Characterization of Type IV-A CRISPR-Cas systems

Dissertation

zur Erlangung des Grades eines

Doktor der Naturwissenschaften

(Dr. rer.nat.)

des Fachbereichs Biologie der Philipps-Universität Marburg

Vorgelegt von

Xiaohan Guo

aus Xiamen (China)

Marburg, 2023

Die vorliegende Dissertation wurde von August 2017 bis November 2019 am Max-Planck-Institut für terrestrische Mikrobiologie in Marburg und von Dezember 2019 bis Mai 2022 an der Philipps-Universität Marburg unter Leitung von Prof. Dr. Lennart Randau angefertigt.

Vom Fachbereich Biologie der Philipps-Universität Marburg (Hochschulkennziffer 1180) als Dissertation angenommen am _____

Erstgutachter(in): Prof. Dr. Lennart Randau

Zweitgutachter(in): Prof. Dr. Martin Thanbichler

Tag der Disputation: 26. Juni 2023

Erklärung

Ich versichere, dass ich meine Dissertation mit dem Titel „Characterization of Type IV-A CRISPR-Cas systems“ selbstständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfsmittel bedient habe.

Diese Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den 11. April 2023

Xiaohan Guo

The results of the small RNA-Seq analysis and *in vivo* targeting experiments involving Type IV-A CRISPR-Cas systems in this work have been published in:

- **Guo, X**., Sanchez-Londono, M., Gomes-Filho, J. V., Hernandez-Tamayo, R., Rust, S., Immelmann, L. M., Schäfer, P., Wiegel, J., Graumann, P. L., & Randau, L. (2022). Characterization of the self-targeting Type IV CRISPR interference system in *Pseudomonas oleovorans*. *Nature Microbiology* 2022 7:11, 7(11), 1870–1878. doi: 10.1038/s41564-022-01229-2

Table of Contents

Summary 1

Zusammenfassung 3

1.Introduction 1

1.1 Prokaryotes Defense Systems 1

1.1.1 CRISPR-Cas systems 1

1.1.2 Other prokaryotes defense mechanisms. 2

1.2 Classification of CRISPR-Cas systems 4

1.3 Adaptation of CRISPR-Cas systems 6

1.3.1 Naive adaptation 6

1.3.2 Primed adaptation 7

1.3.3 PAM recognition during spacer acquisition 7

1.4 Interference mechanisms of class 1 CRISPR-Cas systems 9

1.4.1 Type I CRISPR-Cas systems 10

1.4.2 Type III CRISPR-Cas systems 12

1.4.3 Type IV CRISPR-Cas systems 14

1.5 *Pseudomonas* Type IV Pili 17

1.6 IncP-9 Plasmid family 19

1.7 CRISPR beyond adaptive immunity 20

2. Results 22

2.1 Bioinformatic analyses 22

2.2 Biogenesis of crRNAs and physiology of *Pseudomonas oleovorans* DSM1045 25

2.2.1 The physiology of *Pseudomonas oleovorans* DSM1045 25

2.2.2 crRNA biogenesis of *Pseudomonas oleovorans* DSM1045 26

2.3 Spacer acquisition in a Type IV-A CRISPR array 29

2.4 Investigation of the biological function of the *Pseudomonas oleovorans* Type IV-A CRISPR-Cas system 34

2.5 Self-targeting activity of recombinant Type IV-A crRNPs 41

3. Discussion 44

3.1 Cross-talk of adaptation in a Type IV-A CRISPR array 44

3.2 Type IV-A CRISPR interference activity 46

3.2.1 DinG 46

3.2.2 Type IV-A CRISPR-Cas mediated immunity against phages 47
3.2.3 Type IV-A mediated CRISPR interference 47
3.3 CRISPR arrays and crRNA biogenesis of *Pseudomonas oleovorans* 51
3.4 Investigation of PAM requirements 51
3.5 Evolutionary advantage of Type IV-A CRISPR-Cas systems 52
3.5.1 Evolutionary advantage of Type IV-A CRISPR-Cas systems for carrier plasmids 52
3.5.2 Evolutionary advantage of Type IV-A CRISPR-Cas systems for bacteria hosts 53
3.6 Applications and outlook 55
4. Material and methods 58
4.1 Materials, instruments and source of supplies 58
4.1.1 Chemicals, Kits and enzymes 58
4.1.2 Instruments 59
4.1.3 Strains and culture conditions 60
4.1.3.1 *Escherichia coli* growth conditions 60
4.1.3.2 *Pseudomonas oleovorans* DSM1045 growth conditions 61
4.1.4 Oligonucleotides, plasmids and constructed recombinant vectors 61
4.2 Working with DNA 64
4.2.1 Quantification of DNA 64
4.2.1.1 Spectrophotometric quantification 64
4.2.1.2 Fluorometric quantification 64
4.2.2 Agarose gel electrophoresis of DNA 64
4.2.3 Purification of DNA 64
4.2.3.1 PCR Purification 64
4.2.3.2 Gel extraction from agarose gels 64
4.2.3.3 Gel extraction from polyacrylamide gels 65
4.2.4 Polymerase chain reaction (PCR) 65
4.2.5 Modification of DNA 66
4.2.5.1 Restriction 66
4.2.5.2 Ligation 66
4.2.5.3 Gibson Assembly 66
4.2.6 Conjugation of *P. oleovorans* 67
4.3 Working with RNA 67

4.3.1 RNA extraction 67
4.3.2 Spectrophotometric quantification of RNA 67
4.3.3 Fluorometric quantification 68
4.3.4 Agarose gel electrophoresis of RNA 68
4.3.5 Illumina RNA Sequencing 68
4.4 *In vivo* assays 68
4.4.1 Fluorescence-activated cell sorting 68
4.4.2 Transformation efficiency assays 69
4.4.3 FACS measurements 69
4.4.4 Bacteriophage plaque assay 69
4.4.6 CRISPRi assays 70
5. Reference 71
6. Supplementary Material 88

Summary

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas (CRISPR associated) systems are present in many prokaryotes and provide adaptive immunity against viruses and other mobile genetic elements. They can recognize new foreign genetic material through adaptation, which involves the integration of short segments of foreign DNA into an extending CRISPR array. This process allows the system to evolve continually and to provide a robust defense against a wide range of invaders. Depending on the involved Cas proteins, CRISPR-Cas systems are classified into different types and two major classes. Type I systems use a ribonucleoprotein complex termed Cascade (CRISPR-associated complex for antiviral defense) to scan for invading DNA and bind to the sequence that matches its crRNA (CRISPR RNA). Once the target is identified, the nuclease Cas3 is recruited and degrades the DNA in a process known as interference.

The Type IV CRISPR-Cas system is a member of Class 1 and has three subtypes. This study focuses specifically on the Type IV-A system, which is characterized by the absence of a DNA nuclease, the presence of a CRISPR array, the lack of an adaptation module, and an association with the helicase DinG. To examine the biological function of the Type IV-A CRISPR-Cas system, *Pseudomonas oleovorans* was selected as a model organism and found to contain a Type IV-A CRISPR-Cas system on its megaplasmid. Bioinformatic analyses of its CRISPR array showed that one of the crRNAs targets the gene *pilN* in the genome, providing evidence of adaptation of host sequences.

A 5'-AAG-3' PAM sequence was found to be shared between the Type IV-A and Type I-E CRISPR-Cas systems, suggesting the possibility of cross-talk at the level of adaptation. Deep sequencing of genomic DNA of *P. oleovorans* cells electroporated with pre-spacers revealed spacer rearrangements and the presence of new spacers. *In vivo* assays in *P. oleovorans* and *E. coli* recombinant systems demonstrated that the Type IV-A system conducts PAM-dependent interference. This system exhibits anti-plasmid activity by targeting an open reading frame within one of the multiple cloning sites. Furthermore, it displays resistance to lambda phage infection by targeting gene *E*, which encodes a crucial head protein. Observation of efficient *lacZ* reporter targeting provided the first evidence that the Type IV-A

system conducts interference without DNA degradation. Therefore, the Type IV-A CRISPR-Cas system holds great promise as a powerful transcription modulation tool with a natural CRISPRi-like mechanism.

Zusammenfassung

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas (CRISPR-assoziierte) Systeme sind in vielen Prokaryoten vorhanden und sorgen für eine adaptive Immunität gegen Viren und andere mobile genetische Elemente. Sie können neues fremdes genetisches Material durch Anpassung erkennen, was die Integration kurzer Abschnitte fremder DNA in ein sich erweiterndes CRISPR-Array beinhaltet. Dieser Prozess ermöglicht es dem System, sich ständig weiterzuentwickeln und eine robuste Verteidigung gegen eine Vielzahl von Eindringlingen zu bieten. Je nach den beteiligten Cas-Proteinen werden die CRISPR-Cas-Systeme in verschiedene Typen und zwei Hauptklassen eingeteilt. Systeme des Typs I verwenden einen Ribonukleoprotein-Komplex namens Cascade (CRISPR-assoziierten Komplex für die antivirale Abwehr), um nach eindringender DNA zu suchen und an die Sequenz zu binden, die zu ihrer crRNA (CRISPR-RNA) passt. Sobald das Ziel identifiziert ist, wird die Nuklease Cas3 rekrutiert und baut die DNA in einem Prozess ab, der als Interferenz bekannt ist.

Das CRISPR-Cas-System vom Typ IV gehört zur Klasse 1 und hat drei Untertypen. Diese Studie konzentriert sich speziell auf das Typ IV-A-System, das durch das Fehlen einer DNA-Nuklease, das Vorhandensein eines CRISPR-Arrays, das Fehlen eines Anpassungsmoduls und die Assoziation mit der Helikase DinG gekennzeichnet ist. Um die biologische Funktion des Typ IV-A CRISPR-Cas-Systems zu untersuchen, wurde *Pseudomonas oleovorans* als Modellorganismus ausgewählt und festgestellt, dass es ein einzelnes Typ IV-A CRISPR-Cas-System auf seinem Megaplasmid enthält. Bioinformatische Analysen seines CRISPR-Arrays zeigten, dass eine der crRNAs auf das Gen *pilN* im Genom abzielt, was auf eine Anpassung der Wirtssequenzen hinweist.

Eine 5’-AAG-3’-PAM-Sequenz wurde zwischen den CRISPR-Cas-Systemen des Typs IV-A und des Typs I-E gefunden, was auf die Möglichkeit von Überschneidungen auf der Ebene der Anpassung hindeutet. Die Tiefensequenzierung der genomischen DNA von P. oleovorans-Zellen, die mit Pre-Spacern elektroporiert wurden, ergab Spacer-Umlagerungen und das Vorhandensein neuer Spacer. *In vivo*-Assays in rekombinanten *P. oleovorans*- und *E. coli*-Systemen zeigten, dass das Typ IV-A-System eine PAM-abhängige Interferenz durchführt. Dieses System weist eine Anti-Plasmid-Aktivität auf, indem es auf einen offenen Leserahmen

innerhalb einer der zahlreichen Klonierungsstellen abzielt. Darüber hinaus zeigt es Resistenz gegen Lambda-Phageninfektionen, indem es auf das Gen *E* abzielt, das für ein wichtiges Kopfprotein kodiert.

Die Beobachtung eines effizienten *lacZ*-Reporter-Targetings lieferte den ersten Beweis dafür, dass das Typ IV-A-System Interferenzen ohne DNA-Abbau durchführt. Daher ist das Typ IV-A CRISPR-Cas-System ein vielversprechendes Werkzeug zur Modulation der Transkription mit einem natürlichen CRISPR-ähnlichen Mechanismus.

1.Introduction

1.1 Prokaryotes Defense Systems

Bacteriophages are a constant threat to bacteria. The arms race between bacteriophages and bacteria has resulted in the evolution of diverse anti-phage mechanisms (Azam et al., 2019; Chaudhary, 2017; Chen et al., 2016; Golais et al., 2012). These strategies are effective at different stages of phage infection. Currently, six major categories of phage-defense mechanisms are categorized in bacteria: surface modification, superinfection exclusion, restriction modification, CRISPR, pAgo, and abortive infection (Athukoralage et al., 2022; Castillo et al., 2019; Houte et al., 2016).

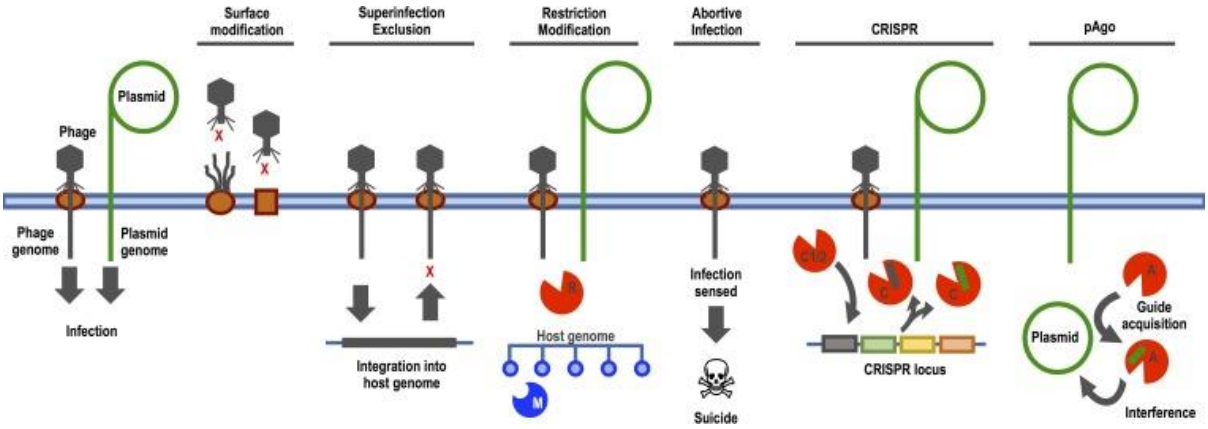


Figure 1. Overview of bacterial anti-phage strategies. Six primary defense mechanisms are described: i) surface modification ii) superinfection exclusion, iii) restriction modification, iv) CRISPR, v) pAgo, vi) abortive infection. Modified from (Houte et al., 2016). The immune mechanism involves protein components that are represented by letters, M for methylase, R for restriction enzyme, C1/2 for Cas1 and Cas2, C for Cas effector-nuclease complex, and A for prokaryotic Argonaute enzyme.

1.1.1 CRISPR-Cas systems

The general architecture of CRISPR arrays with short interspaced repeated DNA segments was first described in 1987 (Ishino et al., 1987). CRISPR-Cas systems were found to be present in 90% of archaea and 40% of known bacterial genomes, and were later determined to be adaptive immune systems in archaea and bacteria (Horvath et al., 2010). This immune system

uses short CRISPR RNAs (crRNAs) that comprise sequences from mobile genetic elements (MGEs) as scaffolds for the assembly of effector complexes and the recognition of targets (Carte et al., 2014). Spacers are stored in an array between short repeats, and there is usually a cluster of *cas* genes next to the array (Carte et al., 2014).

In general, there are three stages of CRISPR-mediated immunity (Figure 3): during the host cell invasion by MGEs, DNA fragments of their genomes are captured by adaptation modules, which are usually formed by the proteins Cas1 and Cas2 (Hu et al., 2021; Savitskaya et al., 2013). These DNA fragments are trimmed and inserted at the beginning of the CRISPR array. In the second stage, the CRISPR array is transcribed as a long precursor crRNA, which endonucleases like Cas6 or RNaseIII then process into mature crRNAs (Charpentier et al., 2015; Hille et al., 2016). The CRISPR ribonucleoprotein complex (crRNP) is an effector complex composed of a single mature crRNA or a diverse set of Cas proteins. These complexes utilize the basepairing potential of its crRNAs to target foreign nucleic acids, effectively conducting interference (Brouns et al., 2008; Louwen et al., 2014). Here, a DNA nuclease is usually recruited to the target to eliminate foreign invader DNA. In addition to the crRNA spacer match with its target protospacer sequence, a short 2-5 bp protospacer adjacent motif (PAM) is required for DNA target selection (Leenay et al., 2016; Westra et al., 2013a; Yoganand et al., 2019). Type III CRISPR-Cas systems can use this motif also for distinguishing self from non-self DNA targets (Estrella et al., 2016; Staals et al., 2014). The PAM is consistent for adaptation and interference to ensure targeting fidelity (Nuñez et al., 2014).

1.1.2 Other prokaryotes defense mechanisms.

Phages use attachment sites on the surface of their bacterial hosts for recognition and entry (Zheng et al., 2004). These attachment sites are often motility organelles, such as the flagelli or pili (Icho et al., 1978; Samuel et al., 1999). Bacteria can mutate, lose, or hide these attachment sites to prevent phages from entering. However, these modifications can come at a cost, such as a reduction in nutrition uptake, which is essential for bacteria in resource-limited environments (Scanlan et al., 2012).

In addition to surface modification, bacteria can also use superinfection exclusion (Sie) mechanisms to prevent phage entry or replication. Sie is a mechanism primarily encoded by

prophages and blocks the entry or replication of phage DNA (Pope et al., 2011; Vostrov et al., 1996). One well-known Sie system, the *cor* gene, inactivates the ferrichrome uptake protein FhuA to prevent phage DNA injection (Uc-Mass et al., 2004). In addition to interfering with phage DNA injection, prophages also encode repressor proteins that maintain the lysogenic life cycle by binding to specific regions of prophage DNA in the host genome. This mechanism confers immunity to phages with similar target sequences.

The most common immune mechanism of bacteria is the restriction-modification (RM) system. These systems consist of two modules: a methyltransferase (MT) and a restriction endonuclease (RE) (Roberts et al., 2003). Self-DNA is protected by MT-catalyzed methylation, while unmodified non-self-DNA is cleaved by RE.

Argonaute (Ago) proteins constitute a critical component of RNA interference (RNAi) pathways, serving as key regulators of gene expression at the post-transcriptional stage (Bartel, 2004; Ketting, 2011). In prokaryotes, pAgo proteins have been demonstrated to execute DNA-guided DNA cleavage (Makarova et al., 2009). Nonetheless, the underlying mechanisms and functions of pAgo proteins remain incompletely elucidated, necessitating additional investigation.

Abortive infection (Abi) systems provide another defense mechanism, inducing programmed cell death to prevent further phage spread. Abi systems are mechanisms that cause programmed cell death upon detecting phage infection, thereby preventing the further spread of phages (Fineran et al., 2009). The Rex system is a well-studied Abi system identified in phage λ-lysogenic *Escherichia coli* (*E. coli*) strains (Bingham et al., 2000). This system contains two proteins, RexA and RexB. RexA is activated during the replication of phage DNA, which leads to the activation of RexB. The activation of RexB decreases the cellular ATP level, thereby aborting ATP-dependent processes (Chopin et al., 2005).

1.2 Classification of CRISPR-Cas systems

In the classification of CRISPR-Cas systems, multiple factors such as genomic loci organization, sequence similarity-based clustering, phylogenetic analysis of conserved genes, and even experimental data are taken into consideration because of the absence of a shared gene among all systems (Figure 2) (Makarova et al., 2019).

To this date, CRISPR-Cas systems have been divided into two classes and six types (Koonin et al., 2019). The main difference between class 1 and class 2 systems is the composition of the effector complexes. Class 2 systems encode a single effector protein for interference, while class 1 systems employ multi-subunit effector complexes. Within class 1 there are three types: Type I, III, and IV. Type I systems use a crRNP termed CRISPR-associated complex for antiviral defense (Cascade) for interference, and the protein Cas3 is responsible for target DNA degradation (Barrangou et al., 2007; Makarova et al., 2015).

Type III CRISPR-Cas systems are capable interfere with both DNA and RNA. The interference of Type III systems on RNA is PAM-independent and does not require strict complementarity between the crRNA and the RNA target (Liu et al., 2017; Tamulaitis et al., 2017). The interference mechanisms of the different class I effector complexes will be detailed in section 1.4.

Studies have identified three subtypes of Type IV CRISPR-Cas systems and have revealed their formation of effector complexes and biological functions (Makarova et al., 2015, 2019). It has been shown that Type IV-A systems form effector complexes similar to those of Type I systems, as demonstrated through recombinant expression in *E. coli* and observed through Transmission electron microscopy images (Özcan et al., 2019). Another study has also reported the anti-plasmid activity of the Type IV-A system through recombinant expression in *E. coli* (Crowley et al., 2019), further supporting similarities between Type IV-A and Type I systems. In this study, we are focusing on the biological function of the Type IV-A CRISPR-Cas system in *Pseudomonas oleovorans* (*P. oleovorans*).

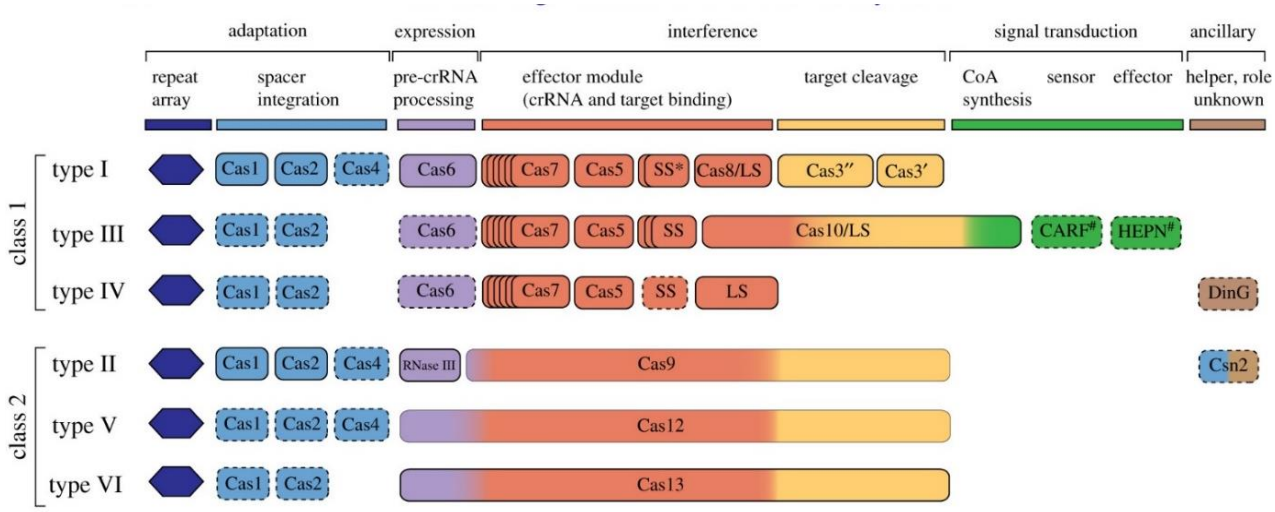


Figure 2. Schematic overview of the classification of CRISPR-Cas systems. Adaptation proteins Cas1 and Cas2 are conserved in all systems with the exception of Type IV systems. In class 1 systems, multiple Cas proteins assemble with matured crRNA as an effector complex. In class 2 systems, a single effector complex is responsible for interference. Modified from (Makarova et al., 2019)

1.3 Adaptation of CRISPR-Cas systems

CRISPR spacer acquisition is a process integrating new DNA sequences from invading MGEs into CRISPR loci (Barrangou et al., 2007; Yosef et al., 2012). The Cas1 and Cas2 proteins are essential for integrating new spacers, as they recognize and bind to the protospacer and facilitate its integration into the CRISPR locus (Erdmann et al., 2012).

1.3.1 Naive adaptation

Naive adaptation describes the process of CRISPR-Cas systems acquiring new spacers from invading viruses or plasmids that they have not previously encountered (Barrangou et al., 2007). This process requires the coordinated activity of multiple Cas proteins, including Cas1 and Cas2, which are both necessary for spacer acquisition (Yosef et al., 2012). In the *E. coli* Type I-E system, it was shown that a single repeat is both necessary and sufficient for naive adaptation. The inserted repeat is identical to the leader-proximal repeat, indicating that this repeat is copied during spacer adaptation (Yosef et al., 2012). The minimal required length of the leader is 40 to 43 bp upstream of the first repeat of the array (Díez-Villaseñor et al., 2013).

During naive adaptation, the Cas1 and Cas2 proteins recognize and bind to the invading DNA, which is then cleaved into short fragments (Nuñez et al., 2016). These fragments or pre-spacers are then integrated into the CRISPR locus by the Cas1-Cas2 complex. Cas4 is an additional protein required in some CRISPR-Cas systems, playing a crucial role in adaptation efficiency. It is found in Type I, II, and V systems and is involved in processing pre-spacers in a PAM-dependent manner (Hudaiberdiev et al., 2017). Cas4 collaborates with Cas1 and Cas2 to form an effector complex Cas4-Cas1-Cas2, where Cas4 specifically cleaves PAM-containing overhangs (Lee et al., 2019). The activity of Cas4 is dependent on the presence of Cas1 and Cas2, which serves to prevent premature integration of unprocessed pre-spacers (Dhingra et al., 2022; Lee et al., 2018).

Naive adaptation involves using double-stranded DNA (dsDNA) as the substrate for spacer acquisition, rather than single-stranded DNA (ssDNA) (Nuñez et al., 2015). The 3’-OH ends of the protospacer are essential to make a nucleophilic attack on one strand of the repeat. Cas1

and Cas2 integrate a spacer with the correct PAM orientation by preferentially using the 3'-OH of the C nucleotide complementary to the G of the AAG PAM (Nuñez et al., 2015).

1.3.2 Primed adaptation

Primed adaptation is a specialized form of CRISPR spacer acquisition that occurs only in the presence of specific Cas proteins and a "priming" spacer targeting an existing protospacer (Datsenko et al., 2012). Priming was revealed to enhance spacer acquisition by 10- to 20-fold over naive adaptation in the Type I-E system (Datsenko et al., 2012; Savitskaya et al., 2013).

The priming process likely evolved to minimize infection by phage escape mutants that would otherwise evade the interference machinery. The efficiency of priming is significantly increased when the priming spacer has mutations in the seed sequence or if the protospacer has a non-cognate PAM, indicating the high specificity and adaptiveness of priming (Datsenko et al., 2012).

Primed adaptation is biased to the strand orientation of the protospacer targeted by the priming spacer in the Type I-E system. Early experiments exclusively resulted in spacers acquired in the same orientation as the first spacer, which was later confirmed in a controlled experiment that monitored primed adaptation from plasmids harboring protospacers in either the forward or reverse orientation (Swarts et al., 2012). Increased spacer acquisition in one strand of the plasmid corroborated the orientation of the protospacer, indicating that acquisition is facilitated from a primed strand (Datsenko et al., 2012).

1.3.3 PAM recognition during spacer acquisition

During spacer acquisition in CRISPR systems, a PAM is necessary for the Cas proteins to identify foreign DNA. However, the PAM sequence must be removed before spacer integration into the CRISPR array to prevent self-targeting. To ensure the absence of PAM sequences in the pre-spacer, CRISPR-Cas systems use different mechanisms to check for its presence.

In the *E. coli* Type I-E system, Cas1 is responsible for PAM recognition (J. Wang et al., 2015). The 3' overhang of the pre-spacer, containing the PAM-complementary 5'-CTT-3' sequence,

is positioned into the PAM recognition pocket of the catalytic Cas1 subunit (J. Wang et al., 2015). Meanwhile, the long C-terminal tail of the non-catalytic subunit acts as the pocket lid to ensure proper binding (J. Wang et al., 2015). Other subtype I-E systems also contain Cas1 proteins with extended C-terminal tails, indicating a similar PAM readout mechanism (Yoganand et al., 2019). In contrast, Cas1 proteins of I-A, B, and C systems lack these tails and instead use an auxiliary Cas4 subunit for PAM readout (Yoganand et al., 2019).

Upon recognition of the PAM sequence, the Cas1 tail undergoes a conformational change that may promote the docking of the PAM-proximal pre-spacer end to the Repeat1-Spacer1 junction, and the PAM-distal end to the Leader-Repeat1 junction (J. Wang et al., 2015).

1.4 Interference mechanisms of class 1 CRISPR-Cas systems

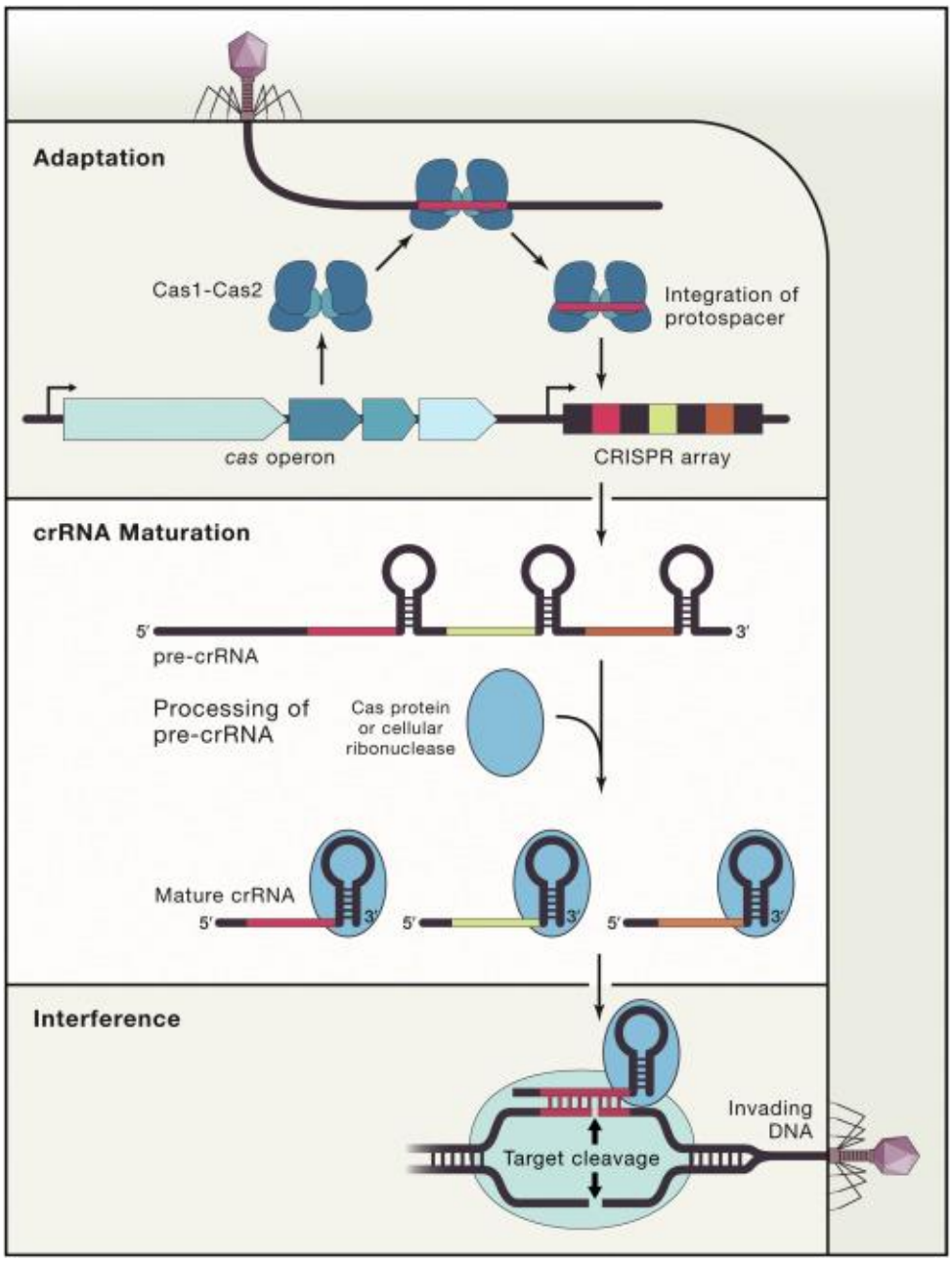


Figure 3. A three-stage model describes CRISPR immunity. Adaptation occurs when bacteria encounter invading phages and their genomic DNA fragments are captured and inserted into the CRISPR array. During the expression stage, CRISPR arrays are first transcribed into long pre-crRNAs then processed by Cas6 or RNase into matured crRNAs. Next, Cas proteins are assembled into a matured crRNA as the effector complex for the interference. Modified from (Hille et al., 2018).

1.4.1 Type I CRISPR-Cas systems

Type I CRISPR-Cas systems are the most abundant in prokaryotes (Koonin et al., 2017a; Makarova et al., 2015). The interference of Type I systems involves a multi-subunit complex and the Cas3 nuclease (Brouns et al., 2008). Although the architecture of Cascade is conserved, the composition of *cas* genes can vary between different subtypes, and the homology of the subunits is normally based on functional similarities rather than sequence similarities (Figure 4) (Koonin et al., 2017a; Makarova et al., 2015). Among the seven subtypes of Type I systems, the Type I-E system found in *E. coli* is the most well studied and often taken as a model for understanding Type I interference (Makarova et al., 2015).

