V0ZqGAA6rFxityjq
pRo_gRNA1	CBS-2507	targeting next to PAM library	this study	-	https://benchling.com/s/seq-zjQqrPwTQKgnAUgrFPFC
pRo_gRNA2	CBS-2865	targeting pGFP_CAATG	this study	-	https://benchling.com/s/seq-VhywyyVf6tpS0Ul6ngCP
pRo_gRNAnt	CBS-2866	non-targeting gRNA	this study	-	https://benchling.com/s/seq-maYQQIIqPHTQdAlHJ6ZR
pRoCascade	CBS-3693	encodes *Rocas6, Rocas8, Rocas7, Rocas5*	this study**	-	https://benchling.com/s/seq-PfWyIfsMvKc57dmj5u0X
pRoCascade_gfp	CBS-3718	pCDFDuet-1 backbone, encoding for *Rocas6*, *Rocas8*, *Rocas7*, *Rocas5*, RoCAST CRISPR repeat, *gfp*	this study	178771	https://benchling.com/s/seq-66MkPMEa7WTnzjJe7sl3
pRoCascade_NT	CBS-3723	pCDFDuet-1*; Rocas6, Rocas8, Rocas7, Rocas5,* RoCAST CRISPR repeat, non-targeting spacer	This study	-	https://benchling.com/s/seq-8OyO9D6pw6Y5DGrODbcB
pRoCascade_T	CBS-3719	pCDFDuet-1*; Rocas6, Rocas8, Rocas7, Rocas5,* RoCAST CRISPR repeat, *ldhL*targeting spacer	This study	-	https://benchling.com/s/seq-RYxUzOeoIjFNizy1Oe7c
pRoDonor	CBS-3724	pUC19; RoCAST left end, *cmR*, RoCAST right end	This study	178773	https://benchling.com/s/seq-K4AlXPc9qV8kkhMXGODB
pRoDonor_extend	CBS-3712	RoCAST predicted left end region, *cmR*, RoCAST predicted right end region	this study	-	https://benchling.com/s/seq-K4AlXPc9qV8kkhMXGODB

Table S6 (continued)

PLASMID LIST					
Name	**Lab number**	**Description**	**Source**	**Addgene number**	**Link**
pRoTarget	CBS-3725	pCDFDuet-1; partial *tRNA-Leu* gene from *Rippkaea orientalis*	This study	178774	https://benchling.com/s/seq-9n9CTDGabZGf395XmgT9
pRoTnsABC	CBS-3694	encodes *RotnsAB, RotnsC*	this study**	-	https://benchling.com/s/seq-XsNaHAH6oklpOs5u09ZK
pRoTnsABCD	CBS-3709	encodes *RotnsAB, RotnsC, RotnsD*	this study	-	https://benchling.com/s/seq-G60ArN3Sdd22hHJxnEHL
pRoTnsABCDQ	CBS-3710	encodes *RotnsAB, RotnsC, RotnsD, RotniQ*	this study	178772	https://benchling.com/s/seq-feihso7G7nudoZC1ovxW
pRoTnsABCQ	CBS-3708	pET24a*; RotnsAB, RotnsC, RotniQ*	This study	-	https://benchling.com/s/seq-f69Eyb7PJvxJipg8w9Vd
pSL0283	CBS-1198	*VchtnsA, VchtnsB, VchtnsC*	Klompe et al. 2019 (PMID: 31189177)	130633	-
pSL0527	CBS-1200	VchCAST right end, *cam,* VchCAST left end	Klompe et al. 2019 (PMID: 31189177)	130634	-
pSL0828	CBS-1202	pCDFDuet-1*; VchtniQ, Vchcas8, Vchcas7, Vchcas6,* VchCAST CRISPR repeat, *ldhL* targeting spacer (protospacer with CC PAM)	Klompe et al. 2019 (PMID: 31189177)	130637	-
pSr_Cas5	CBS-1559	encoding *R. raichei* type I-E *cas5*	this study*	-	https://benchling.com/s/seq-GfqnrEovrErxI9o5PZyz
pSr_Cas6	CBS-1560	encoding *R. raichei* type I-E *cas6*	this study*	-	https://benchling.com/s/seq-jNGZv29Z5WvReUhTmMuc
pSr_Cas7	CBS-1558	encoding *R. raichei* type I-E *cas7*	this study*	-	https://benchling.com/s/seq-c7bWvGuUAKFXAOO0lNt7
pSr_Cas8	CBS-1556	encoding *R. raichei* type I-E *cas8*	this study*	-	https://benchling.com/s/seq-xYk7suZ0bd7rBsG7LUud
pSr_Cse2	CBS-1557	encoding *R. raichei* type I-E *cse2*	this study*	-	https://benchling.com/s/seq-8JOb44hvxPRNZv8tuAiX

Table S6 (continued)

PLASMID LIST					
Name	**Lab number**	**Description**	**Source**	**Addgene number**	**Link**
pSr_gRNA1	CBS-1956	targeting next to PAM library	this study	-	https://benchling.com/s/seq-u19D2y7KCrDvvYyxnIY8
pSth_Cas5	CBS-191	encoding *S. thermophilus* type I-E *cas5*	this study*	178778	https://benchling.com/s/seq-pBdjcWdwV4DxxfW0cYvA
pSth_Cas6	CBS-187	encoding *S. thermophilus* type I-E *cas6*	this study*	178779	https://benchling.com/s/seq-7XhhKpBiBXnD2xY0Zgjx
pSth_Cas7	CBS-193	encoding *S. thermophilus* type I-E *cas7*	this study*	178780	https://benchling.com/s/seq-edC25vBgtrRwC0AxRc7J
pSth_Cas8	CBS-216	encoding *S. thermophilus* type I-E *cas8*	this study*	178776	https://benchling.com/s/seq-jCasBmwBLumysQCuHUbK
pSth_Cse2	CBS-185	encoding *S. thermophilus* type I-E *cse2*	this study*	178777	https://benchling.com/s/seq-HyM3cEuV8BrI8SCK1Foi
pSth_gRNA1	CBS-843	targeting next to PAM library	this study	-	https://benchling.com/s/seq-6lGDKSH5ZpdETgTuSk1S
pSth_gRNA2	CBS-2211	targeting protospacer in *gfp*-promoter	this study	-	https://benchling.com/s/seq-7V63DvnH1oK5EL88BaZR
pSth_gRNAnt	CBS-1409	non-targeting gRNA	this study	-	https://benchling.com/s/seq-P7aZIfClCTry05kdJ3Wq
pVch_IF_Cas_gRNA2	CBS-4187	encoding for V. cholerae type I-F CAST Cascade; targeting protospacer in *gfp*-promoter	this study	-	https://benchling.com/s/seq-OuPeH3BamyV7bUt9GV9t
pVch_IF_Cas_gRNAnt	CBS-2209	encoding for V. cholerae type I-F CAST Cascade; non-targeting gRNA	this study	-	https://benchling.com/s/seq-L12XL8wuX4OYqw42C1A6
pVch_IF_CasQ_gRNA1	CBS-2301	encodes *V. cholerae* type I-F CAST Cascade and TniQ; targeting next to PAM library	this study	-	https://benchling.com/s/seq-e4jlQJX3IPc22BEAksqv
pVch_IF_CasQ_gRNA2	CBS-2803	encodes *V. cholerae* type I-F CAST Cascade and TniQ; targeting protospacer in *gfp*-promoter	this study	-	https://benchling.com/s/seq-fe7Ui12R1Ylxf0MYrQ3t
pVch_IF_CasQ_gRNA3	CBS-2164	encoding V. cholerae type I-F CAST Cascade and TniQ; targeting protospacer in *gfp*-promoter	this study	-	https://benchling.com/s/seq-H4DIXirTPkB7yfEMuBCs
pVch_IF_CasQ_gRNAnt	CBS-2165	encoding V. cholerae type I-F CAST Cascade and TniQ; non-targeting gRNA	this study	-	https://benchling.com/s/seq-9CvYzyJasnyX3QP7YIDg

Table S6 (continued)

PLASMID LIST					
Name	**Lab number**	**Description**	**Source**	**Addgene number**	**Link**
pXalb_IC_Cas3	CBS-072	encoding *X. albilineans* type I-C *cas3*	this study	178766	https://benchling.com/s/seq-D9QNVBlLNFuKdm0hzoUw
pXalb_IC_Cas5	CBS-068	encoding *X. albilineans* type I-C *cas5*	this study	-	https://benchling.com/s/seq-VIzA8gEaVpQMukAHtrRn
pXalb_IC_Cas7	CBS-090	encoding *X. albilineans* type I-C *cas7*	this study	-	https://benchling.com/s/seq-hIFioVI2Na5nzPSWMajx
pXalb_IC_Cas8	CBS-076	encoding *X. albilineans* type I-C *cas8*	this study	-	https://benchling.com/s/seq-Rv7274aiQhGeLJJD8BRj
pXalb_IC_Cascade	CBS-1275	encoding *X. albilineans* type I-C Cascade genes	this study	178765	https://benchling.com/s/seq-04WZGh3avjTcKNX2FHNB
pXalb_IC_gRNA1	CBS-200	targeting protospacer in *gfp*-promoter	this study	-	https://benchling.com/s/seq-uS1WSSoJmtGkhSOIcK0j
pXalb_IC_gRNA2	CBS-202	targeting upstream of *gfp*-promoter	this study	-	https://benchling.com/s/seq-e9jQdofrSkGC7mQnv9Pd
pXalb_IC_gRNA3	CBS-2020	targeting next to PAM library	this study	-	https://benchling.com/s/seq-BCbA4sSrqHvlBdMi8ecd
pXalb_IC_gRNA4	CBS-4193	targeting IC self-target 1 upstream of *gfp*-promoter	this study	-	https://benchling.com/s/seq-RkHRQoAkTRxmLQGpT7jO
pXalb_IC_gRNA5	CBS-4194	targeting IC self-target 2 upstream of *gfp*-promoter	this study	-	https://benchling.com/s/seq-21hGdtYjcUBNNgxT6GUu
pXalb_IC_gRNAnt	CBS-282	non-targeting gRNA	this study	-	https://benchling.com/s/seq-IAYwG0J4488SAWRS89yx
pXalb_IF1_Cas2-3	CBS-044	encoding *X. albilineans* type I-F1 *cas2-3*	this study	178769	https://benchling.com/s/seq-CcOjzRJuc8axhURQMgNC
pXalb_IF1_Cas5	CBS-047	encoding *X. albilineans* type I-F1 *cas5*	this study	-	https://benchling.com/s/seq-n9U9kIIlLB9H0O3xfD41
pXalb_IF1_Cas6	CBS-051	encoding *X. albilineans* type I-F1 *cas6*	this study	-	https://benchling.com/s/seq-JHAfYhlsmBTjb2KccTAU
pXalb_IF1_Cas7	CBS-049	encoding *X. albilineans* type I-F1 *cas7*	this study	-	https://benchling.com/s/seq-BwdrLjlK2d09tblgnbfU

Table S6 (continued)

PLASMID LIST					
Name	**Lab number**	**Description**	**Source**	**Addgene number**	**Link**
pXalb_IF1_Cas8	CBS-091	encoding *X. albilineans* type I-F1 *cas8*	this study	-	https://benchling.com/s/seq-jgf35Fn4wT5zTVWdAcrQ
pXalb_IF1_Cascade	CBS-1274	encoding *X. albilineans* Cascade type I-F1 Cascade genes	this study	178768	https://benchling.com/s/seq-lYgihBHnlIUqJ0xJQ56c
pXalb_IF1_gRNA1	CBS-198	targeting protospacer in *gfp*-promoter	this study	-	https://benchling.com/s/seq-z3VcVNWfTj22uNOMRz5q
pXalb_IF1_gRNA2	CBS-208	targeting upstream of *gfp*-promoter	this study	-	https://benchling.com/s/seq-wNMV1kTSancE5jrYZPU7
pXalb_IF1_gRNA3	CBS-2019	targeting next to PAM library	this study	-	https://benchling.com/s/seq-XCZnf5gqYdOrj53Rbovx
pXalb_IF1_gRNA4	CBS-4195	targeting IF1 self-target 1 upstream of *gfp*-promoter	this study	-	https://benchling.com/s/seq-vpzbVkjjRz9Z5OqP6PZI
pXalb_IF1_gRNA5	CBS-4196	targeting IF1 self-target 2 upstream of *gfp*-promoter	this study	-	https://benchling.com/s/seq-ydCdjLrbq5k6PvgOsnLg
pXalb_IF1_gRNA6	CBS-2130	targeting protospacer in *gfp*-promoter	this study	-	https://benchling.com/s/seq-V9aHBnS0g06tXskYAMhR
pXalb_IF1_gRNAnt	CBS-283	non-targeting gRNA	this study	-	https://benchling.com/s/seq-Q83QEKf0NOVBszmEIrFA
Sth_I-E array_GFP dropout	CBS-1279	Golden Gate GFP dropout vector to generate *S. thermophilus* type I-E single arrays	this study	178775	https://benchling.com/s/seq-CPOILrUwHFLB5UZEtyrj
Xalb_I-C_array_GG	CBS-166	Golden Gate vector to generate *X. albilineans* type I-C single arrays	this study	178767	https://benchling.com/s/seq-4C5zCUNrflHljKpgcdLy
Xalb_I-F1_array_GG	CBS-199	Golden Gate vector to generate *X. albilineans* type I-F1 single arrays	this study	178770	https://benchling.com/s/seq-61VhIdTRGXzhDiAHA5sf

Table S6 (continued)

PRIMER		
Name	**Sequence (5' -> 3')**	**Description**
FW0272	TATCACGAGGCCCTTTCGTC	amplifying pPAM_library for qPCR
FW0273	TCTGAATTGCAGCATCCGGT	amplifying pPAM_library for qPCR
FW0274	GAACTCGCACCTGAATACGC	amplifying pET28a_T7RNAP for qPCR
FW0275	CGGCTTAGGAGGAACTACGC	amplifying pET28a_T7RNAP for qPCR
FW1007	acactctttccctacacgacgctcttccgatctCGTGCGTGTTGACAATTTTAC*C	amplifying LR cr-dependent transposition of RoCAST in TXTL, capital letters are binding to pGFP_CAATG
FW1009	gtgactggagttcagacgtgtgctcttccgatctTGGTGCCCTAAGACTATTTGAC*T	amplifying LR crRNA-dependent transposition of RoCAST in TXTL, capital letters are binding to Cargo from pRoDonor
FW1017	CGACATGTGTGTGCCAATGC	amplifying RTE of RL cr-dependent transposition of RoCAST in TXTL, primer is binding to pGFP_CAATG
FW1018	GCGGTCATGCTAGAATTTTAGTAC	amplifying RTE of LR cr-dependent transposition of RoCAST in TXTL, primer is binding to Cargo from pRoDonor_extend
FW1023	GAGCCGTAGTTATCTTGACACAC	amplifying RL cr-dependent transposition of AvCAST in TXTL, primer is binding to Cargo of pAnava_donor
FW1026	AACTTCAGGGTCAGCTTGC	amplifying RTE of LR cr-dependent transposition of RoCAST in TXTL, primer is binding to pGFP_CAATG
FW207	TTAACTGACCAGCCAGAAAACG	amplifying LTE of LR cr-dependent transposition of RoCAST in TXTL, primer is binding to pGFP_CAATG
FW207	TTAACTGACCAGCCAGAAAACG	amplifying RL cr-dependent transposition of AvCAST in TXTL, primer is binding to pGFP_GTAAT
FW403	AGGGCACGGGCAGCTTGC	amplifying LR and RL transposition of VcCAST in TXTL, primer is binding to p70a_deGFP
FW531	AATTCTGGCGAATCCTTTAATTAACTGAC	amplify p70a_deGFP_PacI to create pPAM_library
FW532	NNNNNAGACGAAAGGGCCTCGTGATAC	amplify p70a_deGFP_PacI to create pPAM_library
FW624	GGCGACACGGAAATGTTGAAT	amplifying pPAM_library for Sanger Sequencing
FW625	GCTGCAACCATTATCACCGC	amplifying pPAM_library for Sanger Sequencing
FW628	acactctttccctacacgacgctcttccgatctTATCACGAGGCCCTTTCGT*C	amplifying pPAM_library for NGS, capital letters are binding to the plasmid

Table S6 (continued)

PRIMER		
Name	**Sequence (5' -> 3')**	**Description**
FW791	gtgactggagttcagacgtgtgctcttccgatctCGTTTTCTGGCTGGTCAGTT*C	amplifying pPAM_library for NGS, capital letters are binding to the plasmid
FW858	CGTCGTGGTATTCACTCCAGAGCG	amplifying LR transposition of VcCAST in TXTL, primer is binding to pSL0527
FW859	CGCTCTGGAGTGAATACCACGACG	amplifying RL transposition of VcCAST in TXTL, primer is binding to pSL0527
FW983	TGGTGCCCTAAGACTATTTGACT	amplifying LTE of LR cr-dependent transposition of RoCAST in TXTL, primer is binding to Cargo from pRoDonor_extend
IMo005	cgcacgatagagattcggg	genome specific primer that binds upstream of the transposition insertion point in *lacZ*
IMo204	tacaccaacgtgacctatcc	genome specific primer that binds downstream of the transposition insertion point in *lacZ*
IMo228	gcagttattggtgccctaagac	amplifying LTE of *in vivo* LR transposition of RoCAST, primer is binding to Cargo from pRoDonor
IMo229	GGTTTCAGAGAATCGAGTGGC	amplifying RTE of RL cr-independent transposition of RoCAST in TXTL, primer is binding to Cargo from pRoDonor
IMo230	cgccacatatcctgatcttcc	amplifying cr-dependent *in vivo* transposition of RoCAST, primer is binding to *lacZ* of *E. coli* genome
IMo234	AGTAGCGAAAGCTGCAAGAG	amplifying LTE of LR and RTE of RLcr-independent transposition of RoCAST in TXTL, primer is binding to pRoTarget
IMo234	agtagcgaaagctgcaagag	amplifying cr-independent *in vivo* transposition of RoCAST, primer is binding to *tRNA-leu* gene of pRoTarget
IMo243	acactctttccctacacgacgctcttccgatctCGCCACATATCCTGATCTTC*C	amplifying LR crRNA-dependent transposition of RoCAST in vivo, capital letters are binding to lacZ of E. coli genome
IMo244	gtgactggagttcagacgtgtgctcttccgatcTGCAGTTATTGGTGCCCTAAGA*C	amplifying *in vivo* transposition of RoCAST, capital letters are binding to Cargo from pRoDonor
IMo323	CTGGCGGTGATAATGGTTG	Target plasmid specific, same direction to protospacer
IMo324	ACAACGCCAGTGAAAAGCTC	Target plasmid specific, opposite direction to protospacer
IMo325	CTGAAGTTTAGACCATGAAGAGGC	VcCAST cargo specific for amplification via the Right end

Table S6 (continued)

PRIMER		
Name	**Sequence (5' -> 3')**	**Description**
IMo326	GGTTGTTTTGTGGTTAAGTTGCTG	VcCAST cargo specific for amplification via the Left end

BACTERIAL STRAINS			
Name	**Identifier**	**Source**	**Description**
One Shot TOP10 Chemically Competent *E. coli*	C404010	ThermoFischer Scientific	Cloning host for all the plasmids made in this study, excluding those containing the p70a promoter
Escherichia coli KL740 cI857+	#14222	*E. coli* Genetic Stock Center	Cloning host for all the plasmids made in this study that contained the p70a promoter
Xanthomonas albilineans CFBP7063	CFBP7063	CIRM-CFBP	Strain used for PCR amplification of its I-C and I-F1 CRISPR-Cas systems
BL21(DE3) Competent *E. coli*	C2527	New England Biolabs	Strain used for the *in vivo* transposition experiments

Supplementary methods S1

Detailed protocol for PAM-DETECT. (Related to STAR Methods).

Reagents

- 100% EtOH
- 3M sodium acetate, pH 5.2
- 70% EtOH
- AMPure XP (Beckman Coulter)
- CutSmart Buffer (NEB)
- KAPA HiFi HotStart ReadyMix PCR Kit (KAPA Biosystems)
- myTXTL Sigma 70 Master Mix Kit (Arbor Bioscience)
- PacI (NEB)
- Proteinase K (20 mg/µL) (Cytiva)

Procedure

PAM assay

1. Prepare 6-µL TXTL reaction on ice with the following composition:

Compound	**Final concentration/amount**
myTXTL	4.5 µL
Cascade encoding plasmid / MasterMix	3 nM (high-Cascade) or 0.25 nM (low-Cascade)
gRNA-encoding plasmid*	1 nM
pPAM_library	1 nM
pET28a_T7RNAP (if necessary)	0.2 nM
IPTG (if necessary)	0.5 nM
Nuclease-free H_2O	add to total volume of 6 µL

*If the gRNA and Cascade are encoded on the same plasmid, add the Cascade/gRNA plasmid at a final concentration of 3 nM or 0.25 nM.

2. Mix carefully and spin down briefly.
3. Incubate at 29°C for 16 h (high-Cascade) or 6 h (low-Cascade).
4. Mix carefully and spin down briefly.
5. Dilute samples 1:400 in nuclease-free H_2O and place on ice.
6. Prepare digestion reaction with the following composition:

Compound	Final concentration/amount
Diluted TXTL sample	500 µL
CutSmart Buffer	1x
PacI**	0.09 units/µL

**Also prepare a “non-digested” control for every TXTL reaction with H_2O instead of PacI.

7. Incubate digestion reaction at 37°C for 1 h.
8. Inactivate digestion reaction at 65°C for 20 min.
9. Prepare Proteinase K reaction by adding proteinase K to the digestion reaction with a final concentration of 0.05 mg/mL.
10. Incubate proteinase K reaction at 45°C for 1 h.
11. Inactivate proteinase K reaction at 95°C for 5 min.

