CRISPR-Cas3:

Studying the molecular interactions that drive adaptation & engineering novel bacterial editing tools

by

Lina Maria Leon

DISSERTATION

Submitted in partial satisfaction of the requirements for degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry and Molecular Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:

Joseph Bondy-Denomy

D23522498F4247C...

Joseph Bondy-Denomy

Chair

DocuSigned by:

D255B261A64948B...

Seemay Chou

Geeta Narlikar

Committee Members

Acknowledgments

Thank you to all who have supported me as I worked towards this moment. Thank you to Dr. Joe Bondy-Denomy for showing me what it takes to be a good, thoughtful scientist. Thank you to my committee members, Dr. Seemay Chou and Dr. Geeta Narlikar, who encouraged me to think bigger, and to always aim higher, even when I felt it was impossible. I strive to be the kind of scientist you are all proud of.

Graduate school would not have been the same without the friends I made along the way: Thank you to the old guard, Adair Borges and Senén Mendoza, for your friendship and support, and to the new generation of graduate students in the group, Erin Huiting and Matt Johnson, for your contagious energy and enthusiasm. To my labmates, thank you for your thoughtful discussions and encouragement. In particular, Caroline Mahendra, for showing me what how to be bold, fearless, and meticulous, and Bálint Csörgő, for showing me how to be fair and thorough.

To my closest friends, Lindsey Backman, Akila Raja, and Nette Ville, thank you for your advice, late night chats, and lighthearted moments that have carried me through this time.

None of this would have been possible without the support of my family. To my mom, dad, and brother, thank you for believing in me. Thank you to my grandparents, Mamuchis, Omi, and Opi, for your constant love. And thank you to my fiancé, Jacob Kimmel, for your love and patience.

Finally, thank you to the women scientists who paved the way for me to be here. I stand on the shoulders of giants.

Contributions

Work in this dissertation has previously been published in *Nature Methods* and *Nucleic Acids Research* in the following volumes:

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

León, L. M., Park, A. E., Borges, A. L., Zhang, J. Y., and Bondy-Denomy, J. (2021). Mobile element warfare via CRISPR and anti-CRISPR in Pseudomonas aeruginosa. Nucleic Acids Research. https://doi.org/10.1093/nar/ gkab006 159202

CRISPR-Cas3: Studying the molecular interactions that drive adaptation & engineering novel bacterial editing tools

Lina M. León

Abstract

Bacteria deploy multiple defenses to prevent mobile genetic element (MGEs) invasion. CRISPR-Cas immune systems feature RNA-guided nucleases that target MGEs, which counter with anti-CRISPR (Acr) proteins. Our understanding of the biology and co-evolutionary dynamics of the common Type I-C CRISPR-Cas subtype has lagged because it lacks an in vivo phage-host model system. Here, we show the anti-phage function of a Pseudomonas aeruginosa Type I-C CRISPR-Cas system encoded on an active conjugative pKLC102 island, and the inhibition of this system by multiple distinct MGEs encoding a diverse repertoire of Type I-C Acr proteins. Seven distinct AcrIC proteins were identified, with four of them, including previously described DNA mimic AcrIF2 (now AcrIF2*), surprisingly also inhibiting other *P. aeruginosa* CRISPR-Cas subtypes (Type I-E or I-F). Dual inhibition comes at a cost, however, as the simultaneous expression of Type I-F and Type I-C systems rendered phages expressing AcrIF2* more sensitive to targeting. This effect was exacerbated by mutagenesis of AcrIF2's acidic residues, which made AcrIF2 defective for Type I-C inhibition, but only when in competition with the Type I-F complex. Like AcrIF2*, five of the AcrIC proteins block DNA binding by the crRNA-guided Cascade complex, while two function downstream of DNA binding, likely preventing Cas3 recruitment or activity. One such inhibitor, AcrIC3, is found encoded alongside bona fide Cas3 inhibitors, AcrIF3 and AcrIE1 in conjugative elements, forming an "anti-Cas3" cluster. Collectively, our findings demonstrate an active battle between an MGE-encoded CRISPR-Cas system and its diverse MGE targets. On the technological angle, CRISPR–Cas enzymes have enabled programmable gene editing in eukaryotes and prokaryotes. However, the leading Cas9 and Cas12a enzymes are limited in their ability to make large deletions. Here, we used the processive nuclease Cas3, together with a minimal Type I-C Cascade-based system for targeted genome engineering in bacteria. DNA cleavage guided by a single CRISPR RNA generated large

deletions (7–424 kilobases) in *Pseudomonas aeruginosa* with near-100 percent efficiency, while Cas9 yielded small deletions and point mutations. Cas3 generated bidirectional deletions originating from the programmed site, which was exploited to reduce the *P. aeruginosa* genome by 837 kb (13.5 percent). Large deletion boundaries were efficiently specified by a homology-directed repair template during editing with Cascade–Cas3, but not Cas9. A transferable 'all-in-one' vector was functional in *Escherichia coli*, *Pseudomonas syringae* and *Klebsiella pneumoniae*, and endogenous CRISPR–Cas use was enhanced with an 'anti-anti-CRISPR' strategy. *P. aeruginosa* Type I-C Cascade–Cas3 (PaeCas3c) facilitates rapid strain manipulation with applications in synthetic biology, genome minimization and the removal of large genomic regions.

Contents

List of Figures

Chapter 1

Mobile element warfare via CRISPR & anti-CRISPR in *P. aeruginosa*

1.1 Introduction

The plasticity and rapid evolution of bacterial genomes is driven by the exchange of genetic material between diverse species. This genetic mobility can be blocked by bacterial immune systems, such as restriction enzymes and CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR associated genes). CRISPR-Cas systems utilize short RNA guides, encoded within a CRISPR array where they are separated by repeat sequences, to direct either a multi-protein (Class 1; Type I, Type III, Type IV) or single protein (Class 2; Type II, V, or VI) effector complex to a matching target on a mobile genetic element (MGE) [1]. The targeting paradigm can be inverted, where a CRISPR-Cas system is encoded by a lytic bacteriophage, targeting the host [2] or targeting other phages [3].

Pseudomonas aeruginosa is an opportunistic human pathogen and also a leading model organism for studies pertaining to bacteriophage-CRISPR interactions and Class 1 CRISPR-Cas biology [4]. Functional Type I-F [5, 6], I-E [7, 8], and now IV-A [9] systems have been described, however, a fourth CRISPR-Cas system encoded by this species, the Type I-C system, has not been well characterized [10]. Type I-C systems are phylogenetically widespread [11], and can be found in

Streptococcus pyogenes, Vibrio, Clostridium, Neisseria, and *Bacillus* species, but are among the least studied subtypes within the adaptive branch of bacterial immunity. A native *Legionella pneumophila* system was used as a model for spacer acquisition and plasmid targeting [12], while remaining studies of Type I-C systems in *Eggerthella lenta* [13], *Desulfovibrio vulgaris* [14], *Bacillus halodurans* [15], and *Xanthomonas oryzae* [16] have been explored heterologously or in vitro, with gaps in our understanding of these systems remaining to be filled. Type I-C systems employ a compact surveillance complex of Cas5, Cas7, and Cas8 with the CRISPR RNA (crRNA) and the trans-acting nuclease-helicase, Cas3, which is recruited to cleave and processively degrade DNA [17]. These systems lack the common Cas6 crRNA-processing RNase, with Cas5 filling that role [18–20].

CRISPR immunity is often simplified to three stages: adaptation, biogenesis and interference, but a fourth, and equally important facet, is MGE counter-evolution. Anti-CRISPR proteins (Acrs) encoded by MGEs disable CRISPR-Cas systems using diverse mechanisms. Strategies range from blocking DNA binding sites (e.g. AcrIF1, AcrIF2, AcrIF10, AcrIIA2, AcrIIA4) [21–23], to blocking DNA cleavage (e.g. AcrIE1, AcrIF3, AcrIIC1) [8, 24, 25] and even acting enzymatically to disable CRISPR-Cas (e.g. AcrVA1, AcrVA5, AcrIII-1) [4, 26] Anti-CRISPR discovery efforts continue to yield new biochemical mechanisms for CRISPR-Cas inhibition, while also providing evidence that MGEs encoding acr genes face CRISPR-Cas challenge in situ. Some Acr proteins are described as "broad-spectrum" due to inhibition of diverged Cas proteins, however, the costs and benefits of this phenotype are yet to be investigated. Here, we describe the MGE targets of the *P. aeruginosa* Type I-C CRISPR-Cas system, which itself is encoded on an MGE, present direct evidence of endogenous Type I-C CRISPR-Cas activity, and report the discovery of seven *Pseudomonas* Type I-C anti-CRISPRs, including four that have dual inhibitory activity.

1.2 Materials and Methods

1.2.1 Microbes

Cell culturing

Pseudomonas aeruginosa strains (PAO1, PA14 and PA4386) and *Escherichia coli* strains (DH5a) were cultured using lysogeny broth (LB) agar or liquid media at 37 °C supplemented with gentamicin, where applicable, to maintain pHERD30T (50 µg/mL for *P. aeruginosa*, 30 µg/mL for *E. coli*). In all *P. aeruginosa* experiments, expression of genes of interest in pHERD30T was induced using 0.1 percent arabinose.

Type I-C CRISPR-Cas expression in PAO1

$PAO1^{IC}$ activity was induced using 1mM IPTG. Construction of this strain is described [27] and may be referred to as LL77 (Targeting crRNA) or LL76 (Non targeting).

Bacterial transformations

P. aeruginosa transformations were performed using standard electroporation protocols [27]. Briefly, overnight cultures were washed twice in an equal volume of 10% glycerol and the washed pellet was concentrated tenfold in 10%glycerol. These electrocompetent cells were transformed with 20 – 200 ng plasmid, incubated shaking in LB for 1 hr at 37°C, plated on LB agar with appropriate selection, and incubated overnight at 37°C. Bacterial transformations for cloning were performed using *E. coli* DH5alpha (NEB) according to the manufacturer's instructions

CRISPRi

CRISPR interference transcriptional repression assays were conducted as in previous work [24]. A Δcas3 strain was lysogenized with a DMS3m phage encoding an Acr of interest. This lysogen was transformed with a plasmid encoding a crRNA targeting the *phzM* promoter. The crRNA and cas genes (in the case of Type I-C) were induced in overnight cultures with 0.25 mM IPTG and 0.05% arabinose. Pyocyanin levels were measured using an acid extraction protocol described

previously [24]. Pyocyanin quantification was normalized to a strain encoding AcrIIA4, which inhibits Cas9, but not the Type I CRISPR-Cas systems included in this study, resulting in cultures lacking pyocyanin.

1.2.2 Phage

Phage maintenance

Pseudomonas aeruginosa DMS3m-like phages (including JBD30 and DMS3m engineered phages) were amplified on PA14 ΔCRISPR, PAO1, or PA4386 Δ*cas3* and stored in SM buffer at 4 °C.

Construction of recombinant DMS3m acr phages

To generate the isogenic panel of DMS3m and JBD30 anti-CRISPR phages, recombination cassettes were generated with up- and down-stream overhangs to aca1 and the acr promoter flanking the Acr of interest, as previously described [28]. These genes were ordered from TWIST or IDT and were assembled into plasmids using Gibson assembly methods. Recombinant phages were generated by infecting cells transformed with the donor constructs and phages were isolated and assessed for resistance to CRISPR-Cas targeting. The presence of the anti-CRISPR gene was confirmed by PCR. To generate the virulent phages used for liquid growth curve assays, the DMS3m c-repressor gene, *gp1*, was mutated using plasmids described previously [28].

Plaque forming unit quantification

Phage plaque forming units (PFU) were quantified by mixing 10 µl of phage with 150 µl of an overnight bacterial culture. The infected cells were aliquoted into 3 mL molten 0.7% top agar and spread on an LB agar plate supplemented with 10 mM MgSO4 and appropriate inducers. After 18 hours of growth at 30 °C or 37 °C, individual plaques were counted. Three biological replicates were done per phage per strain.

Phage spot assays

3 mL of molten 0.7% top agar mixed with 150 µl of bacteria were spread on an LB agar plate supplemented with 10 mM MgSO4 to grow a bacterial lawn. Ten-fold serial dilutions of phage were made in SM buffer and 2 µl of each dilution was spotted on the lawn. Plates were incubated at 30 ˚C or 37 ˚C for 16 hours and imaged.

Efficiency of plaquing (EOP)

EOP was calculated as the ratio of the number of plaque forming units (PFUs) that formed on a targeting strain of bacteria (PAO1^{IC}, PA14 WT, PA4386 WT, PaLML1 plus crRNA plasmid) divided by the number of PFUs that formed on a related non-targeting strain (PAO1, PA14 ΔCRISPR, PA4386 Δ*cas3*, PaLML1 plus NT crRNA). Each PFU measurement was performed in biological triplicate. EOP data are displayed as the mean EOP ± standard deviation.

Escaper phage isolation

High titer phage preparations were mixed with overnight cultures and spread on an agar plate with top agar. Single plaques that formed after overnight propagation were picked with a sterile pipette tip and resuspended in SM buffer. This process was repeated two times under maintained targeting pressure. The escaper phages were ultimately titered and the protospacer region sequenced.