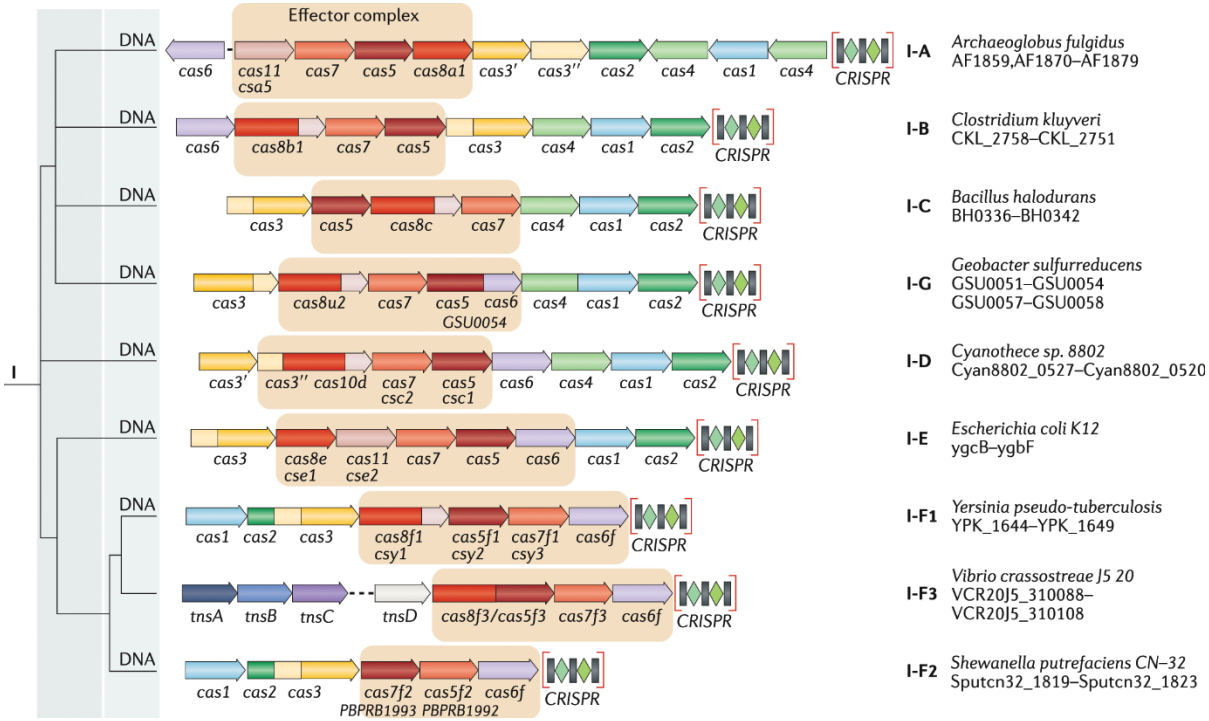


Figure 4. Classification of Type I CRISPR-Cas systems. The diagram shows the *cas* genes encode proteins that form the effector complex, such as *cas6*, *cas7*, *cas5*, and genes encoding small (SS) and large subunits (LS). Cas1 and Cas2, which are adaptation proteins, are conserved in all CRISPR-Cas systems. The signature protein Cas3 is unique to Type I systems and is an additional nuclease recruited for target degradation. In some subtypes, the helicase and nuclease domains of *cas3* are split into multiple genes and fused to other cas genes. This Figure was adapted from (Makarova et al., 2019).

In Type I-E, Cascade comprises Cas5e, Cas6e, Cas7e, Cas8e, and Cas11e subunits (Jackson et al., 2014; Spilman et al., 2013; Xiao, Luo, et al., 2017). Pre- crRNA is processed by Cas6e into mature crRNAs and bound along the backbone of Cascade. The small and large subunits of Cascade, Cas11e and Cas8e, respectively, play critical roles in the formation of the complex and target recognition (Bozic et al., 2019; Savitskaya et al., 2013). The Cascade in Type I-E CRISPR-Cas system is composed of six Cas7e proteins that form a helical backbone structure. The thumb domains of the Cas7e subunits are responsible for the tight connection of the subunits and kink the crRNA at every sixth nucleotide, while the palm domain aligns the five nucleotides in between to enable efficient base pairing with the target DNA (Li et al., 2020a; Mulepati et al., 2014; Xiao, Luo, et al., 2017). Cas11e and Cas8e are the small and large subunits of Cascade, respectively (Cass et al., 2015). The two Cas11e subunits form the "belly" of the complex by interacting directly with the Cas7e backbone, while Cas8e interacts with Cas5e, Cas7e, and Cas11e and forms the "tail". These interactions between subunits and the crRNA are crucial for target recognition and cleavage by the Type I-E system (Bozic et al., 2019; Díez-Villaseñor et al., 2013; Westra et al., 2013a; Yoganand et al., 2019).

During Type I-E interference, recognition of the PAM in the double-stranded target DNA is facilitated by the large subunit of Cascade. The large subunit also initiates the local unwinding of DNA and the binding of crRNA to the complementary DNA strand of the protospacer (Niewoehner et al., 2016; Sashital et al., 2012). The seed sequence, which consists of the first eight PAM-proximal nucleotides of the crRNA, is crucial for protospacer binding of the Cascade complex, except for the sixth nucleotide which does not bind to the target. The non-target strand of the target DNA is bound by two Cas11e subunits, leading to the formation of the R-loop structure. The R-loop structure induces conformational changes in the small and large subunits of Cascade, which enables the recruitment of the nuclease Cas3 for target cleavage (Hochstrasser et al., 2016; Mulepati et al., 2014; Sinkunas et al., 2013). The HD domain of Cas3 nicks the displaced non-target DNA strand and activates its ATP-dependent helicase activity. Then Cas3 degrades the non-target DNA strand in the 3’ to 5’ direction, leaving a 200-300 nucleotides single-stranded DNA gap in the target genome. The complete degradation of the target DNA may be mediated by other host nucleases or the ssDNA

nuclease activity of Cas3 that has been observed *in vitro* (Jore et al., 2011; Sashital et al., 2012).

Cascade, the effector complex, is conserved in structure, but the presence or absence of certain subunits differs between subtypes. The subtypes I-A and I-E have a separate gene for the small subunit, while other subtypes either fuse or replace the small subunit with Cas8. Type I-C lacks a Cas6 homolog, and Type I-Fv has a different replacement for the large and small subunits (Koonin et al., 2017a; Makarova et al., 2015). Type I-F has an interesting variation in the shape of Cascade, where the backbone of the surveillance complex forms an almost closed ring but subsequently "unwinds" upon target DNA recognition. The nuclease Cas3 is the signature protein of Type I systems, but its gene can undergo fusion or fission in different subtypes. The seed sequence is crucial for Type I-E and I-F interference and is likely another common feature among Type I systems (Mulepati et al., 2014; Sinkunas et al., 2013). The type I-G effector displays a pronounced curvature of the crRNA, similar to type I-F systems. Cas8g, the large subunit, is a divergent member of the Cas8 family, with an N-terminal domain distant from the 5′ end of the crRNA (Shangguan et al., 2022). Although high-resolution structures and detailed insight into target recognition and cleavage are still not available for subtypes I-A and I-D, it is likely that they follow the similar mechanisms in other Type I systems (Koonin et al., 2017a, 2014; Makarova et al., 2009).

1.4.2 Type III CRISPR-Cas systems

The effector complexes of Type III CRISPR-Cas systems are similar to the complexes Type I systems, depend on the subtypes these complexes are called Csm (III-A) and Cmr (III-B) (Hochstrasser et al., 2016; Jackson et al., 2014; Makarova et al., 2011; Mulepati et al., 2014). Unlike other CRISPR systems, Type III systems target both RNA and DNA (Deng et al., 2013; Elmore et al., 2016; Estrella et al., 2016), and DNA cleavage only occurs when the target sequence is transcribed.

Mature crRNA is bound by Cas5 at the 5′ repeat handle, it is assembled with a backbone of multiple Cas7, a small and a large subunit as the effector complex of Type III systems (Figure 6). The interference initiated by the complementary of crRNA and nascent RNA, RNA cleavage is conducted by Cas7 and DNA cleavage is carried out by the palm domain of Cas10.

Noteworthy, DNA cleavage only happens inside the transcription bubble (Liu et al., 2017; Pyenson & Marraffini, 2017).

The Csm and Cmr complexes are similar to Cascade and assemble along the mature crRNA bound by Cas5 at the 5′ repeat handle. The backbone of the complex is made up of Cas7-family proteins while Cas11 and Cas10 are the small and large subunits, respectively (Mulepati et al., 2014; Staals et al., 2014). Target cleavage starts with the binding of the Type III effector complex to the target transcript in a crRNA-dependent manner. The Cas7 subunits cut the ssRNA at every sixth nucleotide, and DNA cleavage is done by the palm domain of the Cas10 subunit, which requires transcription of the target (Osawa et al., 2015; Samai et al., 2015; D. W. Taylor et al., 2015).

Type III CRISPR-Cas systems frequently have RNases from the Csm6 or related Csx1 families (Koonin et al., 2017b; Makarova et al., 2011). Both Csm6 and Csx1 degrade foreign transcripts non-specifically and play auxiliary or essential roles in Type III interference, even though they are not part of the effector complex (Deng et al., 2013; Niewoehner et al., 2016; Sheppard et al., 2016). It has been discovered that the Cas10 subunit of the Csm complex cleaves target DNA and converts ATP into cyclic adenylates, which act as second messengers to activate the Csm6 RNase. The production of the messenger by Cas10 depends on the binding of the Csm complex to the target RNA and constitutes a regulatory mechanism that triggers robust interference in the presence of an invader (Kazlauskiene et al., 2017; Niewoehner et al., 2017).

In most Type III systems, the 5′ repeat portion of the crRNA binding to the target serves as a mechanism to inhibit DNA cleavage and distinguish self from non-self (Marraffini et al., 2010). However, the *Pyrococcus furiosus* Type III-B system requires a PAM sequence, known as the RNA-PAM (rPAM), located 3′ of the crRNA-complementary sequence on the target RNA, for DNA cleavage by the Cmr complex (Elmore et al., 2016). A study on the Type III-A system of *S. epidermidis* found no evidence of the necessity of a PAM or rPAM, suggesting subtype or species-related differences in self-versus non-self discrimination in Type III systems (Pyenson et al., 2017).

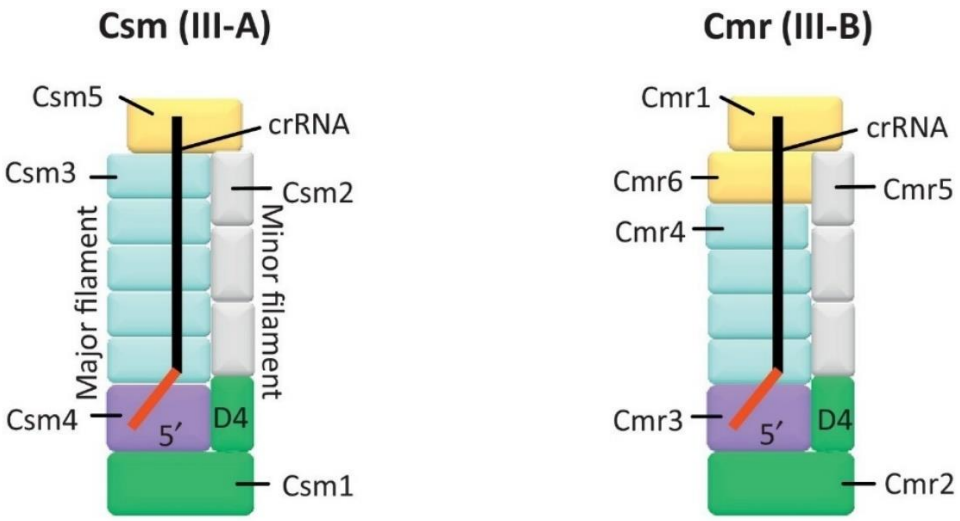


Figure 5. Illustration of the structural arrangement of the Type III complexes, Csm (Type III-A) and Cmr (Type III-B). The schematic architecture of the complexes is shown, with homologous subunits indicated by the same color. The 5´-handle of the crRNA in both complexes is highlighted in red. Adapted from (Tamulaitis et al., 2017).

1.4.3 Type IV CRISPR-Cas systems

Following the identification of Type IV CRISPR-Cas systems, their mechanism and biological role remain to be described in detail. Unlike other class 1 systems, Cas 1, Cas2 and a nuclease conduct target cleavage are absent from Type IV systems (Makarova et al., 2019). Clear evidence of Type IV CRISPR-Cas system mediating interference is still missing.

To date, three subtypes of Type IV CRISPR-Cas systems have been characterized. One of the differences between Type IV-A and Type IV-B is the presence of a CRISPR array (Makarova et al., 2019). Most Type IV-A systems are associated with a CRISPR array, which is missing in Type IV-B systems. DinG (damage-inducible helicase G) is normally associated with Type IV-A systems, and it contains a helicase domain that could be relevant to the biological function of Type IV-A systems. On the other hand, *cysH* is an associated gene specifically to Type IV-B, as *cysH* belongs to the phosphoadenosine phosphosulfate reductase family, the Type IV-B system could be involved in epigenetic silencing (Makarova et al., 2019). Different from Type IV-A and Type IV-B systems, an HD domain is identified in Cas10 of Type IV-C systems. This might hint that Type IV-C systems are capable of target cleavage.

A recent phylogenetic study from Moya-Beltrán et al. proposed that Type IV systems might have evolved from Type III CRISPR-Cas systems (Moya-Beltrán et al., 2021), which might be the result of the co-evolution with plasmids. The investigation by Pinilla-Redondo et al. reveals that Type IV systems are frequently identified in mobile genetic elements (MGEs), with Type IV-A variants predominantly observed in plasmid-like conjugative elements. Additionally, the study highlights a preferential targeting of conjugative plasmids by Type IV systems, potentially influencing the dynamics and distribution of plasmid populations within bacterial community structures (Pinilla-Redondo et al., 2020).

So far, there is still no clear image of the biological function of Type IV systems. In 2018, Özcan et al showed the first evidence of Type IV-A systems as functioning systems. In this study, the model organism *A. aromaticum* EbN1 contains a Type I-C CRISPR-Cas system on its genome and a Type IV-A system on one of its megaplasmids. RNA sequencing revealed crRNA maturation details for both systems, with the Type IV system generating an unusual 7nt 5' -terminal repeat tag. The protein Csf5 was found to be responsible for Type IV crRNA maturation, functioning as a crRNA endonuclease. The study also observed the formation of a Type IV crRNP complex with all four Cas proteins and crRNA, where Csf2 acts as a backbone. Protein-protein and protein-RNA interaction sites were identified within the Type IV crRNP complex, further supporting the formation of a Cascade-like complex guided by a specific Type IV-associated crRNA component (Figure 7) (Özcan et al., 2019). Later, Crowley et al. showed the interference activity of a *Pseudomonas* Type IV-A system (Crowley et al., 2019). According to the study, the associated protein DinG is found to be essential to the interference.

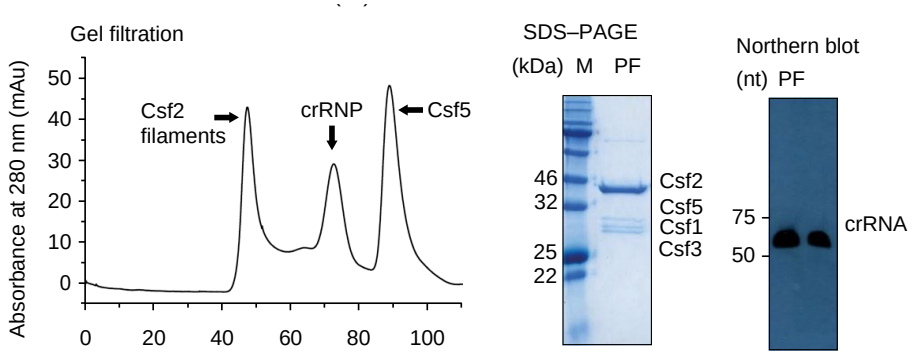


Figure 6. The production of *A. aromaticum* Type IV Cas proteins in *E. coli*. Co-purified RNAs were detected using urea-PAGE and ethidium bromide staining. A crRNP complex was purified and observed under TEM, where crescent-shaped complexes were detected (indicated by arrows) with a scale bar of 50 nm. Mature crRNAs were identified through Northern blot analyses. Adapted from (Özcan et al., 2019).

1.5 *Pseudomonas* Type IV Pili

In this thesis, we have demonstrated the existence of the Type IV-A CRISPR-Cas system in seven *Pseudomonas* strains. This finding could suggest an evolutionary link between the Type IV-A systems and their *Pseudomonas* host organisms. Interestingly, one of the spacers from the Type IV-A system under investigation is base-paired to the sequence of *pilN*, an essential gene of Type IV pili (T4P).

There are two subtypes in T4P family, Type IVa pili (T4aP) and Type IVb pili (T4bP). The distinction between them is based on differences in pilin sequences, leader peptide length, and assembly mechanisms (Ayers et al., 2010; Craig et al., 2004). Moreover, the organization of genes varies between T4aP and T4bP systems. T4aP genes are scattered throughout the genome in multiple operons, while T4bP genes are found in a single operon (Pelicic, 2008). T4bP plays a crucial role in biofilm formation, bacterial colonization, and cell adhesion (Roux et al., 2012), whereas T4aP serves as an important structural component for bacterial twitching motility and DNA uptake (Piepenbrink, 2019).

In addition, T4P serves as the attachment site for T4P-dependent phages. It has been reported that phages regulate T4P to prevent superinfection with other phages. For instance, the temperate *P. aeruginosa* phage D3112 produces a protein that binds to PilB, an assembly and extension ATPase of T4P (Chung et al., 2014). This prevents pili formation and makes the bacterium resistant to T4P-dependent phages. Other prophages in *P. aeruginosa* were found to avoid superinfection through additional mechanisms, including some that modify the receptor on T4P (Bondy-Denomy et al., 2016).

Furthermore, the Type IV secretion system (T4SS), which includes T4P as a component, plays a crucial role in horizontal gene transfer. Mechanisms including natural transformation (Krüger et al., 2011; Stingl et al., 2010), transduction (Scharn et al., 2013; Shen et al., 2013), transport through intercellular 'nanotubes' (Dubey et al., 2011), and conjugation are involved in horizontal gene transfer. During conjugation, several proteins are assembled at the origin of transfer (oriT) of the plasmid to form the relaxosome responsible for DNA processing. In the process the relaxase-mediated nicking and unwinding of the double-stranded plasmid DNA. The resulting single-stranded DNA, with the relaxase attached at oriT, is then recognized

by the coupling protein and transported through the T4SS translocation channel (Arutyunov et al., 2013; de La Cruz et al., 2010; Juhas et al., 2009; Zechner et al., 2012).

1.6 IncP-9 Plasmid family

Type IV CRISPR-Cas systems are commonly found in plasmids and display a preference for targeting plasmid-like elements, suggesting a potential involvement in plasmid competition (Pinilla-Redondo et al., 2020). The Type IV-A system of our study is located in an IncP-9 plasmid (Sevastsyanovich et al., 2008), which is part of a family of catabolic plasmids. These plasmids are relatively large, with sizes ranging from 50 to 500 kilobase pairs, and typically have a low copy number (Dennis, 2005).

In catabolic plasmids, the backbone comprises modularized gene loci that have been minimized to alleviate the metabolic burden on the host (Thomas, 2000). Generally, these plasmids have evolved to reduce their copy number within the host, thereby lowering the energy required for replication (Dennis, 2005).

A portion of the plasmid containing a transposable element indicates a spot where the backbone can be disrupted without affecting the plasmid's normal function (Dennis, 2005; Thomas, 2000). Transposons carrying catabolic pathway genes frequently integrate into the plasmid backbone at these variable regions, which are common integration sites. In many instances, several mobile genetic elements or parts of these elements are found clustered in these areas (Sevastsyanovich et al., 2008).

1.7 CRISPR beyond adaptive immunity

An increasing amount of research is focusing on the benefit–cost ratio of CRISPR-Cas systems. It has been demonstrated that CRISPR-Cas systems can regulate endogenous genes essential to the virulence and pathogenesis of *Francisella novicida* (Sampson et al., 2013).

In addition, CRISPR-Cas systems are reported to respond to stress (Devi et al., 2022). This was first discovered in *E. coli*, where a membrane targeting protein overexpression activated the CRISPR-Cas system (Perez-Rodriguez et al., 2011). CRISPR-Cas systems also respond to stress invoked by phage envelope or induced by environmental factors such as temperature, oxidative stress, and c-di-AMP (Quax et al., 2013; Viswanathan et al., 2007; Young et al., 2012).

Spacers targeting the host chromosome in the CRISPR array may play a role in gene regulation. Computational analysis of the Type I-E CRISPR-Cas system in *E. coli* revealed that CRISPR spacers have a high propensity to target chromosomes, particularly transcriptionally active regions, suggesting that the system may play a role in endogenous gene regulation (Bozic et al., 2019). Experimental evidence has also shown that CRISPR arrays are involved in gene regulation in various organisms, such as *Francisella novicida* and *Pelobacter carbinolicus* (C. L. Jones et al., 2012; Sampson et al., 2013). The Type III system, which targets mRNA instead of DNA, has also been observed to regulate gene expression (Endo et al., 2014; Hale et al., 2012).

In addition, CRISPR-Cas systems and DNA repair machinery are closely associated at various levels (Devi et al., 2022). These two systems typically coincide during the adaptation step, where nucleases generate a nick in the first direct repeat to incorporate a new spacer, which is then sealed by a DNA repair system. The DNA polymerase from the DNA repair pathway is utilized in the spacer acquisition process of the Type I-E system in *E. coli* (Killelea et al., 2017; Liu et al., 2015). Additionally, Cas1 deletion in *E. coli* resulted in defective mutants in the DNA repair pathway, providing experimental evidence of their interdependence (Smith, 2012).

In this thesis, we endeavor to elucidate the biological role of Type IV-A CRISPR systems. The presence of a CRISPR array without the canonical Cas1 and Cas2 proteins implies the existence of an alternative mechanism for spacer acquisition. Furthermore, the lack of an apparent

nuclease domain responsible for target degradation suggests that these systems may serve a function distinct from adaptive immunity. Notably, Type IV-A CRISPR systems are frequently found on plasmids and are present in host organisms that harbor multiple CRISPR-Cas systems, which may potentially indicate a unique biological niche for this system.

2. Results

2.1 Bioinformatic analyses

A bioinformatics investigation was conducted to find a suitable model organism for studying the biological function of Type IV-A CRISPR-Cas systems. Using tblastn and Cas6 (*A. aromaticum* Type IV-A CRISPR-Cas) as a query (Madden et al., 1996), a Cas6-like protein with 66% query coverage and 26% identity was found in *P. putida* KF715 and *P. aeruginosa* DN1. An iterative blast search identified a total of seven *Pseudomonas* strains that host this Type IV-A system (Figure 8).

The CRISPR arrays in these strains contain distinct spacers, indicating acquisition of spacers. A blastn search revealed perfect matches (protospacers) for seven spacers in the NCBI Nucleotide Collection. A multiple sequence alignment of the genomic context of these protospacers showed a conserved 5'-AAG-3' protospacer adjacent motif (PAM) at the 5' end of the target strand (Figure 9).

P. oleovorans DSM1045 was chosen as the model organism for investigating the biological function of Type IV-A CRISPR-Cas systems due to its commercial availability and low biosafety level (BSL1). In this organism, a Type IV-A CRISPR-Cas system containing 20 spacers is located on an IncP-9 family plasmid, while a Type I-E and a Type I-F CRISPR-Cas system are present in the host genome and contain *cas1-cas2* adaptation modules. Notably, the first spacer of the Type IV-A CRISPR array can base-pair with the sequence of the *pilN* gene, which is flanked by a 5'-AAG-3' PAM in the host genome (Table 1). The intactness of *pilN* suggests that base-pairing to crRNA does not associate with DNA degradation. Additionally, two perfect matches of the Type I-E system were identified and both protospacers exhibited a 5'-AAG-3' PAM, which may indicate cross-talk, in which elements such as adaptation modules or nucleases are shared between the Type I-E and Type IV-A systems (Table 1).

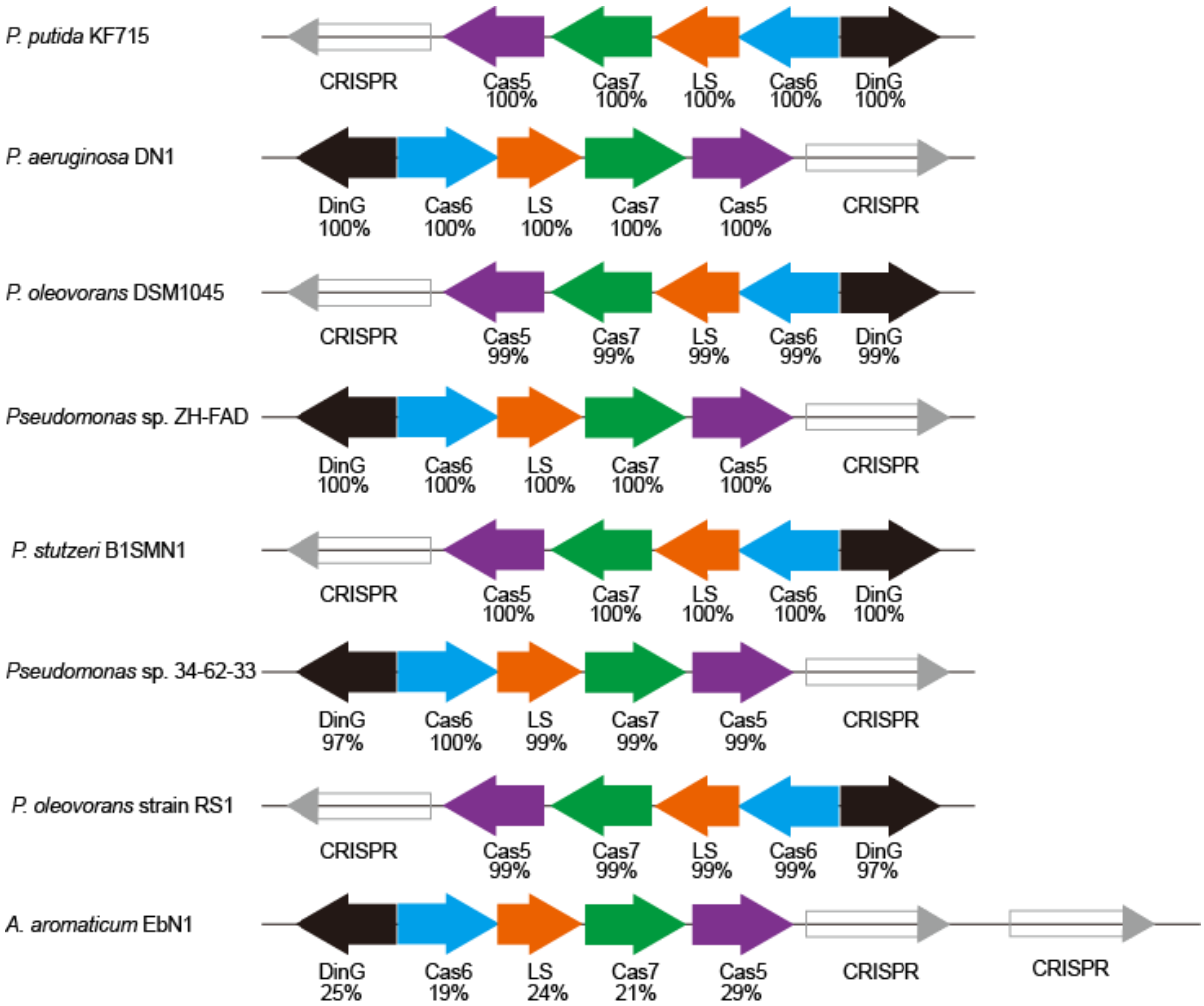


Figure 7. Genomic arrangements of *Pseudomonas* and *A. aromaticum* Type IV CRISPR-Cas loci. Iterative blast using components of the *A. aromaticum* Type IV CRISPR-Cas system identified seven *Pseudomonas* strains with Type IV systems.

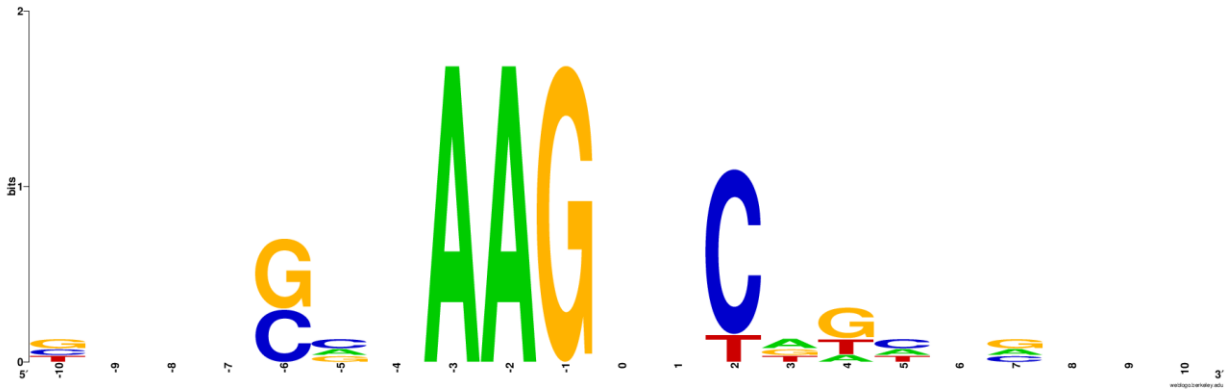


Figure 8. Sequence Logo for the PAM consensus sequence determination. Genomic context upstream (negative number) and downstream (positive number) of 7 protospacers (Table 1) were plotted as WebLogo (Crooks et al., 2004). Position 0 represents the position of the protospacer.

Table 1. Identification of PAM via BLAST search of *P. oleovorans* spacer sequence matches.

Type I-E					
Spacer sequence	**contig**	**position**	**target accession number**	**target position**	**PAM**
AACAGCCAATCAATGGTGCCGTTGAACAGATA	NIUB01000004.1	171430-171461	MN366360.1	15011-15042	ACGTCAACAAGTATCTGTTCAACGGCACCATTGATTGGCTGTTCGATGCC
TTCTCGACGTGGCGGGCGACCAGCAGGAAGCG	NIUB01000004.1	157732-157763	CP046060.1	6883912-6883943	AGTCGCGCAAGCGCTTCCTGCTGGTCGCCCGCCACGTCGAGAAGGTGCCG
Type IV-A					
Spacer sequence	**contig**	**position**	**target accession number**	**target position**	**PAM**
CCTTATCCGCCAAATGCGGCCTCAGCATGATG	Niub01000035.1	17259-17290	CP013124.1/*Pseudomonas mendocina* S5.2	1584335-1584369	GTAAAGGAAAGCATCATGCTGAGGCCGCATTTGGCGGATAAGGGCAGCTA
GAGCCCCTTCTTCGCCAAGAAACTCGCGACCA	Niub01000035.1	38693-38724	CP045916.1/*Pseudomonas aeruginosa* CF39S	6770323-6770354	GGGACGATAAGGAGCCCCTTCTTCGCCAAGAAACTCGCGACCAACGGCCA
AATTTGCCGAGCGCCTTGGATCACAGGAACAC	Niub01000035.1	37596-37627	CP033832.1/ *Pseudomonas aeruginosa* FDAARGOS_505	1747072-1747103	TTTTGGCGAAGGTGTTCCTGTGATCCAAGGCGCTCGGCAAATTCCGGTGG
CGTATCAATAGGCACCGTGCGATCCCGACCAC	Niub01000035.1	38388-38419	CP013124.1/*Pseudomonas mendocina* S5.2	1584270-1584301	CGGCACGAAAGGTGGTCGGGATCGCACGGTGCCTATTGATACGCCAGAGC
TCGCGTCTCGAATCTGATGCGTAACTTGGATG	Niub01000035.1	23923-23954	CP031606.1/*pilN*	509142-509173	CCACTCAGAAGCATCCAAGTTACGCATCAGATTCGAGACGCGATTATTCG
GGTCCCTATATCCCTTAATTTTGAGAAGCTGA	Niub01000035.1	23556-23587	MH547561.1/*tnsE*	20183-20214	GGCGCGCCAAGTCAGCTTCTCAAAATTAAGGGATATAGGGACCACTTCAG
CAGCCAGTGGGTGACAGAATGAACCTGCCCCG	Niub01000035.1	113850-113881	CP032257.1/ *Pseudomonas aeruginosa* AR_0111	3623038-3623069	TCGGCCCTAAGCAGCCAGTGGGTGACAGAATGAACCTGCCCCGTCTAAAC

***** Type IV-A spacer sequences were collected from 4 highly similar systems. The position of PAM is underlined.

2.2 Biogenesis of crRNAs and physiology of *Pseudomonas oleovorans* DSM1045

2.2.1 The physiology of *Pseudomonas oleovorans* DSM1045

P. oleovorans is a versatile microorganism known to thrive on a wide range of carbon sources. Typically, *P. oleovorans* is cultivated in a mineral-rich minimal medium, as documented in several studies (Freitas et al., 2009, 2010; van Beilen et al., 1994). However, to simplify the experimental process, we evaluated the growth of *P. oleovorans* on commercially available lysogeny broth (LB) medium at 37°C or 30°C. Notably, both cultures achieved an optical density (600 nm) of 1 within 8 hours, with no significant differences between the two temperatures. As our primary research focus was not on the metabolism of *P. oleovorans*, we did not further optimize growth conditions and proceeded to cultivate *P. oleovorans* in LB medium at 37°C.