EtOH precipitation

12. Split each sample in two equal parts to ensure EtOH precipitation can be carried out in 1.5 mL tubes.
13. Prepare EtOH precipitation as follows:

Compound	Final amount
Split sample	1 volume
3M sodium acetate, pH 5.2	0.1 volumes
100% EtOH, ice cold	2.5 volume

14. Mix vigorously by vortexing.
15. Place at -80 °C for 20 min or longer.
16. Spin samples at 4°C for 15 min at maximum speed.
17. Discard liquid carefully.
18. Add 200 µL of ice-cold 70% EtOH.
19. Spin samples at 4°C for 10 min at maximum speed.
20. Repeat steps 17-19.
21. Carefully remove liquid completely.
22. Evaporate remaining liquid by drying the pellet at 50°C.
23. Add 10 µL nuclease-free H_2O.
24. Incubate at 65°C for 10 min.
25. Vortex thoroughly.
26. Recombine divided samples.

NGS library preparation

27. Prepare a 50 µL PCR reaction to add Illumina sequencing primer sites (Primers in **Table S6**) to the EtOH-purified samples:

Compound	**Final concentration/amount**
KAPA HiFi HotStart ReadyMix (2x)	1x
Forward primer	1 µM
Reverse primer	1 µM
EtOH-purified sample	15 µL

28. Run the PCR at 62°C annealing temperature and 19 cycles.
29. Purify the PCR reaction with AMPure XP beads.
30. Prepare a 50-µL PCR reaction to add unique dual indices and the flow cell binding site to the amplicons:

Compound	**Final concentration**
KAPA HiFi HotStart ReadyMix (2x)	1x
Forward primer	1 µM
Reverse primer	1 µM
Amplicons purified with AMPure XP	1 ng

31. Run the PCR at appropriate annealing temperature and 18 cycles.
32. Purify the PCR reaction with AMPure XP beads.
33. Samples are now ready for NGS.

Chapter 4: Host-encoded anti-CRISPR proteins block DNA degradation by two extensively self-targeting CRISPR-Cas systems in *Xanthomonas albilineans*

Author contributions:

Conceptualization: F.W. and C.L.B.; Methodology: F.W. and C.L.B.; Software: S.P.C, O.A.; Validation: F.W.; K.W., F.E.; Formal analysis: F.W.; Investigation: F.W., K.W., F.E.; Writing - original draft preparation: F.W.; Writing - review and editing: F.W., C.L.B.; Visualization: F.W.; Supervision: C.L.B.; Funding acquisition: C.L.B.

Abstract

CRISPR-Cas systems store fragments of invader DNA as spacers to recognize and clear those same invaders in the future. Spacers can also be acquired from the host's genomic DNA, leading to lethal self-targeting. While self-targeting can be circumvented through a range of mechanisms, natural examples have been interrogated rarely and normally in hosts with only one self-targeting system. Here, we investigate extensive self-targeting by two CRISPR-Cas systems encoding 24 self-targeting spacers in the plant pathogen *Xanthomonas albilineans*. We show that the native I-C and I-F1 systems are actively expressed based on transcriptomics analyses and proper CRISPR RNA processing. When expressed in *Escherichia coli*, each Cascade complex binds its target to block transcription, while the addition of Cas3 paired with genome targeting induces killing. To explain the lack of lethal self-targeting in *X. albilineans*, we predicted putative anti-CRISPR proteins (Acrs) encoded within the bacterium's genome. Screening of the candidates with cell-free transcription-translation systems and in *E. coli* revealed two Acrs, which we named AcrIC11 and AcrIF12$_{Xal}$, that inhibited Cas3 but not Cascade of the respective system. These findings reveal how a bacterium tolerates extensive self-targeting through two CRISPR-Cas systems and expand the suite of Cas3-inhibiting Acrs.

Introduction

Bacteria and archaea employ a variety of methods to defend against invaders (1). Of these, the only known defenses conferring adaptive immunity are CRISPR-Cas systems. These systems are incredibly diverse, with two classes, six types and more than 30 subtypes defined to-date (2). Despite their diversity, CRISPR-Cas systems utilize three general steps for

adaptive immunity. First, CRISPR-Cas systems acquire short nucleic-acid fragments from invaders that are integrated as so-called spacers in between conserved repeats within CRISPR arrays (3, 4). Second, the CRISPR arrays are transcribed as long premature CRISPR RNAs (crRNAs) that are processed into mature crRNAs (5). Third, mature crRNAs guide the CRISPR effector proteins to a DNA or RNA region complementary to the spacer portion of the crRNA. Targets flanked by a protospacer-adjacent motif (PAM) or targets lacking complementarity with the repeat portion of the crRNA activate the nuclease (6–9). Activation then leads to either cleavage of the target that clears the invader or widespread collateral RNA cleavage that induces cellular dormancy (10–13).

The incorporation of new spacers during the acquisition step is generally biased towards foreign nucleic-acids, although accidental incorporation of genomic fragments as spacers can occur (14, 15). These self-targeting spacers would trigger a genomic attack that should be lethal and therefore selected against (12, 14, 16); nevertheless, these spacers are quite common, with about 20% of bacteria with a CRISPR-Cas system harboring one or multiple self-targeting spacers (15). To-date, several measures have been elucidated to explain how bacteria can evade autoimmunity triggered by self-targeting spacers (17). One measure is mutating *cas* genes to inhibit one or multiple steps of CRISPR-Cas targeting, although this outcome sacrifices the protective function of the CRISPR-Cas system (14, 18–20). Another measure is mutating or deleting the target region or the PAM next to it to avoid complementarity to the spacer and recognition by the CRISPR-Cas system (14, 21, 22). A third measure to avoid autoimmunity is to block expression of one of all *cas* genes to inactivate the CRISPR-Cas system. A final measure is inhibiting targeting by the CRISPR-Cas systems through anti-CRISPR proteins (Acrs), small and diverse proteins often encoded in prophages (15).

While the different measures are clear for how self-targeting can be averted, the exact mechanism in a given bacterium with a self-targeting system can be difficult to determine. Bioinformatic analyses can identify some measures, like mutations of *cas* genes or the target region, but not all, therefore experimental investigations are necessary. Nevertheless, exploration of bacteria with self-targeting spacers revealed new classes of Acrs and uncovered functions of CRISPR-Cas systems that extend beyond adaptive immunity (23–30). However, the few experimental investigations of self-targeting mostly focused on bacteria with a single CRISPR-Cas system and few self-targeting spacers (17). In this study, we investigated self-targeting by two type I CRISPR-Cas systems encoded in the plant pathogen *Xanthomonas albilineans* CFBP7063. Thereby, we discovered two endogenous Acrs that we named AcrIC11 and AcrIF12$_{Xal}$ that inhibit either system's nuclease activity but not DNA binding activity. Our results uncover how *X. albilineans* likely escapes extensive self-targeting through two orthogonal CRISPR-Cas systems and expand the small set of known Cas3-inhibiting Acrs.

Results

The two self-targeting CRISPR-Cas systems in *Xanthomonas albilineans* do not avoid self-targeting through gene repression

Xanthomonas albilineans encodes two CRISPR-Cas systems, a type I-C system and a type I-F1 system, along with six CRISPR arrays. Of these arrays, one is associated with the type I-C system and five are associated with the type I-F1 system (**Figure 1A**). In total, four of the six CRISPR arrays (one type I-C array and three type I-F1 arrays) encode 24 self-targeting spacers directed to the chromosome or one of the three plasmid of *X. albilineans*. Spacers are expected to guide their associated system to complementary targets resulting in target-degradation during the interference step of the CRISPR-Cas immunity, which is performed in two steps in type I CRISPR-Cas systems. In the first step, Cascade (CRISPR-associated complex for antiviral defense), consisting of three to five Cas proteins and the mature crRNA, binds to the target DNA (31). In the second step, the endonuclease Cas3 is recruited to the target bound by the Cascade complex (32) and nicks the non-target strand followed by degradation of the DNA in a 3′-to-5′ direction (33). Our previous work showed that both systems from *X. albilineans* efficiently carried out both steps of type I interference (34).

One measure to evade lethal self-targeting if *cas* genes are functionally encoded is preventing expression of all or some *cas* genes. Therefore, we performed RNA sequencing (RNA-Seq) analysis. We could detect transcripts for all 13 *cas* genes (**Figures 1B** and **S1**), with expression levels ranging between 7 TPM (transcripts per million) for type I-F1 *cas2-3* and 910 TPM for type I-C *cas5*. To compare these values to genes that should be decently expressed, we depicted ten genes that were found to be essential in a member of the *Xanthomonas* species and exhibit TPM levels comparable to the I-C and I-F1 *cas* genes (35) (**Figure S1**). Thus, *X. albilineans* does not appear to protect against lethal self-targeting by actively suppressing transcription of the *cas* genes.

Correct processing of pre-crRNAs to mature crRNAs would be another indication of functional expression of Cas proteins as Cascade proteins are required for processing of crRNAs and the stability of mature crRNAs (36). To examine crRNA processing, we performed RNA-Seq analysis on shorter-length RNAs (**Figure 1C**). Most spacers in the CRISPR arrays gave rise to the expected mature crRNAs for either system with few exceptions (**Figures 1C** and **S1**) (5). Array 4, the type I-C associated array, generally resulted in the expected 11-nt 5′ handle (**Figures 1C** and **S1**) and array 2, array 3 and array 6 showed the expected processing pattern of type I-F systems (8-nt 5′ handle) (**Figures 1C** and **S1**) (5). Interestingly, in array 2, array 4 and array 6 the most abundant crRNAs were self-targeting crRNAs (**Figure 1C**), excluding the possibility of preventing autoimmunity by solely expression of crRNAs targeting

foreign DNA. Therefore, we conclude that mature crRNAs as well as the necessary Cas proteins are produced.

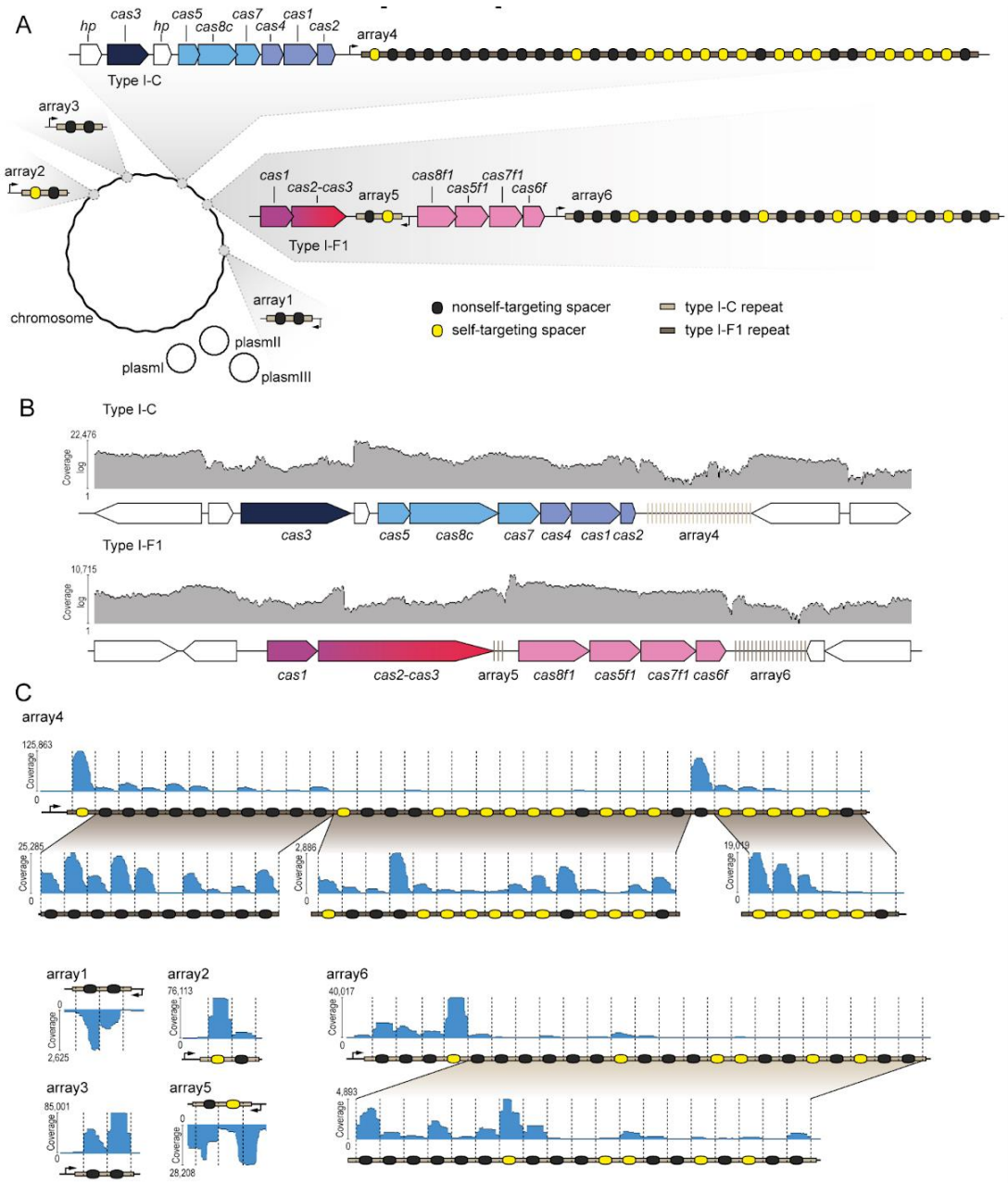


Figure 1: RNA-Seq analysis reveals transcription of *cas* genes and crRNA biogenesis for the two CRISPR-Cas systems in *Xanthomonas albilineans*. (**A**) Overview of the type I-C and type I-F1 CRISPR-Cas systems endogenous to *X. albilineans*. *cas* genes associated with the I-C system and the I-F1 system are shown in different shades of blue and pink, respectively. Spacers complementary to a region in the chromosome of *X. albilineans* or one of its plasmids are shown in yellow, and spacers without complementarity are depicted in black. (**B**) Mapped reads of the type I-C and I-F1 *cas* genes following total RNA-Seq. (**C**) Mapped reads of the mature crRNAs following small RNA-Seq of shorter-length RNAs. Expected processing patterns are indicated with dashed lines.

Both CRISPR-Cas systems bind and degrade target DNA in *E. coli*

Beyond *cas* expression and crRNA processing, we investigated interference as the last step of CRISPR-Cas immunity. While interference could not be assessed in *X. albilineans* due to

issues with plasmid transformation, our prior testing of Cascade and Cas3 with cell-free transcription-translation (TXTL) systems suggested that both CRISPR-Cas systems could enact interference in isolation (34). To assess if interference activity could lead to lethal chromosomal degradation, we assessed DNA targeting by either system in *E. coli*.

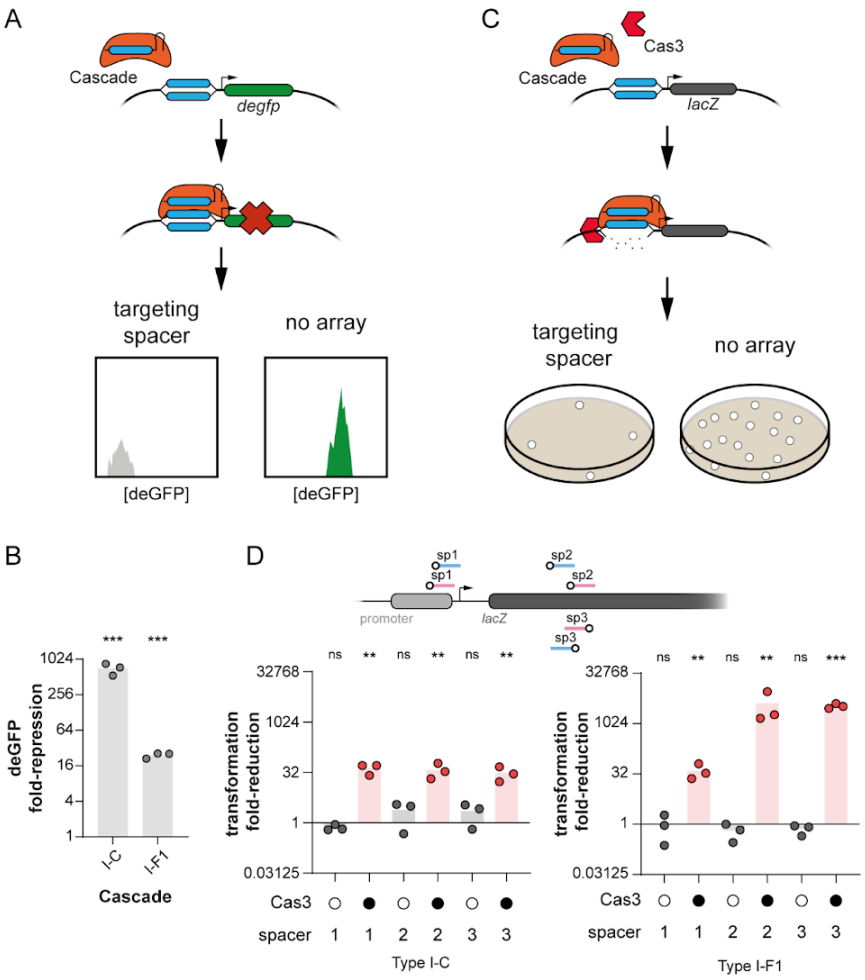


Figure 2: The type I-C and I-F1 CRISPR-Cas systems from *X. albilineans* bind and degrade target DNA in *E. coli*. (**A**) Overview of the DNA binding assay in *E. coli*. Cascade (orange) is guided by its crRNA to the target region (blue) on the deGFP-reporter plasmid complementary to the spacer (blue). Cascade binding to its target covering the promoter of *degfp* inhibits deGFP expression that can be measured by flow cytometry. The experimental setup lacking a CRISPR array (no array) serves as a negative control. (**B**) DNA binding by Cascade from both *X. albilineans* systems in *E. coli*. (**C**) Overview of the DNA degradation assay in *E. coli*. Cascade (orange) is guided by its crRNA to the target region (blue) within the promoter or the coding region of *lacZ* on the *E. coli* chromosome (target locations are shown in D) and recruits Cas3 (red). CRISPR-Cas interference causes DNA degradation which reduces the colony count on agar plates (gray). The experimental setups lacking a CRISPR array (no array) or lacking Cas3 serve as negative controls. (**D**) DNA degradation by Cascade and Cas3 from both *X. albilineans* systems in *E. coli*.
Fold-reduction in B and D is calculated based on a no-array control that is missing a spacer complementary to the *E. coli* genome or the reporter plasmid. The no-array control is the reference for statistical analyses. Bars indicate the mean of triplicate independent experiments. ***: $p < 0.001$. **: $p < 0.01$. *: $p < 0.05$. ns: $p > 0.05$.

As Cascade must bind its DNA target before recruiting the nuclease Cas3 to induce target degradation (32), we first assessed target binding by Cascade. We encoded the associated genes forming Cascade for the I-C system (*cas5*, *cas8c* and *cas7*) or the I-F1 system (*cas8f1*, *cas5f1*, *cas7f1* and *cas6f*) as an operon on a plasmid under a constitutive promoter. The same

plasmid also encoded a constitutively expressed single-spacer array we used in a previous study to target the deGFP-reporter plasmid (34). The targets in the promoter of *degfp* were flanked by a 5′ TTC (I-C system) or 5′ CC PAM (I-F1 system), which we previously identified and validated as strong PAMs *in vitro* (34). Finally, the targeted deGFP-reporter plasmid was added, and deGFP production was measured (**Figure 2A**). Cascade of both CRISPR-Cas systems repressed deGFP expression by ~700-fold (I-C system) and ~25-fold (I-F1 system) (**Figure 2B**). Therefore, either system's Cascade can bind DNA targets *in vivo*.

As target-binding was successfully performed by the I-C and the I-F1 Cascade, we proceeded to test targeted DNA degradation by Cas3. We exchanged the deGFP-reporter plasmid with a plasmid encoding the I-C *cas3* or the I-F1 *cas2-3*. We then tested three different spacers targeting the promoter or the coding region of *lacZ* with a flanking 5′ TTC (I-C system) or 5′ CC (I-F1 system) PAM (**Figures 2C** and **D**). Both CRISPR-Cas systems significantly reduced plasmid transformation compared to the no-array control, indicating chromosomal degradation and cell death (**Figures 2C** and **D**). All three spacers of the type I-C system similarly reduced plasmid transformation, whereas spacer 2 and spacer 3 exhibited a ~80-100 times higher fold change than spacer 1 in the type I-F1 system (**Figure 2D**). As expected, the absence of Cas3 did not significantly reduce plasmid transformation in both systems (**Figure 2D**). Given the lethality of chromosomal targeting with Cascade and Cas3 from either system, additional factors likely exist that protect the *X. albilineans* from lethal self-targeting.

Predicted anti-CRISPR proteins inhibit both CRISPR-Cas systems in TXTL

We hypothesized that lethal self-targeting by both CRISPR-Cas systems is inhibited by the presence of Acrs encoded within the *X. albilineans* genome (37, 38). Acrs present in bacteria are often encoded in prophage regions (37). Therefore, we searched for prophage regions in the genome of *X. albilineans* using three different phage prediction tools (39–42). This search revealed 15 prophage regions in six different chromosomal locations (**Figure 3A** and **Table S1**). Using hidden markov models to identify small proteins encoded in prophage regions next to an HTH-motif containing protein, we identified 17 Acr candidates (initially named Acr_1 through Acr_17) (**Figure 3A** and **Table S2**). The RNA-Seq analyses indicated that a subset of the predicted Acrs is expressed in *X. albilineans*, suggesting at least some of the Acr candidates might actively inhibit one or both CRISPR-Cas systems (**Table S2**).