Liquid culture phage infections

PAO1^{IC} was transformed with plasmids encoding either the Type I-C or Type I-F systems from PaLML1 plus one non-DMS3m targeting spacer ("decoy" surveillance complexes) to determine the effect of CRISPR-Cas system co-expression. A separate Type I-F plasmid with a Cas8 mutation (K247A) was also constructed. *P. aeruginosa* strains were grown overnight and diluted 100x in LB supplemented with 10 mM MgSO4, gentamicin, 1 mM IPTG, and 0.1% arabinose. 140 µl of bacterial culture was infected with 10 µl of serially diluted virulent phage in a 96 well plate. Growth and infection was monitored for 20 hours using the Synergy H1 microplate reader (BioTek) at shaking at 37 ˚C. Phage was extracted after 20 hours by mixing 100 µl of culture from each well

with 20 µl chloroform, shaking at RT for 20 minutes, and centrifugation at 14, 000 x g for 2 minutes.

1.2.3 Bioinformatics

Numerical data were analyzed in Excel and plotted in GraphPad Prism 6.0.

Discovery of acr genes using *aca1* and *aca4*

Anti-CRISPR searches were done as previously described [27].

CRISPR array spacer analysis

Spacers were derived from the van Belkum dataset [10] (18 genomes with 12 non redundant arrays) or from Type I-C containing strains found using BLAST and CRISPRfinder [29](12 non-redundant arrays). Spacers were analyzed using CRISPRTarget [30] using the Genbank-environmental, RefSeq-plasmid, IMG/VR, and PHAST databases.

PAM analysis was done using the Berkeley Web Logo tool by submitting the upstream and downstream regions flanking the protospacer sequence. These 8 nucleotide long flanking sequences are part of the CRISPRTarget output. Every matching protospacer (low cutoff of 20, no redundant matches removed) was utilized for the PAM analysis for n= 4,443.

To determine the types of elements targeted by the spacers in our collection, the cut-off score was increased to 30 and a PAM match score of +5 was used to narrow the total number of hits to matching elements. If a spacer had multiple matches, the match with the highest score was selected as the representative for that spacer OR the match to a phage genome. Only one match was considered per spacer. This reduced the number of spacers to 131.

Matches were placed into the following categories: Myophages, Siphophages, Podophages, plasmids, and assorted prophages. A hit was placed into a phage family, rather than into the prophage category, if the CRISPRTarget output included a link to a specific phage genome. Importantly, this means that being placed into a phage family does not mean that a phage is strictly lytic. Prophages were identified by considering the genes in the protospacer neighborhood.

Lineage Tracing

For the Type I-C encoding strains from this study, WGS reads were imported from NCBI to Benchling, and the repeats were annotated using the Benchling annotation tool. Individual spacers were extracted using the CRISPRCasFinder tool by copying the entire CRISPR array region. Each spacer sequence was assigned a number, such that identical spacers in distinct strains were assigned the same number, allowing the visualization of spacer similarity across different strains. Lineages were manually curated using 18 previously published CRISPR arrays [10] and the additional 12 CRISPR arrays found in this study.

Anti-CRISPR phylogenetic tree generation

BLAST was used to generate the tree of AcrIC5 relatives. The following parameters were selected. Tree method: Fast minimum evolution. Max seq difference: 0.85. Distance: Grishin (protein).

1.3 Results

1.3.1 The MGE-encoded Type I-C system provides immunity in *P. aeruginosa*

Type I-C CRISPR–Cas systems previously described in 20 *P. aeruginosa* genomes [10], an environmental isolate in our lab (PaLML1), and 23 additional genomes found using BLAST, are encoded within pKLC102-like elements (Fig. 1.1A). This conjugative element family can be found as either an integrated island or episome in many gram negative bacteria, and is also known as *P. aeruginosa* pathogenicity island (PAPI-1) in some *P. aeruginosa* strains, including PA14 [31–34]. It is typically ˜100 kb and while it does not always encode a Type I-C system, it is known to carry virulence factors and increase pathogenicity. To determine if Type I-C CRISPR–Cas is active in *P. aeruginosa*, we first took a bioinformatic approach. While the Cas proteins are highly conserved (90–100% sequence identity) across strains, the CRISPR spacers are diverse (Fig. 1.8). Alignments of 3163 protospacers with upstream and downstream regions revealed the consensus PAM (Protospacer adjacent motif) to be 5′-TTC-3′, consistent with previous reports [35] (Fig. 1.1B). Among the 42 strains with CRISPR arrays (two published strains have cas genes without corresponding arrays), we observed

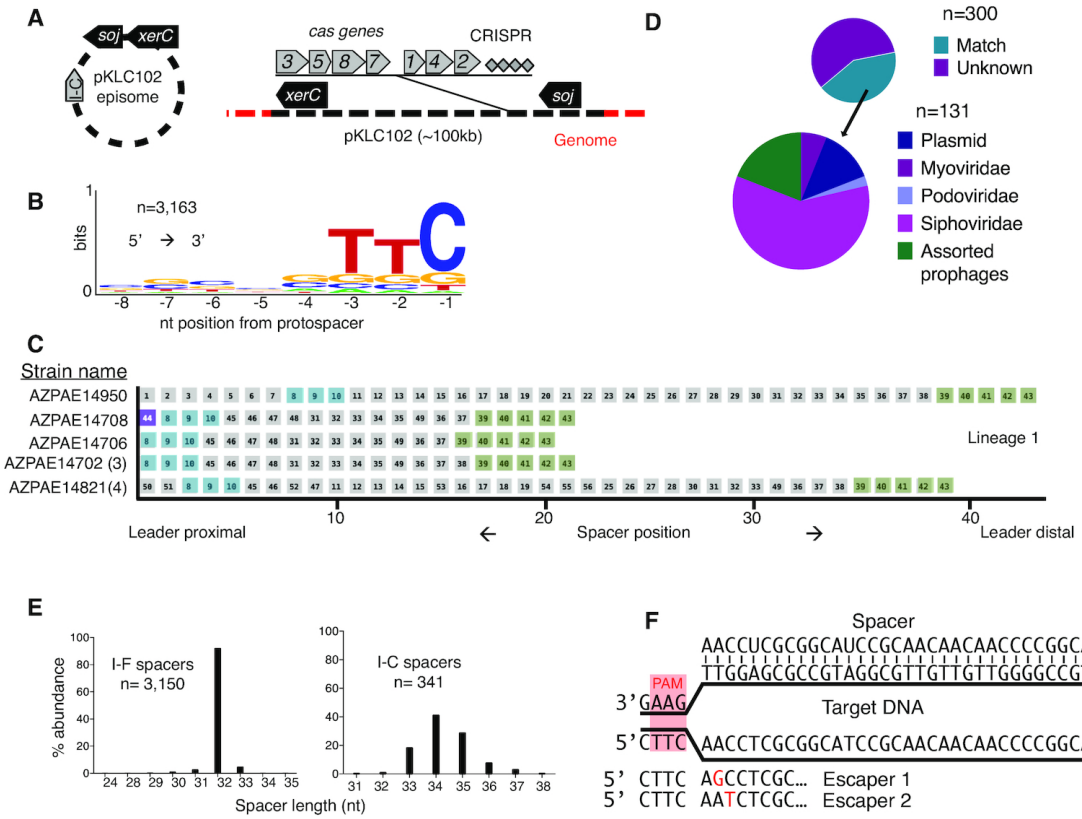


Fig 1.1. The Type I-C system is active in *P. aeruginosa*. (A) *Pseudomonas aeruginosa* Type I-C systems are found on pKLC102 elements. This island can be found integrated into the *P. aeruginosa* genome. Black arrows represent pKLC102 marker genes. *soj* is a chromosome partitioning protein, and xerC is a site-specific recombinase. The island can excise to form an episome. (B) WebLogo of the region upstream of Type I-C targeted regions showing the consensus PAM to be 5′ TTC 3′. (C) CRISPR arrays clustered into lineages based on spacer identity. Lineage 1 is shown here with 5 genomes. Spacer position is marked on the x axis. Spacers that are the same within a lineage are given the same number. Numbers in parentheses following the strain names indicate the number of genomes with the same CRISPR array. The spacer highlighted in purple, #44, is self-targeting. Spacers shared toward the leader proximal region are highlighted in blue. Spacers shared toward the leader distal region are highlighted in green. Shading is meant to facilitate comparisons between related arrays. (D) Of the 300 non-redundant spacers, 163 target sequenced genetic elements. Spacers labeled as unknown (dark purple, top pie chart) did not have any matches in sequence databases used by CRISPR Target. Spacers with matches to independent phage genomes (both lytic and temperate) were categorized into three families (siphoviridae, myoviridae, and podoviridae). Spacers that mapped back to phage-like regions in bacterial genomes were categorized as assorted prophages. (E) Comparison of spacer lengths found in either Type I-F or Type I-C *P. aeruginosa* CRISPR arrays. (F) Schematic of the phage protospacer and CRISPR–Cas spacer. PAM sequence is underlined and highlighted in red. Escaper mutations for phages isolated from PaLML1 are shown below the protospacer sequence with point mutations highlighted in red text.

spacer diversity suggestive of active acquisition (Fig. 1.1C and Fig. 1.8). The CRISPR arrays could be clustered into four broad lineages, with strains grouped if they share at least one spacer with another array (Fig. 1.1C and Fig. 1.8). Some strains have identical CRISPR arrays, which were condensed to one representative per array. Strains that cluster together tend to share most of the spacers towards the leader-distal end of the CRISPR array, suggesting that after diverging, each host continues to expand its array independently. For example, strains in lineage 1 share most of their ˜10–15 leader-distal spacers, and then undergo divergence with a series of unique spacers proximal to the leader (Fig. 1.1C). In lineage 2, the diversity is even more striking, as the strains are grouped together by just two "core" spacers (#74 and #75), but have highly distinct arrays, most notably strain AZPAE14395, with ˜40 unique spacers (Fig. 1.8). Strains in lineage 3 (PaLML1, AZPAE14876, AZPAE12421, etc.), and lineage 4 (WH-SGI-V-07071, and WH-SGI-V-07073) have completely dissimilar spacers (Fig. 1.8), despite having the same frame shift mutation in cas1 that results in an early stop codon, suggesting continued CRISPR dynamics through an unknown mechanism. In total, there are 300 non-redundant spacers in this collection, and 131 (44%) match sequenced elements, with most spacers targeting phages and prophages (114) and some matching plasmids (17) (Fig. 1.1D). Therefore, although pKLC102 is a 'selfish' genetic element, dissection of the Type I-C CRISPR spacer repertoire reveals the immunity module to be 'domesticated', targeting canonical bacterial parasites.

In addition to the Type I-C system encoded by strain PaLML1, it also encodes a Type I-F CRISPR–Cas system. The Type I-C spacers cluster with lineage 3, sharing all but one spacer with two of the published CRISPR arrays. We confirmed that the PaLML1 pKLC102 island is also capable of excision, much like the well-studied PAPI-1 of strain PA14 [21](which lacks a Type I-C CRISPR system), using PCR to amplify the excision junction that forms if the island excises from the chromosome (Fig. 1.9A). To verify CRISPR–Cas function, we transformed PaLML1 with a plasmid encoding a Type I-C crRNA targeting phage DMS3m. Because Type I-C spacer length ranges from 32–37 nt, contrary to Type I-F spacers consistently measuring 32 nt (Fig. 1.1E), we tested spacers of each length (i.e. 32 nt, 33 nt, etc.) in PaLML1 to determine their efficacy. Phage targeting occurred in the presence of the phage-specific crRNAs for the I-F system and all crRNA-lengths for the I-C system (Fig. 1.9B). Phages that escaped Type I-C targeting were also isolated and Sanger

sequencing of the protospacer indicated that these had point mutations at positions +2 and +3 (counting from the PAM) (Fig. 1.1F) and Fig. 1.9C). This suggests that these nucleotides are part of the seed sequence, a region where mutations are not tolerated for accurate base-pairing with the crRNA. In conclusion, active Type I-C systems in *P. aeruginosa* are on a widespread mobile element, have variable CRISPR spacers suggesting activity in situ, and can provide protection against phage.

1.3.2 Discovery of anti-CRISPRs on MGEs that inhibit Type I-C & beyond

Given the diversity of *P. aeruginosa* Type I-C spacers that target MGEs and the robust phage targeting we observed, counter-immunity mechanisms are expected to have manifested due to the threat posed by the Type I-C system. Only one Type I-C anti-CRISPR, AcrIC1, has been previously reported, and it is not found in *P. aeruginosa* [27]. To identify additional candidate anti-CRISPR genes, we used previously established bioinformatics approaches: self-targeting (ST) and guilt-by-association [36]. Because bacterial genome cleavage is a deadly event [37], a sequenced strain with a CRISPR–Cas system that has a spacer targeting its own chromosome is indicative of some CRISPR inactivation mechanism. Additionally, acr genes are often coupled with negative transcriptional regulators known as anti-CRISPR associated (*aca*) genes [38], which can be used to locate candidate acr genes [7, 27, 36] . To test candidate Acrs, we used lab strain PAO1 expressing PaLML1's Cas3–5–8–7 and a phage DMS3m-targeting crRNA from the chromosome, ($PAO1^{IC}$), due to PaLML1's low transformation efficiency.