To determine if *P. oleovorans* DSM1045 is naturally resistant to ampicillin (Amp), an overnight culture of the organism was plated on LB agar containing 50 μg/ml Amp. The presence of visible colonies on the Amp plates after overnight incubation at 37°C indicates that *P. oleovorans* DSM1045 is naturally resistant to Amp.

Initially, plasmids were introduced into *P. oleovorans* via electroporation. However, this method's performance was unstable, likely due to the fragile cell membrane after electric shock. As an alternative, conjugation with *E. coli* WM3064 was tested. Despite the longer process time compared to electroporation, conjugation showed stable performance and good transformation efficiency.

To build a stable multi-plasmid system for genetic manipulation, test transformations of pACYCDuet-1, pCDFDuet-1, pRSFDuet-1, pUCP18, and pHerd30T were performed. Only the shuttle vectors pUCP18 and pHerd30T could be effectively transformed into *P. oleovorans* DSM1045.

2.2.2 crRNA biogenesis of *Pseudomonas oleovorans* DSM1045

An enriched small RNA library was prepared and sequenced to investigate the CRISPR-Cas systems of *P. oleovorans* DSM1045. Cultures of the organism were grown without antibiotics in LB medium at 37°C for 12 hours, and the cells were harvested by centrifugation at 10,000 rpm for 2 minutes. The enriched small RNA was extracted using the Invitrogen™ mirVana™ miRNA Isolation Kit following the manufacturer's guidelines, and a sequencing library was constructed using the NEBNext® Ultra™ RNA Library Prep Kit.

The sequencing was performed on an Illumina MiniSeq system in paired-end mode, and the raw sequencing data was exported in BAM format. Quality scores were assigned to individual nucleotides during sequencing based on a built-in algorithm, and a Python script (FASTQC) was used to filter out low-quality reads, with a Q-score threshold of 28. Adapter sequences were trimmed using Cutadapt, and the resulting reads were aligned to the genome of *P. oleovorans* DSM 1045 using Hisat2.

Analysis of the read coverage of the Type IV-A CRISPR-Cas system revealed the presence of mature crRNAs with an eight nucleotide 5’ repeat tag (5’-GUGAGCGG-3’) and untrimmed 3’ termini containing a sequence with a hairpin structure. A mutation was detected in the repeat (5’-ATATTTCCCGCGTGCGCGGGGGTGAGCGG-3’), where the first nucleotide cytosine was replaced by an adenine, and the spacer8 was found to encode the most abundant crRNA of this system (Table 3). The RNA-seq analysis also confirmed the maturation of crRNAs in the associated Type I-F and Type I-E CRISPR-Cas systems, with mature crRNAs containing eight nucleotide 5 ’ repeat tags (Type I-E: 5’-AUGAACCG-3’, Type I-F: 5’-CUCAGAAA-3’). However, intensity of the mapping to *cas* genes was similar to the background, likely due to the size selection conducted during library preparation (Figure 10).

Mutations were detected in repeats of the second CRISPR array of the Type I-F system, and it is possible that these mutations prevent the processing of pre-crRNAs. In agreement, our deep sequencing data showed no coverage at the locus of the second Type I-F array. Interestingly, the abundance of crRNA suggests that the Type I-F and Type I-E systems are generally more active than the Type IV-A system in *P. oleovorans*. According to the mapping, crRNA22 and crRNA28 of the first Type I-F array, crRNA22 of the first Type I-E array, and

crRNA2 of the second Type I-E array are abundant. To identify these abundant crRNAs, we conducted a blast search but unfortunately found no match in public databases for the identified crRNAs.

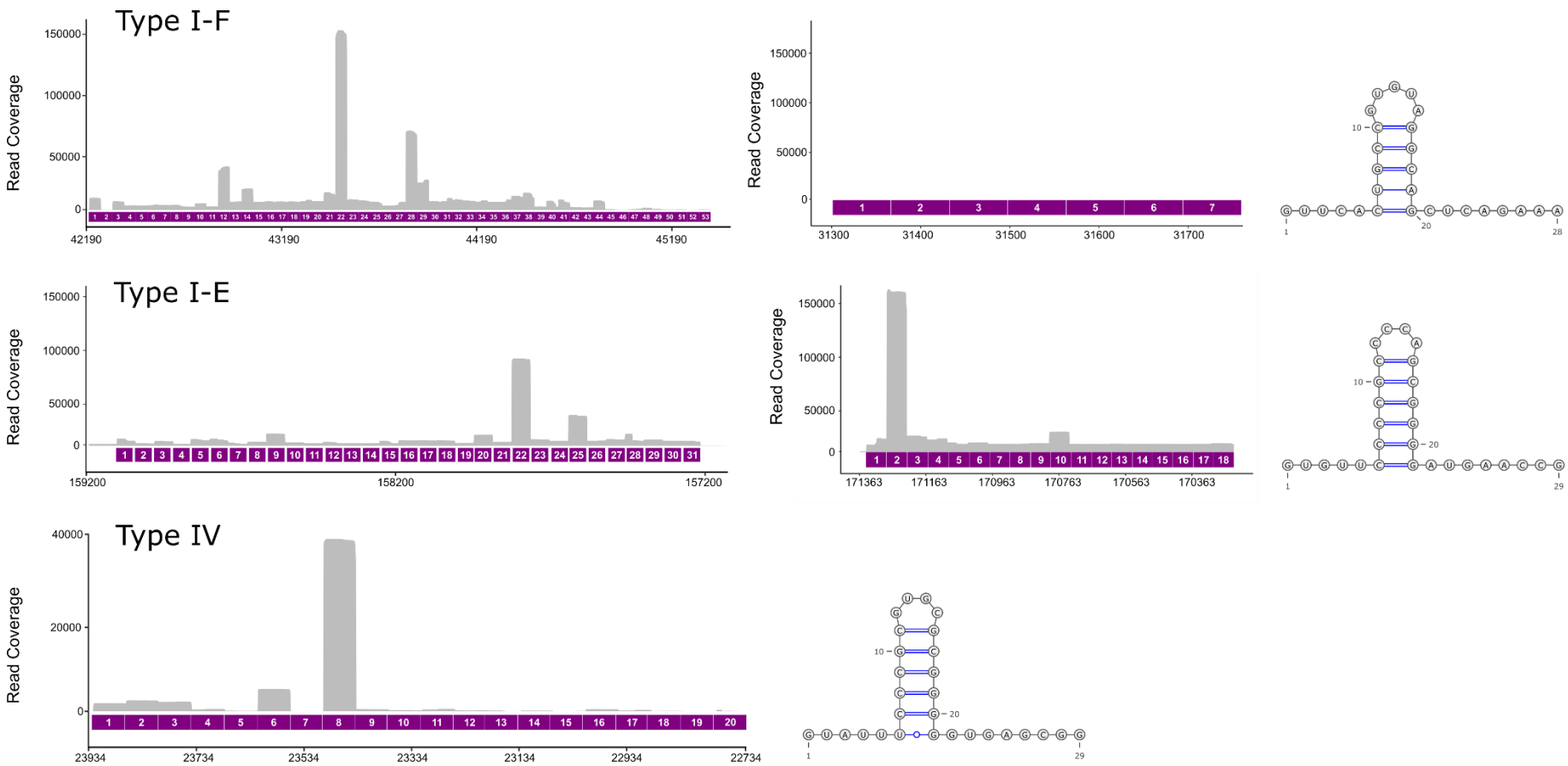


Figure 9. Coverage plots of small RNA-seq analysis of three *P. oleovorans* CRISPR-Cas systems. This plot highlights the transcription and effective processing of individual crRNAs. A high abundance of Type IV-A crRNA8 correlates with a point mutation in repeat 7. Repeat structures are indicated.

2.3 Spacer acquisition in a Type IV-A CRISPR array

One unique feature of Type IV-A CRISPR-Cas systems is the absence of adaptation modules, while CRISPR arrays are often found in Type IV-A systems. This may suggest that an unconventional spacer acquisition pathway is involved.

Type IV-A CRISPR-Cas systems often have a companion system with an adaptation module. Initially, we studied a Type IV-A CRISPR-Cas system from *A. aromaticum*, where a Type I-C system containing Cas1 and Cas2 was found in the chromosome. We hypothesized that there is an adaptation level cross-talk between the Type I-C and Type IV-A systems, a scenario that the Type IV-A system employs Cas1 and Cas2 from the Type I-C system for the spacer acquisition. We reconstructed the Type I-C Cas1, Cas2, and a minimal Type IV-A CRISPR array (leader sequence-repeat-spacer-repeat) in *E. coli* BL21 AI cells to test this hypothesis. To detect potential spacer acquisition, PCR amplificates were produced from genomic DNA of overnight colonies as templates with primers that base-pair to the leader sequence and the first spacer of the minimal CRISPR array, respectively. Despite the presence or absence of Cas1 and Cas2, clear bands at 653 nt, which is equal to the size of the original CRISPR array were observed and no indication of CRISPR array expansion was detected (Figure 11).

We further investigated adaptation with Illumina sequencing for higher-resolution results. The analysis of sequencing data was conducted using QIAGEN CLC Workbench. Obtained reads were first mapped to the original CRISPR array of the Type IV-A system from *A. aromaticum*, and about 200,000 reads without perfect matches were collected. In the second mapping, the collected reads were mapped to a reference sequence consisting of the leader sequence-repeat-(randomized spacer1 with 32 N residues)-repeat-Spacer2 (Figure 12). Most reads with matches fell into Spacer2, while none were mapped as a new spacer (Spacer 1).

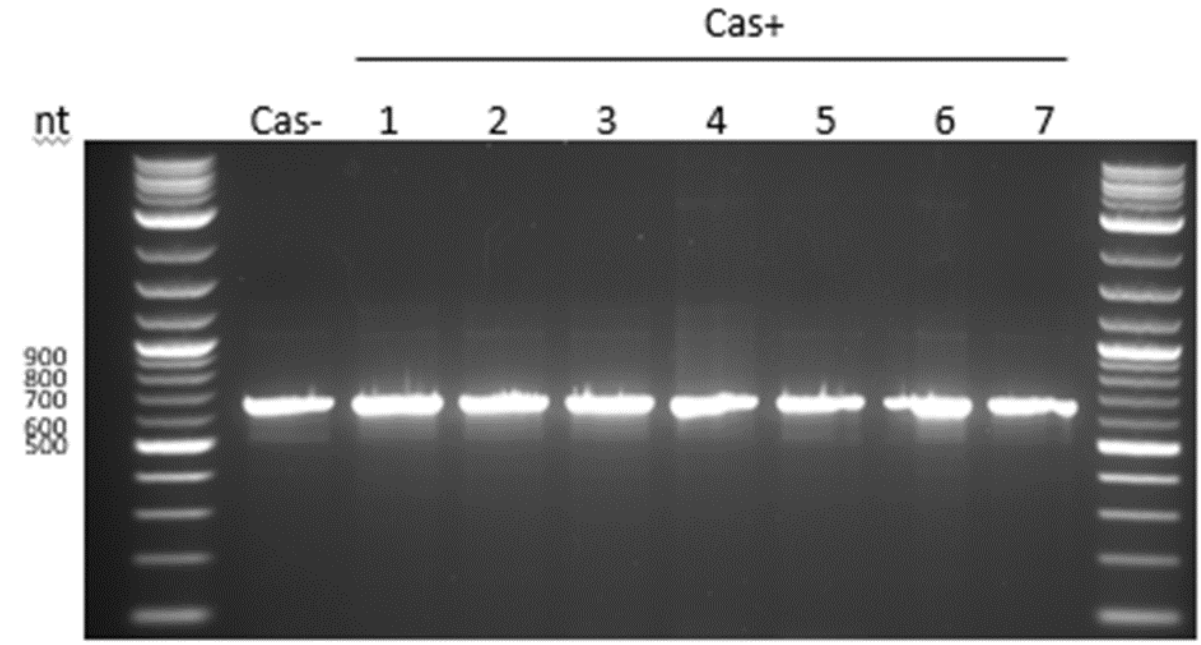


Figure 10. Electrophoresis of PCR amplicons of Type IV-A CRISPR. Colonies of *E. coli* BL21AI cells encoding Type IV-A *csf*1-3 and *csf*5 with (Cas^+) or without (Cas^-) the expression of *A. aromaticum* Type I-C Cas1 and Cas2 were used as templates for PCR amplification. The resulting PCR products were separated on a 1.5% TAE gel.

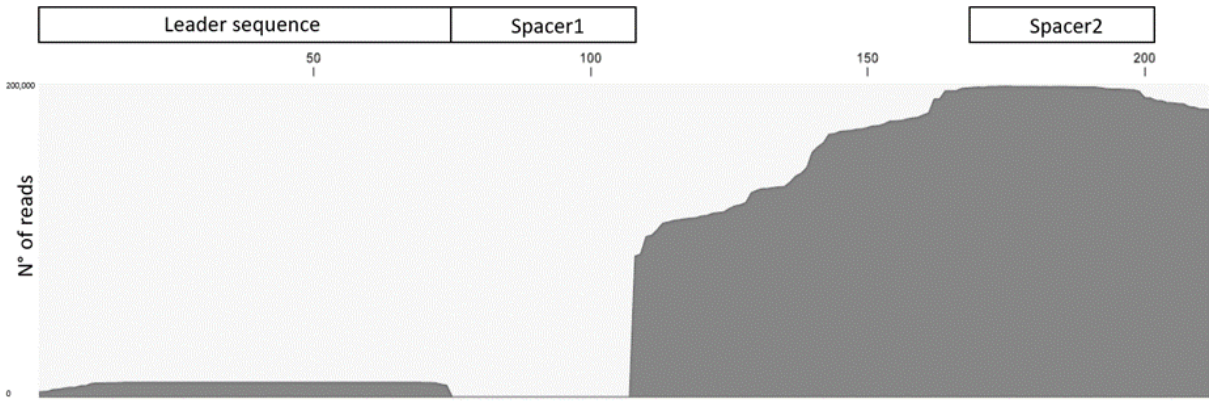


Figure 11. Selective mapping of Illumina sequencing reads from *A. aromaticum*. Obtained reads from Illumina were mapped against the reference sequence: leader sequence-repeat-spacer1-repeat-spacer2. Spacer1 includes 32 nt of N, which enables mapping to any sequence, while spacer2 is the first spacer from the *A. aromaticum* Type IV CRISPR array. Canonical spacer acquisition involves inserting a new spacer in the first position of the CRISPR array, resulting in new spacers appearing in the position of spacer1 in the mapping. None of the reads mapped to spacer1 in this analysis.

Next, we sought to investigate spacer acquisition in *P. oleovorans* DSM 1045. We identified a perfect match of spacer1 in the organism's genome, suggesting that a spacer acquisition event had occurred. Previous research on Type I CRISPR systems has shown that spacers often

originate from fragmented DNA molecules produced during DNA repair (C. Hu et al., 2021; McGinn et al., 2018; Schmidt et al., 2018). Alternatively, spacers can be acquired by transforming synthetic small DNA fragments into cells (Savitskaya et al., 2013). We conducted an adaptation-interference assay to determine if the Type IV-A system in *P. oleovorans* is capable of spacer acquisition.

As a pre-experiment, we introduced 60 nt linear prespacers to *P. oleovorans* cells containing a GFP vector via electroporation (Table 2). According to several studies, electroporation is a sufficient method to introduce linear DNA to bacterial cells (Dower et al., 1988; Kotnik et al., 2015; Lawrenz et al., 2002). If spacer acquisition occurred, we expected to see a downregulation of green fluorescence intensity due to the acquisition of spacers from the provided prespacers (Figure 13). However, we did not observe any sign of spacer acquisition in this assay.

Table 1. Sequences of pre-spacers introduced into cells

Double-stranded	Coding strand	ATGGTTAGCAAAGGTGAAGAACTGTTTACCGGCGTTGTGCCGATTCTGGTGGAACTGGAT
Single-stranded	Coding strand	ATGGTTAGCAAAGGTGAAGAACTGTTTACCGGCGTTGTGCCGATTCTGGTGGAACTGGAT
Single stranded	non-Coding strand	ATCCAGTTCCACCAGAATCGGCACAACGCCGGTAAACAGTTCTTCACCTTTGCTAACCAT

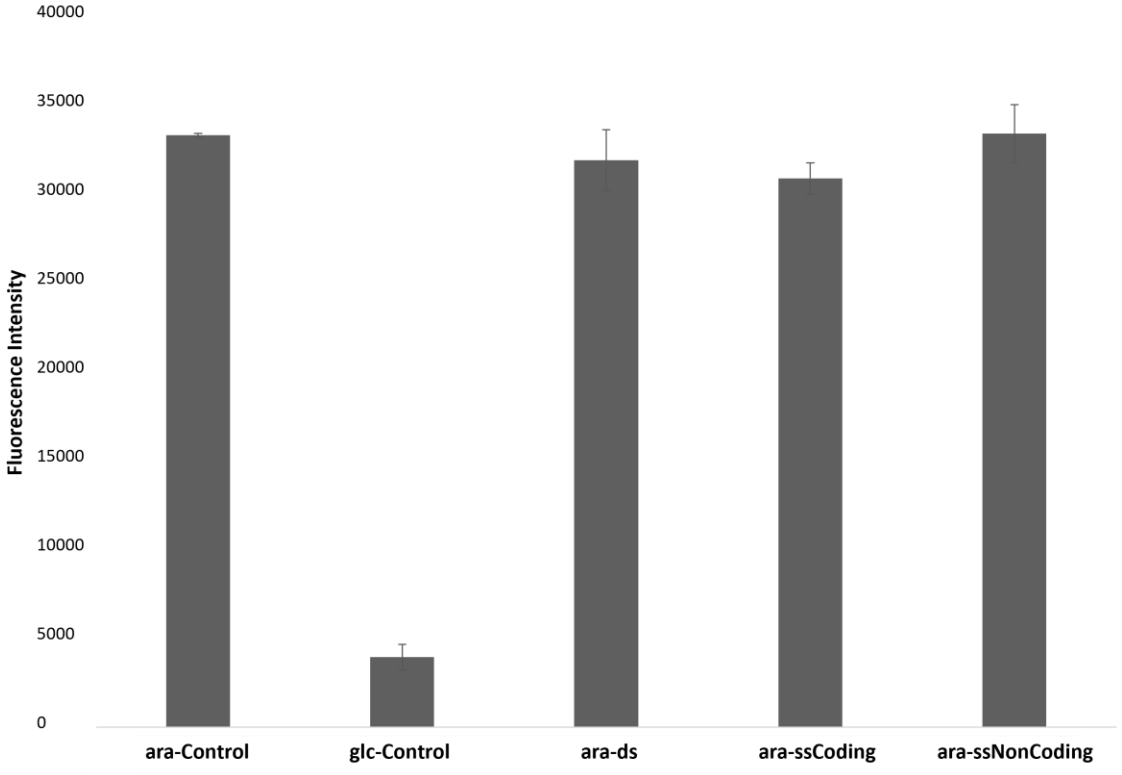


Figure 12. Plate reader measurements of cells harboring different pre-spacers. The assay was conducted with *P. oleovorans* bearing a GFP expression vector electroporated with different synthetic pre-spacers of *gfp* (double-stranded DNA, single-stranded DNA of coding or non-coding strand). Cells without targeting were taken as a negative control, the intensity was adjusted to the value at time 0, and an extra 10000 was added to the axis value to compensate for the negative value in control. *Gfp* is repressed with glucose (glc) to prevent leakage expression and induced with arabinose (ara). This dataset contains three biological replicates.

To further investigate spacer acquisition in *P. oleovorans* DSM 1045, we used a PCR-based method called CAPTRUE to detect insertions between the first repeat and spacer in the CRISPR array (McKenzie et al., 2019). CAPTRUE based on two nested PCR amplifications to low-intensity signals, spacer-specific primers were applied during the amplification. We electroporated linear DNA substrates with putative PAM (5'-AAG-3') in single-stranded or double-stranded forms into cells (Table 1). Genomic DNA was extracted from overnight cultures and used as a template for the CAPTRUE PCR. After PCR enrichment, the amplicons displayed ladder-like patterns on agarose gel (Figure 14). The estimated gap between ladders

was approximately 60 nt, consistent with the spacer size plus a repeat. In addition, deep sequencing of *P. oleovorans* DSM1045 revealed spacer shuffling in the Type IV-A CRISPR array. Analysis of the sequencing data showed a significant number of reads containing Type IV repeats and spacers with different arrangements than the original draft genome, which explains the ladder-like pattern observed on the agarose gel.

Furthermore, we found new Type IV-A spacers from the sequencing data (Table 3). Spacer sequences that were 32 nt long (Table 3) flanked by Type IV repeats were identified. The novel spacer sequences were subjected to BLAST searches against both the NCBI nucleotide collection and the host genome sequence. However, no perfect matches were found in either of these datasets.

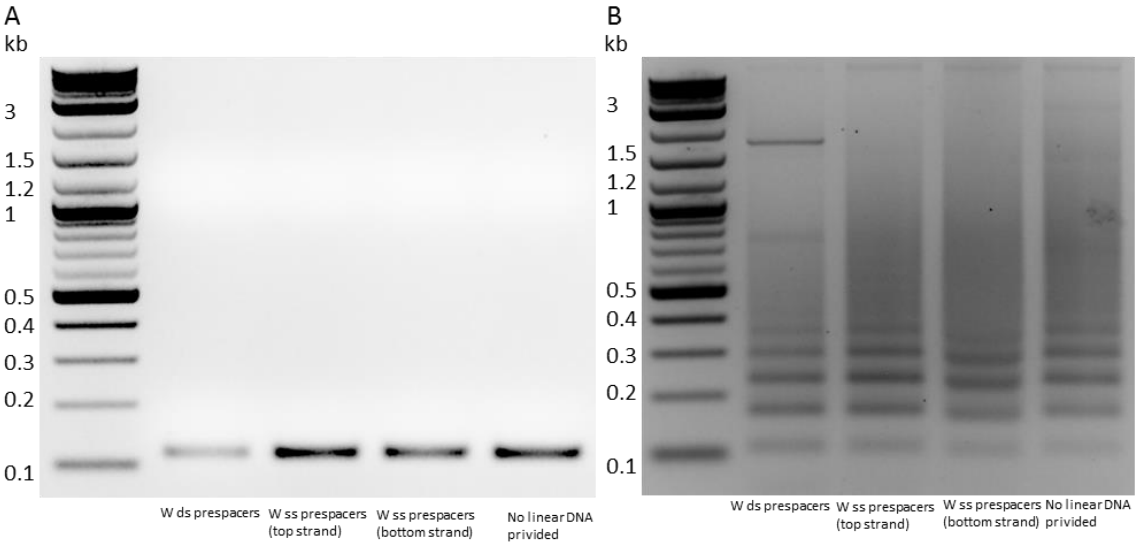


Figure 13. Electrophoresis of CAPTURE PCR products. Genomic DNA was extracted from *P. oleovorans* cells containing a GFP expression vector that were electroporated with various synthetic pre-spacers of gfp, including double-stranded DNA and single-stranded DNA from both the coding and non-coding strands. **a)** Genomic DNA samples were subjected to the first round of PCR amplification using initial primers, which amplified the sequence between the leader sequence and the first spacer of Type IV CRISPR array. Only bands with the expected length of the original sequence were observed on the gel. **b)** Gel extracts from a) were used as templates for the second round of PCR amplification with internal primers, resulting in a ladder pattern of bands on the gel. All PCR products were separated on a 2% TAE gel.

Table 2. Sequencing reads contain new spacers (in red).

5'- AGATCAATATCTATGTGTATTTCCCGCGTGCGCGGGGGTGAGCGGCATCCAAGTTACGCATCAGATTCGAGACGCGA GTATTTCCCGCGTGCGCGGGGGTGAGCGGGAATCAAGTGATAATTAAGGGAGATTCTTACAGTATTTCCCGCGT -3'
5'- ATGCGGAAGATCAATATCTATGTGTATTTCCCGCGTGCGCGGGGGTGAGCGGCATCCAAGTTAATAATAAGATTCG ATACGAGAGTATTTCCCGCGTGCGCGGGGGTGAGCGGGAATGGACTGAAAATGCGAGTAGATACTTACAGTATTT - 3'
5'- GGCCGCTCACCCCCGCGCACGCGGGAAATACTGGAAGAAGGAAATCGCACCAGCCCTGCATAGCCGCTCACCCCCG CGCACGCGGGAAATACTCGGCACAAAAAAGCCCCTCAAGGCAGGCGACCCGCTCCCCCCCGCGCACGCGGGAAAA - 3'
5'- TCGCACTGCTGATCGCCTACTTCGCCCGCTCACCCCCGCGCACGCGGGAAATACCACACGTACCGAGAAGGATTTCAT GGCAACGTCCGCTCACCCCCGCGCACGCGGGAAATACTGTAGGATAGTAACTTAAGATTAAGAAGCTTGCCG -3'
5'- TCGCACTGCTGATCGCCTACTTCGCCCGCTCACCCCCGCGCACGCGGGAAATACACCACGTACCTAGAAGGATTAAAT GGCAAAGACCGCTCACCCCCGCGCACGCGGGAAATAAGCACCAAATTTACCTTTAATATGAGAACATCACCG -3'

2.4 Investigation of the biological function of the *Pseudomonas oleovorans* Type IV-A CRISPR-Cas system

We sought to investigate the activity of the *P. oleovorans* Type IV-A CRISPR-Cas system by using a plasmid expressing sfGFP and different versions of a second plasmid coding for crRNAs targeting selected *gfp* protospacers.

To investigate the impact of *P. oleovorans* Type IV-A CRISPR-Cas system under antibiotic pressure, cells carrying pUCP18 (kanamycin resistance) expressing sfGFP and pHerd30T (gentamicin resistance) containing a crRNA complementary to the sfGFP sequence were cultured in the presence of either gentamicin or kanamycin, or both antibiotics. The optical density at 600 nm was measured after overnight incubation. Distinct differences were only observed between cells grown with both antibiotics (Figure 15).

In parallel, cells were incubated in 24-well plates, and the optical density at 600 nm and the green fluorescence signal was measured using a plate reader. We observed a downregulation of fluorescence intensity and slower growth with the presence of a spacer base pairing with the sequence of *gfp* (Figure 16).

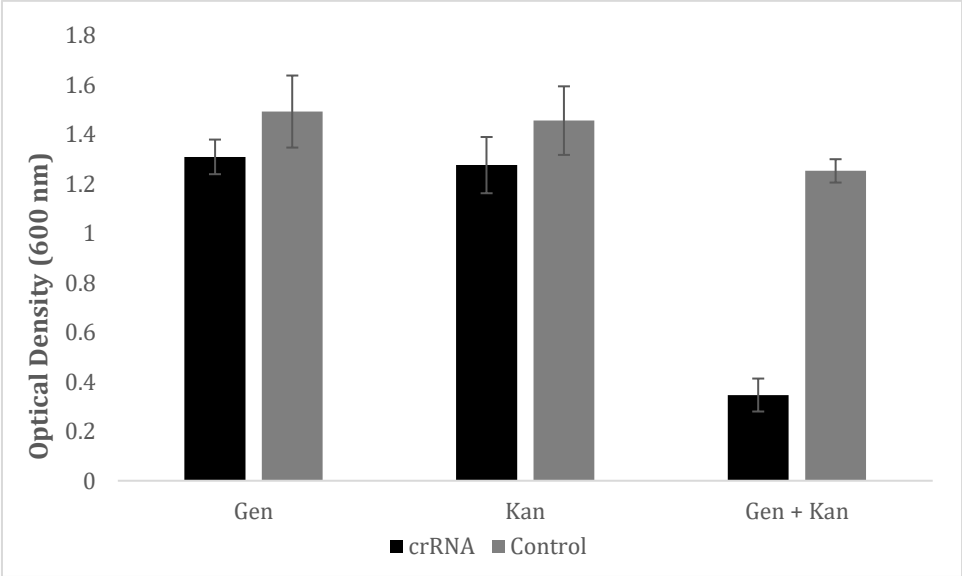


Figure 14. The impact of Type IV-A CRISPR targeting under different selections. OD_{600} measurements of overnight cultures of *P. oleovorans*. Cells bearing a GFP expression vector (Kan) and a vector with crRNA target gfp (Gen) or an empty vector (control), respectively. Cells were grown in the selection of gentamicin, kanamycin, or both antibiotics.

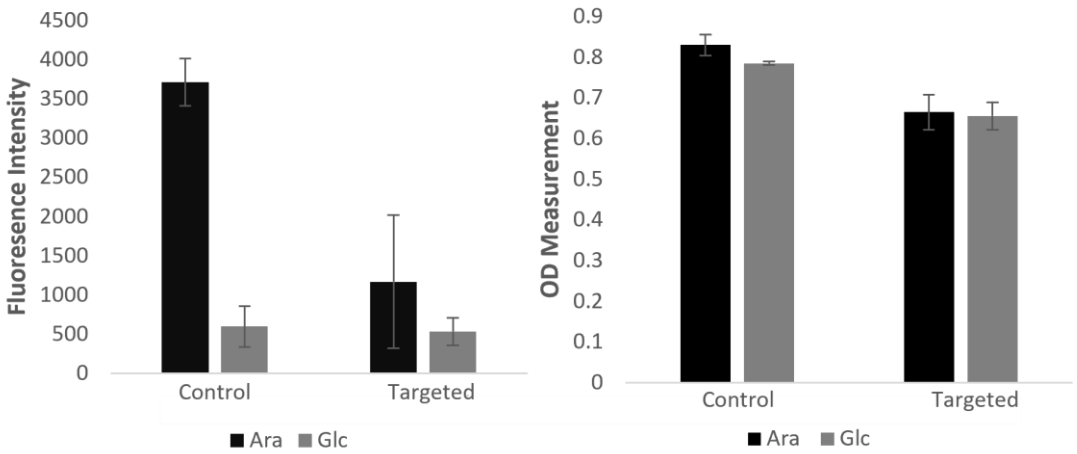


Figure 15. The targeting of Type IV-A CRISPR-Cas systems shows an impact on gene expression and growth. The targeting of the Type IV CRISPR system down regulates GFP expression and shows an impact on growth. *P. oleovorans* cells were growing overnight in 24-well plates. Fluorescence intensity (wave length 508 nm) and OD_{600} were measured with a

plate reader, the measurements were taken after 10 hours of incubation at 37°C. GFP expression was induced with arabinose (in black) or repressed with glucose (in grey).

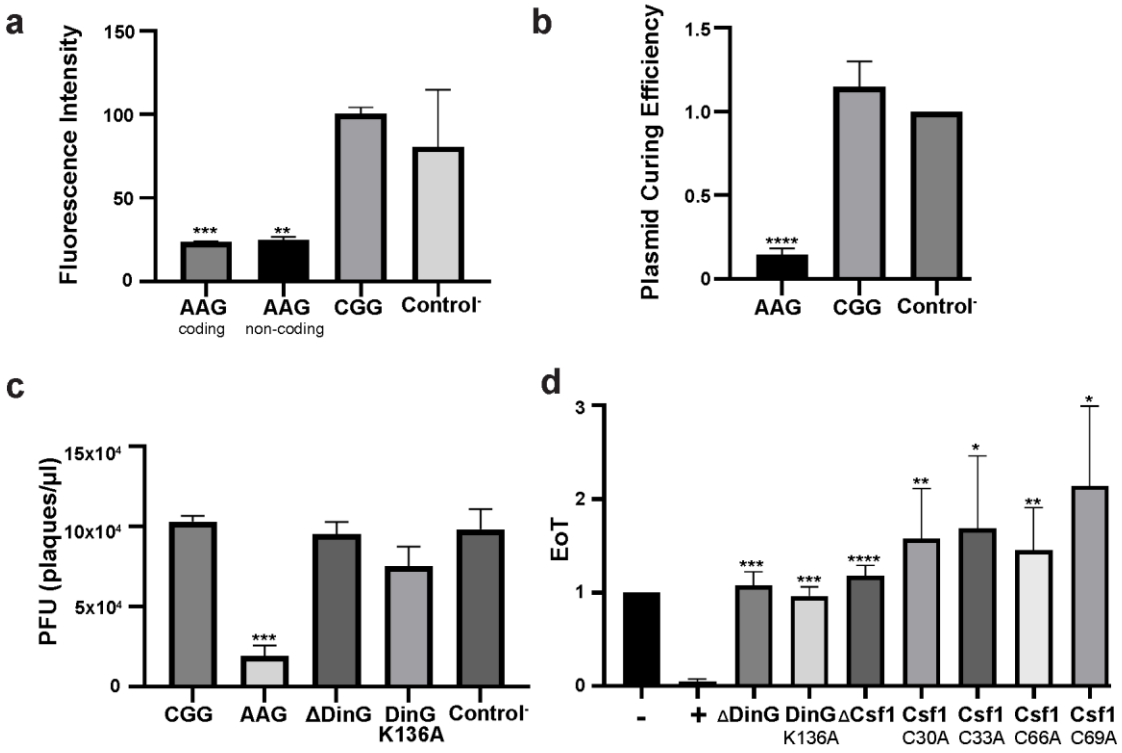


Figure 16. *In vivo* activity of the *P. oleovorans* Type IV-A CRISPR-Cas system. a) Fluorescence intensity measurement of a *gfp*-containing plasmid targeted by engineered crRNAs in *P. oleovorans*. The targeted protospacer strand and its PAM are indicated. A crRNA without a protospacer target served as a control. **b**) Plasmid curing efficiency was calculated in reference to colony forming units of cells containing non-targeting crRNAs. The protospacer PAM is indicated. **c**) Plaque-forming units (PFU) were identified following lytic lambda phage infection of *E. coli* cells producing recombinant Type IV-A crRNPs. Activities of two crRNAs targeting gene *E* with 5'-CGG-3' and 5'-AAG-3' PAMs are compared and DinG deletion and K136A mutants were assayed during AAG protospacer targeting. **d)** EOT assays with an AAG protospacer indicated cas gene deletions and mutations. **a-d)** All experiments were performed in triplicates. P-Values were calculated using an unpaired T-test (* p<0.05, *** pp<0.001, **** p<0.0001). This Figure was published in (Guo et al., 2022).