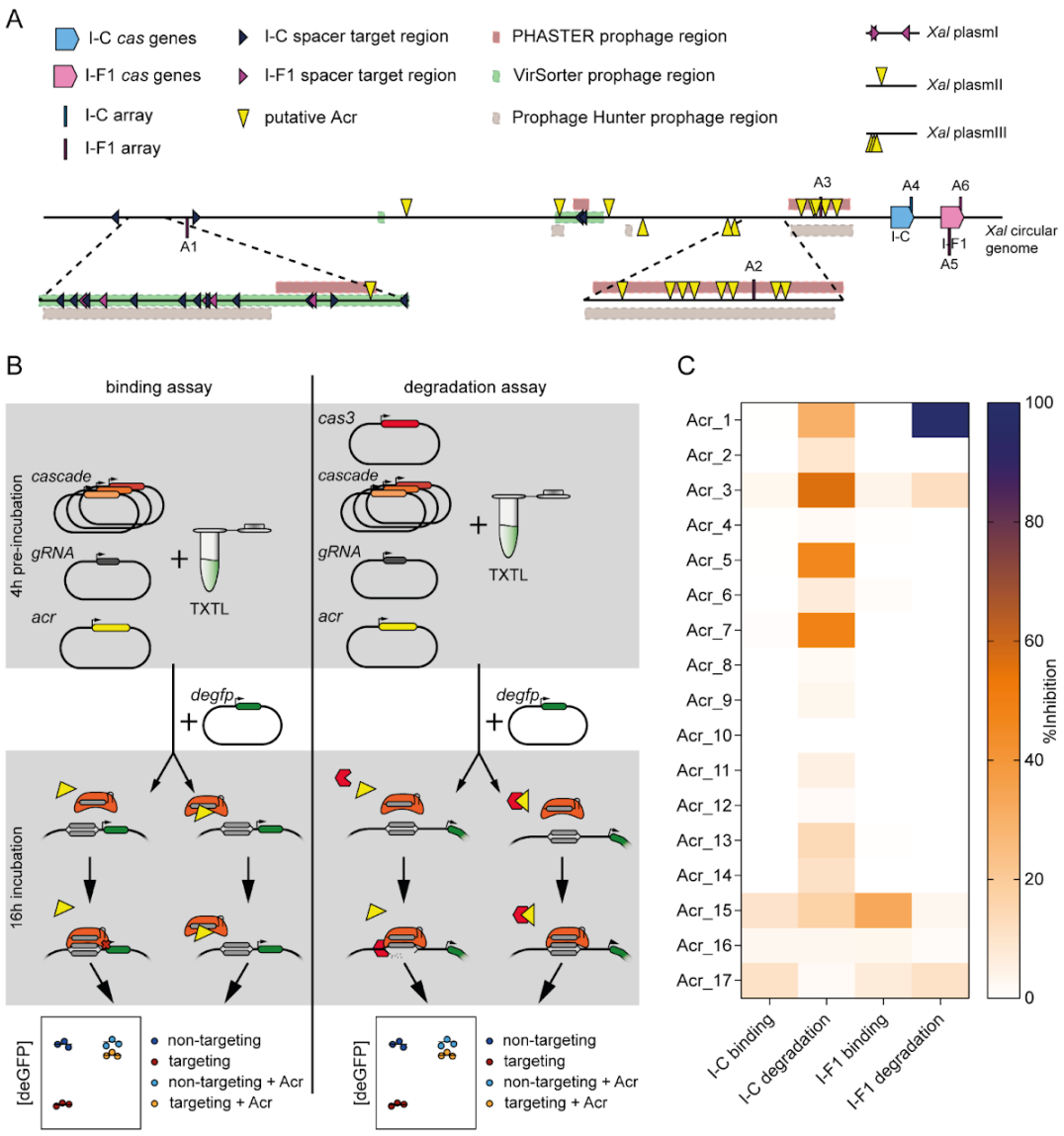


Figure 3: Putative Acrs inhibit DNA binding or DNA degradation via either *X. albilineans* CRISPR-Cas system in TXTL. (**A**) Overview of the genomic organization of CRISPR-Cas systems, putative Acrs and predicted prophages regions in *X. albilineans*. The numbering of the arrays corresponds to those in **Figure 1**. Placement of the arrays, Acr candidates and self-targets indicates whether they are encoded on the top or bottom strand of the chromosome or plasmid. Prophage regions are predicted with VirSorter v1.0.3 (39), Prophage Hunter (40) and PHASTER (41, 42). Amino-acid sequences of all Acr candidates and their genomic location in *X. albilineans* can be found in **Table S2**. (**B**) Overview of testing Acr candidates for their binding and degradation inhibition in TXTL. On the left side, inhibition of binding activity is tested. Inhibition of Cascade-mediated transcriptional repression of deGFP expression indicates a functional Acr. On the right side, inhibition of degradation activity is assessed. Inhibition of DNA degradation by Cas3 recruited by Cascade indicates a functional Acr. Inhibition of DNA degradation while allowing Cascade-mediated DNA-binding classifies an Acrs as degradation-inhibiting Acr. (**C**) Inhibitory activity of putative Acrs in TXTL. Inhibitory activity of Acr candidates was tested in triplicates and the mean inhibitory activity is depicted.

We first subjected the predicted Acrs to TXTL assays we used previously (43) to assess their inhibitory activity. TXTL assays involve adding DNA constructs, resulting in the production of the encoded RNAs and proteins whose activity can be evaluated in the same reaction. We specifically developed two assays to evaluate the extent to which the inhibitory activity of each predicted Acr acted on or upstream of DNA binding, or on or upstream of DNA degradation (**Figure 3B**). The first assay assesses inhibition of Cascade-mediated transcriptional

repression of deGFP expression. Active Acrs prevent binding of Cascade to a target in the *degfp* promoter enabling unhindered deGFP expression. The second assay assesses inhibition of DNA degradation by Cas3 recruited by Cascade. Here, a target upstream of the *degfp* promoter is chosen such that active Acrs prevent plasmid degradation. Inhibitory activity in both assays would indicate an inhibitory mechanism at or upstream of DNA binding, while inhibitory activity in only the second assay would indicate a degradation-inhibiting mechanism.

We tested all 17 putative Acrs with both assays for their activity against the type I-C and the I-F1 CRISPR-Cas systems (**Figure 3C**). Transcriptional repression of *degfp* by the I-C Cascade is not inhibited by any tested Acr candidate, at least not with an inhibitory activity higher than 11%. Type I-C degradation on the other hand was repressed by multiple Acr candidates, with Acr_1, Acr_3, Acr_5 and Acr_7 exhibiting the highest inhibitory activities ranging from 30% (Acr_1) to 57% (Acr_3). Acr_1 fully inhibited degradation by the I-F1 Cas3 but not binding by the I-F1 Cascade and thus can be categorized as a DNA degradation-inhibiting Acr. Acr_15 inhibited repression of deGFP expression in the type I-F1 binding assay by ~30%, although no inhibition was observed in the degradation assay. Overall, our TXTL approach identified Acr_1, Acr_3, Acr_5 and Acr_7 as potential type I-C Acrs, and Acr_1 and Acr_15 as potential type I-F1 Acrs.

Acr_3 and Acr_1 inhibit DNA degradation by the I-C and I-F1 Cas3, respectively, in *E. coli*

Given the fact that Acr_1, Acr_3, Acr_5, Acr_7 and Acr_15 exhibited measurable inhibitory activity in TXTL, we next tested these putative Acrs in *E. coli*. Inhibition of DNA binding by Cascade was investigated by adding each Acr candidate to the DNA binding assay and measuring the Acrs' activity in inhibiting transcriptional repression of deGFP (**Figure 4A**). Acr_3 significantly but modestly reduced deGFP fold-repression by the I-C Cascade (**Figure 4B**). All other tested Acr candidates did not significantly reduce deGFP fold-repression for the I-C or the I-F1 Cascade. The lack of binding-inhibition in *E. coli* was expected for Acr_1, Acr_3, Acr_5 and Acr_7 given our prior TXTL results.

As Cascade bound to target DNA recruits Cas3 to induce DNA degradation, we next assess the inhibitory activity of each Acr candidate in the *E. coli* DNA degradation assay (**Figure 4C**). Thereby, inhibition of Cas3-mediated chromosomal DNA degradation results in elevated colony numbers. Similar to our previous *in vitro* experiments, inhibition of a CRISPR-Cas system in the DNA degradation assay but lacking restorage of deGFP expression in the binding assay categorized the Acr as a degradation-inhibiting Acr. Acr_3 and Acr_1 significantly reduced transformation fold-reduction of the type I-C and type I-F1 system, respectively, compared to a -Acr control (**Figure 4D**). Mirroring our TXTL results, Acr_3

inhibited DNA degradation by 60%, while Acr_1 inhibited DNA degradation by 70% (**Figure 4D**). Furthermore, Acr_15 modestly but significantly reduced plasmid transformation of the I-C Cas3 (17-fold reduction of plasmid transformation compared to 71-fold reduction in the -Acr control), leaving open the question whether Acr_15 represents a bona fide Acr. All other Acr candidates did not suppress degradation of one or both CRISPR-Cas systems.

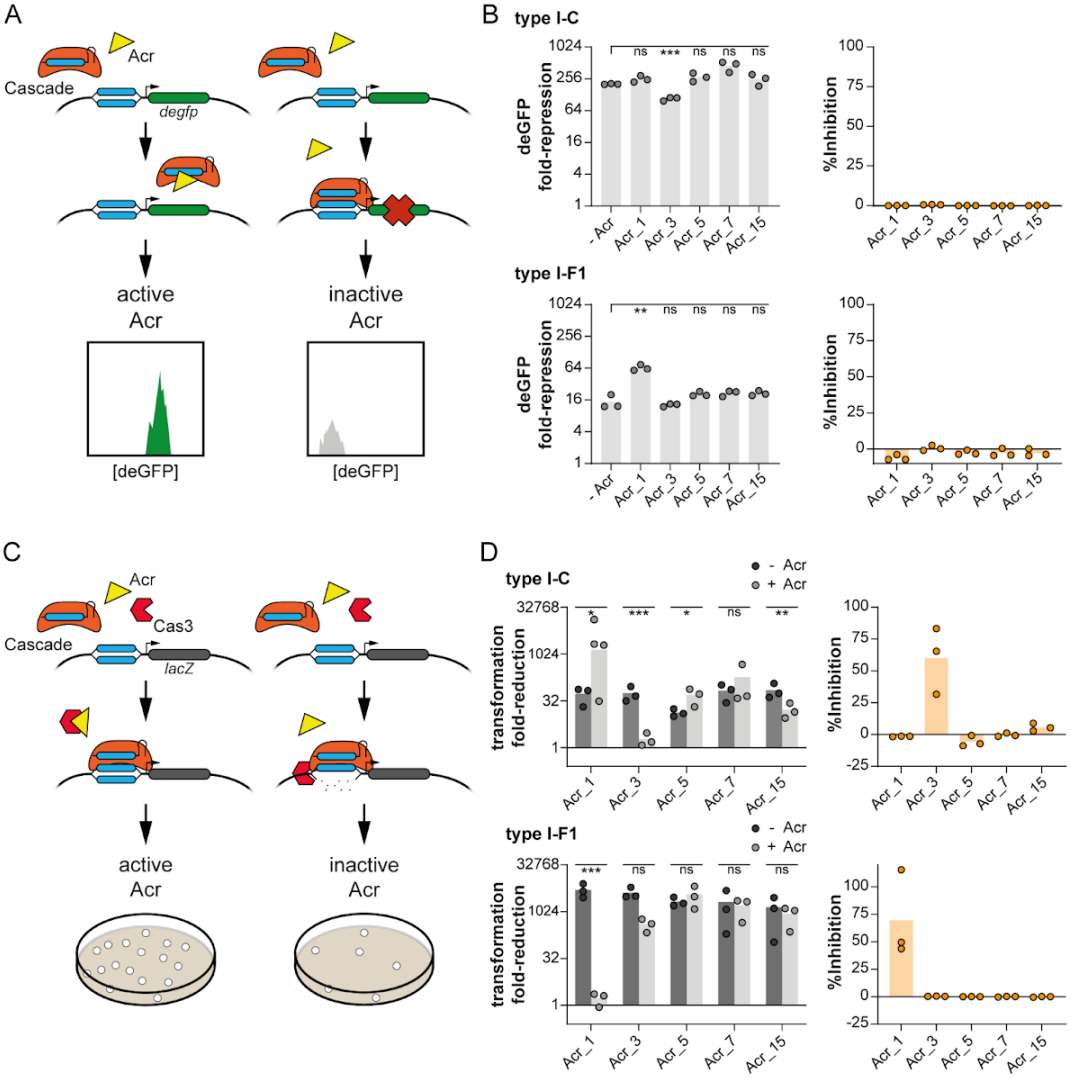


Figure 4: Acr_1 (AcrIF12$_{Xal}$) and Acr_3 (AcrIC11) inhibit DNA degradation but not DNA binding via either *X. albilineans* CRISPR-Cas system in *E. coli*. (**A**) Overview of testing putative Acrs for inhibition of transcriptional repression by Cascade in *E. coli*. Acrs actively inhibiting any step upstream of and including Cascade binding to its target restore deGFP expression. See **Figure 2A** for more details. (**B**) Inhibitory activity of putative Acrs on Cascade-binding in *E. coli*. deGFP repression was measured with flow cytometry. Bars represent the average of three biological replicas. (**C**) Overview of testing putative Acrs for inhibition of DNA degradation in *E. coli.* Acrs actively inhibiting any step upstream of and including Cas3-mediated DNA degradation restore transformation efficiency. See **Figure 2C** for more details. Type I-C spacer 2 and type I-F1 spacer 3 were used here. (**D**) Inhibitory activity of putative Acrs on Cas3-mediated DNA degradation in *E. coli*.
Fold-reduction in B and D is calculated based on a no-array control that is missing a spacer complementary to the *E. coli* genome or the reporter plasmid. The -Acr control is the reference for statistical analyses. Bars in B indicate the mean of biological triplicates and bars D indicate the mean of biological triplicates carried out with technical triplicates. Data points in B represent biological independent experiments and data points in D represent the mean of technical triplicates of a biologically independent sample. ***: $p < 0.001$. **: $p < 0.01$. *: $p < 0.05$. ns: $p > 0.05$.

After validating the inhibitory activity of Acr_3 and Acr_1 in TXTL and *E. coli*, we asked how both Acrs are related to formerly identified Acrs. Acr_3 does not share high amino-acid similarity to any previously characterized Acr and thus, we renamed Acr_3, following the common nomenclature, to AcrIC11. Acr_1 exhibits a 44.8% amino-acid identity with the previously published AcrIF12 (24) (**Figure S2A**), therefore, we renamed Acr_1 to AcrIF12$_{Xal}$. AcrIF12 was discovered next to an anti-CRISPR-associated gene 4 (*aca4*) by the "guilt-by-association" method in *Pseudomonas aeruginosa* (24). The mechanism of AcrIF12 is unknown, it is solely reported that this Acrs does not strongly bind to Cascade or Cas3 in isolation (44). To test, if AcrIF12 is also active against the *X. albilineans* type I-F1 system, we subjected AcrIF12 to our degradation-assay in TXTL (**Figures 3B** and **S2B**). AcrIF3, which was previously shown to inhibit type I-F degradation in *P. aeruginosa* (45–48), did not show activity against the type I-F1 system of *X. albilineans* in our assay (**Figure S2B**). AcrIF12 yielded an inhibitory activity of ~30%. Interestingly, the inhibitory activity of AcrIF12 was consistent with all three tested Acr plasmid concentrations and the same phenomenon was observed with AcrIF12$_{Xal}$ (**Figure S2B**). We wondered if AcrIC11 will give similar results and what the range of AcrIF12$_{Xal}$'s consistent inhibitory activity is. The inhibitory activity of AcrIC11 dropped already from 57% to 11% by using half the amount of Acr plasmid, whereas AcrIF12$_{Xal}$ only dropped from ~100% to 80% when diluting the Acr plasmid 500-fold (**Figure S2C**). Such inhibitory activities over a wide range of Acr concentrations have been associated with catalytic Acrs (44, 49, 50).

Discussion

In this study, we identified two degradation-inhibiting Acrs endogenous to *X. albilineans*, which we named AcrIC11 and AcrIF12$_{Xal}$. By blocking DNA degradation by Cas3, both Acrs are expected to prevent lethal self-targeting by the two CRISPR-Cas systems in *X. albilineans*. The possibility also remains that additional Cascade-inhibiting Acrs are encoded in the genome of *X. albilineans*. AcrIC11 and AcrIF12$_{Xal}$ add to a growing number of Acrs that inhibit Cas3 but not Cascade by two general mechanisms (24, 45–48, 51–56). AcrIF3 and AcrIE1 directly bind Cas3, while AcrIC3 is suggested to do the same (45–47, 51, 53). In contrast, AcrIE2 and AcrIF5 bind Cascade and likely block Cas3 recruitment while preserving Cascade-induced DNA-binding (55, 56). The mechanisms employed by AcrIC1, AcrIF16 and AcrIF17 to block DNA degradation remain unknown.

Elucidating the exact mechanisms by which AcrIC11 and AcrIF12$_{Xal}$ inhibit DNA degradation could reveal new mechanisms of action. In particular, the inhibitory mechanism of AcrIF12$_{Xal}$ and its homolog AcrIF12 likely differs from already known type I degradation-inhibiting mechanisms based on two observations. First, AcrIF12 did not co-elute with Cascade

nor Cas3 *in vitro* in a previous study (44), ruling out direct binding with either. Second, we showed that AcrIF12$_{Xal}$ maintained its inhibitory activity even when diluted (**Figure S2C**), suggesting that AcrIF12$_{Xal}$ and AcrIF12 could function as multi-turnover proteins. Elucidating the inhibitory mechanism of AcrIF12$_{Xal}$ and AcrIF12 therefore could reveal unique means by which Acrs inhibit Cas3-mediated DNA degradation.

Inhibition of DNA degradation by AcrIC11 and AcrIF12$_{Xal}$ still allows for DNA binding and bears the potential to transform each respective CRISPR-Cas system into a gene regulator. By silencing deGFP expression, we demonstrated that Cascade-mediated gene repression is possible even when AcrIC11 or AcrIF12$_{Xal}$ are present (**Figure 2B**). Gene regulation by self-targeting spacers can be beneficial as was shown previously in *Francisella novicida* which utilize scaRNAs (small CRISPR/Cas-associated RNAs) to facilitate immune escape during host invasion (28). Interestingly, of the six most highly expressed self-targeting crRNAs (array 2: spacer 1; array 4: spacer 1, spacers 28-30; array 6: spacer 4), five are complementary to regions within the first predicted prophage (**Figures 1C** and **3A**, see (34) for spacer targets). Array 4: spacer 29 is complementary to a genomic region not associated with a prophage, but is expected to not lead to type I-C targeting as the spacer exhibits 9 mismatches to the target region and the flanking PAM (GGG) was shown to be non-functional (34). The remaining five highly expressed spacers target a putative n6 adenine-specific DNA methyltransferase protein (XALC_0178), two hypothetical proteins (XALC_0182 and XALC_0183), a hypothetical secreted protein (XALC_0189), a hypothetical phage-related protein (XALC_0224) and a putative phage integrase protein (XALC_0242) (57). Cascade-mediated binding to the coding regions of these genes might reduce their expression and could contribute to a stable lysogenic form of the prophage (15).

The genomic location of AcrIC11 and AcrIF12$_{Xal}$ provides hints about the history of *X. albilineans*. AcrIF12$_{Xal}$ is encoded in the first predicted prophage that does also harbor many self-targets (16 in total). The location of AcrIF12$_{Xal}$ would suggest that AcrIF12$_{Xal}$ facilitated prophage integration by hindering DNA degradation by the type I-F1 system. In contrast, AcrIC11 is encoded on plasmid II that does not harbor any self-targets. We suspect that plasmid II was present in *X. albilineans* before integration of the AcrIF12$_{Xal}$-bearing prophage, as the prophage contains multiple targets of the I-C system that would be blocked by the action of AcrIC11. Self-targeting spacers could also be acquired after prophage integration, although this seems unlikely given that many self-targeting spacers are located at the 3′ end of their respective CRISPR arrays (58). Overall, elucidating the order of events could shed light on how prokaryotes come to possess self-targeting spacers and the impact on the evolutionary trajectory of each microorganism.

Material and methods

Plasmid construction

pXalb_IC_Cascade_GG was produced by Gibson Assembly (GA) using pXalb_IC_Cascade (34) as backbone and adding two type I-C repeats interspaced by *mrfp1* that can be excised with the restriction enzyme SapI. J23108 was used as a promoter driving array expression. pXalb_IC_Cascade_sp1-4 were produced with GoldenGate using pXalb_IC_Cascade_GG as backbone and SapI (NEB) as restriction enzyme. Inserts were ordered from IDT as single-stranded oligos, phosphorylated by T4 PNK (NEB) and annealed by heating to 95°C for 5 min and gradually cooling to room temperature.

pXalb_IF1_Cascade_sp1 was created by GA using pXalb_IF1_Cascade (34) as backbone and adding two type I-F1 repeats interspaced by spacer 1. J23108 was used as a promoter driving array expression. pXalb_IF1_Cascade_sp2-4 were created by Site Directed Mutagenesis (SDM) on pXalb_IF1_Cascade_sp1. pXalb_IC_Cascade_NT and pXalb_IF1_Cascade_NT were created by SDM on pXalb_IC_Cascade_GG and pXalb_IF1_Cascade_sp1, respectively. pXalb_IC_Cas3_J23105 and pXalb_IF1_Cas2-3_J23105 were created by GA using pXalb_IC_Cas3 and pXalb_IF1_Cas2-3 (34) for nuclease amplification and pCB705 (43) as backbone, and changing kanamycin resistance to ampicillin resistance. pXalb_noCas3 was produced with SDM on pXalb_IC_Cas3_J23105. p70a_deGFP_sc101 was created by changing the origin of replication (ori) of p70a_deGFP to sc101 with GA using pCB705 (43) as source for the ori.

pAcr_1-17_T7, pAcrIF12_T7 and pAcrIF3_T7 were created by GA using pET28a as backbone and double stranded DNA fragments containing *E. coli* codon optimized Acr sequences ordered from IDT as inserts. pAcr_1/3/5/7_J23105 and pAcr_15_J23115 were created by SDM on pAcr_1/3/5/7/15_T7, respectively.