Strain AZPAE14708 encodes a spacer targeting its own type VI secretion gene, *tagQ*, with a perfect protospacer and PAM match (Fig. 1.2A and Fig. 1.9D). This spacer is absent in other strains within lineage 1 that share spacer content with AZPAE14708 ((Fig. 1.1C)). To identify candidate acr genes, we used acr-associated gene 1 (*aca1*) as an anchor, and found a locus with only the Type I-F inhibitor *acrIF2* (Fig. 1.2A). Surprisingly, expression of AcrIF2 from a phage during infection completely inhibited the Type I-C system (Fig. 1.2B). The dual inhibitory activity was surprising given the evolutionary distance between the I-F and I-C systems [11] (no significant pairwise identity, Fig. 1.9E). Two additional AcrIF2 (hereafter, AcrIF2* to indicate dual specificity) homologues from *Pseudoxanthomonas* and *Stenotrophonomonas*, both associated with *aca1*, with ˜50% sequence identity, were tested and both displayed dual I-C and I-F activity (Fig. 1.9F). Strains

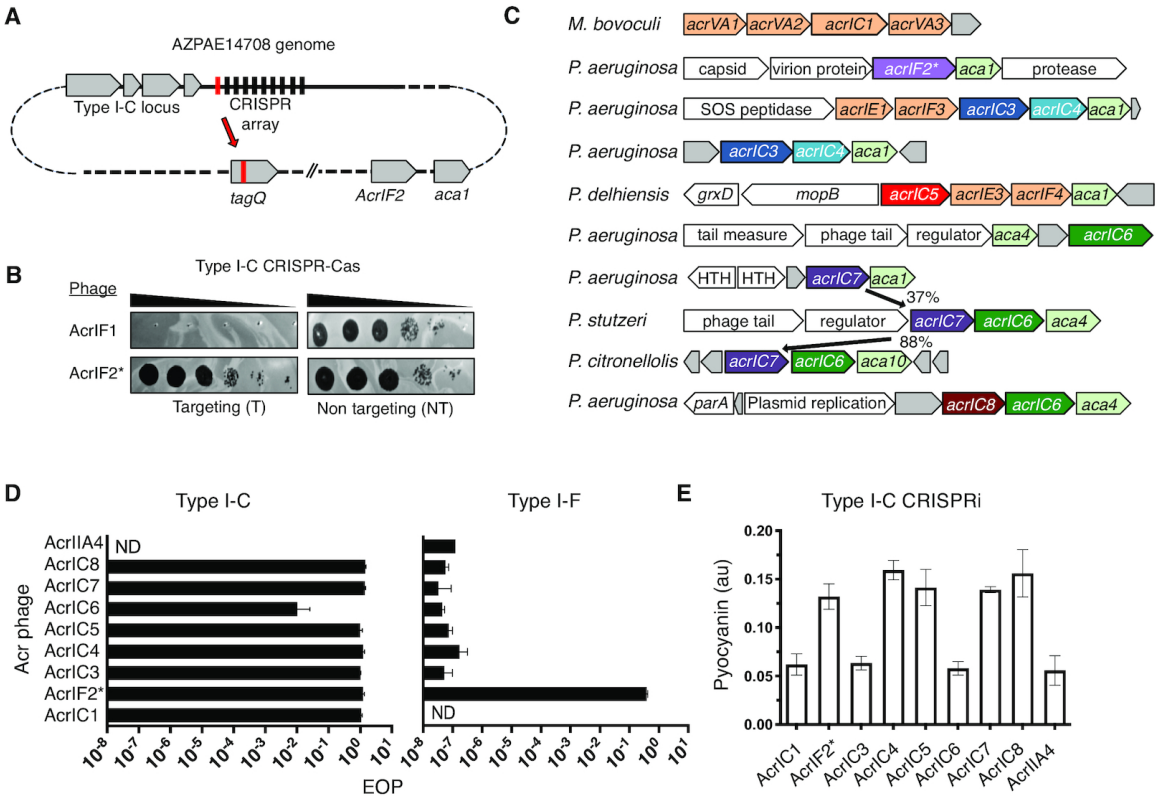


Fig 1.2. Discovery of Type I-C anti-CRISPR proteins. (A) Schematic of the self-targeting *P. aeruginosa* strain AZPAE14708 showing the first spacer (in red) targeting *tagQ* and the *aca1* locus encoding *acrIF2**. (B) A strain expressing the Type I-C CRISPR system in PAO1^{IC} was challenged by phage encoding either AcrIF1 or AcrIF2 in a spot titration plaque assay with ten-fold serial dilutions. (C) Gene neighborhood maps of MGEs where new Type I-C acrs (colored, bolded arrows) were identified. Previously discovered Acrs (orange), annotated MGE genes (white), and hypothetical genes (gray), are shown. (D) Efficiency of plaquing (EOP) calculations for an isogenic panel of phages expressing *acrIC* genes tested in PAO1^{IC} or PA14 (Type I-F). Each strain was infected in triplicate and plaque counts were averaged and normalized against a strain lacking the indicated CRISPR–Cas system. ND, none detected (E) Transcriptional repression via the Type I-C CRISPR system (CRISPRi, strain: PAO1$^{IC}\Delta$*cas3*) and the impact of the *acrIC* genes. Levels of the pigment pyocyanin are quantified at high levels when CRISPRi is inhibited and low levels when CRISPRi is functional. Each measurement is an average of biological triplicate. A prophage encoding AcrIIA4 was used to lysogenize this same strain as a negative control.

from these genera also encode Type I-C and Type I-F systems.

*acrIF2** is very narrowly distributed and thus we reasoned that more Type I-C Acrs likely exist. Using *aca1* and *aca4* as marker genes, 27 aca-associated candidates were tested (Fig. 1.7), revealing six more Type I-C inhibitors in a series of distinct MGEs including plasmids, transposons, conjugative elements, and phages (Fig. 1.2C and Fig. 1.7). Many of the MGEs frequently targeted by pKLC102-encoded Type I-C spacers harbor one or more of these seven new inhibitors (Fig. 1.10A). An additional gene was identified that solely inhibited the *P. aeruginosa* Type I-E system, *acrIE9* (discussed below). This collection consisted of genes associated with *aca1* (*acrIC3, acrIC4 and acrIC5*) or *aca4* (*acrIC6, acrIC7 and acrIC8*). Each of the new *acr* genes were identified in *P. aeruginosa*, except *acrIC7*, which was first identified in *P. stutzeri* ($acrIC7_{\mathrm{Pst}}$) adjacent to *aca4* (Fig. 1.10B). A homologue was found in *P. citronellolis* ($acrIC7_{\mathrm{Pci}}$, 88% amino acid sequence identity), adjacent to a new helix-turn-helix transcriptional regulator, which we have named *aca10*. In both instances, *acrIC6* is also present in the locus. An *aca1*-adjacent distant *acrIC7* homologue was also found in *P. aeruginosa* (37% sequence identity), although it did not confer Type I-C anti-CRISPR activity (Fig. 1.10B and C).

A panel of isogenic DMS3m phages was engineered to express each individual *acr* gene, including a negative control (Cas9 anti-CRISPR, AcrIIA4). Efficiency of plaquing (EOP) was assessed during infection of $\mathrm{PAO1^{IC}}$ (Fig. 1.2D). Each phage had an EOP = 1 when infecting cells expressing the Type I-C system, except AcrIC6, which appeared to be quite weak (EOP = 0.01). Only AcrIF2 had activity against the Type I-F system, with an EOP = 1, compared to EOP = 10-7 for all other Acr proteins.

To determine how the new Acrs inhibit the Cas machinery, we tested whether they alleviate CRISPR transcriptional interference (CRISPRi) in a Δ*cas3* background, a readout for inhibited DNA-binding by Cascade. A colorimetric assay was adapted from previous work [6], using a Type I-C crRNA to repress transcription of the *phzM* gene. If CRISPRi is functional, the surveillance complex blocks *phzM* transcription, turning the *P. aeruginosa* culture yellow. If DNA-binding is inhibited, the culture is a natural blue-green (Fig. 1.10D). Five of the proteins, AcrIF2*, IC4, IC5, $\mathrm{IC7_{Pst}}$ and IC8, blocked CRISPRi. Expression of AcrIC1 (a previously discovered protein from *Moraxella*) and AcrIC3, however, did not interfere with CRISPRi, suggesting that they bind to Cas3

or Cascade in a way that prevents Cas3 recruitment or DNA cleavage, while allowing Cascade-DNA binding (Fig. 1.2E). AcrIC6 did not block CRISPRi but given its weak activity, we are hesitant to interpret this negative result.

1.3.3 Multi-system inactivation by AcrIF2*

Most AcrIC proteins identified here are acidic proteins that block DNA-binding, thus we focused on the well-studied AcrIF2* as a model protein for ecological and mechanistic experiments. AcrIF2* directly prevents the Type I-F CRISPR surveillance complex from binding to DNA [22–24]. While many MGEs encode distinct inhibitors of Type I-C, I-E and I-F systems, AcrIF2* can be found encoded alone and thus we wondered whether a phage expressing this protein can inhibit both systems simultaneously. It has previously been reported that a phage concentration threshold, inversely proportional to Acr strength, is needed to inhibit CRISPR–Cas targeting [28, 39]. To measure the strength of this bi-functional Acr protein, we infected cells expressing the Type I-C system and a targeting spacer ($PAO1^{IC}$) from the chromosome, plus a variable "decoy" complex with a non-targeting spacer from a plasmid, with our Acr phages (Fig. 1.3A). The AcrIF2*-expressing phage causes collapse of the culture at low infectious doses (Fig. 1.3B), similar to AcrIC1, an anti-CRISPR that only inhibits Type I-C (Fig. 2.11D–F). However, when the Type I-F system with a non-targeting crRNA was co-expressed in this strain, the AcrIF2* encoding phage required an initial MOI of 2×10^{-2} to lyse the culture, compared to 2×10^{-5} with empty vector (Fig. 1.3C), while the AcrIC1 phage was unaffected (Fig. 2.11D). Interestingly, when cells expressed the Type I-F system with a Cas8 mutation (K247A) which was previously shown to reduce AcrIF2* binding [22], an intermediate AcrIF2* phage concentration was sufficient for culture collapse (Fig. 1.3D). Phage output from these experiments was also quantified. In the presence of the decoy Type I-F system, phage output at the two lowest MOIs was reduced by ˜100–1000-fold compared to decoy Type I-F $Cas8^{K247A}$, Type I-C, or the non-targeting control (Fig. 1.13A–D). As a control, over-expression of more Type I-C complexes had little impact on concentration thresholds (Fig. 1.3E and Fig. 1.13C), and a phage encoding AcrIIA4, a Cas9 inhibitor, was ineffective at causing culture lysis, except at the highest MOI under all conditions tested (Fig. 1.12G–I). Therefore, it appears that during phage infection there is a cost to being a dual inhibitor, where the ability of AcrIF2* to inhibit the Type

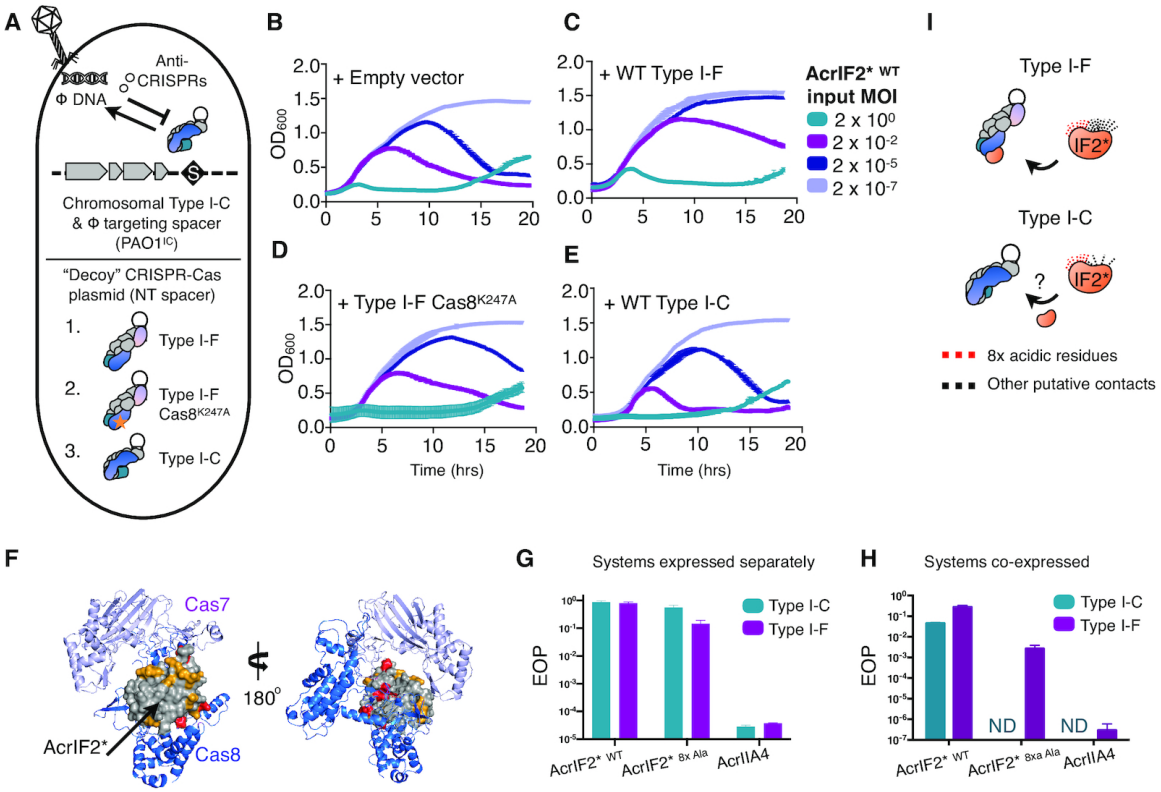


Fig 1.3. Fitness costs to encoding a dual anti-CRISPR protein (A) Schematic of the experiment performed in panels 3b-3e. The PAO1^{IC} strain, with the I-C system and phage-specific crRNA integrated into the chromosome was transformed with plasmids encoding 'decoy' CRISPR–Cas surveillance complexes, which contain a non- DMS3m targeting crRNA. A Type I-F Cas8^{K247A} mutant, which loses affinity for the AcrIF2 inhibitor was included as a control. (B–E) Liquid infection assay with PAO1^{IC} transformed with indicated 'decoy' surveillance complex plasmids and infected with a virulent DMS3m phage expressing AcrIF2* WT. Strains were grown in a plate reader over 20 h with OD600 tracked over time. (F) Color-coded structure of AcrIF2* bound to the Type I-F surveillance complex (PDB: 5UZ9). The Type I-F surveillance complex is shown as a ribbon with Cas8 (blue), and one Cas7 monomer (lilac), AcrIF2* (grey space-filling model), and mutated amino acids (red) and remaining acidic residues (yellow) shown. (G) Quantification of the efficiency of plaquing (EOP) on PAO1^{IC} or PA14 for phages expressing the indicated *acr* gene. (H) Quantification of the efficiency of plaquing (EOP) on PaLML1 for phages expressing the indicated *acr* gene. (I) Schematic representation of AcrIF2 binding to the Type I-F or Type I-C surveillance complex. Red lines are representative of the eight mutated acidic residues. Black lines represent other non-covalent contacts. Because AcrIF2* has stronger activity against the Type I-F system, more black lines are used to represent its additional contacts that mediate activity against Type I-F versus Type I-C.