We also quantified the GFP signal using Fluorescence Activated Cell Sorting (FACS). The FACS analysis showed a reduction in green fluorescence signal for protospacer targets in the *gfp* coding and non-coding strands in the presence of a 5'-AAG-3' PAM (Figure 17a). However, a protospacer in the *gfp* coding strand with a 5'-CGG-3' PAM and a non-targeting crRNA did not significantly reduce the GFP signal, suggesting PAM-dependent CRISPR interference. To study plasmid curing in this system, we transformed *P. oleovorans* with a target plasmid and a crRNA-production plasmid and cultivated the cells for 12 hours without antibiotic selection. We then transferred the cells into a medium with antibiotics to select for the presence of both plasmids. We observed PAM-dependent plasmid curing, indicating that the native Type IV-A crRNP can interfere with plasmids (Figure 17b).

To further analyze this targeting mechanism, we designed a heterologous *E. coli* BL21-AI system for the production of recombinant *P. oleovorans* Type IV-A crRNPs. The *csf*1-5 genes were encoded on pETDuet-1, while a minimal CRISPR array containing a single spacer-repeat-spacer unit was provided on pRSFDuet-1. Throughout the process, IPTG (1 mM) and arabinose (0.2% m/v) were added to induce gene expression, which was assumed to be continuously active.

We transformed target plasmids pACYCDuet-1 into *E. coli* using electroporation and calculated the efficiency of transformation (EOT). We only observed a reduced transformation efficiency for plasmids carrying a protospacer with a 5'-AAG-3' PAM, indicating that the recombinant Type IV-A system facilitates PAM-dependent interference (Figure 17b and d). Base pairing between PAM sequences and crRNA repeats is used to identify self-targets and prevent autoimmunity (Jia et al., 2019; Leenay et al., 2016; Westra et al., 2013a). In our system, the -1 position of the 5'-AAG-3' PAM can base pair with the Type IV-A crRNA repeat. We mutated the -2 and -3 PAM nucleotides in the target plasmid and observed changes in the EOT for the different constructs to gain insights into the PAM specificity in this interaction. We found that the presence of 5'-GTG-3' and 5'-AAN-3' PAM sequences significantly reduced EOT (Figure 18). Larger EOT error bars, which correlate with highly variable sizes of obtained colonies, may indicate competition between plasmid replication and interference in these assays.

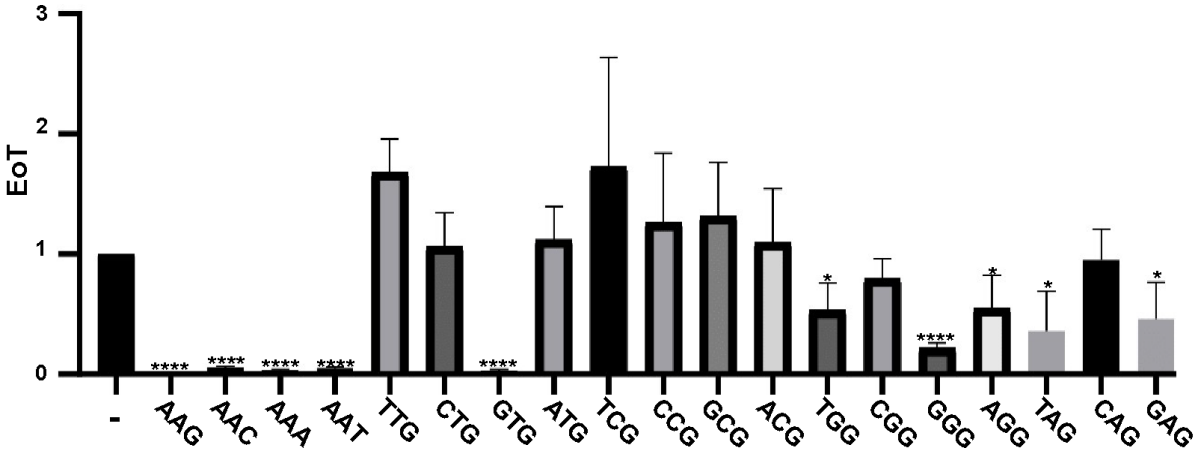


Figure 17. PAM efficiency scanning. EOT assays of recombinant Type IV-A crRNPs targeting protospacers on an electroporated plasmid with the indicated PAM sequence in *E. coli*. The AAG-PAM (see b) served as a positive control, and a crRNA without a protospacer matched a negative control. All experiments were performed in triplicates. P-Values were calculated using an unpaired T-test (* $p<0.05$, **** $p<0.0001$).

Next, we analyzed the function of Cas protein mutants in a recombinant system. The putative helicase DinG is found in all Type IV-A CRISPR-Cas systems, but its role is unknown. We found a conserved walker-A motif indicative of ATP binding by performing multiple sequence alignments of several Type IV-A associated DinG helicases. The efficiency of transformation (EOT) assays showed that a null mutation in the walker-A motif of DinG (K136A) abolished interference, suggesting that ATP-dependent helicase activity of DinG is crucial for CRISPR interference (Figure 17c and d). These results agree with the requirement for DinG activity in EOT assays using recombinant *P. aeruginosa* crRNPs.

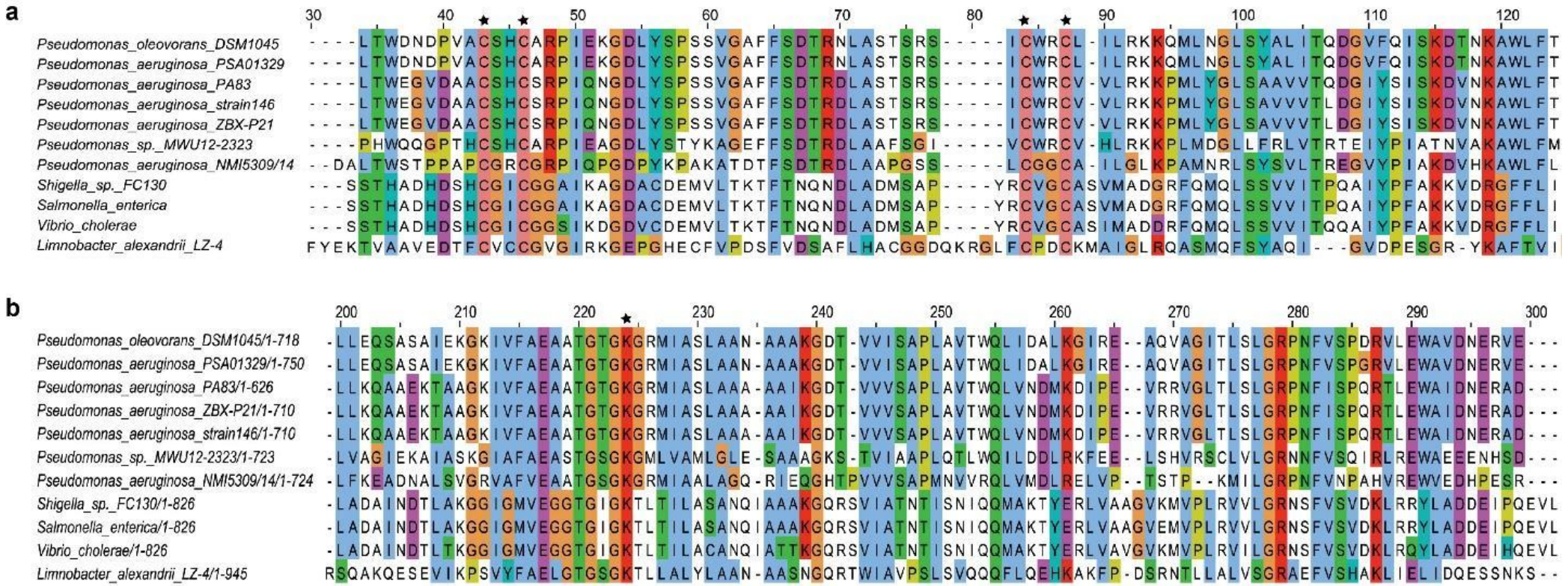


Figure 18. Multiple sequence alignments of conserved Csf1 and DinG sections. The Clustal X default color scheme is applied (Jeanmougin et al., 1998), and the positions of point mutations investigated in this study are labeled with an asterisk. **a**) Multiple sequence alignment of Type IV-A Csf1, cysteine residues at positions 43, 46, 84, and 87 are conserved. **b**) Multiple sequence alignment of Type IV-A associated DinG proteins; a variant walker-A motif with consensus sequence TGXGK is identified.

In Type I CRISPR-Cas systems, LS is responsible for PAM recognition and R-loop binding (Cass et al., 2015; Li et al., 2020b). However, the function and structure of LS in Type IV-A effector complexes are not yet known. We identified a cysteine-rich motif by performing multiple sequence alignments of eleven Type IV-A LS (Figure 19a). Transformation assays revealed that a single null mutation in each cysteine of this motif abolished CRISPR interference (Figure 17d). This cysteine-rich motif may form a zinc-finger-like structure involved in R-loop binding.

We conducted phage targeting assays to determine if Type IV-A systems show immunity against bacteriophages. Vectors carrying expression constructs for a non-targeting crRNA, and a crRNA targeting lambda phage gene *E* with 5'-AAG-3' or 5'-CGG-3' PAM, were respectively transferred into *E. coli* BL21 cells producing recombinant Type IV-A effectors. In addition, plasmid encode for gene *E* targeting crRNA with 5'-AAG-3' PAM was transferred to recombinant *E. coli* cells with DinG deletion or with a null mutation in DinG (Figure 17c). After transformation, all cells were infected with lytic lambda phages. The recombinant Type IV-A systems showed PAM-dependent immunity against lytic lambda phages, and the Walker-A domain of DinG appears to be essential for phage immunity. The target gene *E* encodes a capsid protein essential for lambda phages assembly (Rajagopala et al., 2011). The impact on infection efficiency could result from the blockage of gene *E* or the impedance of circular phage DNA replication.

2.5 Self-targeting activity of recombinant Type IV-A crRNPs

We have demonstrated the activity of the Type IV-A system against plasmids and lambda phage. *In silico* analysis indicated that the system lacks a nuclease for target degradation and is likely to exhibit dCas9-like activity. However, differentiating between stringent knock-down and knock-out effects, and confirming the absence of any hidden nuclease activity, remains a challenging task.

To investigate the targeting mechanism, we employed a blue-white screening system. This system utilizes 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (x-gal), an organic compound composed of galactose linked to a substituted indole (Horwitz et al., 1964). The reporter gene *lacZ* encodes for β-galactosidase, an enzyme that can cleave the colorless X-gal to produce 5-bromo-4-chloro-indoxyl. Subsequently, 5-bromo-4-chloro-indoxyl dimerizes and oxidizes to form an insoluble blue pigment 5,5'-dibromo-4,4'-dichloro-indigo. Blue colonies indicate that *lacZ* is being expressed and translated without disruption, whereas white colonies indicate the opposite (Julin, 2018).

In an earlier study, the activity of Type I-C CRISPR-Cas systems was evaluated using *lacZ* blue-white screening (Csörgő et al., 2020). In the study, white colonies correlated with large genomic deletions around the *lacZ* gene as a result of the activity of the DNA nuclease Cas3. We used this Type I-C system activity as a control and designed a comparable set-up to monitor genomic *lacZ* targeting by our Type IV-A system (Figure 20b and c).

We transferred plasmids coding for a crRNA that paired with protospacers in either the coding or non-coding strand of *lacZ* into *E. coli* cells producing Type IV crRNPs. After transformation, we selected three individual colonies from each transformation for overnight incubation in liquid LB medium at 37°C. The next day, cells were transferred to agar plates containing X-gal and incubated for 10 hours to let the color sediment.

In the negative control, a non-targeting crRNA resulted in all blue colonies. However, crRNAs with *lacZ* protospacer targets resulted in significant and comparable amounts of white colonies for both the Type I-C and Type IV-A CRISPR-Cas systems. The targeting by Type IV-A systems in both DNA strands generated white colonies, further supporting a DNA targeting

mechanism. Interestingly, heterogeneity of color was observed in the colonies targeted by Type IV-A activity.

To further investigate whether DNA degradation was involved in the CRISPR activity, we PCR-amplified the *lacZ* gene from selected blue and white colonies. The electrophoresis revealed that the blue colonies targeted by Type I-C systems had intact *lacZ*, while the white colonies appeared to have a deletion. In contrast, white colonies targeted by Type IV-A showed intact *lacZ*, indicating that the Type IV-A-mediated CRISPR interference of *lacZ* did not involve DNA degradation (Figure 20c).

Sanger sequencing results further confirmed that the genomic context of *lacZ* remained untouched, as all sequenced colonies contained an intact *lacZ*. In addition, we selected three blue and three white colonies from the plates showing Type IV-A CRISPR activity for passaging. These colonies were cultivated in liquid LB medium for 10 hours at 37°C and then transferred to agar plates containing X-gal. Cells were incubated for another 12 hours on the plate to allow for color reaction. Interestingly, the following passages of these colonies showed a similar distribution of blue and white, regardless of the color picked for passaging. Our findings suggest that the observed activities against phages and plasmids do not require DNA degradation for the Type IV-A CRISPR activity.

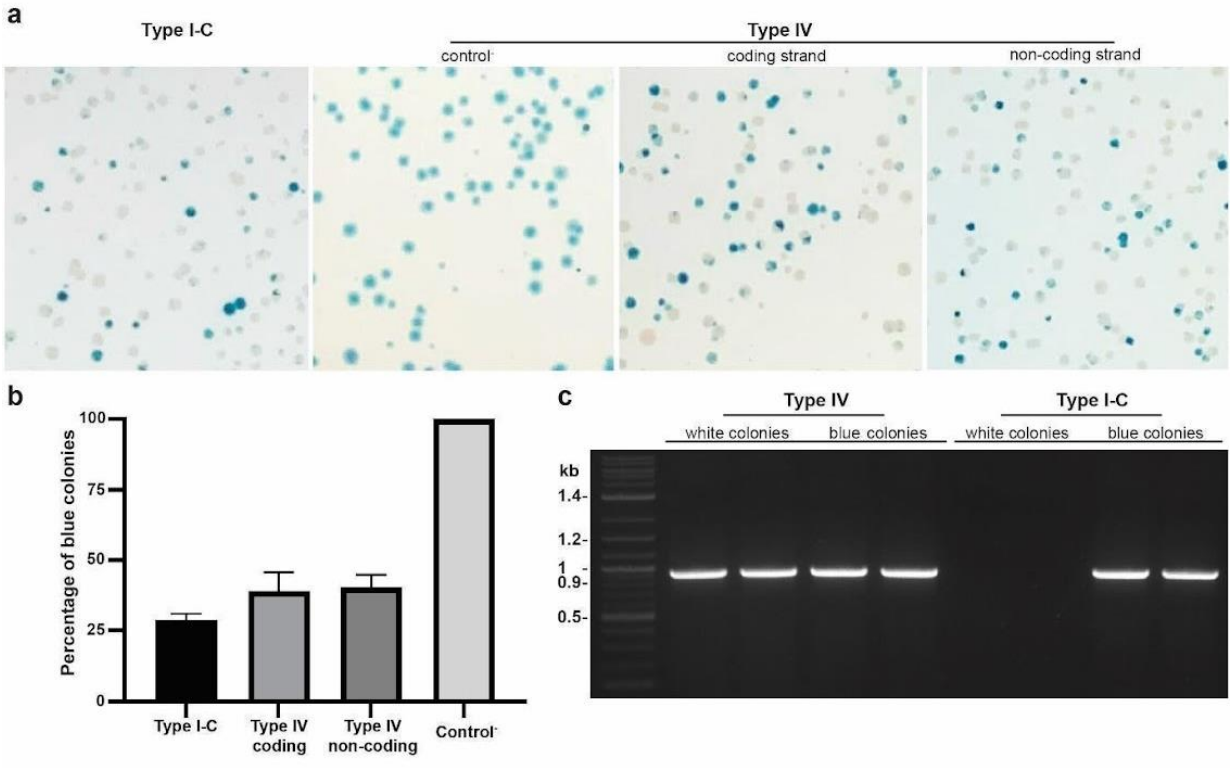


Figure 19. Self-targeting CRISPR interference of recombinant Type IV-A crRNPs in *E. coli*. a) Representative blue-white screening of *E. coli* BL21-AI cells producing either recombinant Type I-C or Type IV-A crRNPs. Individual crRNAs are targeting *lacZ* and the protospacer strand is indicated. A crRNA without a lacZ target served as a negative control, raw data can be seen in Supplementary Figure 1. Twenty-three white colonies and two blue colonies were sampled for sanger sequencing, but no mutations were detected in any of the colonies. **b**) Quantification of the observed percentage of blue colonies for all (n=9852) counted colonies on triplicate plates. **c**) Representative PCR amplification of a 900 bp lacZ product. All investigated colonies revealed the absence of *lacZ* following Type I-C CRISPR-Cas genome editing while all white colonies with Type IV-A crRNPs maintained *lacZ*. This Figure was published in (Guo et al., 2022).

3. Discussion

3.1 Cross-talk of adaptation in a Type IV-A CRISPR array

Adaptation mechanisms of Type IV-A CRISPR-Cas systems have not been reported. One distinct feature of these systems is the absence of Cas1 and Cas2 (Budhathoki et al., 2020; Fagerlund et al., 2017; Nuñez et al., 2014; Xiao, Ng, et al., 2017). However, the presence of CRISPR arrays with unique spacers suggests the existence of an alternative adaptation pathway. Furthermore, the diverse spacers found among hosts and the self-targeting spacer in the *P. oleovorans* Type IV-A system provide evidence for the recent adaptation events.

It is hypothesized that a cross-talk at the adaptation level between different CRISPR-Cas systems exits. Type IV-A systems are often associated with other genome-encoded CRISPR-Cas systems which are proposed to provide a compatible adaptation machinery. However, initial investigation in a heterologous system containing Type I-C Cas1 and Cas2 and a minimal CRISPR array from *A. aromaticum* did not provide evidence for adaptation.

A possible explanation for this lack of evidence of adaptation is the absence of adaptor components that enable I-C Cas1 and Cas2 to recognize the Type IV-A CRISPR array. Additionally, the low sequence similarity between the leader sequence of the CRISPR array and the crRNA hairpin region in Type I-C and Type IV-A systems may require the involvement of components from Type IV-A to bridge the gap between the two systems.

P. oleovorans possesses three CRISPR-Cas systems (Type I-F, Type I-E, and Type IV-A). A BLAST search of the Type IV-A spacers against NCBI nucleotides collection revealed seven Type IV-A spacers target transposon or plasmid elements. The genomic context of these targets showed a consensus 5'-AAG-3' PAM motif at the 5' end of the non-targeting strand (Figure 9). In addition, two targets of Type I-E crRNAs were identified, and both displayed a 5'-AAG-3' PAM. This suggests that the Type I-E adaptation module may be utilized to acquire spacers by the Type IV-A CRISPR array. It is possible that additional helper proteins, which adapt the Type I-E adaptation module to be compatible with the Type IV-A CRISPR array, are required and may be components of the Type IV-A systems.

Deep sequencing of the of *P. oleovorans* transcriptome revealed the presence of new Type IV-A spacers that were 32 nt in length and flanked by Type IV-A repeats on both ends (Table 2). The identified spacers were obtained from uninterrupted single reads, but their origin remains unknown. BLAST searches against both the NCBI nucleotide collection and the *P. oleovorans* genome failed to provide a match. Possible explanations for the origin of these spacers include uptake from a contaminated environment, unsequenced regions of the host genome, or the employment of DNA repair mechanisms. The observed rearrangement of spacers suggests a shuffling mechanism within the CRISPR array, a hypothesis that is also supported by the distinct ladder-like pattern observed on the agarose gel of PCR products obtained from CAPTURE analyses of the *P. oleovorans* Type IV-A CRISPR array.

In addition, Pinilla-Redondo et al. reported that the rearrangement of spacers is an occurrence across multiple Type IV-A CRISPR systems (Pinilla-Redondo et al., 2020). This may be a mechanism to diversify immunological memory, as over time, some spacers become obsolete and diversification prevents "on-duty" spacers from being pruned from the end of the CRISPR array (Garrett, 2021).

One theory for the shuffling is that transposons located within the IncP-9 family plasmid hosting the Type IV-A system may be responsible, as there may be a history of co-evolution between the components within the same plasmid. A TnpB-like transposon locus (GenBank: NIUB01000035.1, location: 11,857 - 19,798) was identified about 5kb upstream of the Type IV-A locus. TnpB is known to have a close evolutionary relationship to Cas9, as both proteins contain RuvC-like nuclease domains (Koonin et al., 2014). Additionally, RuvC is a holliday junction resolvase that makes a cleavage at the bottom of the junction. The stem-loop of repeats is a typical holliday junction structure, so the TnpB transposon may be responsible for shuffling spacers by resolving stem-loops and then rearranging them in a different configuration.

Phylogenetic analyses have shown that the adaptation modules of CRISPR-Cas systems evolved from transposons known as Casposons. Koonin et al. proposed that the origin of CRISPR-Cas occurred through the merge of a Casposon and an ancestral Type IV-like immune system (Koonin et al., 2017a). Considering the genomic context of the Type IV-A CRISPR-Cas system, it is plausible that we have observed an intermediate phase in the merging of the

Holliday junction-specific element TnpB and Type IV-A systems. The merge could potentially lead to further co-evolution, resulting in the formation of an independent adaptive immune system.

3.2 Type IV-A CRISPR interference activity

We investigated the PAM-dependent activity of the Type IV-A CRISPR-Cas system against phages and plasmids using a *gfp* targeting assay. Our results showed only a minor difference in fluorescence intensity between the targeting of coding and non-coding strands, suggesting that DNA targeting plays a role in the interference process. If RNA targeting were involved, complementarity between the coding strand targeting crRNA and the mRNA would have impeded translation, significantly downregulating fluorescence intensity. Based on these findings, we continued to investigate the activity of this system and its critical proteins.

3.2.1 DinG

The *dinG* gene was found to be essential for interference in EOT assays (Figure 17c and d) and the absence of DinG or null mutations in the walker-A domain abolished CRISPR activity. DinG is annotated as an ATP-dependent helicase, but its function in the Type IV-A system is unclear. The Type IV-A DinG is known to contain an ATP-dependant walker-A motif and a DEAH-box (Fig 15b) (Walker et al., 1982). Studies have shown that the DEAH-box of DinG is essential for CRISPR activity (Crowley et al., 2019), and in this study, we have also shown that the walker-A motif is essential.

Previous research found that Csf5 and DinG interact in a cross-linking coupled mass spectrometry analysis, indicating possibly transient interactions with the Type IV-A complex (Özcan et al., 2019). Csf1, Casf2, Csf3, and Csf4 together with matured crRNA form a stable stand-alone complex (Figure 7) (Özcan et al., 2019). Based on this information, one possible scenario for the activity of the Type IV-A CRISPR system is that the stand-alone complex scans for a PAM-motif and complementary sequence to crRNA, and upon recognition, it opens up a seed bubble and then progresses to the formation of a full R-loop similar to the interference of Type I CRISPR-Cas systems (Xiao, Luo, et al., 2017). Subsequently, DinG comes in to stabilize the R-loop structure. In the absence of DinG, the complexes might retain their ability to scan

for targets, however, to exhibit Type IV-A activity, the effector complexes need to remain persistently bound to the target, a process where DinG is likely to play a crucial role.

In Type I CRISPR-Cas systems, R-loop formation creates a bulge in the non-target DNA strand, which triggers cleavage by Cas3 (Xiao, et al., 2017). Without disruption, the R-loop formation in fully base-paired targets remains stable, requiring more than five consecutive PAM-distal mismatches to induce R-loop collapse (Aldag et al., 2022). The lack of CRISPRi activity in Type I CRISPR-Cas systems with HD-mutant Cas3 (Spilman et al., 2013) is likely due to the vulnerability of R-loop formation to RNA polymerase. This hypothesis may also explain why DinG is critical for Type IV-A system activity. In the absence of a protein to stabilize the R-loop, Type IV-A complexes could be easily removed by RNA polymerase.

3.2.2 Type IV-A CRISPR-Cas mediated immunity against phages

The Type IV CRISPR array contains spacers that match viral sequences, suggesting their potential for immunity against targeted phages. To study the phage immunity, we transferred plasmids encoding a crRNA targeting the lambda phage gene *E*, along with a 5'-AAG-3' or 5'-CGG-3' PAM, into *E. coli* cells producing recombinant Type IV-A crRNPs. We then infected the cells with lytic lambda phages and measured plaque formation. Our results showed that the PAM was necessary for immunity against the lytic lambda phages (Figure 17c). Additionally, we found that deletion of the *dinG* gene or introduction of a null mutation (DinG K136A) eliminated the reduction in plaque formation, indicating that DinG is required for this activity.

In the lytic life cycle of lambda phage, the phage chromosome is injected in a linear form and then circularized before replication (Casjens et al., 2015). Targeting the Type IV-A system in the phage genome may impede the progression of the DNA replication fork in the phage chromosome.

3.2.3 Type IV-A mediated CRISPR interference

In our study, we investigated the *Pseudomonas* Type IV-A CRISPR system and its interference activity. We used Fluorescence Activated Cell Sorting (FACS) to quantify the green fluorescence signal of protospacer targets in the *gfp* coding and non-coding strands. Our

results revealed PAM-dependent CRISPR interference, as the green fluorescence signal was reduced only in the presence of a 5'-AAG-3' PAM.

To further explore this, we studied plasmid curing in *P. oleovorans* by transforming the cells with a target plasmid and a crRNA-production plasmid, and cultivating them without antibiotic selection. After 12 hours, we transferred the cells to a medium with antibiotics to select for both plasmids. The results demonstrated PAM-dependent plasmid curing, indicating native Type IV-A CRISPR activity can interfere with plasmids.

We then expressed the recombinant Type IV-A system in *E. coli* and found that plasmids carrying a protospacer with a 5' -AAG-3' PAM had reduced transformation efficiency, which aligns with our hypothesis that the Type IV-A system facilitates PAM-dependent interference (Figure 17).

Despite the evidence supporting the interference activity of the Type IV-A system, the lack of a putative nuclease raises questions about the interference mechanism. To investigate this, we constructed a *lacZ* reporter system and analyzed the color indication of individual colonies, linking phenotypes to genotypes. In the absence of interference, all visible colonies were blue using a non-targeting control. As a positive control, we tested a Type I-C system and observed a mix of blue and white colonies, with white indicating interference.

We then analyzed white colonies affected by Type IV-A activities. Gel electrophoresis showed no deletions (Figure 20c), and Sanger sequencing confirmed no mutations in *lacZ*, suggesting that nuclease activity is not involved in Type IV-A interference. Additionally, we observed a similar color distribution in passaged colonies (Supplementary 2)., further supporting our conclusion.

Interestingly, we observed color heterogeneity within the same colonies when *lacZ* was targeted by the Type IV-A system. This could be related to the timing and duration of CRISPR activity or the nucleoid status of the cells. In some cases, colonies were divided into two sections (Supplementary Figure 1 and 2), possibly due to different nucleoid statuses. Chromatin status has been previously reported to affect CRISPR-Cas systems in eukaryotes (Knight et al., 2015; Singh et al., 2018).

Over 50% of colonies escaped Type IV-A activity (Figure 20b), which may also be related to the cells' nucleoid status. We did not detect any mutations in the colonies that escaped CRISPR activity, and the phenotype was reset after passaging. One possible explanation for this phenomenon is the involvement of inter-cell signaling mechanisms, such as quorum sensing (Puskas et al., 1997; J. P. Ward et al., 2004), which regulate relevant processes. This is supported by the observation that the blue phenotype, indicating escape from CRISPR activity, is only inherited within colonies but reset after passaging.

Our study revealed that the Type IV-A CRISPR-Cas system exhibits dCas9-like activity, despite the differences in their molecular structure. While the Type IV-A system relies on multi-subunit effector complexes, dCas9 is a single effector protein complex (Brezgin et al., 2019). Nevertheless, both systems share similar attributes. dCas9, or catalytically dead Cas9, is derived from Cas9 by introducing null mutations into the HNH and RuvC1 domains (Qi et al., 2013). Like the Type IV-A system, dCas9 binds to DNA without altering the sequence. However, dCas9 displays a bias between template and non-template strands, whereas we did not observe strand bias in Type IV-A systems (Figure 17a and Figure 20b). This observation suggests that the two systems exhibit different binding affinities to the non-base-pairing strand.

The binding mechanism of Type IV-A system remain to be studied, however when targeted to the host chromosome, we did not observe an evident growth defect resulting from Type IV-A activity. One possible explanation is that Type IV-A effector complexes act as roadblocks for DNA replication and RNA transcription (Alberts, 1987; Doi et al., 2021; Whinn et al., 2019). The *lacZ* targeting assays show that Type IV-A-mediated CRISPR interference does not degrade target DNA while displaying a clear knockdown of *lacZ*, likely due to transcription blockage. The Type IV-A effector complex may remain bound to the target DNA and block the action of RNA polymerase. During replication, the complex may as well block the replisome's progression, leading to prolonged stalling of the replication fork and eventually the fork collapse.

There are several pathways for coping with collapsed forks in prokaryotes (Yeeles et al., 2013), but these repair mechanisms are less efficient when dealing with mobile genetic elements, which are usually much smaller than chromosomes. This may explain why Type IV-A activity

does not affect host fitness when targeted to the chromosome, but does show anti-plasmid activity.

3.3 CRISPR arrays and crRNA biogenesis of *Pseudomonas oleovorans*

Small RNA libraries from *P. oleovorans* DSM1045 were analyzed using Illumina RNA-seq technology. The results revealed that each CRISPR-Cas system exhibits crRNA maturation. The abundance of individual crRNAs varied among the different CRISPR arrays, but the large number of mature crRNA sequences suggests that all three native systems are active. Further analysis revealed that mature crRNAs had 8 nt long 5'-terminal repeat tags (Type IV: 5'-GUGAGCGG-3', Type I-E: 5'-AUGAACCG-3', Type I-F: 5'-CUCAGAAA-3'), indicating that the long CRISPR array transcript was processed at the base of hairpin structures of each repeat (Figure 10).

Mutations were detected in four repeats in the Type IV-A system, but none of these mutations occurred at the cleavage site. Despite these mutations, all twenty spacers were still processed into mature crRNA, indicating that the Cas6 enzyme is tolerant to mutations (Table 3).

Table 3. List of DNA sequences of mutated *Pseudomonas* Type IV-A repeat

	DNA sequence
Consensus	5'-GTATTTCCCGCGTGCGCGGGGGTGAGCGG-3'
Repeat8	5'-ATATTTCCCGCGTGCGCGGGGGTGAGCGG-3'
Repeat14	5'-GTATTTCCCGCGTGCGCGGGGGTAAGCGG-3'
Repeat15	5'-GTGTTGTCCACCTGCGTGGAGGTATTCGG-3'
Repeat16	5'-GTGTTCCCCGCGC.CGTAGGGGTAAGCGG-3'
Repeat17	5'-GTATTCCCTGCGCAAGCGGGGGTGAACGG-3'

3.4 Investigation of PAM requirements

Discriminating self from non-self is a crucial aspect of immune systems, and PAM recognition is a fundamental mechanism used to avoid self-targeting in CRISPR-Cas adaptive immune systems (Westra et al., 2013b). Through *in silico* analysis, we identified the 5'-AAG-3' PAM for the *Pseudomonas* Type IV-A system.

We further explored the PAM requirement by conducting PAM scanning assays, which revealed two structurally distinct PAMs, 5'-GTG-3' and 5'-AAN-3', indicating flexibility in the PAM recognition pocket (Figure 18). Given that the interference of the system does not require DNA degradation, it is reasonable to assume a relaxed PAM requirement.

3.5 Evolutionary advantage of Type IV-A CRISPR-Cas systems

3.5.1 Evolutionary advantage of Type IV-A CRISPR-Cas systems for carrier plasmids

Type IV-A CRISPR-Cas systems are commonly found in plasmids. Studies have demonstrated that identical Type IV-A systems are frequently present in multiple strains of the same species (Newire et al., 2020; Pinilla-Redondo et al., 2020)likely due to the horizontal gene transfer of these plasmid vehicles. Furthermore, it has been established that Type IV-A systems exhibit bias in targeting plasmids that are present intracellularly (Newire et al., 2020; Pinilla-Redondo et al., 2020). The co-evolution of carrier plasmids and Type IV-A systems aligns with the "plasmid warfare" scenario, supported by the observed Type IV-A CRISPR anti-plasmid activity.