All constructed plasmids were verified with Sanger sequencing and can be found in **Table S3**.

RNA-sequencing

X. albilineans CFBP7063 was grown in TSB medium to an OD_{600} of 1.0 and 2 mL were pelleted. Total RNA was extracted with Direct-zol RNA Miniprep Plus (Zymo Research) including the in-column DNase I treatment according to manufacturer's instructions. An additional DNase I treatment with TURBO DNA-*free* Kit (Thermo Fisher) was performed and the RNA was cleaned with RNA Clean & Concentrator (Zymo Research). The RNA sample was split into two parts, where one part was used for sequencing of total RNA and the second part was used to sequence shorter-length RNAs.

For total RNA-sequencing, ribosomal RNA was depleted and library was prepared using NEBNext Ultra II Directional RNA Library Preparation Kit (NEB). Next-generation sequencing was performed with 50 bp paired-end reads with 25 million reads on an Illumina NovaSeq 6000 sequencer. Sequencing quality was assessed with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and sequencing data was cleaned with Cutadapt (60). Reads were mapped to the *X. albilineans* CFBP7063 genome (FP565176.1) using RNA STAR (61) and visualized with Geneious Prime 2019.1.3 (https://www.geneious.com). Htseq-count (62) was used to determine the amount of reads per gene for calculation of TPM.

For RNA-sequencing of shorter-length RNAs, the cleaned RNA was treated with 2 U/µL T4 PNK (NEB) in 1x T4 DNA Ligase Reaction Buffer (NEB) and 1 U/µL SUPERase•In RNase Inhibitor (Thermo Fisher) for 40 min at 37°C. An additional clean up with RNA Clean & Concentrator (Zymo Research) was added. RNAs with a length of 15-100 nts were selected and the library was prepared using NEBNext Small RNA Library Preparation Kit (NEB). Next-generation sequencing was performed with 150 bp paired-end reads with 30 million reads on an Illumina NovaSeq 6000 sequencer. Sequencing quality was assessed with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and sequencing data was cleaned with Cutadapt (60). Bowtie2 (63, 64) was used to align sequencing data to *X. albilineans* CFBP7063 genome (FP565176.1) and Geneious Prime 2019.1.3 (https://www.geneious.com) was used to visualize the alignment.

Total RNA-Seq and small RNA-Seq were performed in biological duplicates.

Cascade binding assay in *E. coli*

To assess the binding ability of the type I-C CRISPR-Cas system, *E. coli* MG1655 containing p70a_deGFP_sc101 and pXalb_IC_Cascade_s4 or pXalb_IC_Cascade_NT were used. *E. coli* MG1655 with pXalb_IC_Cascade_s4 only were used as negative control. To determine binding ability of type I-F1 CRISPR-Cas system, *E. coli* MG1655 containing p70a_deGFP_sc101 and pXalb_IF1_Cascade_s4 or pXalb_IF1_Cascade_NT were used. *E. coli* MG1655 with pXalb_IF1_Cascade_s4 only were used as negative control.

Cells were grown in appropriate selection medium at 37°C for 16 h. After back diluting cells to OD_{600}=0.02 cells were grown at 37°C to OD_{600}=0.8. Cells were diluted 1:25 in 1xPBS and deGFP fluorescence was measured by flow cytometry using the Accuri C6 Plus analytical flow cytometer (BD Biosciences). Gating on living cells was applied and 30,000 events were measured. Final fluorescence values were calculated by subtracting fluorescence obtained from the negative control. Fold-reduction was calculated by the ratio of no-array over the targeting final fluorescence values. Significance was calculated between the no-array and the

targeting fluorescence values using Welch's t-test. P > 0.05 is shown as ns, P < 0.05 is shown as *, P < 0.01 is shown as ** and P < 0.001 is shown as ***.

Cas3 degradation assay in *E. coli*

To assess degradation ability of the type I-C system, electrocompetent *E. coli* MG1655 containing type I-C Cascade and a targeting array (pXalb_IC_Cascade_sp1-3) or a no-array control (pXalb_IC_Cascade_NT) were prepared and electroporated with 50 ng pXalb_IC_Cas3_J23105. 50 ng pXalb_noCas3 were electroporated as a no-nuclease control. After a one hour recovery in SOC medium at 29°C, samples were diluted 1:100 in LB medium with 34 µg/mL chloramphenicol (Cm) and incubated at 29°C for 16 h. Following, 1:5 dilutions series of the cultures were prepared in 1xPBS and 5 µL spot dilutions were plated on LB plates with 34 µg/mL Cm and 100 µg/mL ampicillin (Amp). The plates were incubated at 29°C for 24 h before calculation of colony forming units (CFU) values.

Degradation ability of the type I-F1 system was studied with electrocompetent *E. coli* MG1655 containing the type I-F1 Cascade and a targeting array (pXalb_IF1_Cascade_sp1-3) or a no-array control (pXalb_IF1_Cascade_NT) that are electroporated with 50 ng pXalb_IF1_Cas2-3_J23105. 50 ng pXalb_noCas3 was electroporated as a no-nuclease control. After a one hour recovery in SOC medium at 37°C, samples were diluted 1:100 in LB medium with 34 µg/mL Cm and incubated at 37°C for 16 h. Following, 1:5 dilutions series of the cultures were prepared in 1xPBS and 5 µL spot dilutions were plated on LB plates with 34 µg/mL Cm and 100 µg/mL Amp. The plates were incubated at 37°C for 16 h before calculation of CFU values.

Transformation fold-reduction was calculated by the ratio of non-array CFU values over targeting CFU values. Significance was calculated between the $\log_{10}$(CFU) values obtained by the no-array samples and the targeting samples using Welch's t-test. P > 0.05 is shown as ns, P < 0.05 is shown as *, P < 0.01 is shown as ** and P < 0.001 is shown as ***.

Prophage prediction

Prophage regions in the genome of *X. albilineans* CFBP7063 were predicted using VirSorter v1.0.3 (39), Prophage Hunter (40) and PHASTER (41, 42). Predicted prophage regions are listed in **Table S1**.

Acr prediction

All Acrs were predicted by Scott O. Collins and Omer S. Alkhnbashi.

Acr activity in TXTL Cascade binding assay

The Cas proteins required for Cascade formation that were used in TXTL experiments were encoded on separate plasmids. Therefore, a MasterMix with the required Cas protein encoding plasmids in their stoichiometric amount was prepared beforehand. For the type I-C system, we used a stoichiometry of $Cas5_1$-$Cas8c_1$-$Cas7_7$ and for the type I-F1 system, we used the stoichiometry $Cas8f1_1$-$Cas5f1_1$-$Cas7f1_6$-$Cas6f_1$.

To test if and to what extent predicted Acrs lead to inhibition of binding activity in TXTL, we further developed our previously used TXTL deGFP repression assays (34). Therefore, we prepared 3 µL TXTL reactions containing the following: 2.25 µL myTXTL Sigma 70 Master Mix, 0.2 nM p70a_T7RNAP, 0.5 mM IPTG, 1nM pXalb_IC/IF1_gRNA1/nt, 0.5 nM I-C or I-F1 Cascade MasterMix and 1 nM or 0.125 nM pAcr_X_T7 (1 nM: Acr_1-14 and Acr_16; 0.125 nM: Acr_15 and Acr_17). Acr_15 and Acr_17 were added in lower concentrations to avoid unspecific deGFP-inhibition that we observed at a concentration of 1 nM. Reactions without Acr-containing plasmids were used as "-Acr" controls. The TXTL reactions were incubated in a 96-well V-bottom plate at 29°C for 4 h to ensure the formation of a ribonucleoprotein complex. Furthermore, the incubation time leads to expression of the Acrs and allows for inhibition of first steps during CRISPR-Cas activity. After the incubation time, 1 nM p70a_deGFP reporter plasmid is added to the TXTL mix, the reaction is incubated at 29°C for an additional 16 h and fluorescence endpoints are measured with BioTek Synergy H1 plate reader (BioTek) at 485/528 nm excitation/emission (65). The crRNAs encoded in pXalb_IC/IF1_gRNA1 are designed to target within the *degfp* promoter region 3′ of a TTC or a CC PAM for the type I-C or the type I-F1 system, respectively, to ensure active targeting leads to inhibition of deGFP expression. All reactions were prepared with the liquid handling machine Echo525 (Beckman Coulter). Inhibition was calculated with the following equation:

$$\%\mathrm{Inhibition} = 100 * \frac{\frac{\mathrm{deGFP(t, Acr)}}{\mathrm{deGFP(nt, Acr)}} - \frac{\mathrm{deGFP(t, -Acr)}}{\mathrm{deGFP(nt, -Acr}}}{1 - \frac{\mathrm{deGFP(t, -Acr)}}{\mathrm{deGFP(nt, -Acr)}}}$$

"nt" represent values with a non-targeting spacer and "t" represent values with a targeting spacer.

Acr activity in TXTL Cas3 degradation assay

To test Acrs for their inhibitory activity on type I-C or type I-F1 degradation in TXTL, we extended our previously used degradation assay (34) similar to the above described test to check inhibition of Cascade binding. We shifted the target region from the *degfp* promoter to an upstream sequence (flanked by a 5′ TTC or 5′ CC PAM for the type I-C and the type I-F1 system, respectively). Cas3 was added to the TXTL reaction to enable degradation of the

reporter plasmid and thereby reduction deGFP production while Cascade binding without degradation would not impair deGFP expression. Inhibition of a CRISPR-Cas system by an Acr in the degradation test but not in the binding test indicates specific inhibition of DNA degradation by the Acr.

For the initial test analyzing Acr_1-17 3 µL TXTL reactions were prepared. The TXTL reactions including type I-C Cas proteins included the following: 2.25 µL myTXTL Sigma 70 Master Mix, 0.2 nM p70a_T7RNAP, 0.5 mM IPTG, 1 nM pXalb_IC_gRNA2/nt, 0.5 nM pXalb_IC_Cas3, 1 nM I-C Cascade MasterMix and 1 nM or 0.125 nM pAcr_X_T7 (1 nM: Acr_1-14 and Acr_16; 0.125 nM: Acr_15 and Acr_17). Reactions including type I-F1 Cas proteins were composed of: 2.25 µL myTXTL Sigma 70 Master Mix, 0.2 nM p70a_T7RNAP, 0.5 mM IPTG, 1 nM pXalb_IF1_gRNA2/nt, 0.25 nM pXalb_IF1_Cas2-3, 0.5 nM I-F1 Cascade MasterMix and 1 nM or 0.125 nM pAcr_X_T7 (1 nM: Acr_1-14 and Acr_16; 0.125 nM: Acr_15 and Acr_17). TXTL reactions were pre-incubated at 29°C for 4 h. The reporter plasmid p70a_deGFP was added to the reaction to a final concentration of 1 nM and incubated at 29°C for additional 16 h. Fluorescence endpoints are measured with BioTek Synergy H1 plate reader (BioTek) at 485/528 nm excitation/emission (65). All reactions were prepared with the liquid handling machine Echo525 (Beckman Coulter). Inhibition was calculated with the following equation:

$$\%\mathrm{Inhibition} = 100 * \frac{\frac{\mathrm{deGFP(t, Acr)}}{\mathrm{deGFP(nt, Acr)}} - \frac{\mathrm{deGFP(t, -Acr)}}{\mathrm{deGFP(nt, -Acr}}}{1 - \frac{\mathrm{deGFP(t, -Acr)}}{\mathrm{deGFP(nt, -Acr)}}}$$

“nt” represent values with a non-targeting spacer and “t” represent values with a targeting spacer.

Experiments to assess the inhibitory range of Acr_3 (AcrIC11) were performed as described above with final Acr plasmid concentrations (pAcr_3_T7) ranging from 1 nM to 0.25 nM. Inhibitory range of Acr_1 (AcrIF12$_{Xal}$) was investigated with 5 µL TXTL reactions. Thereby, a “homemade TXTL” (66) was used. Type I-F1 Cas proteins, crRNA and Acr_1 were pre-expressed in half the reaction volume. Fresh homemade TXTL including the reporter plasmid was added after the incubation time to prolong activity of the TXTL mix. 2.5 µL pre-expression reactions contained the following: 0.83 µL TXTL extract, 1.04 µL TXTL buffer, 0.4 nM p70a_T7RNAP, 1 mM IPTG, 2 nM pXalb_IF1_gRNA2/nt, 0.5 nM pXalb_IF1_Cas2-3, 2 nM I-F1 Cascade MasterMix and 2 nM - 2^{-8} nM pAcr_1_T7. TXTL reactions were pre-incubated at 29°C for 4 h. The following 2.5 µL reaction was added after incubation time: 0.83 µL TXTL extract, 1.04 µL TXTL buffer and 2 nM p70a_deGFP. Both reactions combined resulted in final plasmid concentrations of: 0.2 nM p70a_T7RNAP, 0.5 mM IPTG, 1 nM pXalb_IF1_gRNA2/nt, 0.25 nM pXalb_IF1_Cas2-3, 1 nM I-F1 Cascade MasterMix, 1 nM - 2^{-9} nM pAcr_1_T7 and 1

nM p70a_deGFP. The 5 µL reactions were incubated at 29°C for 14 h. All reactions were prepared by hand.

Reactions comparing the inhibitory activity of Acr_1 (AcrIF12$_{Xal}$), AcrIF12 and AcrIF3 were performed in 5 µL TXTL reactions as described above. pAcr_1_T7, pAcrIF12_T7 or pAcrIF3_T7 was added at final concentrations of 4 nM - 1 nM.

Acr activity in *E. coli* Cascade binding assay

To test the inhibition of Cascade binding by Acrs in *E. coli*, we adapted our flow cytometry assay assessing binding ability. *E. coli* MG1655 containing the reporter plasmid p70a_deGFP_sc101, pAcr_1/3/5/7_J23105, pAcr_15_J23115 or pET28a ("-Acr" control) and pXalb_IC_Cascade_s4 or pXalb_IC_Cascade_NT were used to investigate the type I-C system. *E. coli* MG1655 with pXalb_IC_Cascade_s4 only were used as negative control. To determine binding ability of the type I-F1 CRISPR-Cas system, *E. coli* MG1655 containing p70a_deGFP_sc101, pAcr_1/3/5/7_J23105 or pAcr_15_J23115 and pXalb_IF1_Cascade_s4 or pXalb_IF1_Cascade_NT were used. *E. coli* MG1655 with pXalb_IF1_Cascade_s4 only were used as negative control.

Cells were grown in appropriate selection medium at 37°C for 16 h. After back diluting cells to OD_{600}=0.02 cells were grown at 37°C to OD_{600}=0.8. After cells were diluted 1:25 in 1xPBS, deGFP fluorescence was measured by flow cytometry using the Accuri C6 Plus analytical flow cytometer (BD Biosciences). Gating on living cells was applied and 30,000 events were measured. Final fluorescence values were calculated by subtracting fluorescence obtained from the negative control. deGFP fold-repression was calculated by the ratio of no-array over the targeting final fluorescence values. Significance was calculated between the -Acr samples and the Acr-containing samples using Welch's t-test. P > 0.05 is shown as ns, P < 0.05 is shown as *, P < 0.01 is shown as ** and P < 0.001 is shown as ***. Inhibition was calculated with the following equation:

$$\%\text{Inhibition} = 100 * \frac{\frac{\text{deGFP(T, Acr)}}{\text{deGFP(NT, Acr)}} - \frac{\text{deGFP(T, }-\text{Acr)}}{\text{deGFP(NT, }-\text{Acr}}}{1 - \frac{\text{deGFP(T, }-\text{Acr)}}{\text{deGFP(NT, }-\text{Acr)}}}$$

"NT" represent no-array values and "T" represent targeting final values.

Acr activity in *E. coli* Cas3 degradation assay

To test the activity of Acrs in degradation inhibition in *E. coli*, we adapted our transformation assay assessing degradation ability. For the type I-C system, electrocompetent *E. coli* MG1655 containing type I-C Cascade, a targeting array (pXalb_IC_Cascade_sp2) or a no-array control (pXalb_IC_Cascade_NT), and pAcr_1/3/5/7_J23105, pAcr_15_J23115 or pET28a ("-Acr"

control) were prepared and electroporated with 50 ng pXalb_IC_Cas3_J23105. After a one hour recovery in SOC medium at 29°C, samples were diluted 1:100 in LB medium with 34 µg/mL Cm and 50 µg/mL kanamycin (Kan) and incubated at 29°C for 16 h. Following, 1:5 dilutions series of the cultures were prepared in 1xPBS and 5 µL spot dilutions were plated on LB plates with 34 µg/mL Cm, 50 µg/mL Kan and 100 µg/mL Amp. The plates were incubated at 29°C for 24 h before calculation of CFU values.

Degradation ability of the type I-F1 system was studied with electrocompetent *E. coli* MG1655 containing the type I-F1 Cascade, a targeting array (pXalb_IF1_Cascade_sp3) or a no-array control (pXalb_IF1_Cascade_NT), and pAcr_1/3/5/7_J23105, pAcr_15_J23115 or pET28a ("-Acr" control) that are electroporated with 50 ng pXalb_IF1_Cas2-3_J23105. After a one hour recovery in SOC medium at 37°C, samples were diluted 1:100 in LB medium with 34 µg/mL Cm and 50 µg/mL Kan and incubated at 37°C for 16 h. Following, 1:5 dilutions series of the cultures were prepared and 5 µL spot dilutions were plated on LB plates with 34 µg/mL Cm, 50 µg/mL Kan and 100 µg/mL Amp. The plates were incubated at 37°C for 16 h before calculation of CFU values.

Transformation fold-reduction was calculated by the ratio of no-array over the targeting CFU values. Significance was calculated between the values obtained by the -Acr samples and the Acr-containing samples using Welch's t-test. $P > 0.05$ is shown as ns, $P < 0.05$ is shown as *, $P < 0.01$ is shown as ** and $P < 0.001$ is shown as ***. Inhibition was calculated with the following equation:

$$\%\text{Inhibition} = 100 * \frac{\frac{\text{CFU(T, Acr)}}{\text{CFU(NT, Acr)}} - \frac{\text{CFU(T, -Acr)}}{\text{CFU(NT, -Acr}}}{1 - \frac{\text{CFU(T, -Acr)}}{\text{CFU(NT, -Acr)}}}$$

"NT" represent no-array values and "T" represent targeting final values.

Amino-acid sequence alignment

Acr_1 and AcrIF12 amino-acid sequences were aligned with Clustal-Omega 1.2.4. (67).

References

1. Bernheim,A. and Sorek,R. (2020) The pan-immune system of bacteria: antiviral defence as a community resource. *Nat. Rev. Microbiol.*, **18**, 113–119.
2. Makarova,K.S., Wolf,Y.I., Iranzo,J., Shmakov,S.A., Alkhnbashi,O.S., Brouns,S.J.J., Charpentier,E., Cheng,D., Haft,D.H., Horvath,P., *et al.* (2019) Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.*, **18**, 67–83.
3. Sternberg,S.H., Richter,H., Charpentier,E. and Qimron,U. (2016) Adaptation in CRISPR-Cas systems. *Mol. Cell*, **61**, 797–808.
4. Mojica,F.J.M., Díez-Villaseñor,C., García-Martínez,J. and Soria,E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.*, **60**, 174–182.
5. Charpentier,E., Richter,H., van der Oost,J. and White,M.F. (2015) Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiol. Rev.*, **39**, 428–441.