I-C system is weakened by the presence of the Type I-F system, while monotypic AcrIC1 retained its potency.AcrIF2* inhibits the Type I-F complex by interacting with key PAM-binding residues on Cas7 and Cas8, as revealed by previous cryo-EM and crystallography studies (Fig. 1.3F). We therefore next sought to determine whether AcrIF2* acidic residues in this interface are required for inhibition or impact the observed competition defect. Of AcrIF2*'s 96 residues, 24% are acidic, giving it an overall negative charge (pI = 4.0). Despite Cas proteins from Type I-C and I-F having completely distinct sequences (Fig. 1.9E), this negative surface charge could allow AcrIF2* to block both the I-C and I-F DNA recognition motifs. We selected eight AcrIF2* residues (D30, E36, D76, E77, E82, E85, E91, E94) that sit within ˜5 Åof a basic residue on Type I-F Cas7/Cas8 (Fig. 1.3F and Fig. 1.11A) and incrementally mutagenized them. All of the plasmid-expressed mutants, up to an 8x Ala mutant (*acrIF2*^{*8xAla}) surprisingly maintained Acr activity against the Type I-F and I-C systems, while more dramatic mutations (e.g. the eight selected residues mutated to lysine or the 8 residues mutated to glutamine or asparagine) lost all inhibitory function (Fig. 1.11B). When the 8x Ala mutant was expressed from the endogenous phage *acr* locus, it also retained function against the Type I-C and Type I-F systems individually (Fig. 1.3G), which was surprising given the presumed reliance on these negatively charged residues. However the 8x Ala mutant phage was inactivate against the Type I-C system (EOP ¡ 10^{-4}) when infecting a strain expressing both Type I-F and I-C (PaLML1), again consistent with a dual inhibition cost (Fig. 1.3H). Activity against the I-F system was only partially weakened (EOP = 0.02, Fig. 1.3G, H, and Fig. 1.11C), however. The more sensitive phage concentration threshold infection assay that revealed the competition cost for WT AcrIF2* also indicated that the mutant had a very weak anti-CRISPR phenotype, only providing protection to the phage at the highest MOI when faced with the targeting I-C and decoy I-F complexes (Fig. 1.12B and C). Output PFUs were again assessed, revealing that mutant AcrIF2* phage output was reduced ˜100 000-fold when competing with WT Type I-F or Type I-F Cas8^{K247A} decoy systems, but less so with excess Type I-C, compared to the NT control (Fig. 1.13E–H). Interestingly, this reduction in mutant AcrIF2* output phage was the same for WT and Type I-F Cas8^{K247A} decoy systems (Compare Fig. 1.13E and F), suggesting that mutant AcrIF2* retains similar binding affinity for WT Cas8 versus Cas8^{K247A}. Overall, AcrIF2* strength is completely context-dependent, exhibiting a conditional defect in the presence of two competing surveillance

complex binding targets. This defect, however, is minimized by the excessive negativity of the protein, as many of these residues are not required for function, per se, but help to buffer the defect generated by *in vivo* competition. The weakened activity of the mutated AcrIF2* protein against the Type I-C system is consistent with non-identical, but perhaps overlapping, AcrIF2* binding interfaces. We posit that WT AcrIF2* may make more contacts with the Type I-F surveillance complex vs. the Type I-C surveillance complex, stabilizing its interaction with the former (Fig. 1.3I).

1.3.4 Broad-spectrum inhibitory activity by the I-C anti-CRISPRs

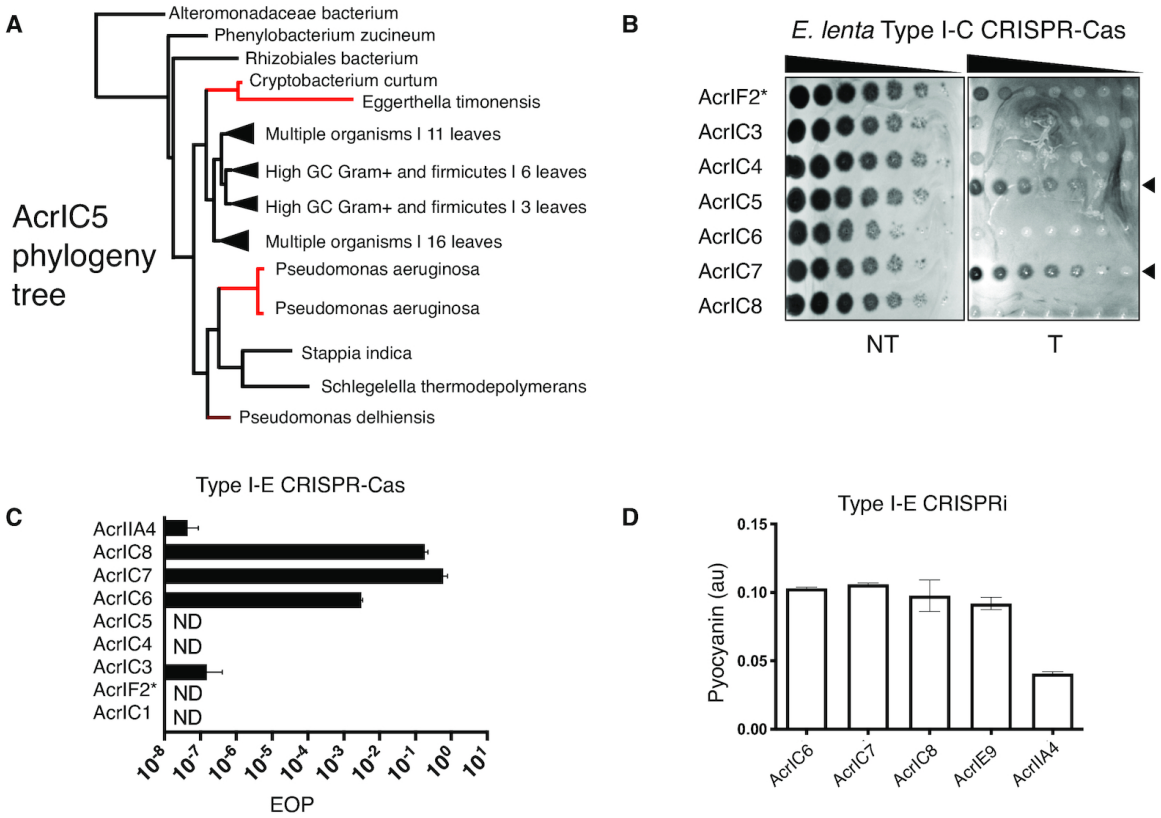


Fig 1.4. Multi-system inhibition by a subset of the novel inhibitors. (A) Phylogenetic tree of AcrIC5 protein showing its broad distribution. Lines highlighted in red denote strains of relevance in this work. Tree was generated using BLAST pairwise alignments (B) Plaque assay of acr-encoding engineered JBD30 phages tested against the *E. lenta* Type I-C system expressed heterologously in *P. aeruginosa*. Phage was serially diluted 10× for each spot. (C) EOP calculations for an isogenic panel of phages encoding the indicated acr gene, infecting a strain expressing the Type I-E CRISPR–Cas system (PA4386). Each bar is the average of infections done in biological triplicate normalized to the number of plaques on PA4386 Δ*cas3*. (D) Type I-E CRISPRi, conducted as in Figure 3 (host: PA4386 Δ*cas3*) with the Acr proteins that inhibit Type I-E function assayed. AcrIIA4 is a negative control.

We next surveyed the phylogenetic distribution of the new acr genes reported here. Most of the Acr proteins were limited to a single genus: AcrIC1 to *Moraxella*, and AcrIF2*, AcrIC3, AcrIC4 and AcrIC7 were only found in *Pseudomonas*. AcrIC5 orthologues were found distributed across Proteobacteria, Firmicutes, and Actinobacteria (Fig. 1.4A), and AcrIC8 orthologues were found sparingly in *Pseudomonas*, *Spirochetes*, and *Rhizobiales*. AcrIC6 can be found broadly in various classes (Alpha-, Beta- and Gamma-proteobacteria) with many homologues in *Salmonella enterica*. We took note of Actinobacterial AcrIC5 homologues in the human-associated microbes *Cryptobacterium curtum* and *Eggerthella timonensis*, given that an active *Eggerthella lenta* Type I-C CRISPR–Cas system was described recently [13]. We tested whether a phage encoding the *Pseudomonas* AcrIC5 homologue could inhibit the *E. lenta* I-C system heterologously expressed in *P. aeruginosa* and observed strong anti-CRISPR function (Fig. 1.4B), despite cas gene sequence identities between 35 and 55% (Fig. 1.14A). Surprisingly, AcrIC7 also inhibited the *E. lenta* I-C system, despite no identified homologues outside of the *Pseudomonas* genus.

The broad-spectrum activity of AcrIF2* (I-F and I-C), AcrIC5 (I-C_{Pae} and I-C_{Ele}), and AcrIC7 (I-C_{Pae} and I-C_{Ele}), motivated us to test the new Acr proteins against another system found in *P. aeruginosa*, Type I-E. Type I-C, Type I-F and Type I-E systems are phylogenetically distinct subtypes, with I-F and I-E systems sharing a more recent common ancestor. AcrIC7*$_{Pst}$, AcrIC7*$_{Pci}$, AcrIC7*$_{Pae}$ and AcrIC8*, inhibited the Type I-E system well, while AcrIC6* was again, a weak anti-CRISPR (Fig. 1.4C, AcrIC8 locus map in Fig. 1.14B). The new Type I-E Acr proteins (AcrIC6*, AcrIC7*$_{Pst}$, AcrIC8* and AcrIE9) all inhibited Type I-E CRISPRi (Fig. 1.4D), indicating that they block DNA binding. Curiously, AcrIC7$_{Pae}$ only inhibited the I-E subtype, unlike its dual I-C/I-E inhibiting homologues (Fig. 1.10C and Fig. 1.14C). Searching through sequenced genomes revealed that *P. stutzeri* and *P. aeruginosa* encode both I-C and I-E subtypes, while *P. citronellolis* encodes only Type I-F systems.