In general, Type IV-A bearing plasmids are classified as megaplasmids and belong to IncP-9 family, capable of carrying multiple copies of themselves within a single cell, providing them with a competitive advantage over smaller plasmids (Hall et al., 2022). Catabolic plasmids, such as those in the IncP-9 family, are characterized by high modularity, with the plasmid backbone containing essential metabolic gene cassettes. These plasmids typically contain a transposon element region that allows for size flexibility, with fewer essential components being eliminated to reduce the fitness costs of the plasmid (Dennis, 2005; Thomas, 2000). Notably, the genomic context of the Type IV-A system is characterized by a high concentration of transposon elements, which suggests that the system is not located in a dispensable genomic island of the IncP-9 plasmid.

Our findings also suggest that there may be a functional association between the Type IV-A system and the Type IV-A secretion system, as the first spacer of the Type IV-A targets and downregulates an essential gene *pilN* of T4SS.

One critical aspect of megaplasmid biology is the competition among different megaplasmids within a bacterial population. This competition is driven by the limited resources available within a cell, such as energy and replication machinery (Baltrus et al., 2021; Hall et al., 2022; Morton et al., 2014; Platt et al., 2014; Romanchuk et al., 2014; Zhang et al., 2011). In order to ensure survival, megaplasmids must outcompete one another for these resources. The

utilization of Type IV-A anti-plasmid activity to eliminate potential competition would grant a significant advantage to the hosting plasmid.

Plasmid competition occurs when two or more plasmids coexist within a bacterial cell or population and compete for resources and replication opportunities. Plasmids can carry various benefits to their host bacteria, such as antibiotic resistance, virulence factors, and metabolic advantages. However, carrying and maintaining plasmids also imposes a fitness cost on the host cell, as the plasmids consume energy and resources that could otherwise be used for growth and reproduction (Dennis, 2005; Friehs, 2004; Thomas, 2000). Plasmids use several mechanisms to ensure their stable maintenance and transmission within bacterial populations, including the toxin-antitoxin system and entry exclusion.

Targeting *pilN* may represent a mechanism to prevent the entry of other mobile genetic elements, analogous to the phenomena of entry exclusion or phage superinfection (Garcillán-Barcia et al., 2008; Pope et al., 2011). *PilN* is an essential component of the Type IV pilus (T4P) and also part of the Type IV secretion system, which is important to prokaryotic conjugation machinery (de Smet et al., 2017).

Our RT-qPCR analysis revealed a significant upregulation of *pilN* transcription in the CRISPR knockout strain of *P. oleovorans* compared to the wildtype (Guo et al., 2022). This suggests that targeting the Type IV-A system downregulates *pilN* expression and likely inhibits conjugation and T4P-dependant phage invasion.

3.5.2 Evolutionary advantage of Type IV-A CRISPR-Cas systems for bacteria hosts

According to a popular theory, CRISPR-Cas systems evolved by repeatedly incorporating mobile genetic element genes (Koonin et al., 2019; Makarova et al., 2019). Evidence suggests that class I CRISPR-Cas systems have evolved from a Type III-like ancestral system (Makarova et al., 2019; Moya-Beltrán et al., 2021). It is believed that Type IV systems evolved from Type III systems through a series of gene loss events (Koonin et al., 2019; Moya-Beltrán et al., 2021). Type IV-C is thought to be the ancestral Type IV system. The recruitment of Cas6 and a CRISPR array is a major event in the evolution of Type IV systems, which marked the emergence of Type IV-A systems (Figure 22) (Moya-Beltrán et al., 2021).

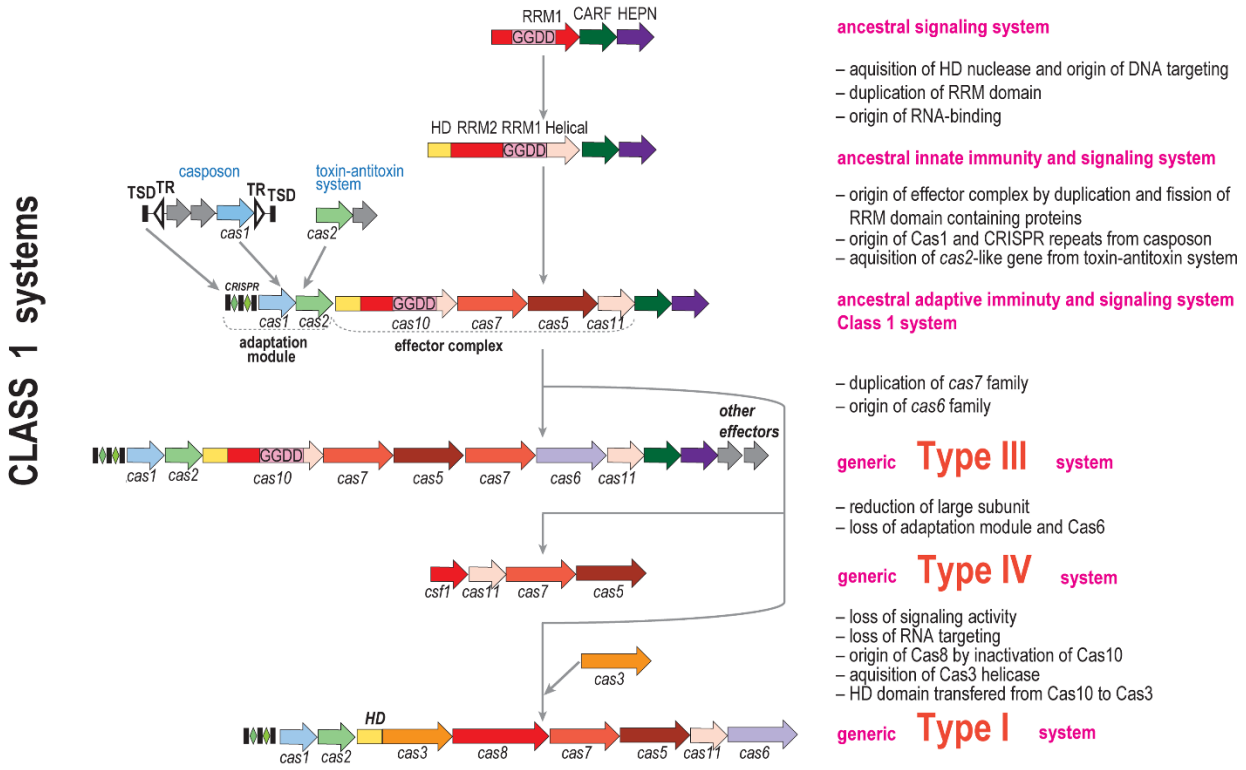


Figure 20. Origins and evolution of Class 1 CRISPR-Cas systems. The diagram illustrates the hypothetical evolutionary pathways for Class 1 CRISPR systems and their derivatives, with genes represented by block arrows (not drawn to scale) and color-coded to indicate protein and domain families. The evolutionary events that may have played a role in each step are briefly explained to the right of the diagram. Adapted from (Koonin et al., 2022).

As previously discussed, Type IV-A systems play a role in megaplasmid competition. However, Type IV-A systems can also confer an advantage to the hosting bacteria as an independent gene cassette. The environment selection process performs real-time calculations of different genetic elements' fitness costs and benefits. Elements with a high fitness cost and do not offer equivalent benefits would be subject to selection pressure. (Baltrus, 2013; Knöppel et al., 2014; Vogwill et al., 2015). Our small RNA-seq analyses revealed a significant percentage of matured Type IV-A crRNAs (Figure 10). A previous study has shown that matured crRNAs are associated with Cas proteins (Özcan et al., 2019), and considering the number of crRNAs detected in the host, the fitness cost of carrying an active Type IV-A system is not negligible.

A hypothetical scenario is that Type IV-A systems are involved in the regulation of the host at transcription and perhaps translation levels, as we cannot rule out potential RNA interactions of the systems. The *pilN* targeting spacer in *P. oleovorans* supports this scenario, as this

targeting likely regulates the genesis of T4P on the cell membrane, enabling the host to evade phage invasion.

The detection of new spacers from deep sequencing (Table 2) may suggest RNA interactions. These spacers are likely generated internally, as no external source of protospacers was introduced before sequencing. However, the complementary sequences of these spacers could not be located in the host genome, which might imply potential interactions with RNA secondary structure.

Type IV-A systems are known to be carried by plasmids and are typically associated with Type I systems (Moya-Beltrán et al., 2021; Pinilla-Redondo et al., 2020; H. N. Taylor et al., 2021). In such a context, additional canonical CRISPR immunity may no longer be essential, and a less aggressive immune system may be advantageous (Pinilla-Redondo et al., 2020). While the invasion of mobile genetic elements (MGEs) comes with a fitness cost, it also brings genetic diversity and beneficial genes (Baltrus, 2013; Knöppel et al., 2014; Vogwill et al., 2015). The biological functions we have observed so far suggest the possibility of harnessing MGEs through the regulation of Type IV-A activity.

3.6 Applications and outlook

Our study has provided new insights into the potential utility of Type IV-A CRISPR-Cas systems as inhibitors of DNA replication and RNA transcription. This novel approach complements existing techniques, such as the use of nuclease-dead Cas9 (dCas9) (Whinn et al., 2019). However, further research is needed to compare these two systems' behaviors systematically.

Despite their large size, which may present challenges for pharmaceutical delivery, Type IV-A systems may still be valuable tools for regulating gene expression in bacteria, archaea, and plants. Further studies are needed to investigate the specific mechanisms of action and potential applications of Type IV-A systems in these organisms. We have presented a hypothesis for the potential role of DinG in DNA metabolism, but further experimental data are required to confirm its actual function. In order to fully understand the mechanism of action of Type IV-A systems, it will be necessary to conduct studies to investigate how these complexes recognize, unwind and bind to target DNA. Additionally, research is needed to

determine the binding affinity and binding endurance of the effector complex, as well as its specificity.

We have not yet identified evidence of Type IV-A systems interacting with RNA. However, it is important to note that collateral RNA activity has been observed in Cas12 and Cas13 (Bot et al., 2022; Nguyen et al., 2020; Tong et al., 2023; Yang et al., 2022). Consequently, additional research is necessary to investigate the potential for nonspecific RNase activity in Type IV-A systems.

Based on our results from *in vivo* experiments, Type IV-A CRISPR-Cas systems are capable of interfering with DNA sequences based on their associated crRNA. So far targeting DNA required 32 bp complementarity between crRNA and DNA, along with a 5′-AAN-3′ PAM sequence.

Studies are still needed to determine the minimal requirement of complementarity length between crRNA and DNA. In a similar system dCas9, the requirement for complementarity is 5-10 bp (Boyle et al., 2017), *in vivo gfp* targeting assays can be conducted to determine the minimal active length of crRNA and complementarity requirement of Type IV-A systems. The binding event of dCas9 disrupts the transcription process, resulting in its premature termination (D. L. Jones et al., 2017; Qi et al., 2013). Gaining structural insights into Type IV-A effector complexes binding to DNA targets is essential for a deeper understanding of the binding mechanism.

Furthermore, by targeting various regions within the -10 to -35 promoter core, the system may achieve distinct levels of repression, enhancing the flexibility of potential targets. Type IV-A systems can also be designed to target diverse operators, such as repressors, activators, or enhancers, and to inhibit their influence on gene expression. Numerous molecular tools have been developed to modulate gene expression, visualize specific DNA sequences, or even modify epigenetic marks by fusing dCas9 with other functional domains (Y. Hu et al., 2022; Morgan et al., 2017). Similar engineering strategies could be employed with Type IV-A systems as a complement to the dCas9 toolbox. For example, fusing a Type IV-A system with a Cas7 oligomer (Kalwani et al., 2020; Özcan et al., 2021) may enhance transcriptional blockade. Finally, there is a need for fine-tuning the targeting mechanism, understanding how

the complex copes with multiple copy number plasmid targets, and how target escape occurs. These studies will be crucial in order to fully understand the potential applications of Type IV-A systems in metabolism and regulation.

In summary, our study demonstrates the potential application of Type IV-A CRISPR-Cas systems in impeding DNA replication and regulating RNA transcription. This innovative approach offers complementation to the established dCas9 method. Our hypothesis for the role of DinG in the Type IV-A activity requires additional experimental validation. A comprehension of the molecular mechanisms underlying Type IV-A systems is also crucial for future studies. Moreover, a thorough exploration of adaptation mechanisms and the potential of RNA activity is essential. Overall, our results indicate that Type IV-A systems provide a promising direction for future research and development as a gene regulation tool.

4. Material and methods

4.1 Materials, instruments and source of supplies

4.1.1 Chemicals, Kits and enzymes

The chemicals, kits and enzymes used in this work were obtained from the companies listed in Table 3.

Table 4. List of special chemicals and reagents used in this work.

Label	Manufacturer
2-log DNA ladder (0.1-10.0 kb)	New England Biolabs, Frankfurt
ATP	Thermo Fisher Scientific Ltd. & Co. KG, Bonn
Antibiotics (kanamycin, ampicillin, spectinomycin, chloramphenicol)	Sigma-Aldrich, Taufkirchen ; Roth GmbH, Karlsruhe
Bovine serum albumin (BSA)	Sigma-Aldrich, Taufkirchen
Diethyl Pyrocarbonate (DEPC)	AppliChem GmbH, Darmstadt
Dimethyl Sulfoxide (DMSO)	Sigma-Aldrich, Taufkirchen
DNA Oligonucleotides	Eurofins MWG operon, Ebersberg
dNTP Mix	New England Biolabs, Frankfurt
Ethidium Bromide	Roth GmbH, Karlsruhe
Gelpilot DNA Loading Dye (5x)	Qiagen GmbH, Hilden
Glycogen	Roche Diagnostics GmbH, Mannheim
Isopropyl-ß-thiogalactopyranoside (IPTG)	Roth GmbH, Karlsruhe
Low Molecular Weight Marker	Affymetrix/USB™
Low Molecular Weight DNA Ladder	New England Biolabs GmbH, Frankfurt
Low Range ssRNA Ladder	New England Biolabs GmbH, Frankfurt
NEBNext® Small RNA Library Prep Set	New England Biolabs GmbH, Frankfurt
NEB Builder HiFi DNA Assembly Master Mix	New England Biolabs GmbH, Frankfurt
NTPs (ATP/GTP/CTP/UTP)	Jena Bioscience GmbH, Jena
Phusion DNA polymerase	Thermo Fisher Scientific Ltd. & Co. KG, Bonn
QIAprep Spin Miniprep Kit	Qiagen GmbH, Hilden
QIAquick Gel Extraction Kit	Qiagen GmbH, Hilden
QIAquick PCR Purification Kit	Qiagen GmbH, Hilden
Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific Ltd. & Co. KG, Bonn
Qubit RNA HS Assay Kit	Thermo Fisher Scientific Ltd. & Co. KG, Bonn
Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)	New England Biolabs GmbH, Frankfurt
Quick-Load® pBR322 DNA-MspI Digest	New England Biolabs GmbH, Frankfurt
Restriction endonucleases	New England Biolabs GmbH, Frankfurt
RNase If	New England Biolabs GmbH, Frankfurt
RNase Inhibitor (murine)	New England Biolabs GmbH, Frankfurt
Sodium dodecyl sulfate (SDS)	Roth GmbH, Karlsruhe
SYBR Gold® Nucleic acid stain	Thermo Fisher Scientific Ltd. & Co. KG, Bonn
T4 DNA Ligase	New England Biolabs GmbH, Frankfurt
T4 Polynucleotide Kinase	New England Biolabs GmbH, Frankfurt

T7 RNA Polymerase	New England Biolabs GmbH, Frankfurt
Taq DNA Polymerase	New England Biolabs GmbH, Frankfurt
Topo ® TA cloning ®	Thermo Fisher Scientific Ltd. & Co. KG, Bonn

4.1.2 Instruments

Table 5. List of Instruments used in this study

Instrument	Model and Company
Agarose gel electrophoresis	Chambers and Casting trays produced by company technician Philipps-University Marburg Power supply Consort E835; MS Laborgeräte, Dielheim
Aqua bidest, water system	PURELAB Plus, ELGA LabWater, Celle
Autoclave	5075 EL, Tuttnauer Europe B.V., Breda, NL
Bioanalyzer	Agilent 2100 Bioanalyzer, Agilent, Santa Clara, CA, USA
Centrifuges	Centrifuge 5424, Eppendorf AG, Hamburg; Sigma 3-30K, Sigma Laborzentrifugen GmbH, Osterode am Harz; Sorvall Lynx 4000, Thermo Fisher Scientific Ltd. & Co. KG, Bonn
Denaturing polyacrylamide gel electrophoresis	PROTEAN II Electrophoresis Chamber, BioRad Laboratories GmbH, Munich
Hybridization oven	Hybrid Shake 'n' Stack, Thermo Fisher Scientific Ltd. & Co. KG, Bonn
Incubators	KB53, Binder GmbH, Tuttlingen
Magnetic stirrer	IKA® RCT Standard, IKA®-Werke GmbH & Co. KG, Staufen
Magnetic Separation Rack	2-Tube Magnetic Separation Rack; New England Biolabs GmbH, Frankfurt
Microscope	Axioplan 2, Carl Zeiss Microscopy GmbH, Göttingen; CoolSnap HQ camera, Visitron Systems GmbH, Puchheim; FluoArc HBO Lamp, Carl Zeiss Microscopy GmbH, Göttingen
Microfluidizer	Microfluidics LM10, Sysmex Deutschland GmbH, Norderstedt.
MinIon	Oxford Nanopore Technologies, UK
MiniSeq	Illumina , Inc. USA
Nanodrop	NanoDrop® ND-1000 Spectrometer, Thermo Fisher Scientific Ltd. & Co. KG, Bonn
PCR-Cycler	C1000TM Thermal Cycler, BioRad Laboratories GmbH, Munich
Peristaltic pump	Peristaltic Pump P-1, GE Healthcare Europe GmbH, Freiburg
pH-meter	INOLAB pH level 1, WTW, Weilheim
Phosphorimager	Storm 840 phosphorimager, Molecular Dynamics, GE Healthcare Europe GmbH, Freiburg
Qubit Fluorometer	Qubit 2.0, Thermo Fisher Scientific Ltd. & Co. KG, Bonn
Rocker	Gyrorocker SSL3, Sigma-Aldrich, Taufkirchen
SDS polyacrylamide gel electrophoresis	Mini-PROTEAN Tetra Cell, Bio-Rad Laboratories GmbH, Munich; Power supply PowerPac Basic, Bio-Rad Laboratories GmbH, Munich
Spectrophotometer	Ultrospec 3000 *pro*, GE Healthcare Europe GmbH, Freiburg
Thermomixer	Thermomixer Comfort 5350, Eppendorf AG, Hamburg
Thermoshaker	HT Thermotron, Infors AG, Bottmingen, Switzerland
Vortex Mixer	Vortex Genie 2, Scientific Industries, Bohemia, NY, USA

4.1.3 Strains and culture conditions

Table 6. Bacterial strains used in this study

Strain	Description	Source
Escherichia coli K12 DH5α	F- Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(rK–, mK+) *pho*A *sup*E44λ–*thi*-1 *gyr*A96 *rel*A1	(Hanahan, 1983)
Escherichia coli Rosetta2 (DE3) pLysS	F- *ompT hsdS*B(rB- mB-) *gal dcm* (DE3) pLysSRARE2 (CamR)	Novagen, Darmstadt
Escherichia coli BL21(DE3) pLsyS	F-ompT hsdSB(rB- mB-) gal dcm (DE3) pLysS (SpecR)	Novagen, Darmstadt
Pseudomonas oleovorans DSM1045	wt	DSMZ-German Collection of Microorganisms and Cell Cultures GmbH
Escherichia coli WM3064	thrB1004 pro thi rpsL hsdS lacZΔM15 RP4-1360 Δ(araBAD)567 ΔdapA1341::[erm pir]	American Type Culture Collection

4.1.3.1 *Escherichia coli* growth conditions

E. coli cultures were grown in LB medium (1 % tryptone (w/v), 0.5 % yeast extract, 1 % NaCl (w/v), pH 7.2) in a rotatory shaker at 200 rpm at 37°C or on solid medium plates (LB medium containing 1.5 % (w/v) agar-agar). Single colonies were inoculated with a pre-culture (2% (v/v)) which contain LB medium with appropriate antibiotics (spectinomycin 100 µg/ml, kanamycin 50 µg/ml, ampicillin 100 µg/ml and chloramphenicol 34µg/ml) based on plasmid encoded antibiotic resistance gene.

E. coli DH5α was used for cloning procedures. This strain transforms with high efficiency and has a number of features useful for cloning. *E. coli* BL21 (DE3) was used for expression cultures. This strain features the gene for expression of the T7 polymerase as well as spectinomycin resistance. Overexpression by the T7 promoter on transformed plasmids is repressed until IPTG induction from a lac promoter. *E. coli* WM3064 was used for conjugation. *E. coli* WM3064 cultures were grown at 37°C in LB media supplemented with diaminopimelic acid (DAP).

4.1.3.2 *Pseudomonas oleovorans* DSM1045 growth conditions

P. oleovorans DSM 1045 cultures were grown in LB medium (1 % tryptone (w/v), 0.5 % yeast extract, 1 % NaCl (w/v), pH 7.2) in a rotatory shaker at 200 rpm at 37°C or on solid medium plates (LB medium containing 1.5 % (w/v) agar-agar).

4.1.4 Oligonucleotides, plasmids and constructed recombinant vectors

Plasmids and constructed recombinant vectors

Table 7. Plasmids used in this study.

Plasmid	Features	Source
pEMG	Suicide vector used for deletions and insertions in Gram-negative bacteria; oriT, traJ, lacZα, oriV(R6K); KmR	(T. Wang et al., 2019)
pHERD30T	pUCP30T derivative plasmid; ori (pBR322); araC-PBAD cassette; GmR	(Qiu et al., 2008)
pUCP18	pUC derivative plasmid; ori(pMB1); KanR	(West et al., 1994)
pUC19	pUC derivative plasmid; ori(pMB1); AmpR	NEB
pET-Duet1	AmpR	Novagen
pRSF-Duet1	Duet MCS, KanR	Novagen
pCDF-Duet1	Duet MCS, SpecR	Novagen

Table 8. List of spacers used for Type IV-A CRISPR-Cas targeting

Spacer sequence	Assay	Description	Plasmid	Restriction sites for insert
ACCAGAATCGGCACAACGCCGGTAAACAGTT	Fluorescence intensity measurement/ Plasmid curing	PAM: AAG, protospacer on coding strand	pUCP18	EcoRI & NcoI
CTGGTGACCACCCTGACCTATGGCGTTCAGTG	Fluorescence intensity measurement	PAM: AAG, protospacer on non-coding strand	pUCP18	EcoRI & NcoI
CCACGGAACCGGCAGTTTACCGGTGGTGCAAA	Fluorescence intensity measurement/ Plasmid curing	PAM: CGG (non-functional)	pUCP18	EcoRI & NcoI
CGTTGTGCCGATTCTGGTGGAACTGGATGGTG	Fluorescence intensity measurement/ Plasmid curing	non-targeting negative control	pUCP18	EcoRI & NcoI
ACAGGCGGCAGTAAGGCGGTCGGGATAGTTTTG	LacZ targeting assay	PAM: AAG, protospacer on coding strand in lacZ	pCDFDuet-1	EcoRI & NcoI
CTGCCATTGTCAGACATGTATACCCCGTACGTG	LacZ targeting assay	PAM: AAG, protospacer on non-coding strand in lacZ	pCDFDuet-1	EcoRI & NcoI
CATAAAGTGTAAAGCCTGGGGTGCCTAATGAG	LacZ targeting assay	PAM: AAG, protospacer in promoter region of lacZ	pCDFDuet-1	EcoRI & NotI
TGTGAGCGGATAACAATTTCACACAGGAAACA	LacZ targeting assay	PAM: AAT, protospacer in UTR of lacZ	pCDFDuet-1	EcoRI & NotI
ATCGCACTCCAGCCAGCTTTCCGGCACCGCTT	LacZ targeting assay	PAM: AAG, protospacer on non-coding strand in lacZ	pCDFDuet-1	EcoRI & NotI
GAAATCCATTATGTACTATTTAAAAAACACAA	LacZ targeting assay	PAM: AAG, protospacer downstream of lacZ	pCDFDuet-1	EcoRI & NotI
TCAACTCGCACTGTGAGGGTCACATGGGCGTT	LacZ targeting assay/phage assay	non-targeting negative control	pCDFDuet-1	EcoRI & NcoI
TCTATCTCTCACAAATTCCGGGACTGGTAAAC	Phage assay	PAM: AAG, gene *E*, coding strand	pCDFDuet-1	EcoRI & NcoI
AGAAAGTCTATCTCTCACAAATTCCGGGACTG	Phage assay	PAM: CGG, gene *E*, coding strand	pCDFDuet-1	EcoRI & NcoI

CATCCAAGTTACGCATCAGATTCGAGACGCGA	Efficiency of transformation assays	PAM: AAG	pCDFDuet-1	NcoI & BamHI

4.2 Working with DNA

4.2.1 Quantification of DNA

4.2.1.1 Spectrophotometric quantification

The concentration of DNA was determined by measuring the absorbance at 260 and 280 nm. The purity was determined by an ratio of absorbance at 260nm/280nm.

4.2.1.2 Fluorometric quantification

The Qubit fluorometer was used for quantification of low-concentrated DNA. After cDNA library preparation, samples were quantified via the Qubit dsDNA HS Assay Kit following manufacturer's protocol.

4.2.2 Agarose gel electrophoresis of DNA

DNA molecules were separated by length by agarose gel electrophoresis. The concentration of agarose in a gel will depend on the sizes of the DNA fragments to be separated, with most gels ranging between 0.5%-2% in TAE buffer (40 mM Tris-acetate, 1 mM EDTA pH 8) and 0.5 µg/mL ethidium bromide. DNA samples were mixed with 6x DNA Gel Loading Dye (Thermo Scientific™) in a 1:6 ratio. Electrophoresis was performed at 120 V at RT in TAE buffer. The DNA was visualized by UV irradiation at 254 nm.

4.2.3 Purification of DNA

4.2.3.1 PCR Purification

Residues in PCR reactions could affect downstream reaction were removed with the QIAquick PCR Purification Kit according to the manufacturer´s instructions. Theoretically, PCR products in the size range of 100 bp -10 kb can be recovered with this kit.

4.2.3.2 Gel extraction from agarose gels

After cutting out the fragment of interest on agarose gel, DNA molecules were recovered by the QIAquick gel extraction kit (Qiagen GmbH) following the instructions of the manufacturer.

4.2.3.3 Gel extraction from polyacrylamide gels

The gel pieces were then transferred to a Corning™ Costar™ Spin-X™ Gel Breaker tube and centrifuged (14,600 rpm, 2 min, RT) into a 2 ml collection tube. 500 µl gel elution buffer (20 mM Tris-HCl pH 7.5, 250 mM sodium acetate, 1 mM EDTA, 0.25 % SDS) were added on the gel debris and the mixture was incubated overnight on ice while shaking (300 rpm). Following this, the DNA containing gel elution buffer was transferred to a Costar® centrifuge filter tube and centrifuged (14,600 rpm, 2 min, RT) to remove remaining gel debris. The DNA was subsequently purified with EtOH precipitation.

4.2.4 Polymerase chain reaction (PCR)

Two primers flanking the sequence of interest were designed. The elongation process was carried out by Phusion polymerase or Taq polymerase. A standard PCR reaction included the following four main stages: I) Denaturation: Heating the reaction at 95°C results in the melting of dsDNA into ssDNA (template). II) Primer annealing: Annealing or binding of the primers to their complementary DNA. III) Elongation: Extension or elongation of the primer in the 5’ to 3’ direction. DNA polymerase catalyzes the elongation by addition of complementary nucleotides. The above-listed steps were repeated to achieve sufficient amplification (25-30x).

PCR amplifications from genomic or plasmid DNA were performed using the following reaction mixture: ~ 50 ng template DNA, 250 µM dNTPs, 0.2 µM of each primer, 1 x concentrated GC buffer 3% (v/v) DMSO, 1 U Phusion polymerase and adjusted to 50 µl with water. The reaction was performed in a thermal cycler (BioRad) using the following program:

Step 1) 95°C – 60 sec
Step 2) 95°C – 30 sec
Step 3) 55-65°C – 30 sec x 30 – 35
Step 4) 72°C – 30 sec/kb
Step 5) 72°C – 5 min

4.2.5 Modification of DNA

4.2.5.1 Restriction

The restriction digestion of DNA molecules was achieved with appropriate restriction endonucleases in respective buffers according to the manufacturer's instructions. The reaction mixture containing 5-10 U enzyme/µg DNA was incubated at 37 °C for 2 h to digest the DNA.

4.2.5.2 Ligation

T4 DNA ligase was used for ligation of restricted plasmid DNA in the appropriate buffer according to the manufacturer's instructions. In a standard ligation reaction, 0.02 pmol vector DNA was mixed with 0.06 pmol insert DNA (ratio 1:3) and 4 U T4 DNA ligase in the recommended DNA ligase reaction buffer containing ATP. Phosphorylated inverse PCR products were self-ligated by addition of 10 U T4 DNA ligase to the phosphorylation reaction. The reactions were incubated overnight at 16°C and subsequently used for transformation with *E. coli* DH5α.

4.2.5.3 Gibson Assembly

The Gibson Assembly technique the merge of DNA fragment without the need of restriction site. This technique requires a minimum 15 nt overlap for the assembly. During isothermal conditions, a T5 exonuclease degrades dsDNA in 5' to 3' direction, resulting in long 3'-overhangs which bind to the complementary overhangs of the neighboring DNA fragment. DNA polymerase subsequently fills up the single-stranded DNA by incorporating the complementary nucleotides. The resulting gaps are afterwards filled up by a DNA ligase. A self-made Gibson Assembly reaction mix was used, which did not contain the DNA ligase making use of *E. coli* own ligase instead. This Hot Fusion reaction mix was proved to contain a higher assembly efficiency than the original Gibson Assembly mix. A typical reaction contained 100 fmol of PCR product, 1.5 U T5 exonuclease and 20 U Phusion DNA polymerase in pre-assembly buffer (100 mM Tris pH 8.0, 10 mM MgCl2, 200 µM dNTPs, 10 mM DTT and 5 % (v/v) PEG-8000) and was incubated for 1 h at 50°C. The reaction was afterwards transformed into *E. coli*.

4.2.6 Conjugation of *P. oleovorans*

Genetic constructs for genome editions in the *P. oleovorans* strains were delivered via conjugation following an adapted protocol from (Wirth et al., 2020) and using the helper strain *E. coli* WM3064. As this strain is DAP-auxotroph, 0.3 mM DAP was added to the LB media and agar plates, allowing the maintenance of the *E. coli* strain during the first step of conjugation. The absence of DAP allowed the elimination of *E. coli* and visualization of only *P. oleovorans* transformed cells. For the first event of conjugation 1 ml and 0.5 ml of a fresh overnight culture of *P. oleovorans* and *E. coli* WM3064, respectively, were harvested separately by centrifugation at 8000 rpm for 3 min at room temperature. Cell pellets were washed twice with LB supplemented with 0.3 mM DAP (LB-DAP) and finally resuspended in a total volume of 100 µl of LB-DAP. The whole suspension was pipetted as a single drop onto a LB-DAP agar plate. Plates were incubated for five to seven hours at 37ºC. After incubation, cells were harvested by adding 2 ml of plain LB medium to the plate and scraping the agar with an inoculation loop. Cells were washed two times with 1ml of plain LB medium to remove traces of DAP. Finally, cells were resuspended in 1 ml of LB medium and serial dilutions of 10^{-1} and 10^{-2} were plated onto agar plates supplemented with 50 µg/ml kanamycin or 30 µg/ml gentamicin. A negative control was considered as the *P. oleovorans* strain before conjugation. Plates were incubated at 37ºC for 36 hours.

4.3 Working with RNA

4.3.1 RNA extraction

RNA was extracted from cell pellets by *mir*Vana™ miRNA Isolation Kit followed by EtOH precipitation following manufacturer's guidence. 2 volumes of ethanol and 0.3 M Na-acetate were added to the solution and the mixture was incubated at -20 °C for 1 h. The sample was then centrifuged (12,000 x g, 10 min and 4 °C), the supernatant removed and the pellet washed by addition of 1 volume of EtOH and repeated centrifugation. After removal of the supernatant, the pellet was air-dried and then resuspended in DEPC-ddH_2O.

4.3.2 Spectrophotometric quantification of RNA

Quantification and quality control of extracted RNA was performed by Nanodrop spectrophotometer as described for DNA.

4.3.3 Fluorometric quantification

Low-yield RNA preparations were quantified with the Qubit fluorometer (see section 4.2.1.2). Quantification was performed using the Qubit RNA HS Assay Kit according to the manufacturer´s instructions.

4.3.4 Agarose gel electrophoresis of RNA

Larger RNA molecules (>100 nt) were separated on agarose gels (1 %, 1x TBE) as described in section 4.2.2.