6. Leenay,R.T. and Beisel,C.L. (2017) Deciphering, communicating, and engineering the CRISPR PAM. *J. Mol. Biol.*, **429**, 177–191.
7. Marraffini,L.A. and Sontheimer,E.J. (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*, **463**, 568–571.
8. Meeske,A.J. and Marraffini,L.A. (2018) RNA guide complementarity prevents self-targeting in type VI CRISPR systems. *Mol. Cell*, **71**, 791–801.e3.
9. Deveau,H., Barrangou,R., Garneau,J.E., Labonté,J., Fremaux,C., Boyaval,P., Romero,D.A., Horvath,P. and Moineau,S. (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.*, **190**, 1390–1400.
10. Meeske,A.J., Nakandakari-Higa,S. and Marraffini,L.A. (2019) Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. *Nature*, **570**, 241–245.
11. Abudayyeh,O.O., Gootenberg,J.S., Konermann,S., Joung,J., Slaymaker,I.M., Cox,D.B.T., Shmakov,S., Makarova,K.S., Semenova,E., Minakhin,L., *et al.* (2016) C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science*, **353**, aaf5573.
12. Vercoe,R.B., Chang,J.T., Dy,R.L., Taylor,C., Gristwood,T., Clulow,J.S., Richter,C., Przybilski,R., Pitman,A.R. and Fineran,P.C. (2013) Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLoS Genet.*, **9**, e1003454.
13. Garneau,J.E., Dupuis,M.-È., Villion,M., Romero,D.A., Barrangou,R., Boyaval,P., Fremaux,C., Horvath,P., Magadán,A.H. and Moineau,S. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, **468**, 67–71.
14. Stern,A., Keren,L., Wurtzel,O., Amitai,G. and Sorek,R. (2010) Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet.*, **26**, 335–340.
15. Nobrega,F.L., Walinga,H., Dutilh,B.E. and Brouns,S.J.J. (2020) Prophages are associated with extensive CRISPR–Cas auto-immunity. *Nucleic Acids Research*, **48**, 12074–12084.
16. Gomaa,A.A., Klumpe,H.E., Luo,M.L., Selle,K., Barrangou,R. and Beisel,C.L. (2014) Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. *mBio*, **5**.
17. Wimmer,F. and Beisel,C.L. (2019) CRISPR-Cas systems and the paradox of self-targeting spacers. *Front. Microbiol.*, **10**, 3078.
18. Zhang,F., Zhao,S., Ren,C., Zhu,Y., Zhou,H., Lai,Y., Zhou,F., Jia,Y., Zheng,K. and Huang,Z. (2018) CRISPRminer is a knowledge base for exploring CRISPR-Cas systems in microbe and phage interactions. *Commun Biol*, **1**, 180.
19. Guan,J., Wang,W. and Sun,B. (2017) Chromosomal targeting by the type III-A CRISPR-Cas system can reshape genomes in *Staphylococcus aureus*. *mSphere*, **2**.
20. Dy,R.L., Pitman,A.R. and Fineran,P.C. (2013) Chromosomal targeting by CRISPR-Cas systems can contribute to genome plasticity in bacteria. *Mob. Genet. Elements*, **3**, e26831.
21. Selle,K., Klaenhammer,T.R. and Barrangou,R. (2015) CRISPR-based screening of genomic island excision events in bacteria. *Proc. Natl. Acad. Sci. U. S. A.*, **112**, 8076–8081.
22. Cañez,C., Selle,K., Goh,Y.J. and Barrangou,R. (2019) Outcomes and characterization of chromosomal self-targeting by native CRISPR-Cas systems in *Streptococcus thermophilus*. *FEMS Microbiol. Lett.*, **366**.
23. Watters,K.E., Fellmann,C., Bai,H.B., Ren,S.M. and Doudna,J.A. (2018) Systematic discovery of natural CRISPR-Cas12a inhibitors. *Science*, **362**, 236–239.
24. Marino,N.D., Zhang,J.Y., Borges,A.L., Sousa,A.A., Leon,L.M., Rauch,B.J., Walton,R.T., Berry,J.D., Joung,J.K., Kleinstiver,B.P., *et al.* (2018) Discovery of widespread type I and type V CRISPR-Cas inhibitors. *Science*, **362**, 240–242.
25. Li,M., Gong,L., Cheng,F., Yu,H., Zhao,D., Wang,R., Wang,T., Zhang,S., Zhou,J., Shmakov,S.A., *et al.* (2021) Toxin-antitoxin RNA pairs safeguard CRISPR-Cas systems. *Science*, **372**.
26. Li,R., Fang,L., Tan,S., Yu,M., Li,X., He,S., Wei,Y., Li,G., Jiang,J. and Wu,M. (2016) Type I CRISPR-Cas targets endogenous genes and regulates virulence to evade mammalian host immunity. *Cell Res.*, **26**, 1273–1287.
27. Heussler,G.E., Cady,K.C., Koeppen,K., Bhuju,S., Stanton,B.A. and O'Toole,G.A. (2015) Clustered regularly interspaced short palindromic repeat-dependent, biofilm-specific death of *Pseudomonas aeruginosa* mediated by increased expression of phage-related genes. *MBio*, **6**, e00129–15.
28. Ratner,H.K., Escalera-Maurer,A., Le Rhun,A., Jaggavarapu,S., Wozniak,J.E., Crispell,E.K., Charpentier,E. and Weiss,D.S. (2019) Catalytically active Cas9 mediates transcriptional interference to facilitate bacterial virulence. *Mol. Cell*, **75**, 498–510.e5.
29. Cady,K.C. and O'Toole,G.A. (2011) Non-identity-mediated CRISPR-bacteriophage interaction mediated via the Csy and Cas3 proteins. *J. Bacteriol.*, **193**, 3433–3445.
30. Zegans,M.E., Wagner,J.C., Cady,K.C., Murphy,D.M., Hammond,J.H. and O'Toole,G.A. (2009) Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviors of *Pseudomonas aeruginosa*. *J. Bacteriol.*, **191**, 210–219.

31. Jore,M.M., Lundgren,M., van Duijn,E., Bultema,J.B., Westra,E.R., Waghmare,S.P., Wiedenheft,B., Pul,U., Wurm,R., Wagner,R., *et al.* (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.*, **18**, 529–536.
32. Hochstrasser,M.L., Taylor,D.W., Bhat,P., Guegler,C.K., Sternberg,S.H., Nogales,E. and Doudna,J.A. (2014) CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 6618–6623.
33. He,L., St. John James,M., Radovcic,M., Ivancic-Bace,I. and Bolt,E.L. (2020) Cas3 protein—A review of a multi-tasking machine. *Genes*, **11**, 208.
34. Wimmer,F., Mougiakos,I., Englert,F. and Beisel,C.L. (2022) Rapid cell-free characterization of multi-subunit CRISPR effectors and transposons. *Mol. Cell*, 10.1016/j.molcel.2022.01.026.
35. Morinière,L., Lecomte,S., Gueguen,E. and Bertolla,F. (2019) In vitro exploration of the *Xanthomonas hortorum* pv. *vitians* genome using transposon insertion sequencing and comparative genomics to discriminate between core and contextual essential genes. *Microbial Genomics*, **7**.
36. Brendel,J., Stoll,B., Lange,S.J., Sharma,K., Lenz,C., Stachler,A.-E., Maier,L.-K., Richter,H., Nickel,L., Schmitz,R.A., *et al.* (2014) A complex of Cas proteins 5, 6, and 7 is required for the biogenesis and stability of clustered regularly interspaced short palindromic repeats (crispr)-derived rnas (crrnas) in *Haloferax volcanii*. *J. Biol. Chem.*, **289**, 7164–7177.
37. Davidson,A.R., Lu,W.-T., Stanley,S.Y., Wang,J., Mejdani,M., Trost,C.N., Hicks,B.T., Lee,J. and Sontheimer,E.J. (2020) Anti-CRISPRs: protein inhibitors of CRISPR-Cas systems. *Annu. Rev. Biochem.*, **89**, 309–332.
38. Gussow,A.B., Park,A.E., Borges,A.L., Shmakov,S.A., Makarova,K.S., Wolf,Y.I., Bondy-Denomy,J. and Koonin,E.V. (2020) Machine-learning approach expands the repertoire of anti-CRISPR protein families. *Nat. Commun.*, **11**, 3784.
39. Roux,S., Enault,F., Hurwitz,B.L. and Sullivan,M.B. (2015) VirSorter: mining viral signal from microbial genomic data. *PeerJ*, **3**, e985.
40. Song,W., Sun,H.-X., Zhang,C., Cheng,L., Peng,Y., Deng,Z., Wang,D., Wang,Y., Hu,M., Liu,W., *et al.* (2019) Prophage Hunter: an integrative hunting tool for active prophages. *Nucleic Acids Res.*, **47**, W74–W80.
41. Arndt,D., Grant,J.R., Marcu,A., Sajed,T., Pon,A., Liang,Y. and Wishart,D.S. (2016) PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.*, **44**, W16–21.
42. Zhou,Y., Liang,Y., Lynch,K.H., Dennis,J.J. and Wishart,D.S. (2011) PHAST: a fast phage search tool. *Nucleic Acids Res.*, **39**, W347–52.
43. Marshall,R., Maxwell,C.S., Collins,S.P., Jacobsen,T., Luo,M.L., Begemann,M.B., Gray,B.N., January,E., Singer,A., He,Y., *et al.* (2018) Rapid and scalable characterization of CRISPR technologies using an *E. coli* cell-free transcription-translation system. *Mol. Cell*, **69**, 146–157.e3.
44. Niu,Y., Yang,L., Gao,T., Dong,C., Zhang,B., Yin,P., Hopp,A.-K., Li,D., Gan,R., Wang,H., *et al.* (2020) A type I-F anti-CRISPR protein inhibits the CRISPR-Cas surveillance complex by ADP-ribosylation. *Mol. Cell*, **80**, 512–524.e5.
45. Wang,J., Ma,J., Cheng,Z., Meng,X., You,L., Wang,M., Zhang,X. and Wang,Y. (2016) A CRISPR evolutionary arms race: structural insights into viral anti-CRISPR/Cas responses. *Cell Res.*, **26**, 1165–1168.
46. Wang,X., Yao,D., Xu,J.-G., Li,A.-R., Xu,J., Fu,P., Zhou,Y. and Zhu,Y. (2016) Structural basis of Cas3 inhibition by the bacteriophage protein AcrF3. *Nat. Struct. Mol. Biol.*, **23**, 868–870.
47. Bondy-Denomy,J., Garcia,B., Strum,S., Du,M., Rollins,M.F., Hidalgo-Reyes,Y., Wiedenheft,B., Maxwell,K.L. and Davidson,A.R. (2015) Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins. *Nature*, **526**, 136–139.
48. Bondy-Denomy,J., Pawluk,A., Maxwell,K.L. and Davidson,A.R. (2013) Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*, **493**, 429–432.
49. Knott,G.J., Thornton,B.W., Lobba,M.J., Liu,J.-J., Al-Shayeb,B., Watters,K.E. and Doudna,J.A. (2019) Broad-spectrum enzymatic inhibition of CRISPR-Cas12a. *Nat. Struct. Mol. Biol.*, **26**, 315–321.
50. Dong,L., Guan,X., Li,N., Zhang,F., Zhu,Y., Ren,K., Yu,L., Zhou,F., Han,Z., Gao,N., *et al.* (2019) An anti-CRISPR protein disables type V Cas12a by acetylation. *Nat. Struct. Mol. Biol.*, **26**, 308–314.
51. Pawluk,A., Shah,M., Mejdani,M., Calmettes,C., Moraes,T.F., Davidson,A.R. and Maxwell,K.L. (2017) Disabling a type I-E CRISPR-Cas nuclease with a bacteriophage-encoded anti-CRISPR protein. *MBio*, 10.1128/mBio.01751-17.
52. Pinilla-Redondo,R., Shehreen,S., Marino,N.D., Fagerlund,R.D., Brown,C.M., Sørensen,S.J., Fineran,P.C. and Bondy-Denomy,J. (2020) Discovery of multiple anti-CRISPRs highlights anti-defense gene clustering in mobile genetic elements. *Nat. Commun.*, **11**, 5652.
53. León,L.M., Park,A.E., Borges,A.L., Zhang,J.Y. and Bondy-Denomy,J. (2021) Mobile element warfare via CRISPR and anti-CRISPR in *Pseudomonas aeruginosa*. *Nucleic Acids Res.*, **49**, 2114–2125.
54. Pawluk,A., Bondy-Denomy,J., Cheung,V.H.W., Maxwell,K.L. and Davidson,A.R. (2014) A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of *Pseudomonas aeruginosa*. *MBio*, **5**, e00896.

55. Mejdani,M., Pawluk,A., Maxwell,K.L. and Davidson,A.R. (2021) Anti-CRISPR AcrIE2 binds the type I-E CRISPR-Cas complex but does not block DNA binding. *J. Mol. Biol.*, **433**, 166759.
56. Xie,Y., Zhang,L., Gao,Z., Yin,P., Wang,H., Li,H., Chen,Z., Zhang,Y., Yang,M. and Feng,Y. (2022) AcrIF5 specifically targets DNA-bound CRISPR-Cas surveillance complex for inhibition. *Nat. Chem. Biol.*, 10.1038/s41589-022-00995-8.
57. Sayers,E.W., Bolton,E.E., Brister,J.R., Canese,K., Chan,J., Comeau,D.C., Connor,R., Funk,K., Kelly,C., Kim,S., *et al.* (2022) Database resources of the national center for biotechnology information. *Nucleic Acids Res.*, **50**, D20–D26.
58. 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.
59. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10.
60. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
61. Anders,S., Pyl,P.T. and Huber,W. (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.
62. 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 Biol.*, **10**, R25.
63. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
64. Shin,J. and Noireaux,V. (2012) An *E. coli* cell-free expression toolbox: application to synthetic gene circuits and artificial cells. *ACS Synth. Biol.*, **1**, 29–41.
65. Shin,J. and Noireaux,V. (2010) Efficient cell-free expression with the endogenous *E. coli* RNA polymerase and sigma factor 70. *J. Biol. Eng.*, **4**, 8.
66. Sievers,F., Wilm,A., Dineen,D., Gibson,T.J., Karplus,K., Li,W., Lopez,R., McWilliam,H., Remmert,M., Söding,J., *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, **7**, 539.

Supplementary material

Supplementary figures

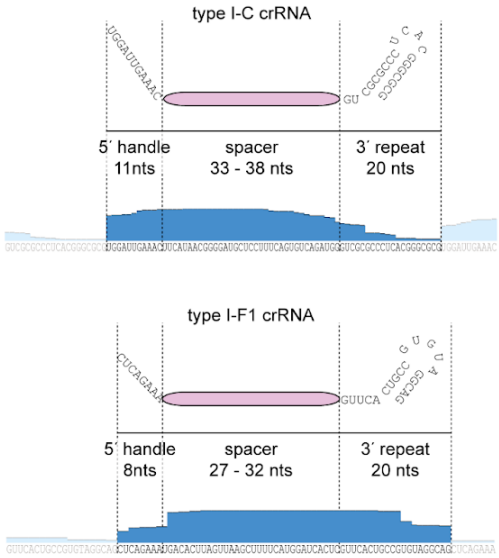
A

gene ID	Name	TPM
XALC_2885	IC *cas3*	31
XALC_2887	IC *cas5*	910
XALC_2888	IC *cas8*	94
XALC_2889	IC *cas7*	146
XALC_2890	IC *cas4*	33
XALC_2891	IC *cas1*	27
XALC_2892	IC *cas2*	89

gene ID	Name	TPM
XALC_3048	IF *cas1*	10
XALC_3049	IF *cas2-3*	7
XALC_3050	IF *cas8*	324
XALC_3051	IF *cas5*	115
XALC_3052	IF *cas7*	87
XALC_3053	IF *cas6*	118

gene ID	Name	TPM
XALC_0251	*birA*	93
XALC_0807	*mreD*	44
XALC_0808	*mrdA*	86
XALC_1005	*nuoG*	964
XALC_1665	*purB*	55
XALC_2038	*lpxB*	116
XALC_2040	*dnaE*	149
XALC_2768	*thyA*	326
XALC_2997	*hemB*	130
XALC_3172	*mpA*	77

B



C

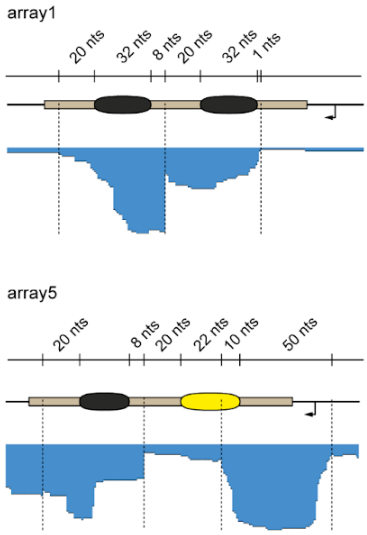


Figure S1: RNA-Seq reveals transcription of *X. albilineans cas* genes and crRNA processing. Related to **Figure 1.** (**A**) Transcripts per million calculated from total RNA-Seq results. TPM of type I-C and type I-F1 cas genes are shown. Right table: essential genes shown for comparison. (**B**) Representation of processed crRNAs. Expected mature crRNAs for the *X. albilineans* type I-C and type I-F1 arrays are shown. (**C**) Processing pattern of CRISPR array 1 and array 5. Small RNA-Seq coverage of type I-F1 array 1 and array 5 are shown in blue. Self-targeting spacers are depicted in yellow ovals and non self-targeting spacers are represented in black ovals. Dashed lines display the location of the processing event.

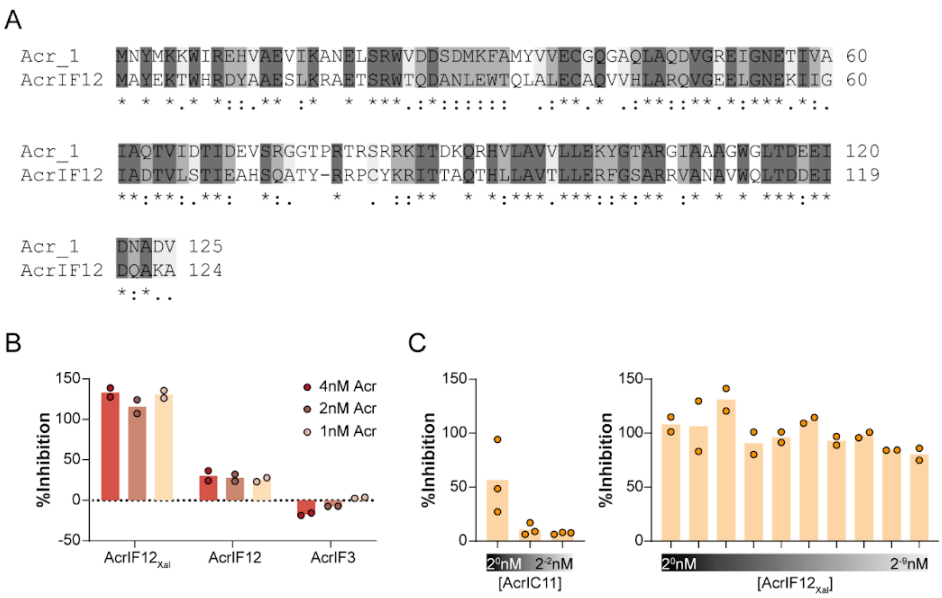


Figure S2: AcrIF12$_{Xal}$ and AcrIF12 share sequence similarity and show inhibitory activity in TXTL throughout a wider range than AcrIC11. Related to **Figures 3** and **4.** (**A**) Sequence alignment of Acr_1 and AcrIF12. Amino acid sequences of Acr_1 and AcrIF12 were aligned with Clustal-Omega 1.2.4. (66). (**B**) Inhibitory activity of AcrIF12$_{Xal}$, AcrIF12 and AcrIF3 on Cas3-mediated DNA degradation in TXTL. Degradation assays were conducted with Acr plasmid concentrations ranging from 4 nM to 1 nM. For more details see **Figure 3B**. (**C**) Inhibitory activity of AcrIC11 and AcrIF12$_{Xal}$ on Cas3-mediated DNA degradation in TXTL with different Acr concentrations. Degradation assays were conducted with Acr plasmid concentrations ranging from 20 nM to 2^{-2} nM or 20 nM to 2^{-9} nM for AcrIC11 or AcrIF12$_{Xal}$, respectively.
Bars in B and C indicate the mean inhibitory activity of triplicate or duplicate independent experiments.

Supplementary tables

Table S1: Genomic location of predicted prophage regions. Prophage regions are predicted with VirSorter v1.0.3 (39), Prophage Hunter (40) and PHASTER (41, 42). Related to **Figure 3.**

Used tool	Genomic region	Comment
VirSorter	204,842-292,722	prophage category 5
Prophage Hunter	217,510-251,216	active
PHASTER	251,524-271,683	incomplete
VirSorter	1,197,275-1,218,686	prophage category 6
Prophage Hunter	1,691,999-1,713,262	ambiguous
VirSorter	1,696,720-1,861,289	prophage category 5
PHASTER	1,790,323-1,798,253	intact
Prophage Hunter	2,174,815-2,185,313	ambiguous
Prophage Hunter	2,830,643-2,860,815	ambiguous
PHASTER	2,837,733-2,874,463	questionable
PHASTER	2,841,444-2,856,796	incomplete
PHASTER	3,098,545-3,132,674	questionable
Prophage Hunter	3,102,291-3,128,520	ambiguous
PHASTER	3,107,687-3,129,023	incomplete
Prophage Hunter	3,120,278-3,132,701	ambiguous

Table S2: Genomic location of putative Acrs and their amino acid sequence. Related to **Figure 3** and **Figure 4.**

Putative Acr	Genomic location	TPM	AA sequence	Comment
Acr_1	265340-265717	316	MNYMKKWIREHVAEVIKANELSRWVDDSDMKFAMYVVECGQGAQLAQDVGREIGNETIVAIAQTVIDTIDEVSRGGTPRTRSRRKITDKQRHVLAVVLLEKYGTARGIAAAGWGLTDEEIDNADV*	
Acr_2	2852041-2852406; 3116105-3116470	0; 0	MPRKAPTPCRHPGCGKLVSDGSGYCADHQRDKVGWHKDRRNAHQRGYGATWQKLRAFVMQRDQGLCQPCKQSGRLTPAVAVDHIVPKSQGGTDHPNNCQAICHRCHVLKTAQESHQGREGA*	
Acr_3	plasmII: 16827-17273	55	MNKETQITASAVVGEDKRLEFLSKHFGVRFARRGEALVFAWLLRLAKVPIEWTRLQYYTLSNSGFYLAPRELRISECELSADAVGIVATMLTLRQLAHESAACVEADSTYPAAKLAVTASVKFAQQYHHLAAYSVKHAESINIYRAID*	
Acr_4	plasmIII: 7731-7997	0	MSTLTVTTRGQVTFRAEVLQHLGIKPGEKIEVYLMPDGRAELKAAKPKGSFRELRGILKHKTNGARLSIEEINDAIAEAGDAAGTGNT*	
Acr_5	2852137-2852406; 3116201-3116470	0; 0	MVGWHKDRRNAHQRGYGATWQKLRAFVMQRDQGLCQPCKQSGRLTPAVAVDHIVPKSQGGTDHPNNCQAICHRCHVLKTAQESHQGREGA*	N-terminal M added; part of Acr2
Acr_6	2852659-2852865	4	MQSQKTARPLNFSRVQNEKLFLDRTVSELSVARDYKADLAQIEQIDATPWTAASHADMTSELKTYARS*	
Acr_7	2852734-2852865	0	MVSELSVARDYKADLAQIEQIDATPWTAASHADMTSELKTYARS*	N-terminal M added; part of Acr6
Acr_8	plasmIII: 8405-8809	45	MNKEAQVTVSVVVAEDERLEFLSNHFGLRFARRGEALVFAWLLRLSKVPIEWTRLQYYTLSNSGFYLAPRELRLSECELSADAVGIVATMLTLRQLAHEAASTPAAAKFAQQYHALAAYSVTHAESINIYRAID*	
Acr_9	plasmIII: 8893-9117	0	MRVFNIAEIEFAINYWRTRIVPDDGALMCAPALSLLQLYGHMIFDRIEAVPESELDAEQGVALSVALYQHELPL*	
Acr_10	3116786-3116941	188	LEKSVSELSAARDYKADLAQIGQIDVTPWTAAAHADMTPAEPVELEPYARS*	

Table S2 (continued)

Putative Acr	Genomic location	TPM	AA sequence	Comment
Acr_11	2852137-2852406; 3116201-3116470	0; 0	VGWHKDRRNAHQRGYGATWQKLRAFVMQRDQGLCQPCKQSGRLTPAVAVDHIVPKSQGGTDHPNNCQAICHRCHVLKTAQESHQGREGA*	Acr5 without N-terminal M; part of Acr2
Acr_12	2845745-2845960	0	MVDAKHAAAALRLPYYWFSDQAMRNKYRIPHYLLGGLVRYRLSELSAWAARSTLVQRSDTSNVGTSTEEAE*	
Acr_13	2845431-2845706	1	MNLITSLRHKLSYLYGEHLPNEIHYHRADGQHVVVALQDATVDQLAFAIQTINTESVALSRHRNALEELHTEVRKRSACGADRIADVAWDN*	
Acr_14	2844724-2844861	2	MADGSAPLPSLTTLPPRDHAMRSLDEFVRVDDGRNHKPAHKSRHT*	
Acr_15	1430699-1431022; 1698203-1698526; 1917087-1917410; 2197118-2197441; 2828433-2828756; 3118350-3118673	0; 2; 0; 6; 2; 3	MSKSNKFSPEVRERAVRMVQEQRGEYQSLWAAIESIAPKIGCVPQTLNEWVKRAEVDAGAREGVTSSEAQRMKELEREVKELRRANEILKLASAFFAQAELDRRLKS*	
Acr_16	2845054-2845317	1	MTTRLPATQIGQLCESKDPGSTTRIALDESELAARWGLSVKTLRRWRQEQLGPVFCKLGARVTYLICDVEAFEQRVSRYSTFARAYP*	
Acr_17	2773986-2774279; 3111383-3111676	0; 0	MQRITRRRYTDDFKAQAIALAESVGLAKAARQLGMSVKTLANWLGASRGGQPLSSPSRKPVSEMESELARLRAENATLKMEREILKKATAFFARESK*	

Table S3: Plasmid list.