1.3.5 Anti-CRISPRs that inhibit DNA cleavage by Cas3

Acr proteins that allow for DNA binding but still block phage DNA cleavage, like AcrIC1 and AcrIC3 (Fig. 1.2E), effectively turn the endogenous CRISPR–Cas machinery into a catalytically dead, transcriptional repression system. acrIC3 can be frequently found flanked by *acrIE1* and

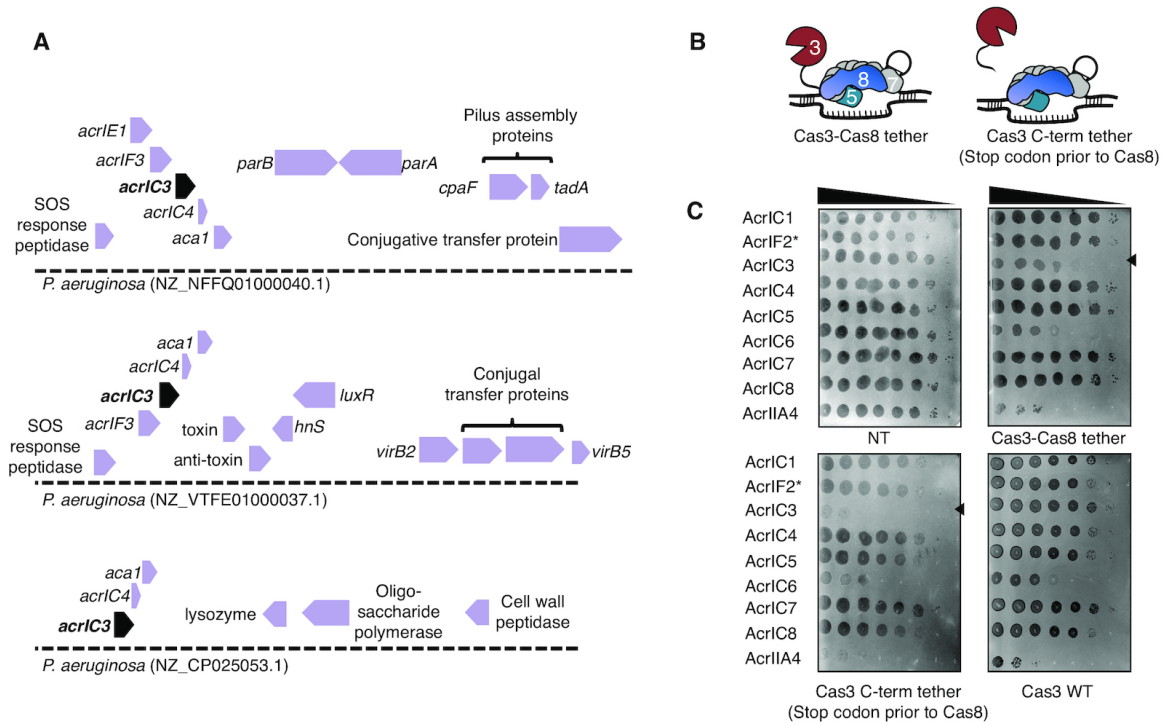


Fig 1.5. *acrIC3* is found on conjugative and plasmid elements. (A) Gene loci showing *acrIC3*. *acrIC3* is found on various MGEs, and is often associated with AcrIE1 and AcrIF3, which are Cas3 interacting proteins. (B) Schematic of the Type I-C mutant where the C-terminus of Cas3 is tethered to the N-terminus of Cas8 with a short linker peptide ('Cas3-Cas8 tether') and a related construct where a stop codon is added after the linker peptide ('Cas3 C-term tether'). (C) Spot titration plaque assay with ten-fold serial dilutions of phage showing the plaquing efficiency of Acr-expressing DMS3m phages on non-targeting (NT), or Type I-C expressing strains, either with Cas3–Cas8 tethered, Cas3 C-term tether, or Cas3 WT.

acrIF3 in *P. aeruginos*a, two Cas3 inhibitors that enable CRISPRi [8,24]. This reveals a remarkable 'anti-Cas3 locus' for all three Type I CRISPR systems in *P. aeruginosa* (Fig. 1.5A). Conjugative transfer, *parA/B* genes, and type IV secretion system genes are found flanking these acr genes. The role that an 'anti-Cas3 island' may play in conjugative transfer from cell to cell is yet to be determined, but this phenomenon may indicate that neutralizing the ssDNAse Cas3 is an effective means to ensure successful transfer, which proceeds through a ssDNA intermediate. When not found with other CRISPRi-enabling inhibitors, *acrIC3* is carried by phages, along with *acrIC4*, which is always paired with *acrIC3*. *acrIC1* is found on *Moraxella* phages and prophages, flanked by*acrVA1, acrVA2 and acrVA3*, genes encoding Cas12 inhibitor proteins.

In an effort to distinguish the inhibitory mechanisms for AcrIC1 and AcrIC3, we constructed a Type I-C complex where the Cas3 C-terminus was tethered to the Cas8 N-terminus with a 13 amino acid sequence (RSTNRAKGLEAVS) (Fig. 1.5B). This construct was inspired by, and designed to mimic, naturally occurring variants of Type I-E systems in *Streptomyces griseus*, which encode Cas3 and Cas8 as a single polypeptide, with the same short linker peptide in between [40]. A control strain with a stop codon immediately following the C-terminal Cas3 tag was also constructed (Fig. 1.5B). A similar fusion of Cas3 to Cas5, which is seen in some Type I-C systems, was inactive when tested (not shown). When the panel of Type I-C Acr-expressing phages infected a strain expressing this minimal system, the fusion efficiently evaded the AcrIC3 protein, targeting this phage by ˜1,000-fold, while all other acr phages, with the exception of AcrIC6, replicated well (Fig. 1.5C). A version of Cas3 just possessing the linker on its C-terminus surprisingly also blocked the activity of AcrIC3, suggesting that AcrIC3 directly interacts with the C-terminus of Cas3, but the linker residues block this interaction. This demonstrates that AcrIC1 and AcrIC3 utilize distinct mechanisms to inhibit the Type I-C system downstream of DNA-binding.

1.4 Discussion

Organisms encoding CRISPR–Cas immune systems are locked in battle with genetic parasites that encode anti-CRISPR proteins capable of disabling CRISPR–Cas activity [4]. However, the Type I-C system in *P. aeruginosa* is also found on a common MGE (pKLC102) that can exist as either an

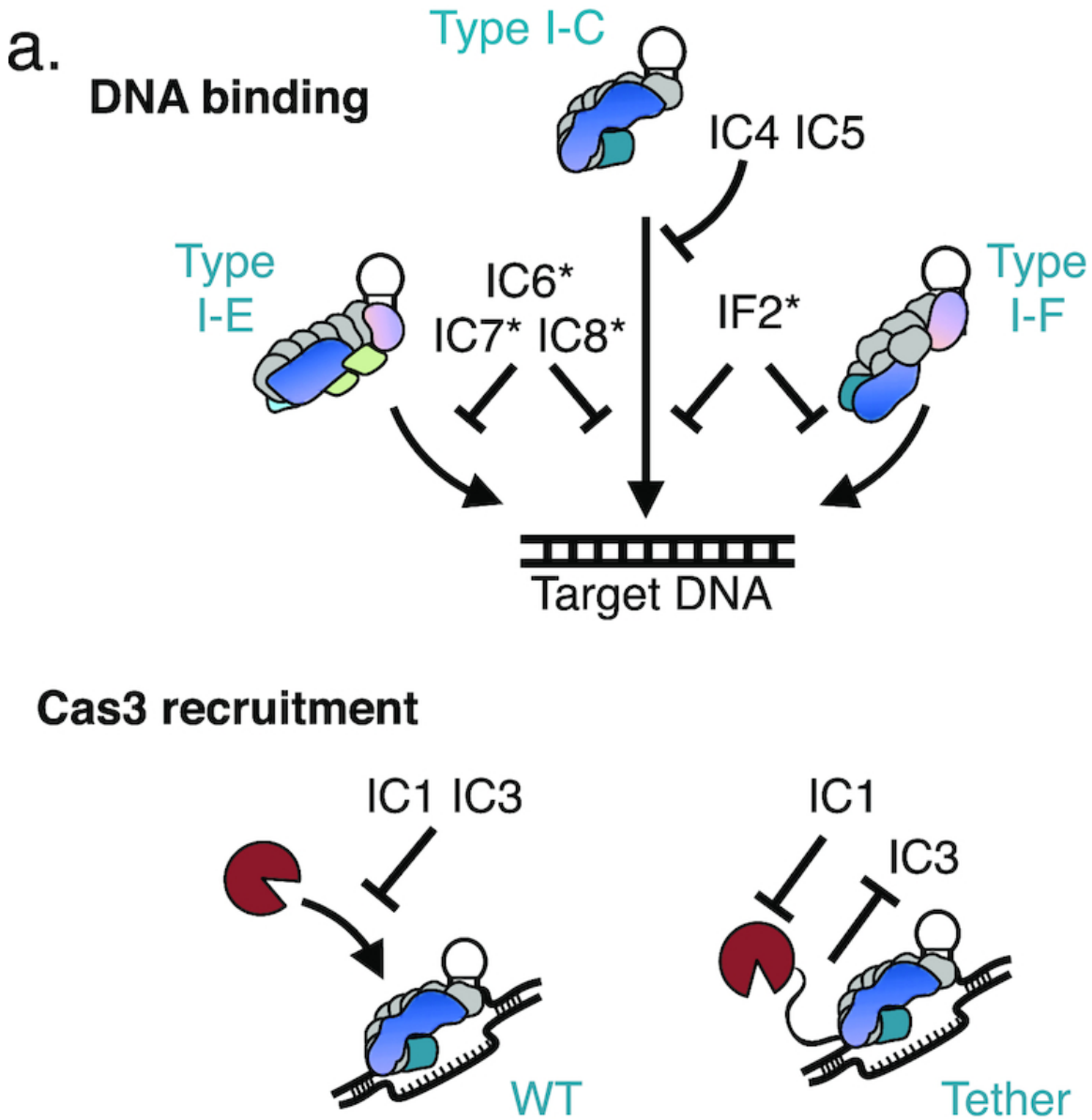


Fig 1.6. The novel I-C inhibitors act through a variety of mechanisms (A) Schematic representation of the three Type I subtypes investigated and the novel cognate anti-CRISPRs that inhibit them.

island or as a plasmid [**?**, 32]. Since mobile elements (here, encoding CRISPR–Cas or anti-CRISPRs) can transfer antibiotic resistance genes, virulence factors, immune systems, and other fitness-altering genetic material to their host [41, 42], this generates an interesting paradigm for CRISPR and anti-CRISPR interactions [43]. Mobile CRISPR–Cas systems can deliver immunity horizontally, transferring not only cas genes, but also a library of spacers against other MGEs. As CRISPR–Cas systems have been identified on plasmids [44] and phages [2, 3], this phenomenon could be highly prevalent.

The role of Acr proteins in the dissemination and maintenance of MGEs in bacterial genomes is just beginning to be explored [45]. The Acrs described in this study were found encoded by diverse MGEs that are frequent targets of the *P. aeruginosa* Type I-C spacer repertoire (Fig. 1.10A). AcrIC1, AcrIF2*, AcrIC5, AcrIC6* and AcrIC7* are commonly found within phages, while AcrIC6* and AcrIC8* are associated with Tn3 family transposases (Fig. 1.14E and F). Acr proteins facilitate the maintenance of prophages in genomes encoding spacers against that phage, which can help maintain CRISPR–Cas by preventing self-targeting, and even weak Acr proteins can overcome kinetic limitations by working cooperatively [28, 46]. Additionally, multi-system inhibition may be commonly exploited by MGEs, since bacteria are not limited to only one CRISPR–Cas subtype (See summary of our data in Fig. 1.6A). Such a tactic conserves genetic real estate, and acts as insurance against the threat of assorted immune systems, but may have a negative impact on fitness, as we demonstrated with AcrIF2*.

Of the eight Type I-C anti-CRISPR proteins, all but one (AcrIC8*) had high acidic amino acid content. This has been a common theme among Acr proteins and inhibitors of other immune systems [47]. Excess acidic residues could help stabilize binding to diverse Cas proteins, provide essential residues for inhibiting more than one system, and buffer against in vivo competition with multiple binding partners. For example, the T7 phage encoded Ocr protein is highly acidic and forms a dimer with a bend similar to B-DNA [48, 49]. Ocr was initially discovered as a Type I restriction enzyme system inhibitor and was more recently shown to inhibit the anti-phage system, BREX [50]. Importantly, systematic mutation of Ocr's acidic residues revealed it to be highly recalcitrant to breakage, similar to AcrIF2*, maintaining inhibitory activity against Type I R-M even with up to 33% of acidic residues mutated [48]. Similarly, Cas9 inhibitors AcrIIA2 and AcrIIA4

are highly acidic, can inhibit diverged Cas9 orthologues [36], and have been subjected to extensive mutagenesis, also appearing to have dispensable acidic residues [21]. AcrIF2* can also be considered a DNA mimic or a DNA competitor, with structural work showing that it partially overlaps with the PAM binding site [22, 23, 51], and our mutagenesis demonstrating that it is also quite resilient. This suggests that DNA mimicry is a potent and flexible anti-immune strategy.

Our work here underscores the importance of studying CRISPR–Cas and Acr mechanisms in vivo, revealing multiple new insights, including broadly inactivating anti-CRISPR proteins encoded by various MGEs and the flexibility of DNA mimicry, a common anti-CRISPR and anti-immune strategy. We propose that these DNA mimics are excessively negative to broaden their inactivation potential and buffer against competition and co-evolution in the DNA-binding pocket for CRISPR–Cas systems. Together with the spacer diversity uncovered, functional phage interference demonstrated, and the discovery of numerous diverse anti-CRISPR proteins encoded by *P. aeruginosa* mobile genetic elements, we conclude that the mobile Type I-C CRISPR–Cas system in *P. aeruginosa* is functional in nature. These observations further bolster our understanding of the importance of CRISPR–Cas to the biology of this species and generate a model organism for future Type I-C CRISPR–Cas work.

Candidate Number	Accession	Anti-CRISPR identity	*aca* association
1	KSR23770.1	AcrIC3	*aca1*
2	KSO29066.1	N/A	*aca1*
3	KSL61975.1	N/A	*aca1*
4	SDK41378.1	AcrIC5	*aca1*
5	CDO85538.1	AcrIC4	*aca1*
6	WP_085056855.1	N/A	*aca1*
7	WP_047296680.1	N/A	*aca1*
8	WP_092238848.1	N/A	*aca1*
9	WP_044274829.1	N/A	*aca1*
10	WP_071574229.1	N/A	*aca1*
11	WP_023657539.1	N/A	*aca1*
12	ABR13386.1	N/A	*aca4*
13	ABR13387.1	N/A	*aca4*
14	SDJ61905.1	N/A	*aca4*
15	OPE29935.1	N/A	*aca4*
16	OPD90261.1	N/A	*aca4*
17	WP_060613673.1	N/A	*aca4*
18	WP_080050315.1	AcrIC6*	*aca4*
19	EWC40192.1	AcrIC7*	*aca4*
20	GCA55691.1	N/A	*aca4*
21	WP_101192668.1	AcrIE9	*aca4*
22	WP_101192667.1	N/A	*aca4*
23	WP_101192666.1	N/A	*aca4*
24	WP_045884682.1	N/A	*aca4*
25	WP_045884679.1	N/A	*aca4*
26	WP_074202337.1	AcrIC8*	*aca4*
27	WP_074202338.1	N/A	*aca4*

Fig 1.7. Gene candidates involved in this study. Candidate genes tested over the course of this work. NCBI accession numbers provided.