4.3.5 Illumina RNA Sequencing

P. oleovorans DSM 1045 cells were grown until OD_{600}=1. To analyse crRNA processing, small RNAs were extracted and enriched with mirVana™ isolation kit, treated with DNase I (NEB), end-repaired with T4 Polynucleotide Kinase (NEB), and submitted to library preparation using a NEBNext® Ultra™ RNA Library Prep Kit for Illumina®, following the manufacturer's instructions. For the transcriptomic analysis of the *P. oleovorans* DSM 1045 Sequencing was performed using an Illumina® MiniSeq™ System in pair-end mode generating150 nt reads. Data quality was analyzed using FastQC, reads were trimmed with Cutadapt, and aligned to the genome of *P. oleovorans* DSM 1045 using Hisat2 (Kim et al., 2019; Martin, 2011; C. M. Ward et al., 2018). Data analysis, coverage plots, and scatter plots were generated using the R packages ggplot2 and DEseq2 (Love et al., 2014; Villanueva et al., 2019). NCBI Sequence Read Archive database

4.4 *In vivo* assays

4.4.1 Fluorescence-activated cell sorting

P. oleovorans cells harboured a pHERD30T vector expressing sfGFP and a pUCP18 vector with a minimal CRISPR array. Different constructs contained spacers targeting the coding strand of gfp with a 5'-AAG-3' PAM or a 5'-CGG-3' PAM, the non-coding strand of gfp with a 5'-AAG-3' PAM, or a non-targeting crRNA. Individual colonies were cultivated in LB medium overnight at 37°C. The cultures were washed twice and diluted 100 times with phosphate-buffered saline solution. Fluorescence intensity measurements were conducted using a BD Fortessa flow cytometer and GFP was excited by the 488 nm laser line. For each sample, 10,000 events

were recorded and the ungated average fluorescence intensity of each measurement was recorded. Data were analysed with BD FACSDiva 8.0.1.

4.4.2 Transformation efficiency assays

E. coli BL21-AI cells were transformed with plasmids enabling production of recombinant Type IV-A crRNPs. A pETDuet-1 contained all five Type IV-A cas genes. Individual mutants (ΔDinG, DinG K136A, ΔCsf1, Csf1 C30A, Csf1 C33A, Csf1 C66A or Csf1 C69A) were created via Quikchange mutagenesis. A second plasmid, pRSFDuet-1 carried a minimal CRISPR array with crRNA1 from *P. oleovorans*. *E. coli* cells producing Type IV-A crRNP variants were then transformed with a target pACYCDuet-1 vector carrying a perfectly complementary protospacer1 against crRNA1 with a 5'-AAG-3' PAM. A pCDFDuet-1 vector with a protospacer1 and a 5'-CGG-3' PAM or a non-base-pairing protospacer served as controls. Transformation efficiency was calculated with the formula: Transformation efficiency = CFU (sample) / CFU (non-matching spacer control). To identify functional PAM elements, protospacer1 sequences were synthesized with different 3 nt PAM combinations and cloned into vector pCDFDuet-1. These vectors were transformed into *E. coli* BL21-AI cells containing recombinant wildtype Type IV-A crRNPs and the transformation efficiency was recorded as described above.

4.4.3 FACS measurements

P. oleovorans cells harbored a pHERD30T vector expressing sfGFP and a pUCP18 vector with a minimal CRISPR array. Different constructs contained spacers targeting the coding strand of *gfp* with a 5'-AAG-3' PAM or a 5'-CGG-3' PAM, the non-coding strand of *gfp* with a 5'-AAG-3' PAM, or a non-targeting crRNA (Table 7). Individual colonies were cultivated in LB medium overnight at 37°C. The cultures were washed twice and diluted 100x with phosphate-buffered saline solution. Fluorescence intensity measurements were conducted using a BD Fortessa Flow Cytometer and GFP was excited by the 488 nm laser line. For each sample, 10,000 events were recorded and the ungated average fluorescence intensity of each measurement was recorded.

4.4.4 Bacteriophage plaque assay

E. coli BL21-AI cells producing all Type IV-A Cas proteins were transformed with a pCDFDuet-1 plasmid carrying a minimal CRISPR array with a spacer targeting the coding or non-coding

strand of lambda phage gene *E*. A pCDFDuet-1 vector with a minimal CRISPR array carrying a random spacer sequence was transformed in control experiments. Individual colonies were inoculated in LB medium for overnight incubation and 100 µl of the overnight culture was pre-incubated with 10 µl of lambda phage (titer 1.2 x 10^7 PFU/ml). After incubation for 10 minutes, cultures were mixed with 3 ml of selective 0.7 % soft LB agar supplemented with 2 mM $MgCl_2$. The mixture was transferred onto a plain LB agar plate containing 1 mM IPTG, spectinomycin (100 µg/ml) and ampicillin (100 µg/ml). Cells and phages were co-incubated at 30°C for 10 hours and plaques were counted.

4.4.6 CRISPRi assays

Genomic *lacZ*-targets in *E. coli* - *E. coli* BL21-AI cells producing all Type IV-A Cas proteins expressed in pETDuet-1 were transformed with a pCDFDuet-1 vector containing a minimal CRISPR array with a spacer targeting the coding or non-coding strand of lacZ (Table 7) or a pCDFDuet-1 vector carrying a random sequence. As control a pCas3cRh vector encoding a complete Type I-C CRISPR-Cas system and a crRNA targeting lacZ were used as detailed before (Csörgő et al., 2020). After transformation, cells were transferred onto LB agar plates containing 0.005 % X-gal, 0.2 % arabinose and 1 mM of IPTG. After overnight incubation at 37°C, images of plates were captured and analyzed with OpenCFU (Geissmann, 2013). For recognition of white colonies, its color filter was set to a hue angle of 0 to 80.

5. Reference

Alberts, B. M. (1987). Prokaryotic DNA replication mechanisms. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *317*(1187), 395–420. doi: 10.1098/RSTB.1987.0068

Aldag, P., Rutkauskas, M., Madariaga-Marcos, J., Songailiene, I., Sinkunas, T., Kemmerich, F. E., Kauert, D. J., Siksnys, V., & Seidel, R. (2022). Dynamic interplay between target search and recognition for the Cascade surveillance complex of type I-E CRISPR-Cas systems. *BioRxiv*, 2022.12.18.520913. doi: 10.1101/2022.12.18.520913

Arutyunov, D., & Frost, L. S. (2013). F conjugation: Back to the beginning. *Plasmid*, *70*(1), 18–32. doi: 10.1016/J.PLASMID.2013.03.010

Athukoralage, J. S., & White, M. F. (2022). Cyclic Nucleotide Signaling in Phage Defense and Counter-Defense. *Annual Review of Virology*, *9*(1), 451–468. doi: 10.1146/ANNUREV-VIROLOGY-100120-010228

Ayers, M., Howell, P. L., & Burrows, L. L. (2010). Architecture of the type II secretion and type IV pilus machineries. *Future Microbiology*, *5*(8), 1203–1218. doi: 10.2217/FMB.10.76

Azam, A. H., & Tanji, Y. (2019). Bacteriophage-host arm race: an update on the mechanism of phage resistance in bacteria and revenge of the phage with the perspective for phage therapy. *Applied Microbiology and Biotechnology 2019 103:5*, *103*(5), 2121–2131. doi: 10.1007/S00253-019-09629-X

Baltrus, D. A. (2013). Exploring the costs of horizontal gene transfer. *Trends in Ecology & Evolution*, *28*(8), 489–495. doi: 10.1016/J.TREE.2013.04.002

Baltrus, D. A., Smith, C., Derrick, M., Leligdon, C., Rosenthal, Z., Mollico, M., Moore, A., & Clark, M. (2021). Genomic Background Governs Opposing Responses to Nalidixic Acid upon Megaplasmid Acquisition in *Pseudomonas*. *MSphere*, *6*(1). doi: 10.1128/mSphere.00008-21

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, *315*(5819), 1709–1712. doi: 10.1126/science.1138140

Bartel, D. P. (2004). MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell*, *116*(2), 281–297. doi: 10.1016/S0092-8674(04)00045-5

Bingham, R., Ekunwe, S. I. N., Falk, S., Snyder, L., & Kleanthous, C. (2000). The major head protein of bacteriophage T4 binds specifically to elongation factor Tu. *The Journal of Biological Chemistry*, *275*(30), 23219–23226. doi: 10.1074/JBC.M002546200

Bondy-Denomy, J., Qian, J., Westra, E. R., Buckling, A., Guttman, D. S., Davidson, A. R., & Maxwell, K. L. (2016). Prophages mediate defense against phage infection through diverse mechanisms. *The ISME Journal 2016 10:12*, *10*(12), 2854–2866. doi: 10.1038/ismej.2016.79

Bot, J. F., van der Oost, J., & Geijsen, N. (2022). The double life of CRISPR–Cas13. *Current Opinion in Biotechnology*, *78*, 102789. doi: 10.1016/J.COPBIO.2022.102789

Boyle, E. A., Andreasson, J. O. L., Chircus, L. M., Sternberg, S. H., Wu, M. J., Guegler, C. K., Doudna, J. A., & Greenleaf, W. J. (2017). High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(21), 5461–5466. doi: 10.1073/pnas.1700557114

Bozic, B., Repac, J., & Djordjevic, M. (2019). Endogenous Gene Regulation as a Predicted Main Function of Type I-E CRISPR/Cas System in *E. coli*. *Molecules*, *24*(4). doi: 10.3390/MOLECULES24040784

Brezgin, S., Kostyusheva, A., Kostyushev, D., & Chulanov, V. (2019). Dead Cas Systems: Types, Principles, and Applications. *International Journal of Molecular Sciences*, *20*(23). doi: 10.3390/IJMS20236041

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., & Van Der Oost, J. (2008). Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science*, *321*(5891), 960. doi: 10.1126/SCIENCE.1159689

Budhathoki, J. B., Xiao, Y., Schuler, G., Hu, C., Cheng, A., Ding, F., & Ke, A. (2020). Real-time observation of CRISPR spacer acquisition by Cas1-Cas2 integrase. *Nature Structural & Molecular Biology 2020 27:5*, *27*(5), 489–499. doi: 10.1038/s41594-020-0415-7

Casjens, S. R., & Hendrix, R. W. (2015). Bacteriophage lambda: early pioneer and still relevant. *Virology*, *0*, 310. doi: 10.1016/J.VIROL.2015.02.010

Cass, S. D. B., Haas, K. A., Stoll, B., Alkhnbashi, O. S., Sharma, K., Urlaub, H., Backofen, R., Marchfelder, A., & Bolt, E. L. (2015). The role of Cas8 in Type I CRISPR interference. *Bioscience Reports*, *35*(3). doi: 10.1042/BSR20150043

Castillo, D., Rørbo, N., Jørgensen, J., Lange, J., Tan, D., Kalatzis, P. G., Svenningsen, S. lo, & Middelboe, M. (2019). Phage defense mechanisms and their genomic and phenotypic implications in the fish pathogen *Vibrio anguillarum*. *FEMS Microbiology Ecology*, *95*(3), 4. doi: 10.1093/FEMSEC/FIZ004

Charpentier, E., Richter, H., van der Oost, J., & White, M. F. (2015). Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiology Reviews*, *39*(3), 428. doi: 10.1093/FEMSRE/FUV023

Chaudhary, K. (2017). BacteRiophage EXclusion (BREX): A novel anti-phage mechanism in the arsenal of bacterial defense system. *Journal of Cellular Physiology*, *233*(2), 771–773. doi: 10.1002/JCP.25973

Chen, D. S., Wu, Y. Q., Zhang, W., Jiang, S. J., & Chen, S. Z. (2016). Horizontal gene transfer events reshape the global landscape of arm race between viruses and homo sapiens. *Scientific Reports 2016 6:1*, *6*(1), 1–8. doi: 10.1038/srep26934

Chopin, M. C., Chopin, A., & Bidnenko, E. (2005). Phage abortive infection in lactococci: variations on a theme. *Current Opinion in Microbiology*, *8*(4), 473–479. doi: 10.1016/J.MIB.2005.06.006

Chung, I. Y., Jang, H. J., Bae, H. W., & Cho, Y. H. (2014). A phage protein that inhibits the bacterial ATPase required for type IV pilus assembly. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(31), 11503–11508. doi: 10.1073/pnas.1403537111

Craig, L., Pique, M. E., & Tainer, J. A. (2004). Type IV pilus structure and bacterial pathogenicity. *Nature Reviews Microbiology 2004 2:5*, *2*(5), 363–378. doi: 10.1038/nrmicro885

Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Research*, *14*(6), 1188–1190. doi: 10.1101/GR.849004

Crowley, V. M., Catching, A., Taylor, H. N., Borges, A. L., Metcalf, J., Bondy-Denomy, J., & 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*(6), 434. doi: 10.1089/CRISPR.2019.0048

Csörgő, B., León, L. M., Chau-Ly, I. J., Vasquez-Rifo, A., Berry, J. D., Mahendra, C., Crawford, E. D., Lewis, J. D., & Bondy-Denomy, J. (2020). A compact Cascade–Cas3 system for targeted genome engineering. *Nature Methods 2020 17:12*, *17*(12), 1183–1190. doi: 10.1038/s41592-020-00980-w

Datsenko, K. A., Pougach, K., Tikhonov, A., Wanner, B. L., Severinov, K., & Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature Communications 2012 3:1*, *3*(1), 1–7. doi: 10.1038/ncomms1937

de La Cruz, F., Frost, L. S., Meyer, R. J., & Zechner, E. L. (2010). Conjugative DNA metabolism in Gram-negative bacteria. *FEMS Microbiology Reviews*, *34*(1), 18–40. doi: 10.1111/J.1574-6976.2009.00195.X

de Smet, J., Hendrix, H., Blasdel, B. G., Danis-Wlodarczyk, K., & Lavigne, R. (2017). *Pseudomonas* predators: understanding and exploiting phage–host interactions. *Nature Reviews Microbiology 2017 15:9*, *15*(9), 517–530. doi: 10.1038/nrmicro.2017.61

Deng, L., Garrett, R. A., Shah, S. A., Peng, X., & She, Q. (2013). A novel interference mechanism by a type IIIB CRISPR-Cmr module in *Sulfolobus*. *Molecular Microbiology*, *87*(5), 1088–1099. doi: 10.1111/MMI.12152

Dennis, J. J. (2005). The evolution of IncP catabolic plasmids. *Current Opinion in Biotechnology*, *16*(3), 291–298. doi: 10.1016/J.COPBIO.2005.04.002

Devi, V., Harjai, K., & Chhibber, S. (2022). CRISPR-Cas systems: role in cellular processes beyond adaptive immunity. *Folia Microbiologica*, *67*(6), 837–850. doi: 10.1007/S12223-022-00993-2

Dhingra, Y., Suresh, S. K., Juneja, P., & Sashital, D. G. (2022). PAM binding ensures orientational integration during Cas4-Cas1-Cas2-mediated CRISPR adaptation. *Molecular Cell*, *82*(22), 4353-4367.e6. doi: 10.1016/J.MOLCEL.2022.09.030

Díez-Villaseñor, C., Guzmán, N. M., Almendros, C., García-Martínez, J., & Mojica, F. J. M. (2013). CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of *Escherichia coli*. *RNA Biology*, *10*(5), 792–802. doi: 10.4161/rna.24023

Doi, G., Okada, S., Yasukawa, T., Sugiyama, Y., Bala, S., Miyazaki, S., Kang, D., & Ito, T. (2021). Catalytically inactive Cas9 impairs DNA replication fork progression to induce focal genomic instability. *Nucleic Acids Research*, *49*(2), 954–968. doi: 10.1093/NAR/GKAA1241

Dower, W. J., Miller, J. F., & Ragsdale, C. W. (1988). High efficiency transformation of E.coli by high voltage electroporation. *Nucleic Acids Research*, *16*(13), 6127–6145. doi: 10.1093/NAR/16.13.6127

Dubey, G. P., & Ben-Yehuda, S. (2011). Intercellular Nanotubes Mediate Bacterial Communication. *Cell*, *144*(4), 590–600. doi: 10.1016/J.CELL.2011.01.015

Elmore, J. R., Sheppard, N. F., Ramia, N., Deighan, T., Li, H., Terns, R. M., & Terns, M. P. (2016). Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR–Cas system. *Genes & Development*, *30*(4), 447–459. doi: 10.1101/GAD.272153.115

Endo, A., Watanabe, T., Ogata, N., Nozawa, T., Aikawa, C., Arakawa, S., Maruyama, F., Izumi, Y., & Nakagawa, I. (2014). Comparative genome analysis and identification of competitive and cooperative interactions in a polymicrobial disease. *The ISME Journal 2015 9:3*, *9*(3), 629–642. doi: 10.1038/ismej.2014.155

Erdmann, S., & Garrett, R. A. (2012). Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. *Molecular Microbiology*, *85*(6), 1044–1056. doi: 10.1111/J.1365-2958.2012.08171.X

Estrella, M. A., Kuo, F. T., & Bailey, S. (2016). RNA-activated DNA cleavage by the Type III-B CRISPR–Cas effector complex. *Genes & Development*, *30*(4), 460–470. doi: 10.1101/GAD.273722.115

Fagerlund, R. D., Wilkinson, M. E., Klykov, O., Barendregt, A., Pearce, F. G., Kieper, S. N., Maxwell, H. W. R., Capolupo, A., Heck, A. J. R., Krause, K. L., Bostina, M., Scheltema, R. A., Staals, R. H. J., & Fineran, P. C. (2017). Spacer capture and integration by a type I-F Cas1-Cas2-3 CRISPR adaptation complex. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(26), E5122–E5128. doi: 10.1073/PNAS.1618421114

Fineran, P. C., Blower, T. R., Foulds, I. J., Humphreys, D. P., Lilley, K. S., & Salmond, G. P. C. (2009). The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(3), 894–899. doi: 10.1073/PNAS.0808832106

Freitas, F., Alves, V. D., Carvalheira, M., Costa, N., Oliveira, R., & Reis, M. A. M. (2009). Emulsifying behaviour and rheological properties of the extracellular polysaccharide produced by *Pseudomonas oleovorans* grown on glycerol byproduct. *Carbohydrate Polymers*, *78*(3), 549–556. doi: 10.1016/J.CARBPOL.2009.05.016

Freitas, F., Alves, V. D., Pais, J., Carvalheira, M., Costa, N., Oliveira, R., & Reis, M. A. M. (2010). Production of a new exopolysaccharide (EPS) by *Pseudomonas oleovorans* NRRL B-14682 grown on glycerol. *Process Biochemistry*, *45*(3), 297–305. doi: 10.1016/J.PROCBIO.2009.09.020

Friehs, K. (2004). Plasmid copy number and plasmid stability. *Advances in Biochemical Engineering/Biotechnology*, *86*, 47–82. doi: 10.1007/b12440

Garcillán-Barcia, M. P., & de la Cruz, F. (2008). Why is entry exclusion an essential feature of conjugative plasmids? *Plasmid*, *60*(1), 1–18. doi: 10.1016/J.PLASMID.2008.03.002

Garrett, S. C. (2021). Pruning and Tending Immune Memories: Spacer Dynamics in the CRISPR Array. *Frontiers in Microbiology*, *12*, 739. doi: 10.3389/FMICB.2021.664299/BIBTEX

Geissmann, Q. (2013). OpenCFU, a New Free and Open-Source Software to Count Cell Colonies and Other Circular Objects. *PLOS ONE*, *8*(2), e54072. doi: 10.1371/JOURNAL.PONE.0054072

Golais, F., Hollý, J., & Vítkovská, J. (2012). Coevolution of bacteria and their viruses. *Folia Microbiologica 2012 58:3*, *58*(3), 177–186. doi: 10.1007/S12223-012-0195-5

Guo, X., Sanchez-Londono, M., Gomes-Filho, J. V., Hernandez-Tamayo, R., Rust, S., Immelmann, L. M., Schäfer, P., Wiegel, J., Graumann, P. L., & Randau, L. (2022). Characterization of the self-targeting Type IV CRISPR interference system in *Pseudomonas oleovorans*. *Nature Microbiology 2022 7:11*, *7*(11), 1870–1878. doi: 10.1038/s41564-022-01229-2

Hale, C. R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A. M., Glover, C. V. C., Graveley, B. R., Terns, R. M., & Terns, M. P. (2012). Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Molecular Cell*, *45*(3), 292. doi: 10.1016/J.MOLCEL.2011.10.023

Hall, J. P. J., Botelho, J., Cazares, A., & Baltrus, D. A. (2022). What makes a megaplasmid? *Philosophical Transactions of the Royal Society B*, *377*(1842). doi: 10.1098/RSTB.2020.0472

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology*, *166*(4), 557–580. doi: 10.1016/S0022-2836(83)80284-8

Hille, F., & Charpentier, E. (2016). CRISPR-Cas: biology, mechanisms and relevance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1707). doi: 10.1098/RSTB.2015.0496

Hille, F., Richter, H., Wong, S. P., Bratovič, M., Ressel, S., & Charpentier, E. (2018). The Biology of CRISPR-Cas: Backward and Forward. *Cell*, *172*(6), 1239–1259. doi: 10.1016/J.CELL.2017.11.032

Hochstrasser, M. L., Taylor, D. W., Kornfeld, J. E., Nogales, E., & Doudna, J. A. (2016). DNA Targeting by a Minimal CRISPR RNA-Guided Cascade. *Molecular Cell*, *63*(5), 840–851. doi: 10.1016/J.MOLCEL.2016.07.027

Horvath, P., & Barrangou, R. (2010). CRISPR/Cas, the Immune System of Bacteria and Archaea. *Science*, *327*(5962), 167–170. doi: 10.1126/SCIENCE.1179555

Horwitz, J. P., Chua, J., Cubby, R. J., Tomson, A. J., da Rooge, M. A., Fisher, B. E., Mauricio, J., & Klundt, I. (1964). Substrates for Cytochemical Demonstration of Enzyme Activity. I. Some Substituted 3-Indolyl-β-D-glycopyranosides. *Journal of Medicinal Chemistry*, *7*(4), 574–575. doi: 10.1021/JM00334A044/ASSET/JM00334A044.FP.PNG_V03

Houte, S. van, Buckling, A., & Westra, E. R. (2016). Evolutionary Ecology of Prokaryotic Immune Mechanisms. *Microbiology and Molecular Biology Reviews: MMBR*, *80*(3), 745. doi: 10.1128/MMBR.00011-16

Hu, C., Almendros, C., Nam, K. H., Costa, A. R., Vink, J. N. A., Haagsma, A. C., Bagde, S. R., Brouns, S. J. J., & Ke, A. (2021). Mechanism for Cas4-assisted directional spacer acquisition in CRISPR–Cas. *Nature 2021 598:7881*, *598*(7881), 515–520. doi: 10.1038/s41586-021-03951-z

Hu, Y., & Li, W. (2022). Development and Application of CRISPR-Cas Based Tools. *Frontiers in Cell and Developmental Biology*, *10*. doi: 10.3389/FCELL.2022.834646

Hudaiberdiev, S., Shmakov, S., Wolf, Y. I., Terns, M. P., Makarova, K. S., & Koonin, E. V. (2017). Phylogenomics of Cas4 family nucleases. *BMC Evolutionary Biology*, *17*(1), 1–14. doi: 10.1186/S12862-017-1081-1/FIGURES/6

Icho, T., & Iino, T. (1978). Isolation and characterization of motile *Escherichia coli* mutants resistant to bacteriophage chi. *Journal of Bacteriology*, *134*(3), 854–860. doi: 10.1128/JB.134.3.854-860.1978

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakatura, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of Bacteriology*, *169*(12), 5429–5433. doi: 10.1128/JB.169.12.5429-5433.1987

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., & Wiedenheft, B. (2014). Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science*, *345*(6203), 1473–1479. doi: 10.1126/science.1256328

Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G., & Gibson, T. J. (1998). Multiple sequence alignment with Clustal X. *Trends in Biochemical Sciences*, *23*(10), 403–405. doi: 10.1016/S0968-0004(98)01285-7

Jia, N., Mo, C. Y., Wang, C., Eng, E. T., Marraffini, L. A., & Patel, D. J. (2019). Type III-A CRISPR-Cas Csm Complexes: Assembly, Periodic RNA Cleavage, DNase Activity

Regulation, and Autoimmunity. *Molecular Cell*, *73*(2), 264-277.e5. doi: 10.1016/J.MOLCEL.2018.11.007

Jones, C. L., Sampson, T. R., Nakaya, H. I., Pulendran, B., & Weiss, D. S. (2012). Repression of bacterial lipoprotein production by *Francisella novicida* facilitates evasion of innate immune recognition. *Cellular Microbiology*, *14*(10), 1531–1543. doi: 10.1111/J.1462-5822.2012.01816.X

Jones, D. L., Leroy, P., Unoson, C., Fange, D., Ćurić, V., Lawson, M. J., & Elf, J. (2017). Kinetics of dCas9 target search in *Escherichia coli*. *Science*, *357*(6358), 1420–1424. doi: 10.1126/SCIENCE.AAH7084

Jore, M. M., Lundgren, M., van Duijn, E., Bultema, J. B., Westra, E. R., Waghmare, S. P., Wiedenheft, B., Pul, Ü., Wurm, R., Wagner, R., Beijer, M. R., Barendregt, A., Zhou, K., Snijders, A. P. L., Dickman, M. J., Doudna, J. A., Boekema, E. J., Heck, A. J. R., van der Oost, J., & Brouns, S. J. J. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature Structural & Molecular Biology 2011 18:5*, *18*(5), 529–536. doi: 10.1038/nsmb.2019

Juhas, M., van der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W., & Crook, D. W. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiology Reviews*, *33*(2), 376–393. doi: 10.1111/J.1574-6976.2008.00136.X

Julin, D. A. (2018). Blue/White Selection. *Molecular Life Sciences*, 72–73. doi: 10.1007/978-1-4614-1531-2_94

Kalwani, P., Rath, D., & Ballal, A. (2020). Novel molecular aspects of the CRISPR backbone protein 'Cas7' from cyanobacteria. *Biochemical Journal*, *477*(5), 971–983. doi: 10.1042/BCJ20200026

Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G., & Siksnys, V. (2017). A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science*, *357*(6351), 605–609. doi: 10.1126/SCIENCE.AAO0100

Ketting, R. F. (2011). The Many Faces of RNAi. *Developmental Cell*, *20*(2), 148–161. doi: 10.1016/J.DEVCEL.2011.01.012

Killelea, T., & Bolt, E. L. (2017). CRISPR-Cas adaptive immunity and the three Rs. *Bioscience Reports*, *37*(4), 20160297. doi: 10.1042/BSR20160297

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology 2019 37:8*, *37*(8), 907–915. doi: 10.1038/s41587-019-0201-4

Knight, S. C., Xie, L., Deng, W., Guglielmi, B., Witkowsky, L. B., Bosanac, L., Zhang, E. T., Beheiry, M. E., Masson, J. B., Dahan, M., Liu, Z., Doudna, J. A., & Tjian, R. (2015). Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science*, *350*(6262), 823–826. doi: 10.1126/SCIENCE.AAC6572

Knöppel, A., Lind, P. A., Lustig, U., Näsvall, J., & Andersson, D. I. (2014). Minor Fitness Costs in an Experimental Model of Horizontal Gene Transfer in Bacteria. *Molecular Biology and Evolution*, *31*(5), 1220–1227. doi: 10.1093/MOLBEV/MSU076

Koonin, E. v., & Krupovic, M. (2014). Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nature Reviews Genetics 2014 16:3*, *16*(3), 184–192. doi: 10.1038/nrg3859

Koonin, E. v., & Makarova, K. S. (2019). Origins and evolution of CRISPR-Cas systems. *Philosophical Transactions of the Royal Society B*, *374*(1772). doi: 10.1098/RSTB.2018.0087

Koonin, E. v., & Makarova, K. S. (2022). Evolutionary plasticity and functional versatility of CRISPR systems. *PLOS Biology*, *20*(1), e3001481. doi: 10.1371/JOURNAL.PBIO.3001481

Koonin, E. V., Makarova, K. S., & Zhang, F. (2017a). Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology*, *37*, 67–78. doi: 10.1016/J.MIB.2017.05.008

Koonin, E. v., Makarova, K. S., & Zhang, F. (2017b). Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology*, *37*, 67–78. doi: 10.1016/J.MIB.2017.05.008

Kotnik, T., Frey, W., Sack, M., Haberl Meglič, S., Peterka, M., & Miklavčič, D. (2015). Electroporation-based applications in biotechnology. *Trends in Biotechnology*, *33*(8), 480–488. doi: 10.1016/J.TIBTECH.2015.06.002

Krüger, N. J., & Stingl, K. (2011). Two steps away from novelty – principles of bacterial DNA uptake. *Molecular Microbiology*, *80*(4), 860–867. doi: 10.1111/J.1365-2958.2011.07647.X

Lawrenz, M. B., Kawabata, H., Purser, J. E., & Norris, S. J. (2002). Decreased Electroporation Efficiency in Borrelia burgdorferi Containing Linear Plasmids lp25 and lp56: Impact on Transformation of Infectious B. burgdorferi. *Infection and Immunity*, *70*(9), 4798–4804. doi: 10.1128/IAI.70.9.4798-4804.2002

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

Lee, H., Zhou, Y., Taylor, D. W., & Sashital, D. G. (2018). Cas4-Dependent Prespacer Processing Ensures High-Fidelity Programming of CRISPR Arrays. *Molecular Cell*, *70*(1), 48-59.e5. doi: 10.1016/j.molcel.2018.03.003

Leenay, R. T., Maksimchuk, K. R., Slotkowski, R. A., Agrawal, R. N., Gomaa, A. A., Briner, A. E., Barrangou, R., & Beisel, C. L. (2016). Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. *Molecular Cell*, *62*(1), 137–147. doi: 10.1016/J.MOLCEL.2016.02.031

Li, Z., Zhang, H., Xiao, R., & Chang, L. (2020a). Cryo-EM structure of a type I-F CRISPR RNA guided surveillance complex bound to transposition protein TniQ. *Cell Research 2020 30:2*, *30*(2), 179–181. doi: 10.1038/s41422-019-0268-y

Li, Z., Zhang, H., Xiao, R., & Chang, L. (2020b). Cryo-EM structure of a type I-F CRISPR RNA guided surveillance complex bound to transposition protein TniQ. *Cell Research 2020 30:2*, *30*(2), 179–181. doi: 10.1038/s41422-019-0268-y

Liu, T., Li, Y., Wang, X., Ye, Q., Li, H., Liang, Y., She, Q., & Peng, N. (2015). Transcriptional regulator-mediated activation of adaptation genes triggers CRISPR *de novo* spacer acquisition. *Nucleic Acids Research*, *43*(2), 1044. doi: 10.1093/NAR/GKU1383

Liu, T., Pan, S., Li, Y., Peng, N., & She, Q. (2017). Type III CRISPR/Cas System: Introduction and Its Application for Genetic Manipulations. *Current Issues in Molecular Biology 2018, Vol. 26, Pages 1-14*, *26*(1), 1–14. doi: 10.21775/CIMB.026.001

Louwen, R., Staals, R. H. J., Endtz, H. P., van Baarlen, P., & van der Oost, J. (2014). The Role of CRISPR-Cas Systems in Virulence of Pathogenic Bacteria. *Microbiology and Molecular Biology Reviews*, *78*(1), 74–88. doi: 10.1128/MMBR.00039-13

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. doi: 10.1186/S13059-014-0550-8

Madden, T. L., Tatusov, R. L., & Zhang, J. (1996). Applications of network BLAST server. *Methods in Enzymology*, *266*, 131–141. doi: 10.1016/S0076-6879(96)66011-X

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J. M., Wolf, Y. I., Yakunin, A. F., van der Oost, J., & Koonin, E. v. (2011). Evolution and classification of the CRISPR–Cas systems. *Nature Reviews Microbiology 2011 9:6*, *9*(6), 467–477. doi: 10.1038/nrmicro2577

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., Horvath, P., Moineau, S., Mojica, F. J. M., Terns, R. M., Terns, M. P., White, M. F., Yakunin, A. F., Garrett, R. A., van der Oost, J., … Koonin, E. v. (2015). An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology 2015 13:11*, *13*(11), 722–736. doi: 10.1038/nrmicro3569

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., Moineau, S., Mojica, F. J. M., Scott, D., Shah, S. A., Siksnys, V., Terns, M. P., Venclovas, Č., White, M. F., Yakunin, A. F., … Koonin, E. v. (2019). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology 2019 18:2*, *18*(2), 67–83. doi: 10.1038/s41579-019-0299-x

Makarova, K. S., Wolf, Y. I., van der Oost, J., & Koonin, E. v. (2009). Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biology Direct*, *4*, 29. doi: 10.1186/1745-6150-4-29

Marraffini, L. A., & Sontheimer, E. J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature 2010 463:7280*, *463*(7280), 568–571. doi: 10.1038/nature08703

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10. doi: 10.14806/EJ.17.1.200

McGinn, J., & Marraffini, L. A. (2018). Molecular mechanisms of CRISPR–Cas spacer acquisition. *Nature Reviews Microbiology 2018 17:1*, *17*(1), 7–12. doi: 10.1038/s41579-018-0071-7

McKenzie, R. E., Almendros, C., Vink, J. N. A., & Brouns, S. J. J. (2019). Using CAPTURE to detect spacer acquisition in native CRISPR arrays. *Nature Protocols 2019 14:3*, *14*(3), 976–990. doi: 10.1038/s41596-018-0123-5