Name	Lab number	Description	Source	Addgene number	benchling link
p70a_deGFP	CBS-338	deGFP-reporter plasmid with p70a promoter	commercially available at arbor bioscience	-	https://benchling.com/s/seq-U3rbd9AzzncflxwpHcQn?m=slm-hQMrBEElEUo4N8eMyOcW
p70a_deGFP_sc101	CBS-4056	deGFP-reporter plasmid with p70a promoter and sc101 ori	this study	-	https://benchling.com/s/seq-7oSWqvFQVCgOtpvnaD3J?m=slm-92qtCvXA7eVhtHgZiGBA
p70a_T7RNAP	CBS-011	expressing T7 RNA-Polymerase	Garamella et al. 2016 (PMID: 26818434)	-	https://benchling.com/s/seq-Mm5rTePyuv6PwZftgKtY?m=slm-dDGstExMbHF6TipFITr1
pAcr_1_J23105	CBS-2709	encoding *X. albilineans* putative *acr_1* with J23105 promoter	this study	-	https://benchling.com/s/seq-I5dMKExbOl60LbowaE5G?m=slm-OaFLCPMXYXB6jlvRbrOH
pAcr_1_T7	CBS-130	encoding *X. albilineans* putative *acr_1* with T7 promoter	this study	-	https://benchling.com/s/seq-7xiycvJ87lJvbQzi7RFI?m=slm-qQVqg48dZka1dWT9b4JM
pAcr_10_T7	CBS-080	encoding *X. albilineans* putative *acr_10* with T7 promoter	this study	-	https://benchling.com/s/seq-Bw59WnGliECP8WfaWp98?m=slm-ml6E93rHJE4loQTYB0z7
pAcr_11_T7	CBS-074	encoding *X. albilineans* putative *acr_11* with T7 promoter	this study	-	https://benchling.com/s/seq-DisT4DLOcjN95KwRcSob?m=slm-wZOCeTRToWwq1W1rMU8T
pAcr_12_T7	CBS-073	encoding *X. albilineans* putative *acr_12* with T7 promoter	this study	-	https://benchling.com/s/seq-7rDDpfsesuTH5ppOrslx?m=slm-glA44vwpvqevllQljeGK
pAcr_13_T7	CBS-066	encoding *X. albilineans* putative *acr_13* with T7 promoter	this study	-	https://benchling.com/s/seq-Jf9I2W34LMwtAyc5a4xW?m=slm-6GHNVtm4HGpId9TBlzSa
pAcr_14_T7	CBS-065	encoding *X. albilineans* putative *acr_14* with T7 promoter	this study	-	https://benchling.com/s/seq-NYvzniE9uJqNwbsEOq9h?m=slm-xgU1gAylhN0eDrKhngC9

Table S3 (continued)

Name	Lab number	Description	Source	Addgene number	benchling link
pAcr_15_J23115	CBS-4317	encoding *X. albilineans* putative *acr_15* with J23115 promoter	this study	-	https://benchling.com/s/seq-seMC6XPJvhbrifekBHZQ?m=slm-85RSJOyhQNtCefMjRlko
pAcr_15_T7	CBS-595	encoding *X. albilineans* putative *acr_15* with T7 promoter	this study	-	https://benchling.com/s/seq-vkO4l0EMmYEyYJ8t7ysG?m=slm-H7d5sdNCodFtyypVZCKR
pAcr_16_T7	CBS-596	encoding *X. albilineans* putative *acr_16* with T7 promoter	this study	-	https://benchling.com/s/seq-h0b16N2kcnEf39UDv7PG?m=slm-piaeXGiVCBEx2yg6wZ9L
pAcr_17_T7	CBS-597	encoding *X. albilineans* putative *acr_17* with T7 promoter	this study	-	https://benchling.com/s/seq-O7fjwVIGui0P3PWhwpW5?m=slm-BE9VpPt1CTDBFJ5wPUnw
pAcr_2_T7	CBS-131	encoding *X. albilineans* putative *acr_2* with T7 promoter	this study	-	https://benchling.com/s/seq-665wewgEX9ygVriMI9Ff?m=slm-4TE6HyxJv8d6VqKQe37m
pAcr_3_J23105	CBS-4236	encoding *X. albilineans* putative *acr_3* with J23105 promoter	this study	-	https://benchling.com/s/seq-TkIFCu5j7HnFTmvRvhGk?m=slm-6nfo3V5XNammk8yliRHC
pAcr_3_T7	CBS-132	encoding *X. albilineans* putative *acr_3* with T7 promoter	this study	-	https://benchling.com/s/seq-v3TIRhaHlGQ8bTTx0NKR?m=slm-P8wwNIcuMMh6dO950zth
pAcr_4_T7	CBS-133	encoding *X. albilineans* putative *acr_4* with T7 promoter	this study	-	https://benchling.com/s/seq-jKTie638BytRoVMIHMJe?m=slm-x2G9jUvDTpIF1csqDd2g
pAcr_5_J23105	CBS-4238	encoding *X. albilineans* putative *acr_5* with J23105 promoter	this study	-	https://benchling.com/s/seq-ZmkqOTcddyBiAFsiN9Y0?m=slm-2szntwy6TXqhT8WuCm2Y
pAcr_5_T7	CBS-134	encoding *X. albilineans* putative *acr_5* with T7 promoter	this study	-	https://benchling.com/s/seq-gJSdrc3khIEO80QNIBaq?m=slm-M7QI4d4UzQkoGMWwJgHJ

Table S3 (continued)

Name	Lab number	Description	Source	Addgene number	benchling link
pAcr_6_T7	CBS-088	encoding *X. albilineans* putative *acr_6* with T7 promoter	this study	-	https://benchling.com/s/seq-7xmzx0rkJCM7sznyEnzK?m=slm-Aflvwwkpzl1tLVYBtwwE
pAcr_7_J23105	CBS-4240	encoding *X. albilineans* putative *acr_7* with J23105 promoter	this study	-	https://benchling.com/s/seq-pb1C7gHWPXnCSNDPFDDE?m=slm-y8tAcgbGqH3wXXki1LJj
pAcr_7_T7	CBS-087	encoding *X. albilineans* putative *acr_7* with T7 promoter	this study	-	https://benchling.com/s/seq-0DFIqDWIzuLeGP8Ode7p?m=slm-pY5PJpIjo9dTJ8lYBeZK
pAcr_8_T7	CBS-083	encoding *X. albilineans* putative *acr_8* with T7 promoter	this study	-	https://benchling.com/s/seq-j3cgisQSAk2o2Cbrtrx8?m=slm-MxKnLvTWVlw7npd5pjkl
pAcr_9_T7	CBS-081	encoding *X. albilineans* putative *acr_9* with T7 promoter	this study	-	https://benchling.com/s/seq-v86H3m3GtizMP2YTPt4c?m=slm-21nxKgU5d5oA2E6NTAxp
pAcrIF12_T7	CBS-2207	encoding *acrIF12* with T7 promoter	this study	-	https://benchling.com/s/seq-zmJ0ouukLb32DSs2uzG5?m=slm-vTPa9EZ2JVqroQp4FRiP
pAcrIF3_T7	CBS-2307	encoding *acrIF13* with T7 promoter	this study	-	https://benchling.com/s/seq-yjNWLtN0dsfTXNnk1lOE?m=slm-a5gm2OBnyFWhnPFWgvjC
pXalb_IC_Cas3	CBS-072	encoding *X. albilineans* type I-C *cas3*	Wimmer et al. 2022 (PMID: 35216669)	178766	https://benchling.com/s/seq-K1gKu7mphxNtFUUozMjl?m=slm-7cDREmJiR8iLiGaUhgKt
pXalb_IC_Cas3_J23105	CBS-4152	encoding *X. albilineans* type I-C *cas3* with J23105 promoter	this study	-	https://benchling.com/s/seq-Gfx1xTzClaIXNp9kmdLf?m=slm-5J3bFWUHHx6rq4AxYDoD
pXalb_IC_Cas5	CBS-068	encoding *X. albilineans* type I-C *cas5*	Wimmer et al. 2022 (PMID: 35216669)	-	https://benchling.com/s/seq-fx2pIzcu4nVVMA1mK2XC?m=slm-ruiHN4wCw5Y2eji2LAEI

Table S3 (continued)

Name	Lab number	Description	Source	Addgene number	benchling link
pXalb_IC_Cas7	CBS-090	encoding *X. albilineans* type I-C *cas7*	Wimmer et al. 2022 (PMID: 35216669)	-	https://benchling.com/s/seq-z409zZTScM4o6n2cjBjq?m=slm-Rxp1Fhfnhj3YEzByvdsl
pXalb_IC_Cas8	CBS-076	encoding *X. albilineans* type I-C *cas8*	Wimmer et al. 2022 (PMID: 35216669)	-	https://benchling.com/s/seq-MS2AJgyRVra9gprLpnD1?m=slm-Cvljzi4BSmNbSDTHnjOc
pXalb_IC_Cascade_GG	CBS-4133	Golden Gate vector for *X. albilineans* type I-C Cascade genes and single spacer array	this study	-	https://benchling.com/s/seq-Kv1NcEDovbONHJ9UTNBe?m=slm-pPU5gwiu6xm0aZA3m4uK
pXalb_IC_Cascade_NT	CBS-4138	encoding *X. albilineans* type I-C Cascade genes without single spacer array	this study	-	https://benchling.com/s/seq-ggvjI7kp7ADmESxk0hnj?m=slm-t6xR2X4eU9ZP3XQKS6Qp
pXalb_IC_Cascade_sp1	CBS-4145	encoding *X. albilineans* type I-C Cascade genes and single spacer array targeting *lacZ* promoter	this study	-	https://benchling.com/s/seq-I4uMvNPQHIOM0lFCY7hM?m=slm-V4rFE7FVhfsKP8wgU0Hl
pXalb_IC_Cascade_sp2	CBS-4149	encoding *X. albilineans* type I-C Cascade genes and single spacer array targeting *lacZ*	this study	-	https://benchling.com/s/seq-bqmlm7fdBxf2iB6CtnbN?m=slm-egwP6qI8YuqGvMlbsTSC
pXalb_IC_Cascade_sp3	CBS-4150	encoding *X. albilineans* type I-C Cascade genes and single spacer array targeting *lacZ*	this study	-	https://benchling.com/s/seq-4A8llev0PvU0eAXRlhfc?m=slm-Bcpt0VQk6fBMFH9aBHCJ
pXalb_IC_Cascade_sp4	CBS-4151	encoding *X. albilineans* type I-C Cascade genes and single spacer array targeting *degfp* promoter	this study	-	https://benchling.com/s/seq-mLbPo1aU7bnkNB1OpoxT?m=slm-TZqzMDmfvfwnI7C9SkuE
pXalb_IC_gRNA1	CBS-200	encoding *X. albilineans* type I-C single spacer array targeting *degfp* promoter	Wimmer et al. 2022 (PMID: 35216669)	-	https://benchling.com/s/seq-Pv7DeNUzx6CEBaNecCIp?m=slm-ja85ll62aGzj7xLi2Ujr
pXalb_IC_gRNA2	CBS-202	encoding *X. albilineans* type I-C single spacer array targeting upstream of *degfp* promoter	Wimmer et al. 2022 (PMID: 35216669)	-	https://benchling.com/s/seq-3r48tGZHk0V7i7Flcnup?m=slm-olhMu91qx7Qi8533VscL

Table S3 (continued)

Name	Lab number	Description	Source	Addgene number	benchling link
pXalb_IC_gRNAnt	CBS-282	encoding *X. albilineans* type I-C single spacer array with non-targeting spacer	Wimmer et al. 2022 (PMID: 35216669)	-	https://benchling.com/s/seq-4aMeAIEV26rY8HuCjjyz?m=slm-htHmKRQOuJvJWyKu8LMq
pXalb_IF1_Cas2-3	CBS-044	encoding *X. albilineans* type I-F1 *cas2-3*	Wimmer et al. 2022 (PMID: 35216669)	178769	https://benchling.com/s/seq-PsjIyfennP6Hx8kpXGNp?m=slm-I6pUfJokJtbBZGNO6PDY
pXalb_IF1_Cas2-3_J23105	CBS-2790	encoding *X. albilineans* type I-F1 *cas2-3* with J23105 promoter	this study	-	https://benchling.com/s/seq-32Fw9m50kR4tceeIyt9O?m=slm-tIEL3CiHz15ln7SgqQuK
pXalb_IF1_Cas5	CBS-047	encoding *X. albilineans* type I-F1 *cas5*	Wimmer et al. 2022 (PMID: 35216669)	-	https://benchling.com/s/seq-p9LFfBPwGUY3KLFp6ixZ?m=slm-HVB4NomaPZdsKzgVIKkS
pXalb_IF1_Cas6	CBS-051	encoding *X. albilineans* type I-F1 *cas6*	Wimmer et al. 2022 (PMID: 35216669)	-	https://benchling.com/s/seq-5XQH6d9XdslJgY95wDid?m=slm-3iqaj2zhjPeANc7IZa7J
pXalb_IF1_Cas7	CBS-049	encoding *X. albilineans* type I-F1 *cas7*	Wimmer et al. 2022 (PMID: 35216669)	-	https://benchling.com/s/seq-fjMzEVX2IJgwc4vQIX4a?m=slm-wL6Ze8kzO6HxhzHXPsuG
pXalb_IF1_Cas8	CBS-091	encoding *X. albilineans* type I-F1 *cas8*	Wimmer et al. 2022 (PMID: 35216669)	-	https://benchling.com/s/seq-zQe3FJmwv0uCaPkNnPaX?m=slm-QoSwY0zAj0C8utcRn0z2
pXalb_IF1_Cascade_NT	CBS-2703	encoding *X. albilineans* type I-F1 Cascade genes without single spacer array	this study	-	https://benchling.com/s/seq-uODY4wu7DdS9z0lFqWz8?m=slm-jdLvnyYCc32on9EphwFx
pXalb_IF1_Cascade_sp1	CBS-2702	encoding *X. albilineans* type I-F1 Cascade genes and single spacer array targeting *lacZ* promoter	this study	-	https://benchling.com/s/seq-G4aWwQbl6jZEtIg3hz0Q?m=slm-hJY0evB7Y6sIl2xhHLDA
pXalb_IF1_Cascade_sp2	CBS-2731	encoding *X. albilineans* type I-F1 Cascade genes and single spacer array targeting *lacZ*	this study	-	https://benchling.com/s/seq-4K2OY4o1Yhz6ANeWCNbI?m=slm-mTMGyVO3TvJv8bL898Sw

Table S3 (continued)

Name	Lab number	Description	Source	Addgene number	benchling link
pXalb_IF1_Cascade_sp3	CBS-2708	encoding *X. albilineans* type I-F1 Cascade genes and single spacer array targeting *lacZ*	this study	-	https://benchling.com/s/seq-xR0hXmOj6UdneDb7btTx?m=slm-MpKLzSBnZCcKmS9VxeH5
pXalb_IF1_Cascade_sp4	CBS-4057	encoding *X. albilineans* type I-F1 Cascade genes and single spacer array targeting *degfp* promoter	this study	-	https://benchling.com/s/seq-8EPQnAjuacDwX339QxYi?m=slm-6V4uL9AtpU3cfEZiMEbf
pXalb_IF1_gRNA1	CBS-198	encoding *X. albilineans* type I-F1 single spacer array targeting *degfp* promoter	Wimmer et al. 2022 (PMID: 35216669)	-	https://benchling.com/s/seq-Z1nvx2yl8EcMirzOFMfB?m=slm-nF6rTxOm5Dk35pNTewXO
pXalb_IF1_gRNA2	CBS-208	encoding *X. albilineans* type I-F1 single spacer array targeting upstream of *degfp* promoter	Wimmer et al. 2022 (PMID: 35216669)	-	https://benchling.com/s/seq-UAtMbI2Bv0caxyUkYgZP?m=slm-f5PtsmP6e6elr7HIjYOm
pXalb_IF1_gRNAnt	CBS-283	encoding *X. albilineans* type I-F1 single spacer array with non-targeting spacer	Wimmer et al. 2022 (PMID: 35216669)	-	https://benchling.com/s/seq-j89yU6KdsH2wesUKlw68?m=slm-Q67JoZxsQZHObXJXU5Ot
pXalb_noCas3	CBS-2753	negative control plasmid missing *cas3* gene	this study	-	https://benchling.com/s/seq-Tz7PI1Lg0PFlkZNG9IGK?m=slm-EAH7240SpU78IMd0b77K

Discussion

Studies investigating type I CRISPR-Cas systems are still underrepresented if compared to their prevalence in nature. In this work, new techniques were developed to contribute to closing this gap. The newly developed method PAM-DETECT (PAM DETermination with Enrichment-based Cell-free TXTL) was generated to facilitate investigation of PAM requirements focusing on multi-component CRISPR-Cas systems (105). This TXTL-based assay was used to unravel PAM characteristics of 17 different CRISPR-Cas systems. However, also PAM-DETECT comes with disadvantages. One potential downside of PAM-DETECT is its reliance on target-binding rather than target-degradation and PAM requirements during degradation might differ from those during binding (141). Nevertheless, we also presented a method to investigate target-degradation in TXTL and PAM recognition during degradation can readily be verified (105). Formerly, our group developed a TXTL-based PAM assay that relies on degradation rather than binding (142). However, this assay relies on depletion of functional PAMs - contrasting PAM-DETECT, which relies on enrichment of recognized PAMs - and therefore requires a high sequencing depth to also detect lowly recognized PAMs. Nevertheless, combining both assays would unravel PAMs that enable binding while not allowing for target cleavage. Binding-specific PAMs could be used to reprogram endogenous CRISPR-Cas systems to function as gene regulators without the need to mutate/delete the nuclease or downregulate it. Partial complementary or shortened spacers were also shown to enable gene regulation by impairing target-cleavage (143–145). Yet, every CRISPR-Cas system needs to be evaluated separately as the spacer requirements that allow target-binding but impair target-cleavage might differ in every system.

PAM-DETECT facilitated studying PAM requirements of multi-component systems. Even though solely type I systems were assessed in this work, the assay is also adaptable to other class 1 systems or class 2 systems. As type III systems recognize RNAs (30, 146), their interrogation with PAM-DETECT is restricted. Nevertheless, type III systems do not rely on PAMs to distinguish between self and non-self, therefore PAM assays are not expedient anyway (32). However, the poorly understood type IV systems could be well suited for interrogation with PAM-DETECT. The assay requires lack of nuclease activity and type IV systems - besides type IV-C - are typically nuclease deficient (15). One potential hurdle impeding investigation is that type IV systems might require additional, not yet known accessory proteins that are not present in TXTL. Type II and type V CRISPR-Cas systems represent class 2 systems that are suitable for PAM-DETECT, while investigation of type VI systems is not possible as they recognize RNA sequences (42). Mutations of Cas9 or Cas12 impairing their nuclease activity are necessary due to type II and type V systems engaging one

single protein to bind and cleave their target (37, 147). Herein, type V-K systems represent exceptions as these transposon-associated systems include a naturally inactivated RuvC-like nuclease domain (101, 103, 106). PAM-DETECT is therefore a promising tool to contribute to the fundamental understanding of type V-K CASTs, a recently discovered subgroup of type V CRISPR-Cas systems (15, 103).