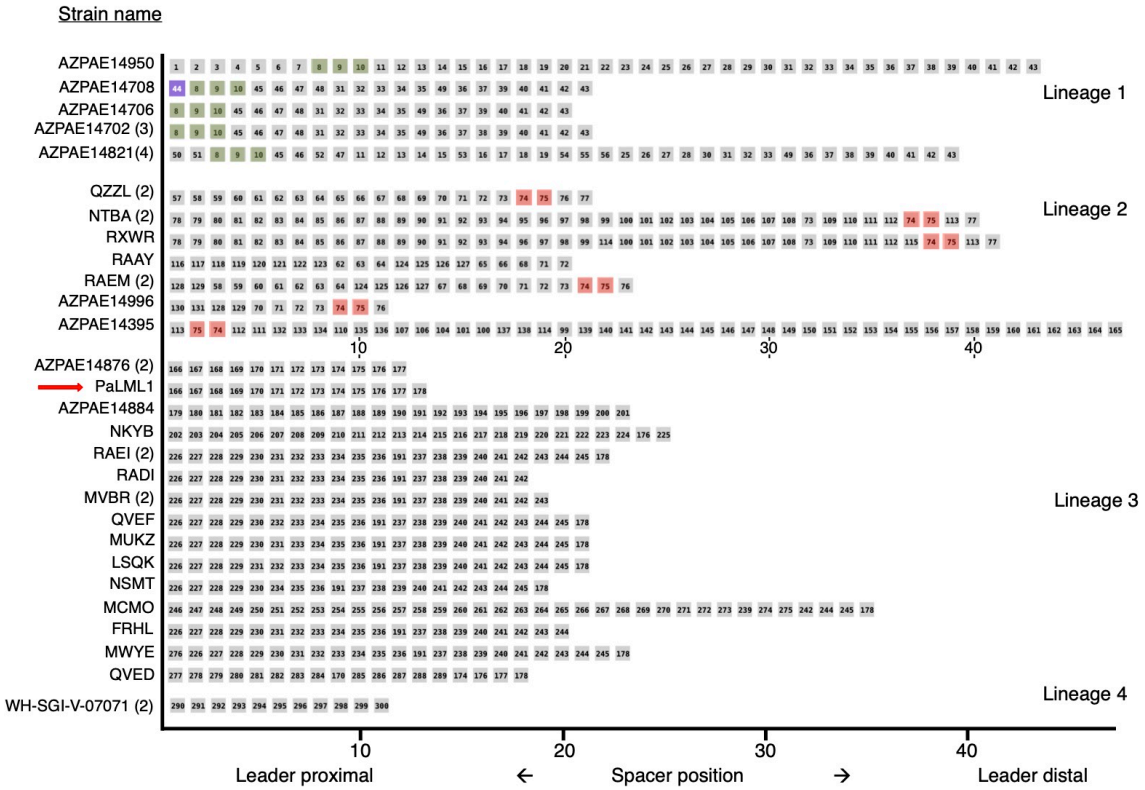


Fig 1.8. Lineage tracing of the Type I-C CRISPR arrays. (A) Full CRISPR array lineage mapping of the 28 unique CRISPR arrays from 42 genomes. Each lineage contains CRISPR arrays that share at least one spacer. Spacers with the same DNA sequence are given the same number. Spacer 44 is a self-targeting spacer and is colored in purple.

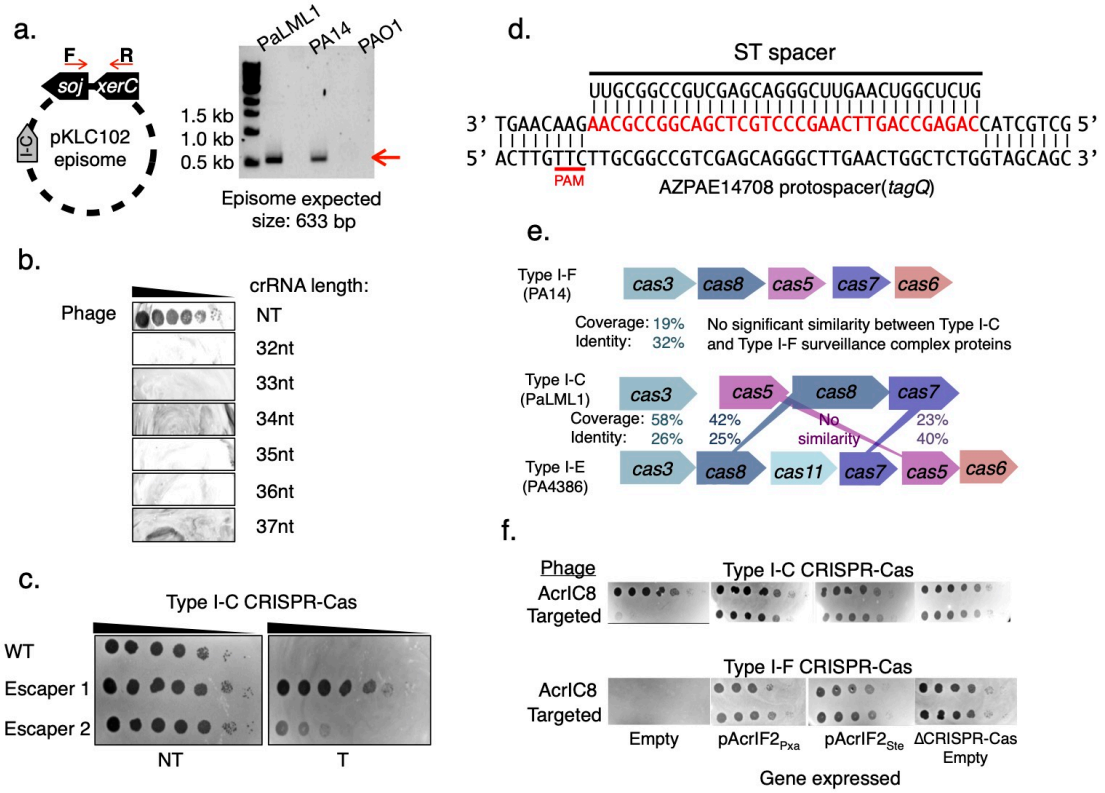


Fig 1.9. Analysis of the *P. aeruginosa* Type I-C system. (A) The pKLC102 island can excise to form an episome. Red arrows represent forward and reverse primer pair used to analyze island excision. Gel depicts PCR products using the forward and reverse primer pair from schematic demonstrating the episomal nature of the pKLC102 element in PaLML1 and PA14 isolates. PAO1 does not harbor pKLC102 elements, resulting in no PCR product. (B) Spot titration plaque assay of CRISPR-Cas sensitive phage serially diluted on a lawn of PaLML1, expressing crRNAs of lengths between 32-37 nt. The targeted phage is DMS3m, which does not have an *acrIC* gene. (C) Spot titration plaque assay of WT (i.e. non escaper) phage and escaper phages 1 and 2 challenged with the Type I-C system in PAO1^{IC}. T = Targeting, NT = Non-targeting. Phages were spotted in 10x serial dilutions. (D) Alignment of self-targeting spacer 1 from AZPAE14708 with corresponding protospacer. PAM is underlined in red. (E) Comparison of Type I-F and Type I-E Cas protein sequences to Type I-C Cas protein sequences for the systems used in this study. (F) Plaque assay testing the activities of two AcrIF2 homologues identified in *Pseudoxanthomonas* and *Stenotrophomonas* genomes. Homologues were expressed from a plasmid in either a strain encoding the Type I-C system (PAO1^{IC}, induced with 1mM IPTG) or the Type I-F system (PA14). A phage encoding a Type I-C Acr (AcrIC8) was used as a positive control, and a phage encoding AcrIIA4 (a Cas9 inhibitor) was used as the targeted phage.

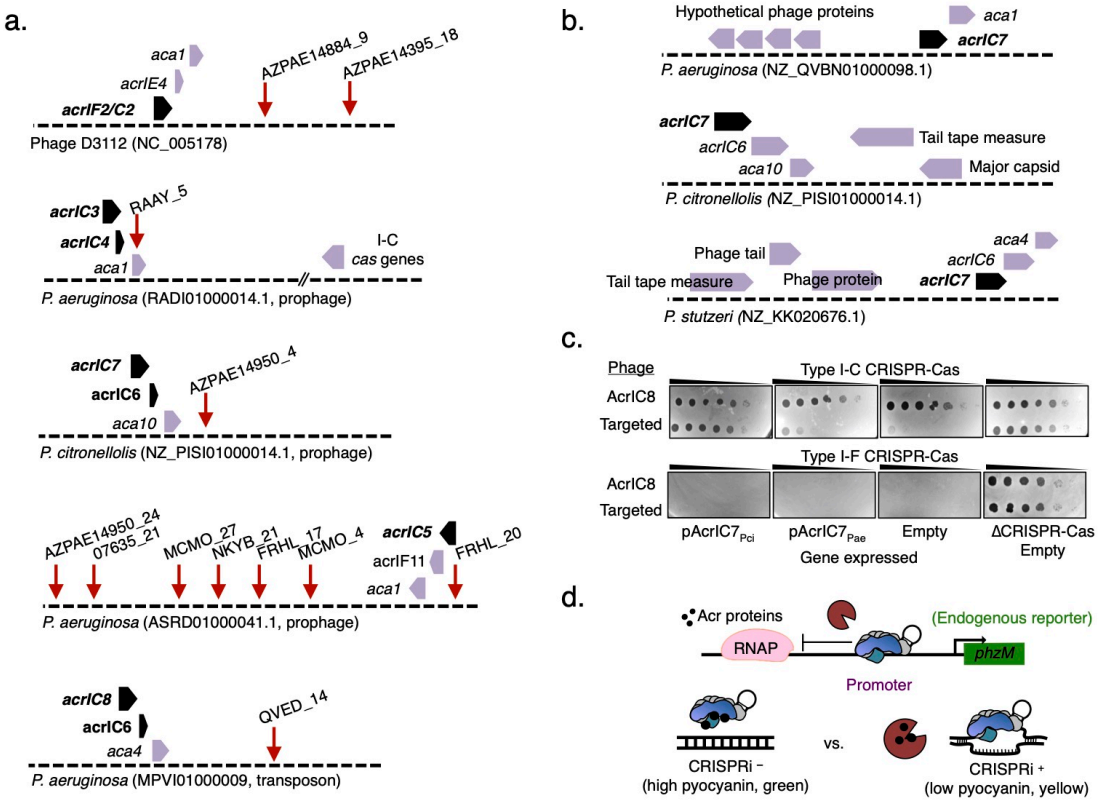


Fig 1.10. The Type I-C system targets acr-encoding elements. (A) Schematic representation of the MGEs targeted by spacers found encoded by the Type I-C CRISPR arrays. Each MGE is shown with the Type I-C acr it encodes (black arrow) and relevant upstream and downstream genes. Spacer matches are indicated by a red arrow and the strain name plus spacer number. Accession number for the MGE is found below the genome annotation. (B) Loci showing typical genetic context of *acrIC7* in three *Pseudomonas* species. Genome accession code in parentheses. (C) Plaque assays of two AcrIC7 homologues expressed from a plasmid in PAO1^{IC} or PA14. Acr activity was assessed by spotting a CRISPR-Cas sensitive phage in 10x serial dilutions (DMS3m expressing AcrIIA4) and an untargeted control (DMS3m expressing AcrIC8). (D) Schematic of the CRISPRi assay used to screen Acr activity. A crRNA is designed to bind upstream of *phzM*, a gene whose expression results in green pigmented *P. aeruginosa* cultures. Acrs that inhibit the surveillance complex from binding target DNA result in a CRISPRi- phenotype. Acrs that bind Cas3 or do not block DNA binding result in a CRISPRi+ phenotype.