Morgan, S. L., Mariano, N. C., Bermudez, A., Arruda, N. L., Wu, F., Luo, Y., Shankar, G., Jia, L., Chen, H., Hu, J. F., Hoffman, A. R., Huang, C. C., Pitteri, S. J., & Wang, K. C. (2017). Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nature Communications*, *8*. doi: 10.1038/NCOMMS15993

Morton, E. R., Platt, T. G., Fuqua, C., & Bever, J. D. (2014). Non-additive costs and interactions alter the competitive dynamics of co-occurring ecologically distinct plasmids. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1779). doi: 10.1098/RSPB.2013.2173

Moya-Beltrán, A., Makarova, K. S., Acuña, L. G., Wolf, Y. I., Covarrubias, P. C., Shmakov, S. A., Silva, C., Tolstoy, I., Johnson, D. B., Koonin, E. v., & Quatrini, R. (2021). Evolution of Type IV CRISPR-Cas Systems: Insights from CRISPR Loci in Integrative Conjugative Elements of *Acidithiobacillia*. *The CRISPR Journal*, *4*(5), 656. doi: 10.1089/CRISPR.2021.0051

Mulepati, S., Héroux, A., & Bailey, S. (2014). Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science (New York, N.Y.)*, *345*(6203), 1479–1484. doi: 10.1126/SCIENCE.1256996/SUPPL_FILE/MULEPATI.SM.PDF

Newire, E., Aydin, A., Juma, S., Enne, V. I., & Roberts, A. P. (2020). Identification of a Type IV-A CRISPR-Cas System Located Exclusively on IncHI1B/IncFIB Plasmids in *Enterobacteriaceae*. *Frontiers in Microbiology*, *11*, 1937. doi: 10.3389/fmicb.2020.01937

Nguyen, L. T., Smith, B. M., & Jain, P. K. (2020). Enhancement of trans-cleavage activity of Cas12a with engineered crRNA enables amplified nucleic acid detection. *Nature Communications 2020 11:1*, *11*(1), 1–13. doi: 10.1038/s41467-020-18615-1

Niewoehner, O., Garcia-Doval, C., Rostøl, J. T., Berk, C., Schwede, F., Bigler, L., Hall, J., Marraffini, L. A., & Jinek, M. (2017). Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. *Nature 2017 548:7669*, *548*(7669), 543–548. doi: 10.1038/nature23467

Niewoehner, O., & Jinek, M. (2016). Structural basis for the endoribonuclease activity of the type III-A CRISPR-associated protein Csm6. *RNA*, *22*(3), 318–329. doi: 10.1261/RNA.054098.115

Nuñez, J. K., Bai, L., Harrington, L. B., Hinder, T. L., & Doudna, J. A. (2016). CRISPR Immunological Memory Requires a Host Factor for Specificity. *Molecular Cell*, *62*(6), 824–833. doi: 10.1016/J.MOLCEL.2016.04.027

Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. v., Davies, C. W., & Doudna, J. A. (2014). Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nature Structural & Molecular Biology 2014 21:6*, *21*(6), 528–534. doi: 10.1038/nsmb.2820

Nuñez, J. K., Lee, A. S. Y., Engelman, A., & Doudna, J. A. (2015). Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature 2015 519:7542*, *519*(7542), 193–198. doi: 10.1038/nature14237

Osawa, T., Inanaga, H., Sato, C., & Numata, T. (2015). Crystal Structure of the CRISPR-Cas RNA Silencing Cmr Complex Bound to a Target Analog. *Molecular Cell*, *58*(3), 418–430. doi: 10.1016/J.MOLCEL.2015.03.018

Özcan, A., Krajeski, R., Ioannidi, E., Lee, B., Gardner, A., Makarova, K. S., Koonin, E. V., Abudayyeh, O. O., & Gootenberg, J. S. (2021). Programmable RNA targeting with the single-protein CRISPR effector Cas7-11. *Nature 2021 597:7878*, *597*(7878), 720–725. doi: 10.1038/s41586-021-03886-5

Özcan, A., Pausch, P., Linden, A., Wulf, A., Schühle, K., Heider, J., Urlaub, H., Heimerl, T., Bange, G., & Randau, L. (2019). Type IV CRISPR RNA processing and effector complex formation in *Aromatoleum aromaticum*. *Nature Microbiology*, *4*(1), 89–96. doi: 10.1038/S41564-018-0274-8

Pelicic, V. (2008). Type IV pili: *e pluribus unum*? *Molecular Microbiology*, *68*(4), 827–837. doi: 10.1111/J.1365-2958.2008.06197.X

Perez-Rodriguez, R., Haitjema, C., Huang, Q., Nam, K. H., Bernardis, S., Ke, A., & DeLisa, M. P. (2011). Envelope stress is a trigger of CRISPR RNA-mediated DNA silencing in *Escherichia coli*. *Molecular Microbiology*, *79*(3), 584–599. doi: 10.1111/J.1365-2958.2010.07482.X

Piepenbrink, K. H. (2019). DNA uptake by type IV filaments. *Frontiers in Molecular Biosciences*, *6*(FEB), 1. doi: 10.3389/FMOLB.2019.00001

Pinilla-Redondo, R., Mayo-Muñoz, D., Russel, J., Garrett, R. A., Randau, L., Sørensen, S. J., & Shah, S. A. (2020). Type IV CRISPR–Cas systems are highly diverse and involved in competition between plasmids. *Nucleic Acids Research*, *48*(4), 2000–2012. doi: 10.1093/NAR/GKZ1197

Platt, T. G., Morton, E. R., Barton, I. S., Bever, J. D., & Fuqua, C. (2014). Ecological dynamics and complex interactions of *Agrobacterium* megaplasmids. *Frontiers in Plant Science*, *5*(November), 1–15. doi: 10.3389/FPLS.2014.00635

Pope, W. H., Jacobs-Sera, D., Russel, D. A., Peebles, C. L., Al-Atrache, Z., Alcoser, T. A., Alexander, L. M., Alfano, M. B., Alford, S. T., Amy, N. E., Anderson, M. D., Anderson, A. G., Ang, A. A. S., Manuel, A., Barber, A. J., Barker, L. P., Barrett, J. M., Barshop, W. D., Bauerle, C. M., … Hatfull, G. F. (2011). Expanding the diversity of mycobacteriophages: insights into genome architecture and evolution. *PloS One*, *6*(1). doi: 10.1371/JOURNAL.PONE.0016329

Puskas, A., Greenberg, E. P., Kaplan, S., & Schaefer, A. L. (1997). A quorum-sensing system in the free-living photosynthetic bacterium *Rhodobacter sphaeroides*.

Journal of Bacteriology, *179*(23), 7530–7537. doi: 10.1128/JB.179.23.7530-7537.1997

Pyenson, N. C., Gayvert, K., Varble, A., Elemento, O., & Marraffini, L. A. (2017). Broad Targeting Specificity during Bacterial Type III CRISPR-Cas Immunity Constrains Viral Escape. *Cell Host & Microbe*, *22*(3), 343-353.e3. doi: 10.1016/J.CHOM.2017.07.016

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell*, *152*(5), 1173–1183. doi: 10.1016/J.CELL.2013.02.022

Qiu, D., Damron, F. H., Mima, T., Schweizer, H. P., & Yu, H. D. (2008). pBAD-Based Shuttle Vectors for Functional Analysis of Toxic and Highly Regulated Genes in *Pseudomonas* and *Burkholderia* spp. and Other Bacteria. *Applied and Environmental Microbiology*, *74*(23), 7422. doi: 10.1128/AEM.01369-08

Quax, T. E. F., Voet, M., Sismeiro, O., Dillies, M.-A., Jagla, B., Coppée, J.-Y., Sezonov, G., Forterre, P., van der Oost, J., Lavigne, R., & Prangishvili, D. (2013). Massive Activation of Archaeal Defense Genes during Viral Infection. *Journal of Virology*, *87*(15), 8419–8428. doi: 10.1128/JVI.01020-13

Rajagopala, S. v., Casjens, S., & Uetz, P. (2011). The protein interaction map of bacteriophage lambda. *BMC Microbiology*, *11*(1), 1–15. doi: 10.1186/1471-2180-11-213

Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J., Blumenthal, R. M., Degtyarev, S. K., Dryden, D. T. F., Dybvig, K., Firman, K., Gromova, E. S., Gumport, R. I., Halford, S. E., Hattman, S., Heitman, J., Hornby, D. P., Janulaitis, A., Jeltsch, A., … Xu, S. Y. (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Research*, *31*(7), 1805–1812. doi: 10.1093/NAR/GKG274

Romanchuk, A., Jones, C. D., Karkare, K., Moore, A., Smith, B. A., Jones, C., Dougherty, K., & Baltrus, D. A. (2014). Bigger is not always better: Transmission and fitness burden of ∼1 MB *Pseudomonas syringae* megaplasmid pMPPla107. *Plasmid*, *73*, 16–25. doi: 10.1016/J.PLASMID.2014.04.002

Roux, N., Spagnolo, J., & de Bentzmann, S. (2012). Neglected but amazingly diverse type IVb pili. *Research in Microbiology*, *163*(9–10), 659–673. doi: 10.1016/J.RESMIC.2012.10.015

Samai, P., Pyenson, N., Jiang, W., Goldberg, G. W., Hatoum-Aslan, A., & Marraffini, L. A. (2015). Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell*, *161*(5), 1164–1174. doi: 10.1016/J.CELL.2015.04.027

Sampson, T. R., Saroj, S. D., Llewellyn, A. C., Tzeng, Y. L., & Weiss, D. S. (2013). A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature 2013 497:7448*, *497*(7448), 254–257. doi: 10.1038/nature12048

Samuel, A. D. T., Pitta, T. P., Ryu, W. S., Danese, P. N., Leung, E. C. W., & Berg, H. C. (1999). Flagellar determinants of bacterial sensitivity to chi-phage. *Proceedings of*

the National Academy of Sciences of the United States of America, *96*(17), 9863–9866. doi: 10.1073/PNAS.96.17.9863

Sashital, D. G., Wiedenheft, B., & Doudna, J. A. (2012). Mechanism of Foreign DNA Selection in a Bacterial Adaptive Immune System. *Molecular Cell*, *46*(5), 606–615. doi: 10.1016/J.MOLCEL.2012.03.020

Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A., & Severinov, K. (2013). High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in *E. coli*. *RNA Biol*, *10*(5), 716–725. doi: 10.4161/RNA.24325

Scanlan, P. D., & Buckling, A. (2012). Co-evolution with lytic phage selects for the mucoid phenotype of *Pseudomonas fluorescens* SBW25. *The ISME Journal*, *6*(6), 1148–1158. doi: 10.1038/ISMEJ.2011.174

Scharn, C. R., Tenover, F. C., & Goering, R. v. (2013). Transduction of staphylococcal cassette chromosome mec elements between strains of *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*, *57*(11), 5233–5238. doi: 10.1128/AAC.01058-13

Schmidt, F., Cherepkova, M. Y., & Platt, R. J. (2018). Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature 2018 562:7727*, *562*(7727), 380–385. doi: 10.1038/s41586-018-0569-1

Sevastsyanovich, Y. R., Krasowiak, R., Bingle, L. E. H., Haines, A. S., Sokolov, S. L., Kosheleva, I. A., Leuchuk, A. A., Titok, M. A., Smalla, K., & Thomas, C. M. (2008). Diversity of IncP-9 plasmids of *Pseudomonas*. *Microbiology*, *154*(10), 2929–2941. doi: 10.1099/MIC.0.2008/017939-0

Shangguan, Q., Graham, S., Sundaramoorthy, R., & White, M. F. (2022). Structure and mechanism of the type I-G CRISPR effector. *Nucleic Acids Research*, *50*(19), 11214–11228. doi: 10.1093/NAR/GKAC925

Shen, S., Horowitz, E. D., Troupes, A. N., Brown, S. M., Pulicherla, N., Samulski, R. J., Agbandje-McKenna, M., & Asokan, A. (2013). Engraftment of a galactose receptor footprint onto adeno-associated viral capsids improves transduction efficiency. *Journal of Biological Chemistry*, *288*(40), 28814–28823. doi: 10.1074/jbc.M113.482380

Sheppard, N. F., Glover, C. V. C., Terns, R. M., & Terns, M. P. (2016). The CRISPR-associated Csx1 protein of *Pyrococcus furiosus* is an adenosine-specific endoribonuclease. *RNA*, *22*(2), 216–224. doi: 10.1261/RNA.039842.113

Singh, D., & Ha, T. (2018). Understanding the Molecular Mechanisms of the CRISPR Toolbox Using Single Molecule Approaches. *ACS Chemical Biology*, *13*(3), 516–526. doi: 10.1021/ACSCHEMBIO.7B00905

Sinkunas, T., Gasiunas, G., Waghmare, S. P., Dickman, M. J., Barrangou, R., Horvath, P., & Siksnys, V. (2013). *In vitro* reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *The EMBO Journal*, *32*(3), 385–394. doi: 10.1038/EMBOJ.2012.352

Smith, G. R. (2012). How RecBCD Enzyme and Chi Promote DNA Break Repair and Recombination: a Molecular Biologist's View. *Microbiology and Molecular Biology Reviews : MMBR*, *76*(2), 217. doi: 10.1128/MMBR.05026-11

Spilman, M., Cocozaki, A., Hale, C., Shao, Y., Ramia, N., Terns, R., Terns, M., Li, H., & Stagg, S. (2013). Structure of an RNA Silencing Complex of the CRISPR-Cas Immune System. *Molecular Cell*, *52*(1), 146–152. doi: 10.1016/J.MOLCEL.2013.09.008

Staals, R. H. J., Zhu, Y., Taylor, D. W., Kornfeld, J. E., Sharma, K., Barendregt, A., Koehorst, J. J., Vlot, M., Neupane, N., Varossieau, K., Sakamoto, K., Suzuki, T., Dohmae, N., Yokoyama, S., Schaap, P. J., Urlaub, H., Heck, A. J. R., Nogales, E., Doudna, J. A., … vanderOost, J. (2014). RNA Targeting by the Type III-A CRISPR-Cas Csm Complex of *Thermus thermophilus*. *Molecular Cell*, *56*(4), 518–530. doi: 10.1016/J.MOLCEL.2014.10.005

Stingl, K., Müller, S., Scheidgen-Kleyboldt, G., Clausen, M., & Maier, B. (2010). Composite system mediates two-step DNA uptake into *Helicobacter pylori*. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(3), 1184–1189. doi: 10.1073/PNAS.0909955107

Swarts, D. C., Mosterd, C., van Passel, M. W. J., & Brouns, S. J. J. (2012). CRISPR Interference Directs Strand Specific Spacer Acquisition. *PLOS ONE*, *7*(4), e35888. doi: 10.1371/JOURNAL.PONE.0035888

Tamulaitis, G., Venclovas, Č., & Siksnys, V. (2017). Type III CRISPR-Cas Immunity: Major Differences Brushed Aside. *Trends in Microbiology*, *25*(1), 49–61. doi: 10.1016/J.TIM.2016.09.012

Taylor, D. W., Zhu, Y., Staals, R. H. J., Kornfeld, J. E., Shinkai, A., van der Oost, J., Nogales, E., & Doudna, J. A. (2015). Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning. *Science*, *348*(6234), 581–585. doi: 10.1126/SCIENCE.AAA4535

Taylor, H. N., Laderman, E., Armbrust, M., Hallmark, T., Keiser, D., Bondy-Denomy, J., & Jackson, R. N. (2021). Positioning Diverse Type IV Structures and Functions Within Class 1 CRISPR-Cas Systems. *Frontiers in Microbiology*, *12*, 1236. doi: 10.3389/FMICB.2021.671522

Thomas, C. M. (2000). Paradigms of plasmid organization. *Molecular Microbiology*, *37*(3), 485–491. doi: 10.1046/J.1365-2958.2000.02006.X

Tong, H., Huang, J., Xiao, Q., He, B., Dong, X., Liu, Y., Yang, X., Han, D., Wang, Z., Wang, X., Ying, W., Zhang, R., Wei, Y., Xu, C., Zhou, Y., Li, Y., Cai, M., Wang, Q., Xue, M., … Yang, H. (2023). High-fidelity Cas13 variants for targeted RNA degradation with minimal collateral effects. *Nature Biotechnology*, *41*(1). doi: 10.1038/S41587-022-01419-7

Uc-Mass, A., Loeza, E. J., de La Garza, M., Guarneros, G., Hernández-Sánchez, J., & Kameyama, L. (2004). An orthologue of the cor gene is involved in the exclusion of temperate lambdoid phages. Evidence that Cor inactivates FhuA receptor functions. *Virology*, *329*(2), 425–433. doi: 10.1016/J.VIROL.2004.09.005

van Beilen, J. B., Wubbolts, M. G., & Witholt, B. (1994). Genetics of alkane oxidation by *Pseudomonas oleovorans*. *Biodegradation*, *5*(3–4), 161–174. doi: 10.1007/BF00696457

Villanueva, R. A. M., & Chen, Z. J. (2019). *ggplot2: Elegant Graphics for Data Analysis (2nd ed.)*. *17*(3), 160–167. doi: 10.1080/15366367.2019.1565254

Viswanathan, P., Murphy, K., Julien, B., Garza, A. G., & Kroos, L. (2007). Regulation of dev, an operon that includes genes essential for *Myxococcus xanthus* development and CRISPR-associated genes and repeats. *Journal of Bacteriology*, *189*(10), 3738–3750. doi: 10.1128/JB.00187-07

Vogwill, T., & Maclean, R. C. (2015). The genetic basis of the fitness costs of antimicrobial resistance: a meta-analysis approach. *Evolutionary Applications*, *8*(3), 284–295. doi: 10.1111/EVA.12202

Vostrov, A. A., Vostrukhina, O. A., Svarchevsky, A. N., & Rybchin, V. N. (1996). Proteins responsible for lysogenic conversion caused by coliphages N15 and phi80 are highly homologous. *Journal of Bacteriology*, *178*(5), 1484–1486. doi: 10.1128/JB.178.5.1484-1486.1996

Walker, J. E., Saraste, M., Runswick, M. J., & Gay, N. J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *The EMBO Journal*, *1*(8), 945. doi: 10.1002/J.1460-2075.1982.TB01276.X

Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M., & Wang, Y. (2015). Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell*, *163*(4), 840–853. doi: 10.1016/J.CELL.2015.10.008

Wang, T., Li, Y., Li, J., Zhang, D., Cai, N., Zhao, G., Ma, H., Shang, C., Ma, Q., Xu, Q., & Chen, N. (2019). An update of the suicide plasmid-mediated genome editing system in *Corynebacterium glutamicum*. *Microbial Biotechnology*, *12*(5), 907. doi: 10.1111/1751-7915.13444

Ward, C. M., To, H., & Pederson, S. M. (2018). ngsReports: An R Package for managing FastQC reports and other NGS related log files. *BioRxiv*, 313148. doi: 10.1101/313148

Ward, J. P., King, J. R., Koerber, A. J., Croft, J. M., Sockett, R. E., & Williams, P. (2004). Cell-signalling repression in bacterial quorum sensing. *Mathematical Medicine and Biology*, *21*(3), 169–204. doi: 10.1093/IMAMMB/21.3.169

West, S. E. H., Schweizer, H. P., Dall, C., Sample, A. K., & Runyen-Janecky, L. J. (1994). Construction of improved *Escherichia-Pseudomonas* shuttle vectors derived from pUC18/19 and sequence of the region required for their replication in *Pseudomonas aeruginosa*. *Gene*, *148*(1), 81–86. doi: 10.1016/0378-1119(94)90237-2

Westra, E. R., Semenova, E., Datsenko, K. A., Jackson, R. N., Wiedenheft, B., Severinov, K., & Brouns, S. J. J. (2013a). Type I-E CRISPR-Cas systems discriminate target from

non-target DNA through base pairing-independent PAM recognition. *PLoS Genetics*, *9*(9). doi: 10.1371/JOURNAL.PGEN.1003742

Westra, E. R., Semenova, E., Datsenko, K. A., Jackson, R. N., Wiedenheft, B., Severinov, K., & Brouns, S. J. J. (2013b). Type I-E CRISPR-Cas Systems Discriminate Target from Non-Target DNA through Base Pairing-Independent PAM Recognition. *PLoS Genetics*, *9*(9). doi: 10.1371/JOURNAL.PGEN.1003742

Whinn, K. S., Kaur, G., Lewis, J. S., Schauer, G. D., Mueller, S. H., Jergic, S., Maynard, H., Gan, Z. Y., Naganbabu, M., Bruchez, M. P., O'Donnell, M. E., Dixon, N. E., van Oijen, A. M., & Ghodke, H. (2019). Nuclease dead Cas9 is a programmable roadblock for DNA replication. *Scientific Reports 2019 9:1*, *9*(1), 1–9. doi: 10.1038/s41598-019-49837-z

Wirth, N. T., Kozaeva, E., & Nikel, P. I. (2020). Accelerated genome engineering of *Pseudomonas putida* by I-SceI—mediated recombination and CRISPR-Cas9 counterselection. *Microbial Biotechnology*, *13*(1), 233. doi: 10.1111/1751-7915.13396

Xiao, Y., Luo, M., Hayes, R. P., Kim, J., Ng, S., Ding, F., Liao, M., & Ke, A. (2017). Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. *Cell*, *170*(1), 48-60.e11. doi: 10.1016/J.CELL.2017.06.012

Xiao, Y., Ng, S., Hyun Nam, K., & Ke, A. (2017). How type II CRISPR–Cas establish immunity through Cas1-Cas2-mediated spacer integration. *Nature 2017 550:7674*, *550*(7674), 137–141. doi: 10.1038/nature24020

Yang, J., Song, Y., Deng, X., Vanegas, J. A., You, Z., Zhang, Y., Weng, Z., Avery, L., Dieckhaus, K. D., Peddi, A., Gao, Y., Zhang, Y., & Gao, X. (2022). Engineered LwaCas13a with enhanced collateral activity for nucleic acid detection. *Nature Chemical Biology 2022 19:1*, *19*(1), 45–54. doi: 10.1038/s41589-022-01135-y

Yeeles, J. T. P., Poli, J., Marians, K. J., & Pasero, P. (2013). Rescuing Stalled or Damaged Replication Forks. *Cold Spring Harbor Perspectives in Biology*, *5*(5). doi: 10.1101/CSHPERSPECT.A012815

Yoganand, K. N., Muralidharan, M., Nimkar, S., & Anand, B. (2019). Fidelity of prespacer capture and processing is governed by the PAM-mediated interactions of Cas1-2 adaptation complex in CRISPR-Cas type I-E system. *The Journal of Biological Chemistry*, *294*(52), 20039. doi: 10.1074/JBC.RA119.009438

Yosef, I., Goren, M. G., & Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Research*, *40*(12), 5569–5576. doi: 10.1093/NAR/GKS216

Young, J. C., Dill, B. D., Pan, C., Hettich, R. L., Banfield, J. F., Shah, M., Fremaux, C., Horvath, P., Barrangou, R., & VerBerkmoes, N. C. (2012). Phage-Induced Expression of CRISPR-Associated Proteins Is Revealed by Shotgun Proteomics in *Streptococcus thermophilus*. *PLOS ONE*, *7*(5), e38077. doi: 10.1371/JOURNAL.PONE.0038077

Zechner, E. L., Lang, S., & Schildbach, J. F. (2012). Assembly and mechanisms of bacterial type IV secretion machines. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1592), 1073–1087. doi: 10.1098/RSTB.2011.0207

Zhang, X., Vrijenhoek, J. E. P., Bonten, M. J. M., Willems, R. J. L., & van Schaik, W. (2011). A genetic element present on megaplasmids allows *Enterococcus faecium* to use raffinose as carbon source. *Environmental Microbiology*, *13*(2), 518–528. doi: 10.1111/J.1462-2920.2010.02355.X

Zheng, Y., Roberts, R. J., & Kasif, S. (2004). Identification of genes with fast-evolving regions in microbial genomes. *Nucleic Acids Research*, *32*(21), 6347. doi: 10.1093/NAR/GKH935

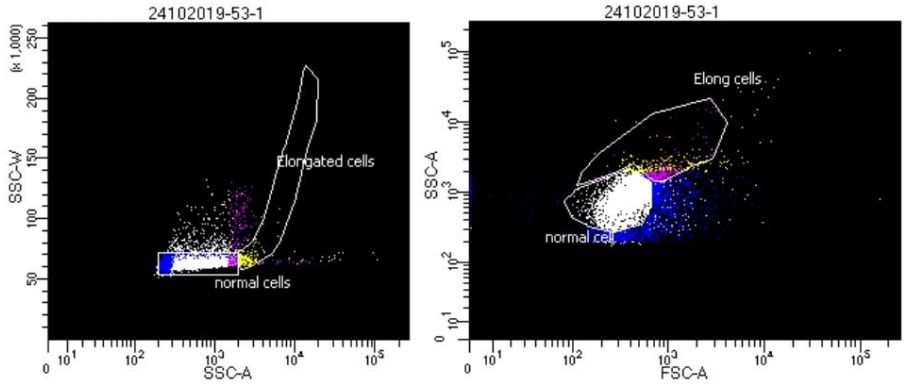
6. Supplementary Material

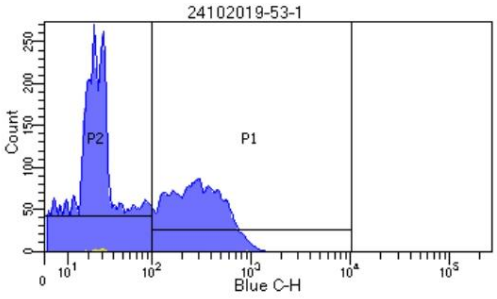
6.1 FACS analysis raw data

Supplementary Table 1. Samples of FACS screening. Fluorescence intensity measurement of a gfp-containing plasmid targeted by engineered crRNAs in *P. oleovorans*.

Tube	Sample
53-1	Control^{-}, randomized spacer sequence
53-2	
53-3	
50-1	5′-CCG-3′ PAM, targeting coding strand
50-2	
50-3	
48-1	5′-AAG-3′ PAM, targeting non-coding strand
48-2	
48-3	
47-1	AAG PAM, targeting coding targeting
47-2	
47-3	

BD FACSDiva 8.0.1





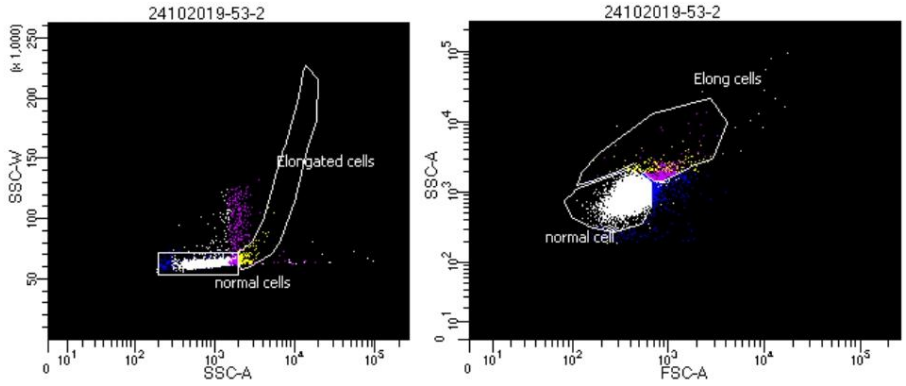
Tube: 53-1

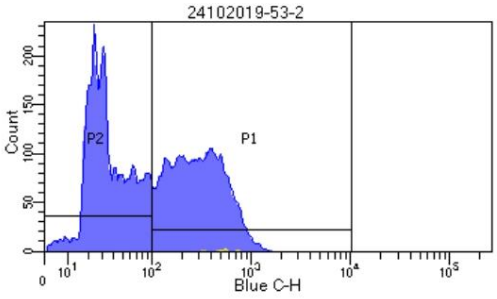
Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
normal cells	9,361	93.6	93.6
normal cell	7,842	78.4	78.4
Elong cells	379	3.8	3.8
Elongated cells	192	1.9	1.9
P1	3,770	37.7	37.7
P2	6,230	62.3	62.3

Experiment Name: Xiaohan
Specimen Name: 24102019
Tube Name: 53-1
Record Date: Oct 24, 2019 6:28:15 PM

Population	#Events	%Parent	Blue C-H Median	Blue C-H rSD
All Events	10,000	####	41	47
normal cells	9,361	93.6	42	49
normal cell	7,842	78.4	56	64
Elong cells	379	3.8	150	187
Elongated cells	192	1.9	172	219
P1	3,770	37.7	275	176
P2	6,230	62.3	24	10

BD FACSDiva 8.0.1



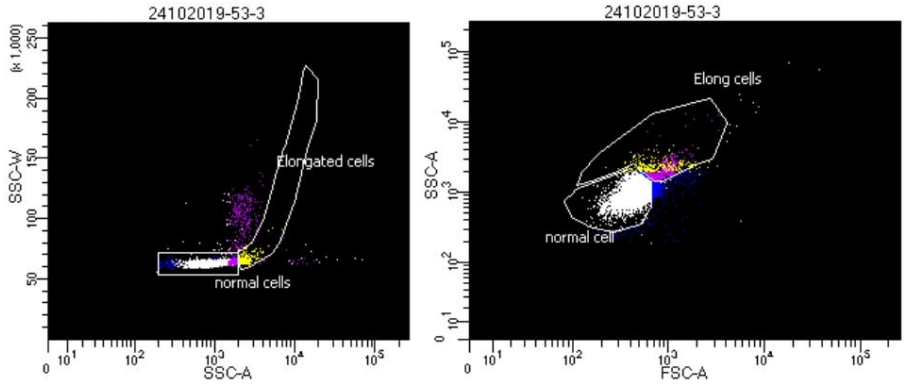


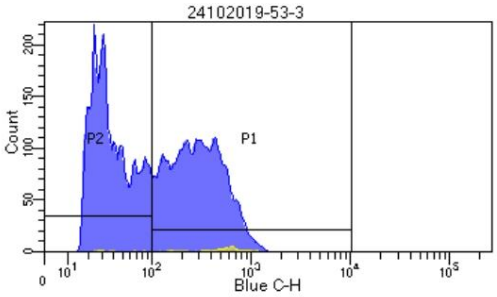
Tube: 53-2

Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
normal cells	9,598	96.0	96.0
normal cell	9,092	90.9	90.9
Elong cells	403	4.0	4.0
Elongated cells	142	1.4	1.4
P1	4,936	49.4	49.4
P2	5,064	50.6	50.6

Experiment Name: Xiaohan
Specimen Name: 24102019
Tube Name: 53-2
Record Date: Oct 24, 2019 6:29:21 PM

Population	#Events	%Parent	Blue C-H Median	Blue C-H rSD
All Events	10,000	####	95	110
normal cells	9,598	96.0	91	104
normal cell	9,092	90.9	90	102
Elong cells	403	4.0	261	295
Elongated cells	142	1.4	361	356
P1	4,936	49.4	287	191
P2	5,064	50.6	29	15





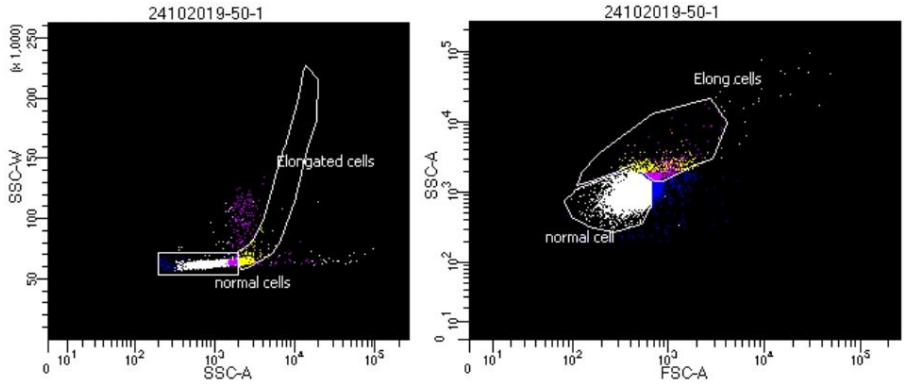
Tube: 53-3

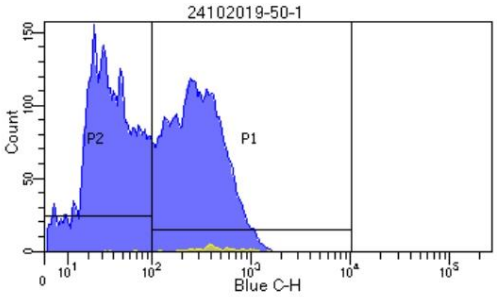
Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
normal cells	9,498	95.0	95.0
normal cell	8,397	84.0	84.0
Elong cells	683	6.8	6.8
Elongated cells	268	2.7	2.7
P1	5,128	51.3	51.3
P2	4,872	48.7	48.7

Experiment Name: Xiaohan
Specimen Name: 24102019
Tube Name: 53-3
Record Date: Oct 24, 2019 6:30:34 PM

Population	#Events	%Parent	Blue C-H Median	Blue C-H rSD
All Events	10,000	####	105	120
normal cells	9,498	95.0	99	111
normal cell	8,397	84.0	93	104
Elong cells	683	6.8	270	316
Elongated cells	268	2.7	314	365
P1	5,128	51.3	282	185
P2	4,872	48.7	31	16