We did not assess type V-K CASTs in this work, however PAM-DETECT was used to unravel PAM requirements of three type I CASTs (105). CASTs do not only depend on a CRISPR-Cas system for their function but also employ transposon associated genes. As type I-B CASTs integrate in a CRISPR-dependent or a CRISPR-independent manner, they require many proteins to function. We therefore seeked to reconstitute transposition in TXTL. The type I-B2.2 CAST from *Rippkaea orientalis* for instance utilizes eight genes in total- even more than the six genes required for type I-E interference. Cascade is formed by *cas6*, *cas8*, *cas7*, and *cas5*, while *tniQ*, *tnsAB*, *tnsC*, and *tnsD* represent transposon associated genes (102, 105). Due to the high number of required proteins and the fact that high expression of RoCascade proteins turned out to be toxic in *E. coli* (105), utilizing TXTL facilitated investigation of transposition. However, the here established TXTL method exhibited some inconsistencies with results derived from our *in vivo* study in *E. coli* (105). Nevertheless, the TXTL-based method successfully defined transposon ends, uncovered which proteins are involved in CRISPR-dependent and CRISPR-independent transposition, and determined the approximate insertion site. Therefore, we recommend TXTL for an initial determination of transposition requirements that can be more extensively studied *in vivo*.

In addition to CASTs, we harnessed TXTL-based methods to assess self-targeting systems as another non-canonical aspect of CRISPR-Cas systems. Thereby, we investigated two extensively self-targeting systems endogenous to the plant pathogen *Xanthomonas albilineans* (105). PAM-DETECT revealed recognition of nearly all PAMs associated with self-targets in *X. albilineans* and both systems were shown to functionally bind and degrade their targets in TXTL (105) and *E. coli*. Therefore, we sought to understand how the bacterium survives lethal autoimmunity. We identified two Acrs (AcrIC11 and AcrIF12$_{Xal}$) that inhibit target degradation while preserving Cascade-binding for both CRISPR-Cas systems. Nevertheless, the exact mechanism of how both Acrs achieve inhibition of Cas3-degradation remains to be solved.

As most identified Acr mechanisms rely on direct binding to Cas proteins, probing for a direct binding partner would be the first step to assess the means of AcrIC11. So far, three type I degradation inhibiting Acrs were found to bind to Cas3 and two Acrs were found to bind Cascade, probably blocking Cas3 recruitment (83, 148–152). A homolog of AcrIF12$_{Xal}$ was shown previously to not bind to Cascade or Cas3 in isolation (81), which is why we conclude

that AcrIF12$_{Xal}$'s inhibitory activity most likely does not rely on binding to a Cas protein. As AcrIF12$_{Xal}$ is active in TXTL even if present in high dilutions, we suspect that AcrIF12$_{Xal}$ is a multi-turnover enzyme (105). The exact stoichiometry, in which this Acr inactivates the type I-F1 system still needs to be assessed *in vitro* to allow for reaction conditions with a defined stoichiometry that are not possible in TXTL. To our knowledge, five enzymatic Acrs have been reported with one being a type I Acr (AcrIF11) (60, 73, 79–81). AcrIF11 ADP-ribosylates Cas8, preventing it from PAM recognition (81) and therefore inhibits not only target degradation but also target binding. Thus, we assume that AcrIF12$_{Xal}$ - if it is an enzymatic Acr - uses a mechanism of action that is distinct from AcrIF11. As the already uncovered means employed by enzymatically active Acrs are diverse (60, 73, 79–81), unraveling the mode of action of a novel enzymatic Acr is challenging. A potential starting point to investigate AcrIF12$_{Xal}$ is to figure out which protein is attacked by the Acr. Therefore, a bacterial two hybrid assay could be employed as this screen detects if two proteins are in close proximity even if they are not directly bound to each other (153, 154). As Cas3 activity relies on interactions with Cas8 (28), it seems probable that an enzymatic activity of AcrIF12$_{Xal}$ would be directed to either Cas3 itself or to Cas8 to prevent activation of Cas3. Subsequently, post-translational modifications of either Cas3 or Cas8 can be assessed with mass spectrometry (155, 156).

Besides the means of AcrIC11 and AcrIF12$_{Xal}$, the reason for two Cas3 proteins being functionally encoded in *X. albilineans* is still unclear. As maintenance of CRISPR-Cas systems was previously associated with fitness costs (157, 158), it would be expected that permanent inactivation of the CRISPR-Cas nucleases would lead to loss of both *cas3* genes. We therefore suggest, *X. albilineans* might have employed its type I-C and type I-F1 systems for a non-canonical defense mechanism. Sensing of phage-infection by *X. albilineans* could trigger downregulation of AcrIC11 and/or AcrIF12$_{Xal}$ expression which would result in host genome degradation. Both type I CRISPR-Cas systems would act as Abi systems and cell death of the infected bacterium would enable altruistic protection of the bacterial community. Overall, TXTL-based methods were established here that can be used as a starting point to look into alternative functions of CRISPR-Cas including self-targeting spacers and CRISPR-Cas system utilizing transposon. The time-saving investigation of CAST PAM requirements and determination of necessary transposition proteins could furthermore boost the investigation of new CASTs and advance their use as biotechnological tools.

References for Introduction and Discussion

1. Valen,V.A.N. and L (1973) A new evolutionary law. *Evol Theory*, **1**, 1–30.
2. Stern,A. and Sorek,R. (2011) The phage-host arms race: shaping the evolution of microbes. *Bioessays*, **33**, 43–51.
3. Labrie,S.J., Samson,J.E. and Moineau,S. (2010) Bacteriophage resistance mechanisms. *Nature Reviews Microbiology*, **8**, 317–327.

4. Dy,R.L., Richter,C., Salmond,G.P.C. and Fineran,P.C. (2014) Remarkable mechanisms in microbes to resist phage infections. *Annu Rev Virol*, **1**, 307–331.
5. Oliveira,P.H., Touchon,M. and Rocha,E.P.C. (2014) The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.*, **42**, 10618–10631.
6. Tock,M.R. and Dryden,D.T.F. (2005) The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.*, **8**, 466–472.
7. Kronheim,S., Daniel-Ivad,M., Duan,Z., Hwang,S., Wong,A.I., Mantel,I., Nodwell,J.R. and Maxwell,K.L. (2018) A chemical defence against phage infection. *Nature*, **564**, 283–286.
8. Lopatina,A., Tal,N. and Sorek,R. (2020) Abortive infection: bacterial suicide as an antiviral immune strategy. *Annu Rev Virol*, **7**, 371–384.
9. Benzer,S. (1955) Fine structure of a genetic region in bacteriophage. *Proceedings of the National Academy of Sciences*, **41**, 344–354.
10. Wong,S., Alattas,H. and Slavcev,R.A. (2021) A snapshot of the λ T4rII exclusion (Rex) phenotype in *Escherichia coli*. *Curr. Genet.*, **67**, 739–745.
11. Parma,D.H., Snyder,M., Sobolevski,S., Nawroz,M., Brody,E. and Gold,L. (1992) The Rex system of bacteriophage lambda: tolerance and altruistic cell death. *Genes Dev.*, **6**, 497–510.
12. Cohen,D., Melamed,S., Millman,A., Shulman,G., Oppenheimer-Shaanan,Y., Kacen,A., Doron,S., Amitai,G. and Sorek,R. (2019) Cyclic GMP–AMP signalling protects bacteria against viral infection. *Nature*, **574**, 691–695.
13. Severin,G.B., Ramliden,M.S., Hawver,L.A., Wang,K., Pell,M.E., Kieninger,A.-K., Khataokar,A., O'Hara,B.J., Behrmann,L.V., Neiditch,M.B., *et al.* (2018) Direct activation of a phospholipase by cyclic GMP-AMP in El Tor. *Proc. Natl. Acad. Sci. U. S. A.*, **115**, E6048–E6055.
14. Lau,R.K., Ye,Q., Birkholz,E.A., Berg,K.R., Patel,L., Mathews,I.T., Watrous,J.D., Ego,K., Whiteley,A.T., Lowey,B., *et al.* (2020) Structure and mechanism of a cyclic trinucleotide-activated bacterial endonuclease mediating bacteriophage immunity. *Mol. Cell*, **77**, 723–733.e6.
15. Makarova,K.S., Wolf,Y.I., Iranzo,J., Shmakov,S.A., Alkhnbashi,O.S., Brouns,S.J.J., Charpentier,E., Cheng,D., Haft,D.H., Horvath,P., *et al.* (2019) Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.*, **18**, 67–83.
16. Mojica,F.J.M., Díez-Villaseñor,C., García-Martínez,J. and Soria,E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.*, **60**, 174–182.
17. Yosef,I., Shitrit,D., Goren,M.G., Burstein,D., Pupko,T. and Qimron,U. (2013) DNA motifs determining the efficiency of adaptation into the *Escherichia coli* CRISPR array. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, 14396–14401.
18. 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.
19. Lee,H. and Sashital,D.G. (2022) Creating memories: molecular mechanisms of CRISPR adaptation. *Trends Biochem. Sci.*, 10.1016/j.tibs.2022.02.004.
20. Charpentier,E., Richter,H., van der Oost,J. and White,M.F. (2015) Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiol. Rev.*, **39**, 428–441.
21. Brouns,S.J.J., Jore,M.M., Lundgren,M., Westra,E.R., Slijkhuis,R.J.H., Snijders,A.P.L., 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.
22. Mojica,F.J.M., Díez-Villaseñor,C., García-Martínez,J. and Almendros,C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, **155**, 733–740.
23. Deveau,H., Barrangou,R., Garneau,J.E., Labonté,J., Fremaux,C., Boyaval,P., Romero,D.A., Horvath,P. and Moineau,S. (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.*, **190**, 1390–1400.
24. Marraffini,L.A. and Sontheimer,E.J. (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*, **463**, 568–571.
25. Meeske,A.J. and Marraffini,L.A. (2018) RNA guide complementarity prevents self-targeting in type VI CRISPR systems. *Mol. Cell*, **71**, 791–801.e3.
26. Makarova,K.S., Wolf,Y.I., Alkhnbashi,O.S., Costa,F., Shah,S.A., Saunders,S.J., Barrangou,R., Brouns,S.J.J., Charpentier,E., Haft,D.H., *et al.* (2015) An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.*, **13**, 722–736.
27. Jore,M.M., Lundgren,M., van Duijn,E., Bultema,J.B., Westra,E.R., Waghmare,S.P., Wiedenheft,B., Pul,U., Wurm,R., Wagner,R., *et al.* (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.*, **18**, 529–536.

28. Hochstrasser,M.L., Taylor,D.W., Bhat,P., Guegler,C.K., Sternberg,S.H., Nogales,E. and Doudna,J.A. (2014) CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 6618–6623.
29. Sinkunas,T., Gasiunas,G., Fremaux,C., Barrangou,R., Horvath,P. and Siksnys,V. (2011) Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.*, **30**, 1335–1342.
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. 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.
32. Kazlauskiene,M., Tamulaitis,G., Kostiuk,G., Venclovas,Č. and Siksnys,V. (2016) Spatiotemporal control of type III-A CRISPR-Cas immunity: coupling DNA degradation with the target RNA recognition. *Mol. Cell*, **62**, 295–306.
33. 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.
34. Niewoehner,O., Garcia-Doval,C., Rostøl,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.
35. Kazlauskiene,M., Kostiuk,G., Venclovas,Č., Tamulaitis,G. and Siksnys,V. (2017) A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science*, **357**, 605–609.
36. Crowley,V.M., Catching,A., Taylor,H.N., Borges,A.L., Metcalf,J., Bondy-Denomy,J. and Jackson,R.N. (2019) A Type IV-A CRISPR-Cas system in *Pseudomonas aeruginosa* mediates RNA-guided plasmid interference *in vivo*. *The CRISPR Journal*, **2**, 434–440.
37. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.
38. Yan,W.X., Hunnewell,P., Alfonse,L.E., Carte,J.M., Keston-Smith,E., Sothiselvam,S., Garrity,A.J., Chong,S., Makarova,K.S., Koonin,E.V., *et al.* (2019) Functionally diverse type V CRISPR-Cas systems. *Science*, **363**, 88–91.
39. Harrington,L.B., Ma,E., Chen,J.S., Witte,I.P., Gertz,D., Paez-Espino,D., Al-Shayeb,B., Kyrpides,N.C., Burstein,D., Banfield,J.F., *et al.* (2020) A scoutRNA is required for some type V CRISPR-Cas systems. *Mol. Cell*, **79**, 416–424.e5.
40. Zetsche,B., Gootenberg,J.S., Abudayyeh,O.O., Slaymaker,I.M., Makarova,K.S., Essletzbichler,P., Volz,S.E., Joung,J., van der Oost,J., Regev,A., *et al.* (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, **163**, 759–771.
41. Chen,J.S., Ma,E., Harrington,L.B., Da Costa,M., Tian,X., Palefsky,J.M. and Doudna,J.A. (2018) CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science*, **360**, 436–439.
42. Abudayyeh,O.O., Gootenberg,J.S., Konermann,S., Joung,J., Slaymaker,I.M., Cox,D.B.T., Shmakov,S., Makarova,K.S., Semenova,E., Minakhin,L., *et al.* (2016) C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science*, **353**, aaf5573.
43. Meeske,A.J., Nakandakari-Higa,S. and Marraffini,L.A. (2019) Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. *Nature*, **570**, 241–245.
44. 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.e10.
45. 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.e12.
46. Westra,E.R., van Erp,P.B.G., Künne,T., Wong,S.P., Staals,R.H.J., Seegers,C.L.C., Bollen,S., Jore,M.M., Semenova,E., Severinov,K., *et al.* (2012) CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell*, **46**, 595–605.
47. Rostøl,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.
48. Watson,B.N.J., Vercoe,R.B., Salmond,G.P.C., Westra,E.R., Staals,R.H.J. and Fineran,P.C. (2019) Type I-F CRISPR-Cas resistance against virulent phages results in abortive infection and provides population-level immunity. *Nat. Commun.*, **10**, 5526.
49. Hampton,H.G., Watson,B.N.J. and Fineran,P.C. (2020) The arms race between bacteria and their phage foes. *Nature*, **577**, 327–336.
50. Malone,L.M., Birkholz,N. and Fineran,P.C. (2021) Conquering CRISPR: how phages overcome bacterial adaptive immunity. *Curr. Opin. Biotechnol.*, **68**, 30–36.
51. Watson,B.N.J., Easingwood,R.A., Tong,B., Wolf,M., Salmond,G.P.C., Staals,R.H.J., Bostina,M. and Fineran,P.C. (2019) Different genetic and morphological outcomes for phages targeted by single or multiple CRISPR-Cas spacers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **374**, 20180090.

52. Zhang,Z., Pan,S., Liu,T., Li,Y. and Peng,N. (2019) Cas4 nucleases can effect specific integration of CRISPR spacers. *J. Bacteriol.*, **201**.

53. Lin,P., Qin,S., Pu,Q., Wang,Z., Wu,Q., Gao,P., Schettler,J., Guo,K., Li,R., Li,G., *et al.* (2020) CRISPR-Cas13 inhibitors block RNA editing in bacteria and mammalian cells. *Mol. Cell*, **78**, 850–861.e5.

54. Marino,N.D., Zhang,J.Y., Borges,A.L., Sousa,A.A., Leon,L.M., Rauch,B.J., Walton,R.T., Berry,J.D., Joung,J.K., Kleinstiver,B.P., *et al.* (2018) Discovery of widespread type I and type V CRISPR-Cas inhibitors. *Science*, **362**, 240–242.

55. He,F., Bhoobalan-Chitty,Y., Van,L.B., Kjeldsen,A.L., Dedola,M., Makarova,K.S., Koonin,E.V., Brodersen,D.E. and Peng,X. (2018) Anti-CRISPR proteins encoded by archaeal lytic viruses inhibit subtype I-D immunity. *Nat Microbiol*, **3**, 461–469.

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

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

58. Rauch,B.J., Silvis,M.R., Hultquist,J.F., Waters,C.S., McGregor,M.J., Krogan,N.J. and Bondy-Denomy,J. (2017) Inhibition of CRISPR-Cas9 with bacteriophage proteins. *Cell*, **168**, 150–158.e10.

59. Pawluk,A., Amrani,N., Zhang,Y., Garcia,B., Hidalgo-Reyes,Y., Lee,J., Edraki,A., Shah,M., Sontheimer,E.J., Maxwell,K.L., *et al.* (2016) Naturally occurring off-switches for CRISPR-Cas9. *Cell*, **167**, 1829–1838.e9.

60. Athukoralage,J.S., McMahon,S.A., Zhang,C., Grüschow,S., Graham,S., Krupovic,M., Whitaker,R.J., Gloster,T.M. and White,M.F. (2020) An anti-CRISPR viral ring nuclease subverts type III CRISPR immunity. *Nature*, **577**, 572–575.

61. Bhoobalan-Chitty,Y., Johansen,T.B., Di Cianni,N. and Peng,X. (2019) Inhibition of type III CRISPR-Cas immunity by an archaeal virus-encoded anti-CRISPR protein. *Cell*, **179**, 448–458.e11.

62. Watters,K.E., Fellmann,C., Bai,H.B., Ren,S.M. and Doudna,J.A. (2018) Systematic discovery of natural CRISPR-Cas12a inhibitors. *Science*, **362**, 236–239.

63. Meeske,A.J., Jia,N., Cassel,A.K., Kozlova,A., Liao,J., Wiedmann,M., Patel,D.J. and Marraffini,L.A. (2020) A phage-encoded anti-CRISPR enables complete evasion of type VI-A CRISPR-Cas immunity. *Science*, **369**, 54–59.

64. Wandera,K.G., Alkhnbashi,O.S., Basset,H.V.I., Mitrofanov,A., Hauns,S., Migur,A., Backofen,R. and Beisel,C.L. (under revision) Anti-CRISPR prediction using deep learning reveals an inhibitor of Cas13b nucleases. *Mol. Cell*.

65. Hynes,A.P., Rousseau,G.M., Lemay,M.-L., Horvath,P., Romero,D.A., Fremaux,C. and Moineau,S. (2017) An anti-CRISPR from a virulent streptococcal phage inhibits *Streptococcus pyogenes* Cas9. *Nat Microbiol*, **2**, 1374–1380.

66. Hynes,A.P., Rousseau,G.M., Agudelo,D., Goulet,A., Amigues,B., Loehr,J., Romero,D.A., Fremaux,C., Horvath,P., Doyon,Y., *et al.* (2018) Widespread anti-CRISPR proteins in virulent bacteriophages inhibit a range of Cas9 proteins. *Nat. Commun.*, **9**, 2919.

67. Uribe,R.V., van der Helm,E., Misiakou,M.-A., Lee,S.-W., Kol,S. and Sommer,M.O.A. (2019) Discovery and characterization of Cas9 inhibitors disseminated across seven bacterial phyla. *Cell Host Microbe*, **26**, 702.

68. Forsberg,K.J., Bhatt,I.V., Schmidtke,D.T., Javanmardi,K., Dillard,K.E., Stoddard,B.L., Finkelstein,I.J., Kaiser,B.K. and Malik,H.S. (2019) Functional metagenomics-guided discovery of potent Cas9 inhibitors in the human microbiome. *Elife*, **8**.

69. Osuna,B.A., Karambelkar,S., Mahendra,C., Christie,K.A., Garcia,B., Davidson,A.R., Kleinstiver,B.P., Kilcher,S. and Bondy-Denomy,J. (2020) *Listeria* phages induce Cas9 degradation to protect lysogenic genomes. *Cell Host Microbe*, **28**, 31–40.e9.

70. Watters,K.E., Shivram,H., Fellmann,C., Lew,R.J., McMahon,B. and Doudna,J.A. (2020) Potent CRISPR-Cas9 inhibitors from genomes. *Proc. Natl. Acad. Sci. U. S. A.*, **117**, 6531–6539.

71. Mahendra,C., Christie,K.A., Osuna,B.A., Pinilla-Redondo,R., Kleinstiver,B.P. and Bondy-Denomy,J. (2020) Broad-spectrum anti-CRISPR proteins facilitate horizontal gene transfer. *Nat Microbiol*, **5**, 620–629.

72. Eitzinger,S., Asif,A., Watters,K.E., Iavarone,A.T., Knott,G.J., Doudna,J.A. and Minhas,F.U.A.A. (2020) Machine learning predicts new anti-CRISPR proteins. *Nucleic Acids Res.*, **48**, 4698–4708.

73. Forsberg,K.J., Schmidtke,D.T., Werther,R., Uribe,R.V., Hausman,D., Sommer,M.O.A., Stoddard,B.L., Kaiser,B.K. and Malik,H.S. (2021) The novel anti-CRISPR AcrIIA22 relieves DNA torsion in target plasmids and impairs SpyCas9 activity. *PLoS Biol.*, **19**, e3001428.

74. Varble,A., Campisi,E., Euler,C.W., Maguin,P., Kozlova,A., Fyodorova,J., Rostøl,J.T., Fischetti,V.A. and Marraffini,L.A. (2021) Prophage integration into CRISPR loci enables evasion of antiviral immunity in *Streptococcus pyogenes*. *Nat Microbiol*, **6**, 1516–1525.

75. Song,G., Zhang,F., Tian,C., Gao,X., Zhu,X., Fan,D. and Tian,Y. (2022) Discovery of potent and versatile CRISPR–Cas9 inhibitors engineered for chemically controllable genome editing. *Nucleic Acids Research*, 10.1093/nar/gkac099.