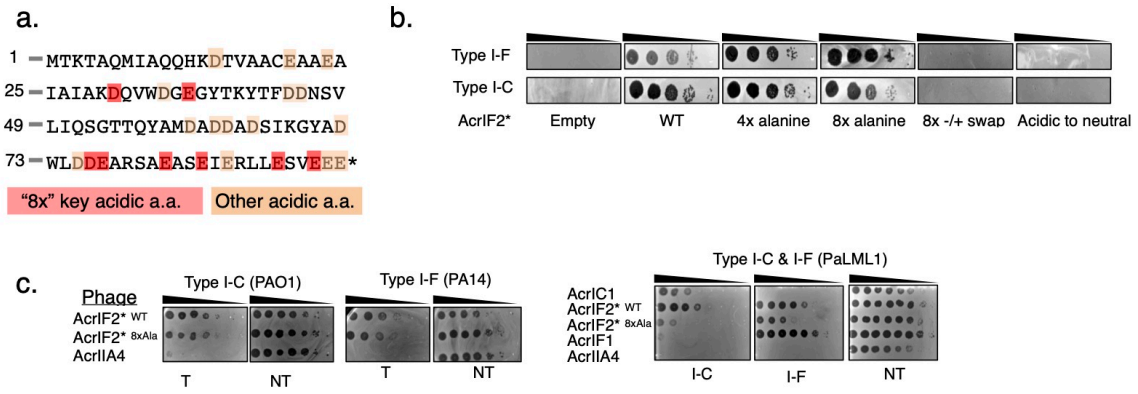


Fig 1.11. Analysis of AcrIF2*, a highly acidic anti-CRISPR. (A) AcrIF2* amino acid sequence shown with 8 key acidic residues (red) and all other acidic residues shaded (orange). These residues were mutated to either alanine, corresponding opposite charge residues, or corresponding neutral residues as indicated in the text. (B) Plaque assays testing the activity of AcrIF2* mutants expressed from a plasmid. A I-F strain (PA14) or IC strain ($PAO1^{IC}$) were transformed with plasmids encoding the mutants indicated under each panel. Ten-fold serial dilutions of a CRISPR-Cas sensitive phage (DMS3m-AcrIIA4) was used to determine the activity of the AcrIF2* mutants. 4x alanine is D76A, D77A, E91A, and E94A. 8x mutant includes all mutated residues from the 4x mutant plus D30A, E36A, E82A, AND E85A. (C) Plaque assays with ten-fold dilutions of the engineered mutant AcrIF2* phage tested in $PAO1^{IC}$ (Type I-C alone), PA14 (PA14 alone), or PaLML1 (both Type I-C and Type I-F co- expressed, with phage-specific crRNA provided on a plasmid and indicated below the appropriate panel). T= Targeting crRNA, NT= Non-targeting.

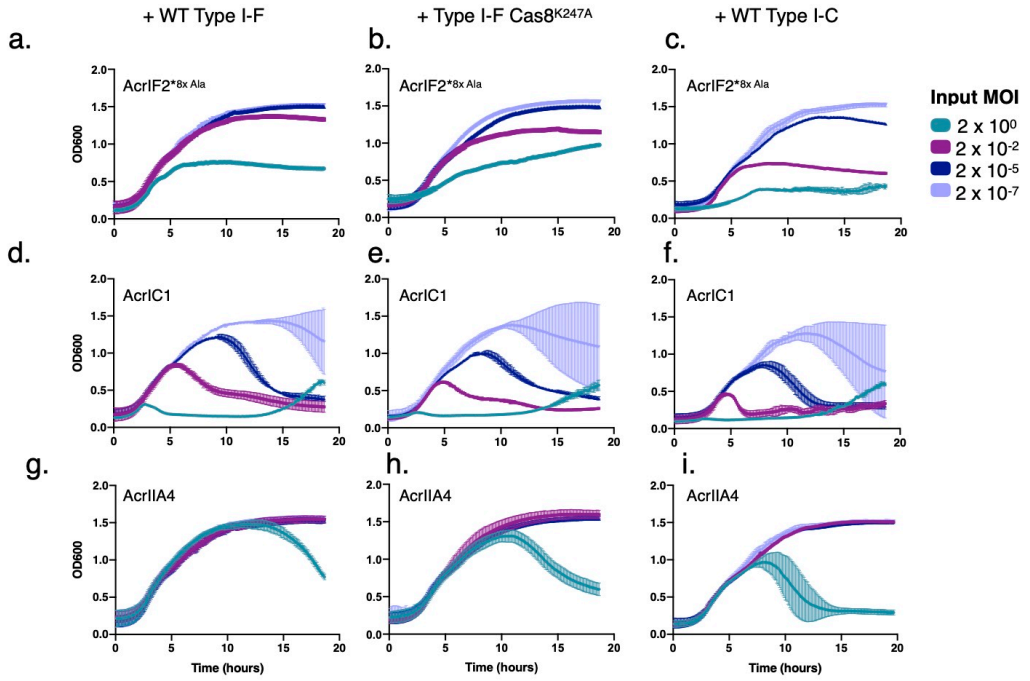


Fig 1.12. Fitness tradeoffs of dual inhibitors. (A)-(I) Liquid infection assay with PAO1^{IC} transformed with indicated "decoy" surveillance complex plasmids (WT Type I-F, WT Type I-C, or Type I-F Cas8^{K247A}) and infected with a virulent DMS3m phage expressing AcrIF2*, the 8x Ala mutant, AcrIC1, or AcrIIA4. Input phage MOI shown in legend.

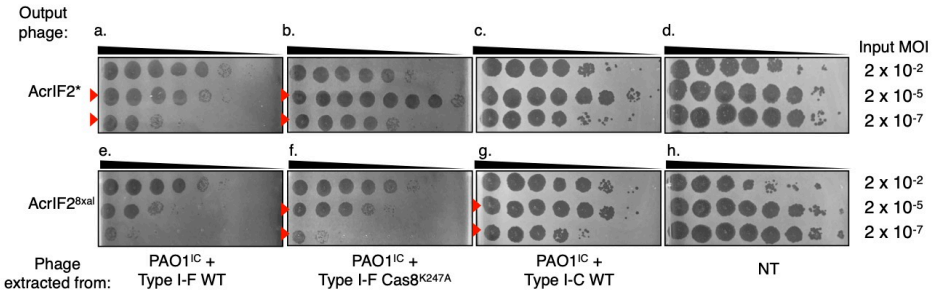


Fig 1.13. Infection thresholds vary for anti-CRISPR encoding phage. (A)-(H) Output phage following liquid infection assay from supplemental figure 5. Phage from cultures infected with virulent DMS3m expressing either AcrIF2*^{WT} or AcrIF2*$^{8x\ Ala}$was collected. 10x serial dilutions were spotted on an indicator strain (PAO1 WT). Input phage MOI shown in legend.

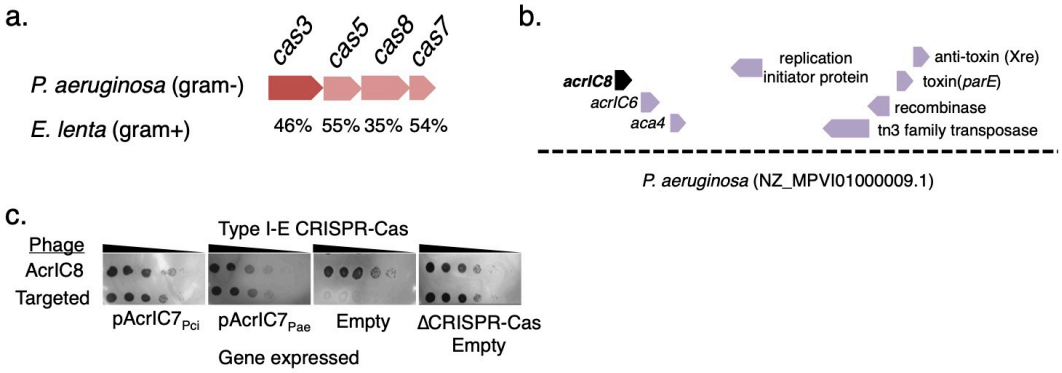


Fig 1.14. Type I-C inhibitors function against multiple systems. (A) Protein percent identity comparison of the *E. lenta* Type I-C CRISPR-Cas system to the *P. aeruginosa* Type I-C CRISPR-Cas system. (B) Loci showing typical genetic context of *acrIC8*. (C) Plaque assays of two AcrIC7 homologues expressed from a plasmid in PA4386. Acr activity was assessed by spotting a CRISPR-Cas sensitive phage in 10x serial dilutions (DMS3m expressing AcrIIA4) and an untargeted control (DMS3m expressing AcrIC8).

Chapter 2

A compact Cascade–Cas3 system for targeted genome engineering

2.1 Introduction

CRISPR–Cas systems are a diverse group of RNA-guided nucleases that defend prokaryotes against viral invaders [11, 52, 53]. Gene-editing applications have focused on single subunit Class 2 CRISPR systems [54] (for example, Cas9 and Cas12a), but Class 1 systems hold great potential for editing technologies, despite consisting of multi-subunit complexes [55]. The signature gene in Class 1 Type I systems is Cas3, a 3′–5′ single-strand DNA helicase-nuclease enzyme that, unlike Cas9 or Cas12a, degrades target DNA processively [56–59]. This property of Cascade (CRISPR-associated complex for antiviral defense)–Cas3 systems raises the possibility of its development as a tool for large genomic deletions, such as the targeted removal of entire genes, gene clusters, islands, prophages or plasmids, a task that Class 2 systems are inefficient at.

Type I systems are the most prevalent CRISPR–Cas systems in nature [11], which has enabled the use of endogenous CRISPR–Cas3 systems for genetic manipulation via self-targeting. This has been accomplished in *Pectobacterium atrosepticum* (Type I-F) [60], *E. coli* (Type I-E) [61, 62], *Sulfolobus islandicus* (Type I-A) [63], Clostridium species (Type I-B) [64] Lactobacillus crispatus (Type I-E) [65], Serratia sp. (Type I-F) [66], *Haloarcula hispanica* (Type I-B) [67], *Streptococcus*

thermophilus (Type I-E) [68], *P. aeruginosa* (Type I-F) [69] and *Zymomonas mobilis* (Type I-F) [70], most frequently being used to generate small deletions with homologous repair templates. This multitude of different Type I systems have been shown to work to various degrees as editing tools in their native hosts, however, no Type I system has been optimized for efficient heterologous editing in bacteria, beyond the demonstration of the toxic effects of self-targeting [71]. Recent studies have repurposed Type I-E systems for DNA cleavage in human cells [72–74], and Type I-F, I-E and I-B systems for transcriptional modulation [75–77].

Here, we repurposed and optimized a Type I-C CRISPR system from *P. aeruginosa* (PaeCas3c) for both endogenous and heterologous genome engineering in four microbial species. Compared to other Type I systems, such as the well studied Type I-E system (six different proteins), Type I-C is streamlined, requiring only four proteins. By targeting the genome with a single CRISPR RNA (crRNA) and selecting only for survival after editing, PaeCas3c is a rapid, counter-selection-free approach to programmable large-scale genome engineering and genome minimization. Cascade–Cas3 is capable of efficient genome-scale deletions currently not achievable using other methodologies. It has the potential to serve as a powerful tool for basic research, discovery and strain optimization.

2.2 Materials and Methods

2.2.1 Bacterial strains, plasmids, DNA oligonucleotides, and media

A previously described [78] environmental strain of *P. aeruginosa* was used as a template to amplify the four cas genes of the Type I-C CRISPR–Cas system genes (cas3, cas5, cas7 and cas8). The genes were cloned into the pUC18-mini-Tn7T-LAC vector [79] using the SacI-PstI restriction endonuclease cut sites in the order cas5, cas7, cas8, cas3 to generate the plasmid pJW31 (Addgene number 136423). This vector was introduced into *P. aeruginosa* PAO1 [80], inserting the cas genes into the chromosome, following previously described methods [81]. Following integration, the excess sequences, including the antibiotic resistance marker, were removed via Flp-mediated excision. The resulting strain, dubbed PAO1$^{\mathrm{IC}}$, allowed inducible expression of the I-C system through induction with isopropylβ-D-1 thiogalactopyranoside (IPTG). This same method was used to integrate the Cas3–Cas8 tether mutant in the order cas5, cas3, cas8, cas7. The linker amino acid sequence is

RSTNRAKGLEAVS. An isogenic strain carrying Cas9 derived from *S. pyogenes* was constructed in the same fashion, resulting in the strain PAO1^{IIA}. For experiments to test the system in *P. syringae*, we used the previously characterized strain DC3000 [82]. *E. coli* editing experiments were conducted with strain K-12 MG1655 [83]. Experiments conducted with *K. pneumoniae* were performed using strain KPPR1 [84].

To construct the Cas3 helicase and nuclease mutant strains, the PAO1^{IC} system was used to introduce point mutations. crRNAs were designed to target Cas3 along with a HDR template that included the desired mutation, and silent mutations to prevent CRISPR–Cas targeting of the final strain.

To achieve genomic self-targeting of the I-C CRISPR–Cas strains, crRNAs designed to target the genome were expressed from the pHERD20T and pHERD30T shuttle vectors [85]. So-called 'entry vectors' pHERD20T-ICcr and pHERD30T-ICcr were first generated by cloning at the EcoRI and HindIII sites an annealed linear dsDNA template carrying the I-C CRISPR–Cas system repeat sequences flanking two BsaI Type IIS restriction endonuclease recognition sites. Additionally, a preexisting BsaI site in a noncoding site of the pHERD30T and pHERD20T plasmids was mutated using whole-plasmid amplification so it would not interfere with the cloning of the crRNAs [78]. Oligonucleotides with repeat-specific overhangs encoding the various spacer sequences were annealed and phosphorylated using T4 polynucleotide kinase and cloned into the entry vectors using the BsaI sites. For experiments using Cas9, sgRNAs were expressed from the same pHERD30T vector, with the sgRNA construct cloned using the same restriction sites as with the I-C crRNAs.