BD FACSDiva 8.0.1





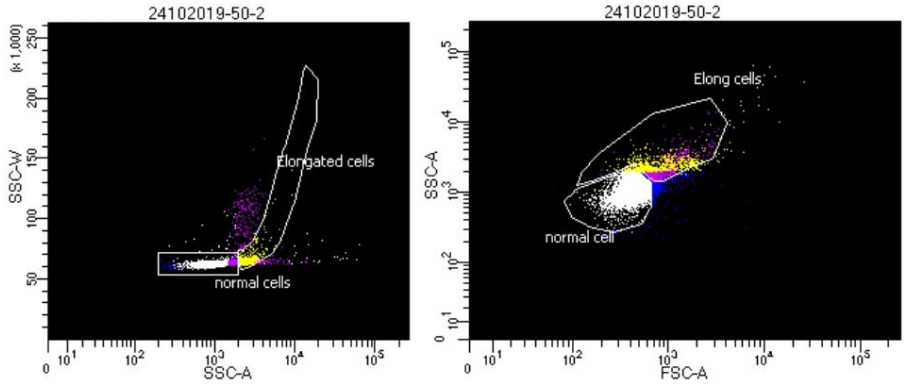
Tube: 50-1

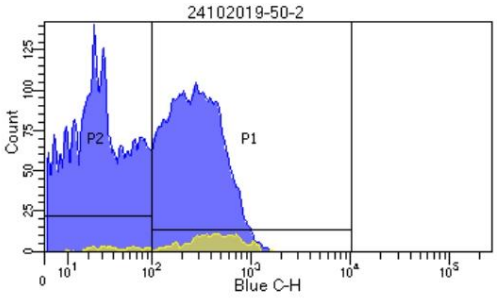
Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
normal cells	9,564	95.6	95.6
normal cell	8,362	83.6	83.6
Elong cells	620	6.2	6.2
Elongated cells	264	2.6	2.6
P1	5,066	50.7	50.7
P2	4,934	49.3	49.3

Experiment Name: Xiaohan
Specimen Name: 24102019
Tube Name: 50-1
Record Date: Oct 24, 2019 6:31:26 PM

Population	#Events	%Parent	Blue C-H Median	Blue C-H rSD
All Events	10,000	####	102	119
normal cells	9,564	95.6	98	113
normal cell	8,362	83.6	92	105
Elong cells	620	6.2	247	296
Elongated cells	264	2.6	314	344
P1	5,066	50.7	277	176
P2	4,934	49.3	33	21

BD FACSDiva 8.0.1





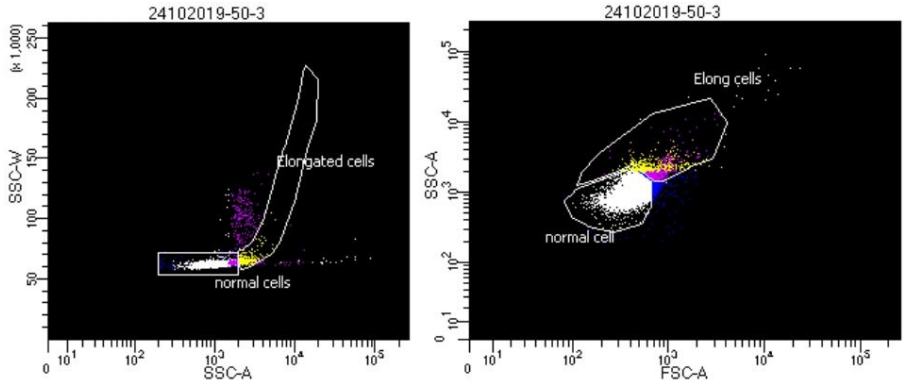
Tube: 50-2

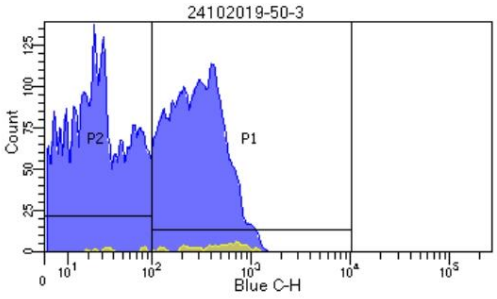
Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
normal cells	9,079	90.8	90.8
normal cell	7,738	77.4	77.4
Elong cells	1,174	11.7	11.7
Elongated cells	682	6.8	6.8
P1	5,081	50.8	50.8
P2	4,919	49.2	49.2

Experiment Name: Xiaohan
Specimen Name: 24102019
Tube Name: 50-2
Record Date: Oct 24, 2019 6:32:44 PM

Population	#Events	%Parent	Blue C-H Median	Blue C-H rSD
All Events	10,000	####	103	130
normal cells	9,079	90.8	88	111
normal cell	7,738	77.4	79	99
Elong cells	1,174	11.7	278	311
Elongated cells	682	6.8	309	340
P1	5,081	50.8	278	181
P2	4,919	49.2	25	22

BD FACSDiva 8.0.1





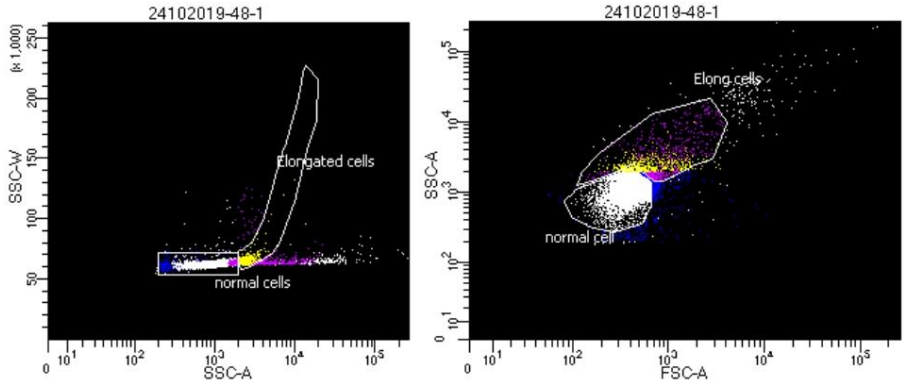
Tube: 50-3

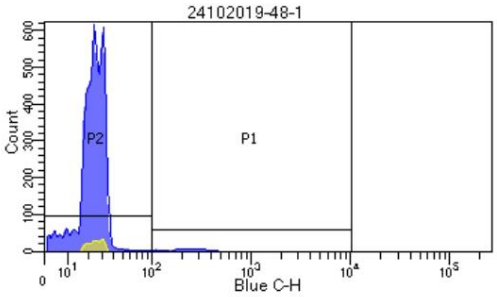
Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
normal cells	9,442	94.4	94.4
normal cell	8,517	85.2	85.2
Elong cells	738	7.4	7.4
Elongated cells	365	3.6	3.6
P1	4,950	49.5	49.5
P2	5,050	50.5	50.5

Experiment Name: Xiaohan
Specimen Name: 24102019
Tube Name: 50-3
Record Date: Oct 24, 2019 6:33:10 PM

Population	#Events	%Parent	Blue C-H Median	Blue C-H rSD
All Events	10,000	####	96	123
normal cells	9,442	94.4	88	113
normal cell	8,517	85.2	82	105
Elong cells	738	7.4	272	322
Elongated cells	365	3.7	274	345
P1	4,950	49.5	285	185
P2	5,050	50.5	24	22

BD FACSDiva 8.0.1



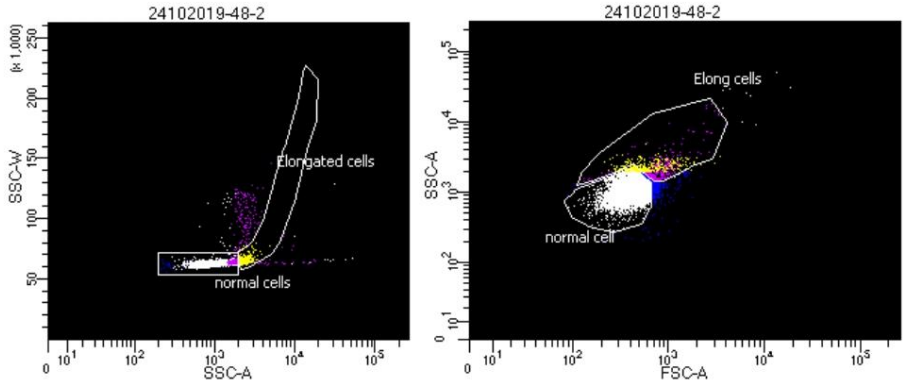


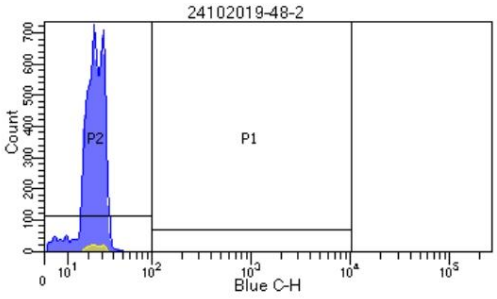
Tube: 48-1

Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
normal cells	9,072	90.7	90.7
normal cell	8,119	81.2	81.2
Elong cells	871	8.7	8.7
Elongated cells	526	5.3	5.3
P1	446	4.5	4.5
P2	9,554	95.5	95.5

Experiment Name: Xiaohan
Specimen Name: 24102019
Tube Name: 48-1
Record Date: Oct 24, 2019 6:37:20 PM

Population	#Events	%Parent	Blue C-H Median	Blue C-H rSD
All Events	10,000	####	23	7
normal cells	9,072	90.7	23	7
normal cell	8,119	81.2	23	7
Elong cells	871	8.7	24	7
Elongated cells	526	5.3	24	7
P1	446	4.5	239	128
P2	9,554	95.5	23	7





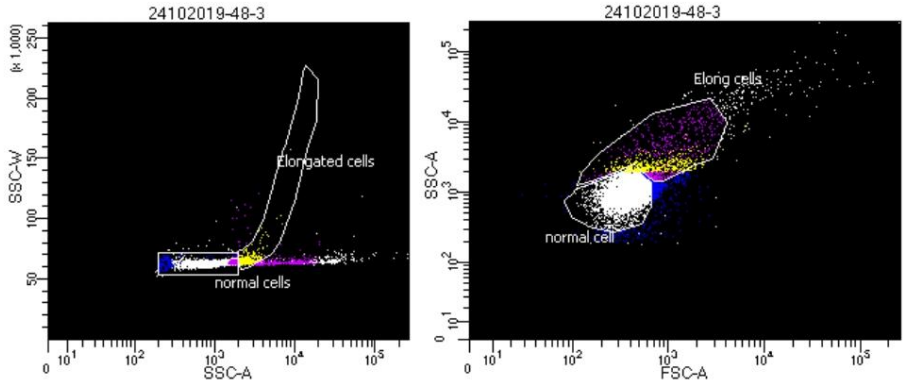
Tube: 48-2

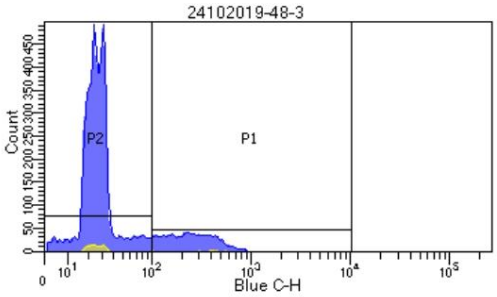
Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
normal cells	9,512	95.1	95.1
normal cell	8,720	87.2	87.2
Elong cells	549	5.5	5.5
Elongated cells	350	3.5	3.5
P1	43	0.4	0.4
P2	9,957	99.6	99.6

Experiment Name: Xiaohan
Specimen Name: 24102019
Tube Name: 48-2
Record Date: Oct 24, 2019 6:37:56 PM

Population	#Events	%Parent	Blue C-H Median	Blue C-H rSD
All Events	10,000	####	23	7
normal cells	9,512	95.1	23	7
normal cell	8,720	87.2	23	7
Elong cells	549	5.5	23	6
Elongated cells	350	3.5	23	6
P1	43	0.4	282	175
P2	9,957	99.6	23	6

BD FACSDiva 8.0.1





Tube: 48-3

Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
normal cells	9,046	90.5	90.5
normal cell	8,051	80.5	80.5
Elong cells	863	8.6	8.6
Elongated cells	503	5.0	5.0
P1	1,968	19.7	19.7
P2	8,032	80.3	80.3

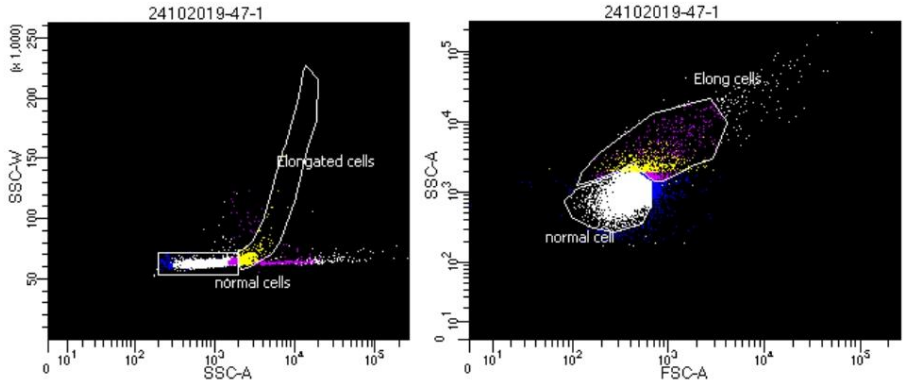
Experiment Name:	Xiaohan
Specimen Name:	24102019
Tube Name:	48-3
Record Date:	Oct 24, 2019 6:38:51 PM

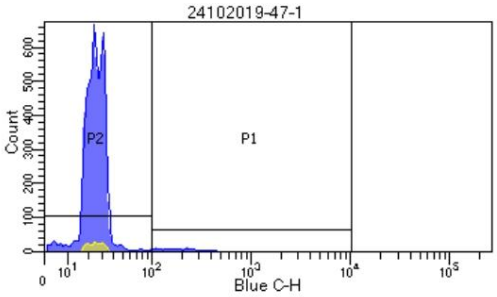
Population	#Events	%Parent	Blue C-H Median	Blue C-H rSD
All Events	10,000	####	27	12
normal cells	9,046	90.5	26	10
normal cell	8,051	80.5	26	10
Elong cells	863	8.6	28	15
Elongated cells	503	5.0	34	27
P1	1,968	19.7	251	159
P2	8,032	80.3	24	7

Global Sheet1

Printed on: Thu Oct 24, 2019 07:02:18 CEST

BD FACSDiva 8.0.1





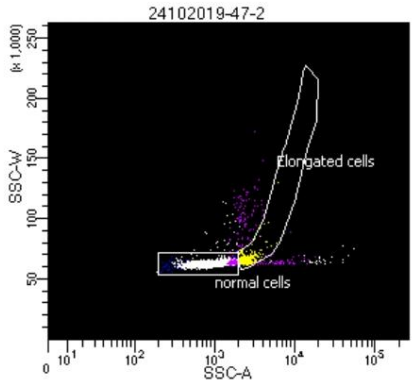
Tube: 47-1

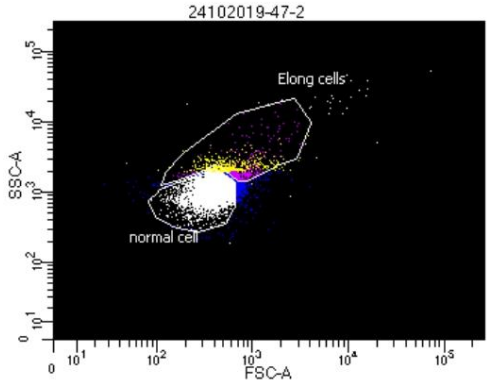
Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
normal cells	9,195	92.0	92.0
normal cell	8,314	83.1	83.1
Elong cells	746	7.5	7.5
Elongated cells	484	4.8	4.8
P1	513	5.1	5.1
P2	9,487	94.9	94.9

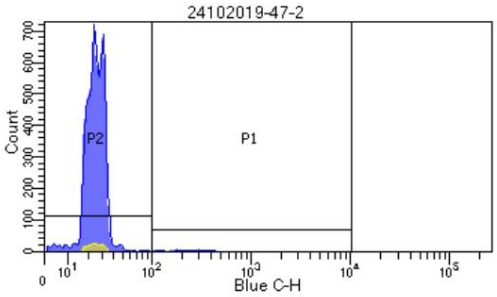
Experiment Name: Xiaohan
Specimen Name: 24102019
Tube Name: 47-1
Record Date: Oct 24, 2019 6:39:42 PM

Population	#Events	%Parent	Blue C-H Median	Blue C-H rSD
All Events	10,000	####	24	7
normal cells	9,195	92.0	24	7
normal cell	8,314	83.1	24	7
Elong cells	746	7.5	25	7
Elongated cells	484	4.8	24	7
P1	513	5.1	217	123
P2	9,487	94.9	23	7

BD FACSDiva 8.0.1





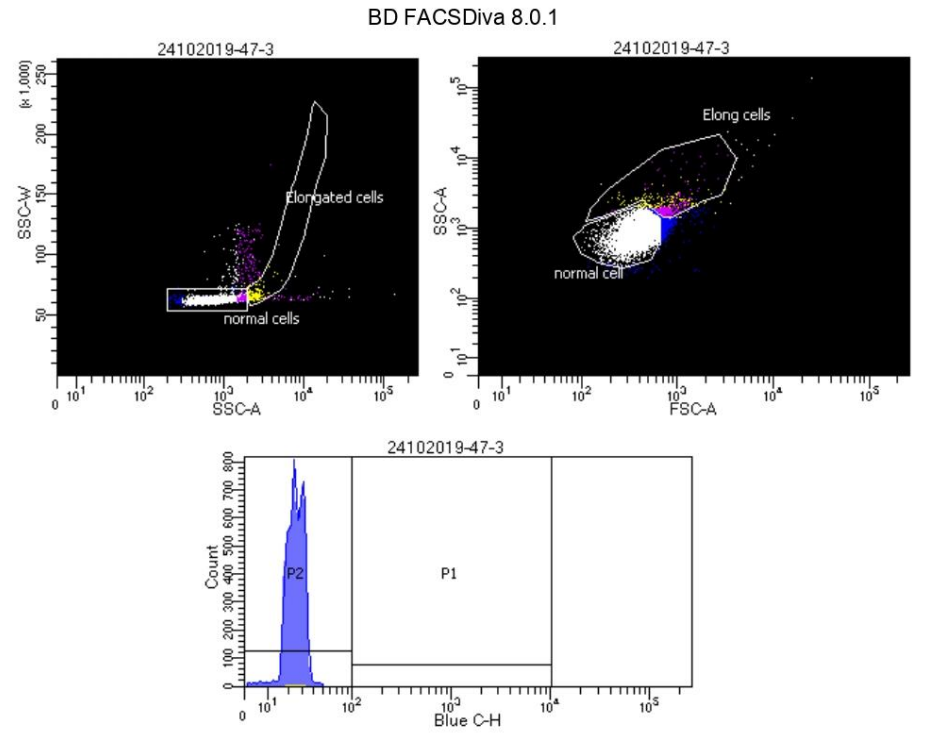


Tube: 47-2

Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
normal cells	9,404	94.0	94.0
normal cell	8,470	84.7	84.7
Elong cells	705	7.0	7.0
Elongated cells	447	4.5	4.5
P1	414	4.1	4.1
P2	9,586	95.9	95.9

Experiment Name: Xiaohan
Specimen Name: 24102019
Tube Name: 47-2
Record Date: Oct 24, 2019 6:40:14 PM

Population	#Events	%Parent	Blue C-H Median	Blue C-H rSD
All Events	10,000	####	24	7
normal cells	9,404	94.0	24	7
normal cell	8,470	84.7	24	7
Elong cells	705	7.0	24	7
Elongated cells	447	4.5	24	7
P1	414	4.1	260	165
P2	9,586	95.9	23	6



Tube: 47-3

Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
normal cells	9,633	96.3	96.3
normal cell	8,793	87.9	87.9
Elong cells	392	3.9	3.9
Elongated cells	152	1.5	1.5
P1	20	0.2	0.2
P2	9,980	99.8	99.8

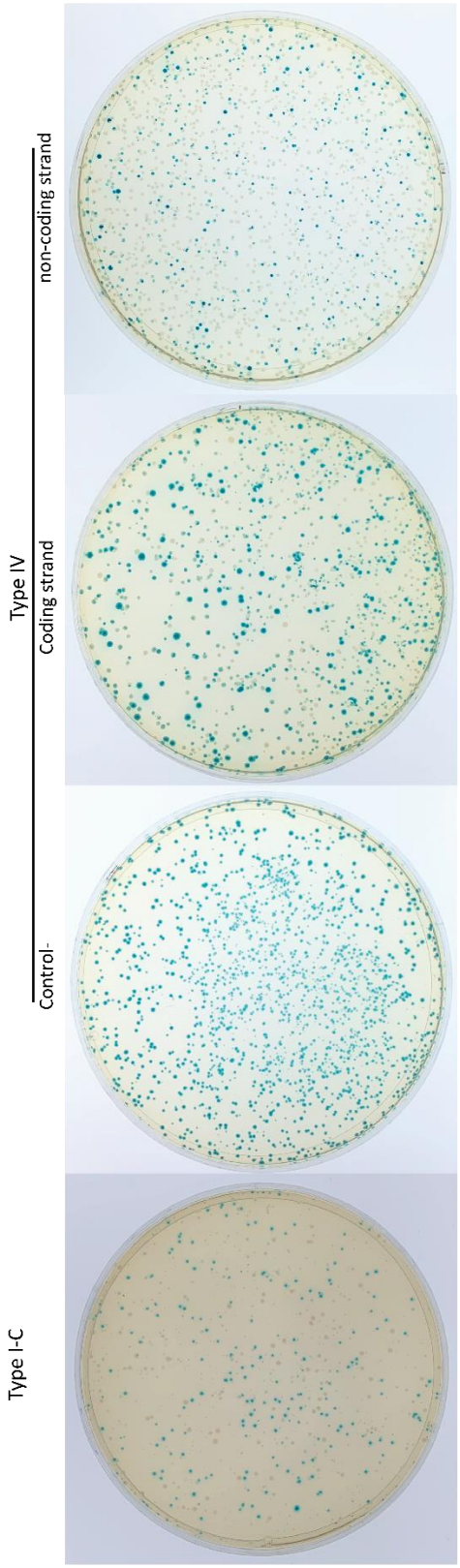
Experiment Name: Xiaohan
Specimen Name: 24102019
Tube Name: 47-3
Record Date: Oct 24, 2019 6:40:42 PM

Population	#Events	%Parent	Blue C-H Median	Blue C-H rSD
All Events	10,000	####	23	6
normal cells	9,633	96.3	23	6
normal cell	8,793	87.9	23	6
Elong cells	392	3.9	23	6
Elongated cells	152	1.5	23	6
P1	20	0.2	178	73
P2	9,980	99.8	23	6

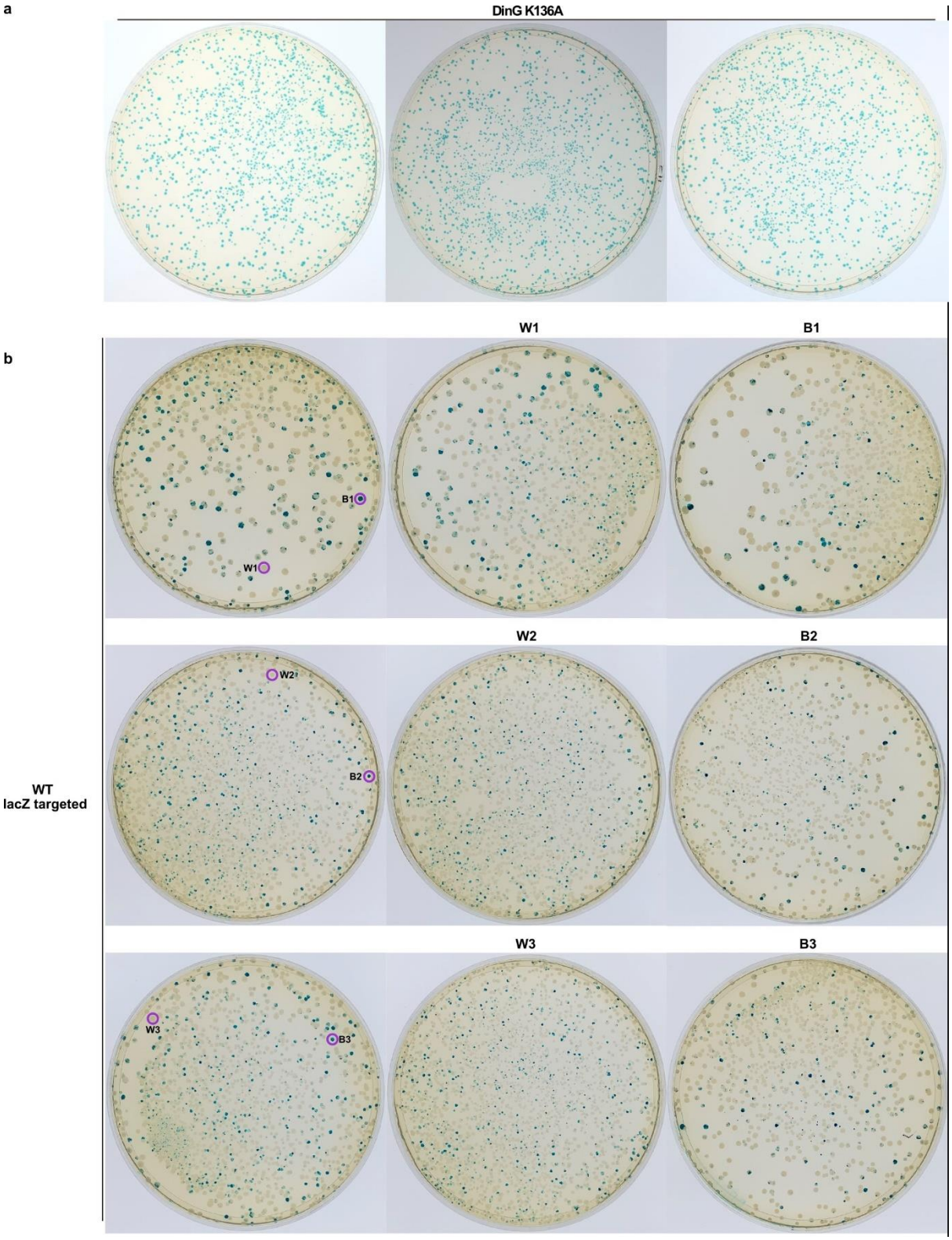
Global Sheet1 Printed on: Thu Oct 24, 2019 07:02:22 CEST

Supplementary Figure 1. FACS analysis of green fluorescence intensity in *P. oleovorans* cells expressing sfGFP. A total of 10,000 events were recorded for each sample, and cells were categorized as elongated or normal based on size and shape. The threshold for blue C-H was set at 10^2, with cells above this threshold considered to be GFP-positive. Detailed information for each sample can be found in Supplementary Table 1.

6.2 LacZ targeting assay raw data



Supplementary Figure 2. Raw data of LacZ targeting assay (Figure 20a). Raw images of the plates.



Supplementary Figure 3. Blue white screening of *lacZ* targeted *E. coli* colonies. a. *LacZ* (genome) targeted by the type IV-A CRISPR-Cas system with point mutation in DinG (K136A). b. *LacZ* (genome) targeted by the type IV-A CRISPR-Cas system (left), followed by plates derived from white colonies (W1-3, circled in purple) and from blue colonies (B1-3, circled in purple).

Abbreviations

ATP adenosine triphosphate

BSA bovine serum albumin

cDNA complementary DNA

ddH2O two times destilled water

DEPC diethylpyrocarbonate

DMSO dimethyl sulfoxide

DNA deoxyribonucleic acid

DNase desoxyribonuclease

dNTP deoxyribonucleotide triphosphate

ds downstream

dsDNA double-stranded DNA

DTT dithiothreitol

EDTA ethylene-diamine-tetraacetic acid

g gram

h hour

IPTG isopropyl β-D-1-thiogalactopyranoside

kb kilobase

l liter

LB lysogeny broth

M molar (mol/l)

min minutes

MW molecular weight

µ micro (10^{-6})

n nano (10^{-9})

nt nucleotides

ORF open reading frame

PCR polymerase chain reaction

pH negative logarithm of the hydrogen ion

RNA ribonucleic acid

RNase ribonuclease

RNP ribonucleoprotein complex

rpm rounds per minute

rRNA ribosomal RNA

RT room temperature

s seconds

crRNA CRISPR RNA

ssDNA single-stranded DNA

ssRNA single stranded RNA

TAE tris-acetate-EDTA-buffer

TBE tris-borate-EDTA-buffer

U unit (enzyme activity)

us upstream

UV ultraviolet

w/o without

% (v/v) percent by volume

% (w/v) percent by weight

> higher than

< lower than

Δ deletion

Contributions

The results presented in this work were independently conducted by me, with no assistance beyond that listed here. The contributions of other individuals involved in this work are as follows:

Dr. José Vicente Gomes-Filho: operated the Illumina® MiniSeq Sequencing System and provided assistance with the preparation of small RNA and DNA sequencing libraries (Figure 10 and 12, Table 2). He also analyzed the small RNA-seq data and created the associated Figure (Figure 10).

Julia Wiegel: assisted in cloning 9 constructs for the PAM scanning library (Figure 18).

Selina Rust: constructed the ΔCsf1, Csf1 C30A, Csf1 C33A, Csf1 C66A, and Csf1 C69A strains and performed the corresponding transformation efficiency assay with the recombinant *E. coli* Type IV-A CRISPR-Cas system (Figure 17d).

Pascal Schäfer: constructed the recombinant *E. coli* Type I-E CRISPR-Cas system and performed the lacZ targeting assay for the Type I-E system (Figure 20a).

Curriculum vitae

Xiaohan Guo
Address: Mari Andriessenrade 41, 2907 MA, Capelle, Netherlands
Phone: +31(0)647603740, Email: guoxiaohan92@outlook.com

Education

• PhD, Max Planck Institute for Terrestrial Microbiology, 2017 - Current (expected December 2022)
• MS, Biotechnology, Wageningen University & Research, 2014 - 2016
• BS, Biotechnology, China Agricultural University, 2010 - 2014

Experience

Max Planck Institute for Terrestrial Microbiology

PhD Researcher August 2017 – Present
(maternity leave from April 2020 – December 2020)
• Published paper in Nature Microbiology on the characterization of the self-targeting Type IV CRISPR interference system in *Pseudomonas oleovorans*
• Conducted *in vitro* and *in vivo* investigations of essential Cas proteins
• Performed RNA-Seq and whole genome sequencing and analysis (Illumina & Nanopore)
• Identified the biological function of the Type IV CRISPR system
• Conducted bioinformatics analysis and functional prediction of Type IV CRISPR-Cas systems

Thatchtec BV

Junior Researcher February 2017 — June 2017
• Played a key role in a project: "Integrated solution for innovative biodegradation control of Agricultural Plastic Mulches"
• Setup a lab from scratch
• Isolated and characterized of bacteria strains from soil samples
• Contributed to the development of a new biodegradation control solution

University of Toronto

Research Assistant April 2016 — October 2016
• Identified crucial residues of a dual-functional anti-CRISPR protein AcrF6pae
• Investigated virus/host interaction
• Conducted null mutation analysis of AcrF6pae

Wageningen University & Research

Research Assistant September 2015 — April 2016
• Optimized succinic acid production of *Actinomyces* Strain Am4 from starch waste by two-fold
• Cultivated anaerobic bacteria
• Engineered chemostat bioprocesses

Institute of Biophysics, Chinese Academy of Sciences

Research Assistant March 2014 — June 2014
• Validated a high-throughput screening method for anti-CD146 antibodies
• Cultivated and maintained HUVEC cell cultures for antibody screening
• Performed RT-qPCR analysis to validate screening results

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Prof. Dr. Lennart Randau, for his invaluable advice, ongoing support, and patience during my PhD studies. I also extend my genuine appreciation to my thesis advisory committee, Prof. Dr. Tobias Erb and Prof. Dr. Martin Thanbichler, for their valuable advice and practical suggestions.

I would like to take this opportunity to thank my examiners, Prof. Dr. Martin Thanbichler, Prof. Dr. Peter Graumann and Dr. Georg Hochberg, for their invaluable contributions in reviewing this thesis.

I am also grateful to my colleagues and collaborators for their contributions to this publication: Mariana Sanchez-Londono, Dr. José Vicente Gomes-Filho, Rogelio Hernandez-Tamayo, Selina Rust, Leah M. Immelmann, Pascal Schäfer, Julia Wiegel, and Prof. Dr. Peter Graumann. I would also like to express my special thanks to all the current and former members of the Landau group for creating a friendly and supportive work environment. I would also like to thank my former students, Alessia Girschik, Selina Rust, Nathalie Klein and Pascal Schäfer, for their contributions.

I would also like to extend my appreciation to the Max Planck Society, IMPRS, Philipps-University Marburg, and DFG for the funding, facilities, and resources provided.

Lastly, I would like to thank my family and friends for their support, especially during the difficult time of COVID-19 social distancing. I could not have made it through without them.