76. Lee,J., Mir,A., Edraki,A., Garcia,B., Amrani,N., Lou,H.E., Gainetdinov,I., Pawluk,A., Ibraheim,R., Gao,X.D., *et al.* (2018) Potent Cas9 inhibition in bacterial and human cells by AcrIIC4 and AcrIIC5 anti-CRISPR proteins. *MBio*, **9**.
77. Meeske,A.J., Johnson,M.C., Hille,L.T., Kleinstiver,B.P. and Bondy-Denomy,J. Lack of Cas13a inhibition by anti-CRISPR proteins from *Leptotrichia* prophages. 10.1101/2021.05.27.445852.
78. Jia,N. and Patel,D.J. (2021) Structure-based functional mechanisms and biotechnology applications of anti-CRISPR proteins. *Nat. Rev. Mol. Cell Biol.*, **22**, 563–579.
79. Dong,L., Guan,X., Li,N., Zhang,F., Zhu,Y., Ren,K., Yu,L., Zhou,F., Han,Z., Gao,N., *et al.* (2019) An anti-CRISPR protein disables type V Cas12a by acetylation. *Nat. Struct. Mol. Biol.*, **26**, 308–314.
80. Knott,G.J., Thornton,B.W., Lobba,M.J., Liu,J.-J., Al-Shayeb,B., Watters,K.E. and Doudna,J.A. (2019) Broad-spectrum enzymatic inhibition of CRISPR-Cas12a. *Nat. Struct. Mol. Biol.*, **26**, 315–321.
81. Niu,Y., Yang,L., Gao,T., Dong,C., Zhang,B., Yin,P., Hopp,A.-K., Li,D., Gan,R., Wang,H., *et al.* (2020) A type I-F anti-CRISPR protein inhibits the CRISPR-Cas surveillance complex by ADP-ribosylation. *Mol. Cell*, **80**, 512–524.e5.
82. Pawluk,A., Staals,R.H.J., Taylor,C., Watson,B.N.J., Saha,S., Fineran,P.C., Maxwell,K.L. and Davidson,A.R. (2016) Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nat Microbiol*, **1**, 16085.
83. León,L.M., Park,A.E., Borges,A.L., Zhang,J.Y. and Bondy-Denomy,J. (2021) Mobile element warfare via CRISPR and anti-CRISPR in *Pseudomonas aeruginosa*. *Nucleic Acids Res.*, **49**, 2114–2125.
84. Pinilla-Redondo,R., Shehreen,S., Marino,N.D., Fagerlund,R.D., Brown,C.M., Sørensen,S.J., Fineran,P.C. and Bondy-Denomy,J. (2020) Discovery of multiple anti-CRISPRs highlights anti-defense gene clustering in mobile genetic elements. *Nat. Commun.*, **11**, 5652.
85. Yin,Y., Yang,B. and Entwistle,S. (2019) Bioinformatics identification of anti-CRISPR loci by using homology, guilt-by-association, and CRISPR self-targeting spacer approaches. *mSystems*, **4**.
86. Birkholz,N., Fagerlund,R.D., Smith,L.M., Jackson,S.A. and Fineran,P.C. (2019) The autoregulator Aca2 mediates anti-CRISPR repression. *Nucleic Acids Res.*, **47**, 9658–9665.
87. Stanley,S.Y., Borges,A.L., Chen,K.-H., Swaney,D.L., Krogan,N.J., Bondy-Denomy,J. and Davidson,A.R. (2019) Anti-CRISPR-associated proteins are crucial repressors of anti-CRISPR transcription. *Cell*, **178**, 1452–1464.e13.
88. Lee,S.Y., Kim,G.E. and Park,H.H. (2022) Molecular basis of transcriptional repression of anti-CRISPR by anti-CRISPR-associated 2. *Acta Crystallogr D Struct Biol*, **78**, 59–68.
89. Davidson,A.R., Lu,W.-T., Stanley,S.Y., Wang,J., Mejdani,M., Trost,C.N., Hicks,B.T., Lee,J. and Sontheimer,E.J. (2020) Anti-CRISPRs: Protein Inhibitors of CRISPR-Cas Systems. *Annu. Rev. Biochem.*, **89**, 309–332.
90. Mohanraju,P., Saha,C., van Baarlen,P., Louwen,R., Staals,R.H.J. and van der Oost,J. (2022) Alternative functions of CRISPR–Cas systems in the evolutionary arms race. *Nature Reviews Microbiology*, 10.1038/s41579-021-00663-z.
91. Wimmer,F. and Beisel,C.L. (2019) CRISPR-Cas systems and the paradox of self-targeting spacers. *Front. Microbiol.*, **10**, 3078.
92. Li,R., Fang,L., Tan,S., Yu,M., Li,X., He,S., Wei,Y., Li,G., Jiang,J. and Wu,M. (2016) Type I CRISPR-Cas targets endogenous genes and regulates virulence to evade mammalian host immunity. *Cell Res.*, **26**, 1273–1287.
93. Heussler,G.E., Cady,K.C., Koeppen,K., Bhuju,S., Stanton,B.A. and O'Toole,G.A. (2015) Clustered regularly interspaced short palindromic repeat-dependent, biofilm-specific death of *Pseudomonas aeruginosa* mediated by increased expression of phage-related genes. *MBio*, **6**, e00129–15.
94. Cady,K.C. and O'Toole,G.A. (2011) Non-identity-mediated CRISPR-bacteriophage interaction mediated via the Csy and Cas3 proteins. *J. Bacteriol.*, **193**, 3433–3445.
95. Zegans,M.E., Wagner,J.C., Cady,K.C., Murphy,D.M., Hammond,J.H. and O'Toole,G.A. (2009) Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviors of *Pseudomonas aeruginosa*. *J. Bacteriol.*, **191**, 210–219.
96. Ratner,H.K., Escalera-Maurer,A., Le Rhun,A., Jaggavarapu,S., Wozniak,J.E., Crispell,E.K., Charpentier,E. and Weiss,D.S. (2019) Catalytically active Cas9 mediates transcriptional interference to facilitate bacterial virulence. *Mol. Cell*, **75**, 498–510.e5.
97. Bozic,B., Repac,J. and Djordjevic,M. (2019) Endogenous gene regulation as a predicted main function of type I-E CRISPR/Cas system in *E. coli*. *Molecules*, **24**, 784.
98. Li,M., Gong,L., Cheng,F., Yu,H., Zhao,D., Wang,R., Wang,T., Zhang,S., Zhou,J., Shmakov,S.A., *et al.* (2021) Toxin-antitoxin RNA pairs safeguard CRISPR-Cas systems. *Science*, **372**.
99. Yang,S., Zhang,Y., Xu,J., Zhang,J., Zhang,J., Yang,J., Jiang,Y. and Yang,S. (2021) Orthogonal CRISPR-associated transposases for parallel and multiplexed chromosomal integration. *Nucleic Acids Res.*, **49**, 10192–10202.

100. Rybarski,J.R., Hu,K., Hill,A.M., Wilke,C.O. and Finkelstein,I.J. (2021) Metagenomic discovery of CRISPR-associated transposons. *Proc. Natl. Acad. Sci. U. S. A.*, **118**.

101. Peters,J.E., Makarova,K.S., Shmakov,S. and Koonin,E.V. (2017) Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proc. Natl. Acad. Sci. U. S. A.*, **114**, E7358–E7366.

102. Saito,M., Ladha,A., Strecker,J., Faure,G., Neumann,E., Altae-Tran,H., Macrae,R.K. and Zhang,F. (2021) Dual modes of CRISPR-associated transposon homing. *Cell*, **184**, 2441–2453.e18.

103. Strecker,J., Ladha,A., Gardner,Z., Schmid-Burgk,J.L., Makarova,K.S., Koonin,E.V. and Zhang,F. (2019) RNA-guided DNA insertion with CRISPR-associated transposases. *Science*, **365**, 48–53.

104. Klompe,S.E., Vo,P.L.H., Halpin-Healy,T.S. and Sternberg,S.H. (2019) Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature*, **571**, 219–225.

105. Wimmer,F., Mougiakos,I., Englert,F. and Beisel,C.L. (2022) Rapid cell-free characterization of multi-subunit CRISPR effectors and transposons. *Mol. Cell*, 10.1016/j.molcel.2022.01.026.

106. Faure,G., Shmakov,S.A., Yan,W.X., Cheng,D.R., Scott,D.A., Peters,J.E., Makarova,K.S. and Koonin,E.V. (2019) CRISPR–Cas in mobile genetic elements: counter-defence and beyond. *Nature Reviews Microbiology*, **17**, 513–525.

107. Balderston,S., Clouse,G., Ripoll,J.-J., Pratt,G.K., Gasiunas,G., Bock,J.-O., Bennett,E.P. and Aran,K. (2021) Diversification of the CRISPR toolbox: applications of CRISPR-Cas systems beyond genome editing. *CRISPR J*, **4**, 400–415.

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

109. Pardee,K., Green,A.A., Takahashi,M.K., Braff,D., Lambert,G., Lee,J.W., Ferrante,T., Ma,D., Donghia,N., Fan,M., *et al.* (2016) Rapid, low-cost detection of Zika virus using programmable biomolecular components. *Cell*, **165**, 1255–1266.

110. Baylis,F. and McLeod,M. (2017) First-in-human phase 1 CRISPR gene editing cancer trials: are we ready? *Curr. Gene Ther.*, **17**, 309–319.

111. Zheng,Y., Li,J., Wang,B., Han,J., Hao,Y., Wang,S., Ma,X., Yang,S., Ma,L., Yi,L., *et al.* (2020) Endogenous type I CRISPR-Cas: from foreign DNA defense to prokaryotic engineering. *Front Bioeng Biotechnol*, **8**, 62.

112. Pyne,M.E., Bruder,M.R., Moo-Young,M., Chung,D.A. and Perry Chou,C. (2016) Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *Clostridium*. *Scientific Reports*, **6**.

113. Ungerer,J. and Pakrasi,H.B. (2016) Cpf1 is a versatile tool for CRISPR genome editing across diverse species of cyanobacteria. *Sci. Rep.*, **6**, 39681.

114. Xu,Z., Li,Y., Li,M., Xiang,H. and Yan,A. (2021) Harnessing the type I CRISPR-Cas systems for genome editing in prokaryotes. *Environ. Microbiol.*, **23**, 542–558.

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

116. Gomaa,A.A., Klumpe,H.E., Luo,M.L., Selle,K., Barrangou,R. and Beisel,C.L. (2014) Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. *mBio*, **5**.

117. Yosef,I., Manor,M., Kiro,R. and Qimron,U. (2015) Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria. *Proc. Natl. Acad. Sci. U. S. A.*, **112**, 7267–7272.

118. Luo,M.L., Mullis,A.S., Leenay,R.T. and Beisel,C.L. (2015) Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. *Nucleic Acids Res.*, **43**, 674–681.

119. Rath,D., Amlinger,L., Hoekzema,M., Devulapally,P.R. and Lundgren,M. (2015) Efficient programmable gene silencing by Cascade. *Nucleic Acids Res.*, **43**, 237–246.

120. Stachler,A.-E. and Marchfelder,A. (2016) Gene Repression in *Haloarchaea* using the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas I-B system. *J. Biol. Chem.*, **291**, 15226–15242.

121. Dolan,A.E., Hou,Z., Xiao,Y., Gramelspacher,M.J., Heo,J., Howden,S.E., Freddolino,P.L., Ke,A. and Zhang,Y. (2019) Introducing a spectrum of long-range genomic deletions in human embryonic stem cells using type I CRISPR-Cas. *Mol. Cell*, **74**, 936–950.e5.

122. Csörgő,B., León,L.M., Chau-Ly,I.J., Vasquez-Rifo,A., Berry,J.D., Mahendra,C., Crawford,E.D., Lewis,J.D. and Bondy-Denomy,J. (2020) A compact Cascade–Cas3 system for targeted genome engineering. *Nat. Methods*, **17**, 1183–1190.

123. Morisaka,H., Yoshimi,K., Okuzaki,Y., Gee,P., Kunihiro,Y., Sonpho,E., Xu,H., Sasakawa,N., Naito,Y., Nakada,S., *et al.* (2019) CRISPR-Cas3 induces broad and unidirectional genome editing in human cells. *Nat. Commun.*, **10**, 5302.

124. Cameron,P., Coons,M.M., Klompe,S.E., Lied,A.M., Smith,S.C., Vidal,B., Donohoue,P.D., Rotstein,T., Kohrs,B.W., Nyer,D.B., *et al.* (2019) Harnessing type I CRISPR–Cas systems for genome engineering in human cells. *Nat. Biotechnol.*, **37**, 1471–1477.

125. Garside,E.L., Schellenberg,M.J., Gesner,E.M., Bonanno,J.B., Sauder,J.M., Burley,S.K., Almo,S.C., Mehta,G. and MacMillan,A.M. (2012) Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases. *RNA*, **18**, 2020–2028.

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

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

128. van Erp,P.B.G., Jackson,R.N., Carter,J., Golden,S.M., Bailey,S. and Wiedenheft,B. (2015) Mechanism of CRISPR-RNA guided recognition of DNA targets in *Escherichia coli*. *Nucleic Acids Res.*, **43**, 8381–8391.

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

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

131. Gao,F., Zheng,K., Li,Y.-B., Jiang,F. and Han,C.-Y. (2022) A Cas6-based RNA tracking platform functioning in a fluorescence-activation mode. *Nucleic Acids Research*, 10.1093/nar/gkac014.

132. Luo,M.L., Jackson,R.N., Denny,S.R., Tokmina-Lukaszewska,M., Maksimchuk,K.R., Lin,W., Bothner,B., Wiedenheft,B. and Beisel,C.L. (2016) The CRISPR RNA-guided surveillance complex in *Escherichia coli* accommodates extended RNA spacers. *Nucleic Acids Res.*, **44**, 7385–7394.

133. Hochstrasser,M.L., Taylor,D.W., Kornfeld,J.E., Nogales,E. and Doudna,J.A. (2016) DNA targeting by a minimal CRISPR RNA-guided Cascade. *Mol. Cell*, **63**, 840–851.

134. Garamella,J., Marshall,R., Rustad,M. and Noireaux,V. (2016) The all *E. coli* TX-TL toolbox 2.0: A platform for cell-free synthetic biology. *ACS Synthetic Biology*, **5**, 344–355.

135. Marshall,R., Maxwell,C.S., Collins,S.P., Jacobsen,T., Luo,M.L., Begemann,M.B., Gray,B.N., January,E., Singer,A., He,Y., *et al.* (2018) Rapid and scalable characterization of CRISPR technologies using an *E. coli* cell-free transcription-translation system. *Mol. Cell*, **69**, 146–157.e3.

136. Caschera,F. and Noireaux,V. (2014) Synthesis of 2.3 mg/ml of protein with an all *Escherichia coli* cell-free transcription–translation system. *Biochimie*, **99**, 162–168.

137. Wandera,K.G., Collins,S.P., Wimmer,F., Marshall,R., Noireaux,V. and Beisel,C.L. (2020) An enhanced assay to characterize anti-CRISPR proteins using a cell-free transcription-translation system. *Methods*, **172**, 42–50.

138. Liao,C., Slotkowski,R.A., Achmedov,T. and Beisel,C.L. (2019) The *Francisella novicida* Cas12a is sensitive to the structure downstream of the terminal repeat in CRISPR arrays. *RNA Biol.*, **16**, 404–412.

139. Khakimzhan,A., Garenne,D., Tickman,B., Fontana,J., Carothers,J. and Noireaux,V. (2021) Complex dependence of CRISPR-Cas9 binding strength on guide RNA spacer lengths. *Phys. Biol.*, 10.1088/1478-3975/ac091e.

140. Liao,C., Ttofali,F., Slotkowski,R.A., Denny,S.R., Cecil,T.D., Leenay,R.T., Keung,A.J. and Beisel,C.L. (2019) Modular one-pot assembly of CRISPR arrays enables library generation and reveals factors influencing crRNA biogenesis. *Nat. Commun.*, **10**, 2948.

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

142. Maxwell,C.S., Jacobsen,T., Marshall,R., Noireaux,V. and Beisel,C.L. (2018) A detailed cell-free transcription-translation-based assay to decipher CRISPR protospacer-adjacent motifs. *Methods*, **143**, 48–57.

143. Wu,X., Scott,D.A., Kriz,A.J., Chiu,A.C., Hsu,P.D., Dadon,D.B., Cheng,A.W., Trevino,A.E., Konermann,S., Chen,S., *et al.* (2014) Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.*, **32**, 670–676.

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

145. Du,K., Gong,L., Li,M., Yu,H. and Xiang,H. (2022) Reprogramming the endogenous type I CRISPR-Cas system for simultaneous gene regulation and editing in *Haloarcula hispanica*. *mLife*, 10.1002/mlf2.12010.

146. Hale,C.R., Majumdar,S., Elmore,J., Pfister,N., Compton,M., Olson,S., Resch,A.M., Glover,C.V.C.,3rd, Graveley,B.R., Terns,R.M., *et al.* (2012) Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Mol. Cell*, **45**, 292–302.

147. Zhang,X., Wang,J., Cheng,Q., Zheng,X., Zhao,G. and Wang,J. (2017) Multiplex gene regulation by CRISPR-ddCpf1. *Cell Discov*, **3**, 17018.

148. Mejdani,M., Pawluk,A., Maxwell,K.L. and Davidson,A.R. (2021) Anti-CRISPR AcrIE2 binds the type I-E CRISPR-Cas complex but does not block DNA binding. *Journal of Molecular Biology*, **433**, 166759.

149. Pawluk,A., Shah,M., Mejdani,M., Calmettes,C., Moraes,T.F., Davidson,A.R. and Maxwell,K.L. (2017) Disabling a type I-E CRISPR-Cas nuclease with a bacteriophage-encoded anti-CRISPR protein. *mBio*, **8**.

150. Bondy-Denomy,J., Garcia,B., Strum,S., Du,M., Rollins,M.F., Hidalgo-Reyes,Y., Wiedenheft,B., Maxwell,K.L. and Davidson,A.R. (2015) Multiple mechanisms for CRISPR–Cas inhibition by anti-CRISPR proteins. *Nature*, **526**, 136–139.

151. Wang,X., Yao,D., Xu,J.-G., Li,A.-R., Xu,J., Fu,P., Zhou,Y. and Zhu,Y. (2016) Structural basis of Cas3 inhibition by the bacteriophage protein AcrF3. *Nat. Struct. Mol. Biol.*, **23**, 868–870.

152. Xie,Y., Zhang,L., Gao,Z., Yin,P., Wang,H., Li,H., Chen,Z., Zhang,Y., Yang,M. and Feng,Y. (2022) AcrIF5 specifically targets DNA-bound CRISPR-Cas surveillance complex for inhibition. *Nat. Chem. Biol.*, 10.1038/s41589-022-00995-8.

153. Battesti,A. and Bouveret,E. (2012) The bacterial two-hybrid system based on adenylate cyclase reconstitution in *Escherichia coli*. *Methods*, **58**, 325–334.

154. Karimova,G., Pidoux,J., Ullmann,A. and Ladant,D. (1998) A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proceedings of the National Academy of Sciences*, **95**, 5752–5756.

155. Jensen,O.N. (2004) Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr. Opin. Chem. Biol.*, **8**, 33–41.

156. Witze,E.S., Old,W.M., Resing,K.A. and Ahn,N.G. (2007) Mapping protein post-translational modifications with mass spectrometry. *Nature Methods*, **4**, 798–806.

157. Vale,P.F., Lafforgue,G., Gatchitch,F., Gardan,R., Moineau,S. and Gandon,S. (2015) Costs of CRISPR-Cas-mediated resistance in *Streptococcus thermophilus*. *Proc. Biol. Sci.*, **282**, 20151270.

158. Iranzo,J., Lobkovsky,A.E., Wolf,Y.I. and Koonin,E.V. (2013) Evolutionary dynamics of the prokaryotic adaptive immunity system CRISPR-Cas in an explicit ecological context. *J. Bacteriol.*, **195**, 3834–3844.

Acknowledgments/Danksagung

This thesis would not be possible without the help and support of many people. Firstly, I want to thank my thesis committee Prof. Chase Beisel, Prof. Cynthia Sharma, Jun. Prof. Neva Caliskan and Prof. Frederik Börnke for the support in the last four years. Special thanks to Chase for giving me the opportunity to do my PhD in the Beisel lab and for supporting my scientific development by enabling conference visits or joining the Cold Spring Harbor Summer Course.

Furthermore, many thanks to everyone from the RSYN lab. It was an interesting journey seeing the lab grow from only five members to more than 20 in such a short time. Thanks for the many fun conversations and the uncountable amounts of cake I enjoyed the last four years. Special thanks to my co-first author Ioannis for support in the weird and frustrating last breaths of our paper. My biggest thanks obviously go to my wonderful friends and office mates from the office that should be renamed from “office on the second floor” to “main office”. Daphne, Elena, Frank, Kathi, and Sandra (I had to choose the alphabetic order as I cannot decide whom to name first), you made my PhD life fun! Thank you for many happy days and all the support and gossip. You also made sure that the social distancing times were not so lonely. Thank you office hams! I hope we can all stay in contact, and I will obviously visit you guys and fill the snack area that helps to ease the many frustrating days that every PhD journey holds ready.

Ein großer Dank geht auch an meine Würzburger Freunde, dich mich seit dem ersten Tag in meinem Studium begleiten. Wir haben uns alle durch unsere Höhen und Tiefen unterstützt, was unsere Freundschaft in den letzten Jahren nochmal vertieft hat. Ich hoffe wir verlieren uns nie aus den Augen, machen unsere jährlichen Treffen in Würzburg wahr und trinken das eine oder andere Gläschen (oder Fläschchen) Wein. Ausdrücklich zu nennen ist dabei mein Partner Julian, der mich immer unterstützt hat, auch wenn ich vielleicht in manchen gestressten Momenten etwas weniger liebenswert war. Du hast die gleiche anstrengende Zeit parallel durchgemacht und hattest trotzdem immer ein offenes Ohr und tröstende Worte.

Natürlich geht auch ein großer Dank an meine Familie. Vielen Dank an meine Mama Michaela und meinen Papa Ferdinand für die unendliche Unterstützung all die Jahre in meinem Studium und darüber hinaus. Ohne euch wäre ich nicht die Person ich die jetzt bin und hätte weniger erreicht. Vielen Dank auch an meine beiden Brüderlein Florian und Hermann, ohne die mein Leben um einiges langweiliger wäre.

Appendix: Curriculum Vitae