The all-in-one vector pCas3cRh (Addgene number 133773) is a derivative of the pHERD30T-IC plasmid, with the four I-C system genes cloned downstream of the crRNA site. This was achieved by amplifying the genes *cas3, cas5, cas8 and cas7* in two fragments with a junction within cas8 designed to eliminate an intrinsic BsaI site with a synonymous point mutation. The amplified fragments were cloned into pHERD30T-IC using the Gibson Assembly protocol [86]. Finally, to guard against potential leaky toxic expression, we replaced the araC-ParaBAD promoter with the rhamnose-inducible rhaSR-PrhaBAD system [87]. The sequence for rhaSR-PrhaBAD was amplified from the pJM230 template [87], provided by the laboratory of J.B. Goldberg (Emory University) and cloned into the pHERD30T-IC plasmid to replace araC-ParaBAD using the Gibson Assembly

(New England Biolabs). Without induction, transformation efficiencies of targeting constructs of assembled pCas3cRh were on average 5–10-fold lower when compared to nontargeting controls (Fig. 2.12C), indicating residual leakiness of the I-C system.

The aca1-containing vector pICcr-aca1 is a derivative of the pHERD30T-ICcr plasmid, with aca1 cloned downstream of the crRNA site under the control of the pBAD promoter. The aca1 gene was cloned from *P. aeruginosa* phage DMS3m.

All oligonucleotides used in this study were obtained from Integrated DNA Technologies. For a complete list of all DNA oligonucleotides and a short description, see Supplementary Table 4.

2.2.2 Bacterial transformations

Transformations of *P. aeruginosa*, *E. coli*, *P. syringae* and *K. pneumoniae* strains were conducted using standard electroporation protocols. Then, 10 ml of overnight cultures were centrifuged and washed twice in an equal volume of 300 mM sucrose (20% glycerol for *E. coli*) and suspended in 1 ml of 300 mM sucrose (20% glycerol for *E. coli*). Next, 100-μl aliquots of the resulting competent cells were electroporated using a Gene Pulser Xcell Electroporation System (BioRad) with 50–200ng of plasmid with the following settings: 200 ohms, 25 μF, 1.8 kV, using 0.2-mm gap width electroporation cuvettes (BioRad). Electroporated cells were incubated in antibiotic-free Super optimal broth with catabolite repression media for 1h at 37°C (28°C for P. syringae), then plated onto LB agar (King's medium B agar for P. syringae) with the selecting antibiotic and grown overnight at 37°C (28°C for P. syringae). Cloning procedures were performed in commercial *E. coli* DH5α cells (New England Biolabs) or *E. coli* XL1-Blue (QB3 Macrolab Berkeley), according to the manufacturer's protocols.

Construction of recombinant DMS3m acr phages

The isogenic DMS3m^{acrIIA4} and *acrIC1* phages were constructed using previously described methods [28]. A recombination cassette, pJZ01, was constructed with homology to the DMS3m acr locus. Using the Gibson Assembly (New England Biolabs), either *acrIC1* or *acrIIA4* were cloned upstream of *aca1*, and the resulting vectors were used to transform PAO1^{IC}. The transformed strains were

infected with wild-type DMS3m, and recombinant phages were screened for. Phages were stored in SM buffer at 4°C.

2.2.3 Isolation of PAO1^{IC} lysogens

PAO1^{IC} was grown overnight at 37°C in LB media. Then 150μl of overnight culture was added to 4ml of 0.7% LB top agar and spread on 1.5% LB agar plates supplemented with 10mM MgSO4. 5μl of phage, expressing either *acrIC1* or *acrIIA4* were spotted on the solidified top agar and plates were incubated at 30°C overnight. Following incubation, bacterial growth within the plaque was isolated and spread on a 1.5% LB agar plate. After an overnight incubation at 37°C, single colonies were assayed for the prophage. Confirmed lysogens were used for genomic targeting experiments.

2.2.4 Genomic targeting and measurement of growth rates

P.aeruginosa

Genomic self-targeting of *P. aeruginos*a PAO1^{IC} was achieved by electroporating cells with pHERD30T (or pHERD20T) expressing the self-targeting spacer of choice. Cells were plated onto LB agar plates containing the selective antibiotic, without inducers, and grown overnight. Single colonies were then grown in liquid LB media containing the selective antibiotic, as well as IPTG to induce the genomic expression of the I-C system genes, and arabinose to induce the expression of the crRNA from the plasmid. The aca1-containing crRNA plasmids do not need additional inducers, as the pBAD promoter controls aca1. Cultures were grown at 37°C in a shaking incubator overnight to saturation, then plated onto LB agar plates containing the selecting antibiotic, as well as the inducers, and incubated overnight again at 37°C. The resulting colonies were then analyzed individually using colony PCR for any differences at the targeted genomic site compared to a wild-type cell. gDNA was isolated by resuspending one colony in 20 μl of H2O, followed by incubation at 95°C for 15 min. Then 1–2 μl of boiled sample was used for PCR. The primers used to assay the targeted sites were designed to amplify genomic regions 1.5–3 kb in size. In the event of a PCR product equal to or smaller than the wild-type fragment (as was often observed when analyzing Cas9-targeted cells), Sanger sequencing (Quintara Biosciences) was used to determine any

modifications of the targeted sequences. In some cases, additional analysis of the crRNA-expressing plasmids of the surviving colonies was also performed, by isolating and reintroducing the plasmids into the original I-C CRISPR–Cas strain, where functional self-targeting could be determined based on a substantial increase in the lag time of induced cultures, characteristic of self-targeting events. In cases where a HDR template was used, homology arms ranging in size of 500–600 bp were cloned using a nested PCR-based approach where the two different arms were stitched together via 25-bp overlaps. These fragments were then cloned into the pHERD30T plasmid expressing self-targeting crRNAs at the NheI restriction sites. Genomic targeting was performed as described above. Surviving cells were analyzed using colony PCRs amplifying the desired deletion junction (verified with Sanger sequencing), as well as the wild-type target site. Editing efficiencies were counted as the number of colonies producing a desired deletion junction fragment from the total number of analyzed colonies.

Growth dynamics of various strains were measured using a Synergy 2 automated 96-well plate reader (Biotek Instruments) and the accompanying Gen5 software (Biotek Instruments). Individual colonies were picked and grown overnight in 300 μl volumes of LB in 96-well deep-well plates at 37°C. The grown cultures were then diluted 100-fold into 100 μl of fresh LB in a 96-well clear microtiter plate (Costar) and sealed with Microplate sealing adhesive (Thermo Scientific). Small holes were punched in the sealing adhesive for each well for increased aeration. Doubling times were calculated as described previously [88].

E. coli

Genomic self-targeting of *E. coli* was conducted in a similar fashion to *P. aeruginosa*, except using the pCas3cRh all-in-one vector. Electrocompetent *E. coli* cells were transformed with pCas3cRh expressing a crRNA targeting the genome. Individual transformants were selected and grown in liquid LB media containing the selecting antibiotic (gentamicin) overnight without any inducers added. The overnight cultures were then plated in the presence of inducer and X-gal to screen for functional lacZ (LB agar + 15 μg ml^{-1} gentamicin + 0.1% rhamnose + 1 mM IPTG + 20 μg ml^{-1} X-gal) and blue/white colonies were counted the next day.

P. syringae

Electrocompetent *P. syringae* cells were also transformed with pCas3cRh plasmids targeting selected genomic sequences. Initial transformants were plated onto King's medium B agar + 100 μg ml^{-1} rifampicin + 50 μg ml^{-1} gentamicin plates and incubated at 28°C overnight. Single colony transformants were then selected and inoculated in King's medium B liquid media supplemented with rifampicin, gentamicin and 0.1% rhamnose inducer, and grown to saturation in a shaking incubator at 28°C. Cultures were finally plated onto King's medium B agar plates with rifampicin, gentamicin and rhamnose and incubated at 28 °C. Individual colonies were finally assayed with colony PCR to determine the presence of deletions at the targeted genomic sites.

To test bacterial growth in planta, we used the *Arabidopsis thaliana* ecotype Columbia (Col-0), which has previously been shown to be susceptible to infection by P. syringae DC3000. Plants were grown for 5–6 weeks in 9h light/15h darkness and 65% humidity. For each inoculum, we measured bacterial growth in ten individual Col-0 plants. Four leaves from each plant were infiltrated at an optical density (OD600) of 0.0002, and cored with a no. 3 borer. The four cores from each plant were then ground, resuspended in 10 mM MgCl2 and plated in a dilution series on selective media for colony counts at both the time of infection and 3d postinfection.

To test bacterial growth in vitro, we used both King's medium B and plant apoplast mimicking minimal media. Overnight cultures were prepared from single colonies of each strain, washed and diluted to OD600=0.1 in 96-well plates using either King's medium B or minimal media. Plates were incubated with shaking at 28°C. OD600 was measured over the course of 24–25h using an Infinate 200 Pro automated plate reader (Tecan). Statistical analysis determined significantly different groups based on analysis of variance (ANOVA) analysis on the day 0 group of values and the day 3 group of values. Significant ANOVA results ($P<0.01$) were further analyzed with a Tukey's honestly significant difference post hoc test to generate adjusted P values for each pairwise comparison. A significance threshold of 0.01 was used to determine which treatment groups were significantly different.

K. pneumoniae

Electrocompetent *K. pneumoniae* cells were transformed with pCas3cRh plasmids targeting selected genomic sequences. Individual transformants were selected and grown in liquid LB media (containing 50 μg ml^{-1} gentamicin, as well as 0.1% rhamnose inducer) overnight. Various dilutions from saturated cultures were then plated the next day onto LB agar plates containing the selective antibiotic (gentamicin 50 μg ml^{-1}) and 0.1% rhamnose inducer. Individual colonies were then assayed for deletions using colony PCR.

WGS

The gDNA for WGS analysis was isolated directly from bacterial colonies using the Nextera DNA Flex Microbial Colony Extraction kit (Illumina) according to the manufacturer's protocol. The gDNA concentration of the samples was determined using a DS-11 Series Spectrophotometer/Fluorometer (DeNovix) and all fell into the range of 200–500ng per ul. Library preparation for WGS analysis was done using the Nextera DNA Flex Library Prep kit (Illumina) according to the manufacturer's protocol starting from the tagment genomic DNA step. Tagmented DNA was amplified using Nextera DNA CD Indexes (Illumina). Samples were placed overnight at 4°C following the tagmented DNA amplification step, then continued the next day with the library clean up steps. Quality control of the pooled libraries was performed using a 2100 Bioanalyzer Instrument (Agilent Technologies) with a High Sensitivity DNA Kit (Agilent Technologies). Most samples were sequenced using a MiSeq Reagent Kit v.2 (Illumina) for a 150bp paired-end sequencing run using the MiSeq sequencer (Illumina). *P. syringae* and Cas9-generated *P. aeruginosa* deletion strains were sequenced using a NextSeq 500 Reagent Kit v.2 (Illumina) for a 150bp paired-end sequencing run using the NextSeq 500 sequencer (Illumina).

Genome sequence assembly was performed using Geneious Prime software v.2019.1.3. Paired read data sets were trimmed using the BBDuk (decontamination using kmers) plugin using a minimum Q value of 20. The genome for the ancestral PAO1$^{\mathrm{IC}}$ strain was de novo assembled using the default automated sensitivity settings offered by the software. The consensus sequence of PAO1$^{\mathrm{IC}}$ assembled in this manner was then used as the reference sequence for mapping all of the PAO1$^{\mathrm{IC}}$ strains with

multiple deletions. As a control, the sequences were also mapped to the reference *P. aeruginosa* PAO1 sequence (NC002516) to verify deletion border coordinates. Coverage of these sequenced strains ranged from 66- to 143-fold, with an average of 98.3-fold. The sequenced *P. aeruginosa* environmental strains were also mapped to the PAO1 (NC_002516) reference, while the sequenced *E. coli* strains were mapped to the *E. coli* K-12 MG1655 reference sequence (NC_000913). Finally, sequenced *P. syringae* strains were mapped to the *P. syringae* DC3000 (NC_ 004578) reference sequence, along with the pDC3000A endogenous 73.5kb plasmid sequence (NC_004633). All of the remaining sequenced strains had >100-fold coverage. All deletion junction sequences were manually verified by the presence of multiple reads spanning the deletions, containing sequences from both end boundaries.

WGS data were visualized using the BLAST Ring Image Generator [89] tool65 using BLAST+ v.2.9.0. In several cases, short sequences were aligned inside previously determined large deletions at redundant sequences such as transposase genes. Such misrepresentations created by the BLAST Ring Image Generator were manually removed to reflect the actual sequencing data.

2.3 Results

2.3.1 Implementation and optimization of genome editing with CRISPR-Cas3

Type I-C CRISPR–Cas systems use three cas genes (*cas5, cas8 and cas7*) to produce the crRNA-guided Cascade surveillance complex [18, 90] that can recruit Cas3 (Fig. 2.1A). A previously constructed *P. aeruginosa* PAO1 strain ($PAO1^{IC}$) [78] with inducible cas genes (*cas5-8-7-3*) and plasmid-expressed crRNAs targeting the genome was used to conduct genome manipulation (see Fig. 2.6A for a comparison to other previously identified I-C systems). Introduction of crRNA-expressing plasmids under noninducing conditions was not noticeably toxic, indicating tight regulation of the constructs (see Methods). Induction of genome-targeting crRNAs caused a transient growth delay (Fig. 2.1B), but survivors were isolated after extended growth. By targeting *phzM*, a gene required for production of a blue-green pigment (pyocyanin), we observed yellow cultures (Fig. 2.1C) derived from 10/18 and 6/18 surviving colonies, from two independent *phzM*-targeting crRNAs. PCR of genomic DNA confirmed that the yellow cultures had lost this region, while

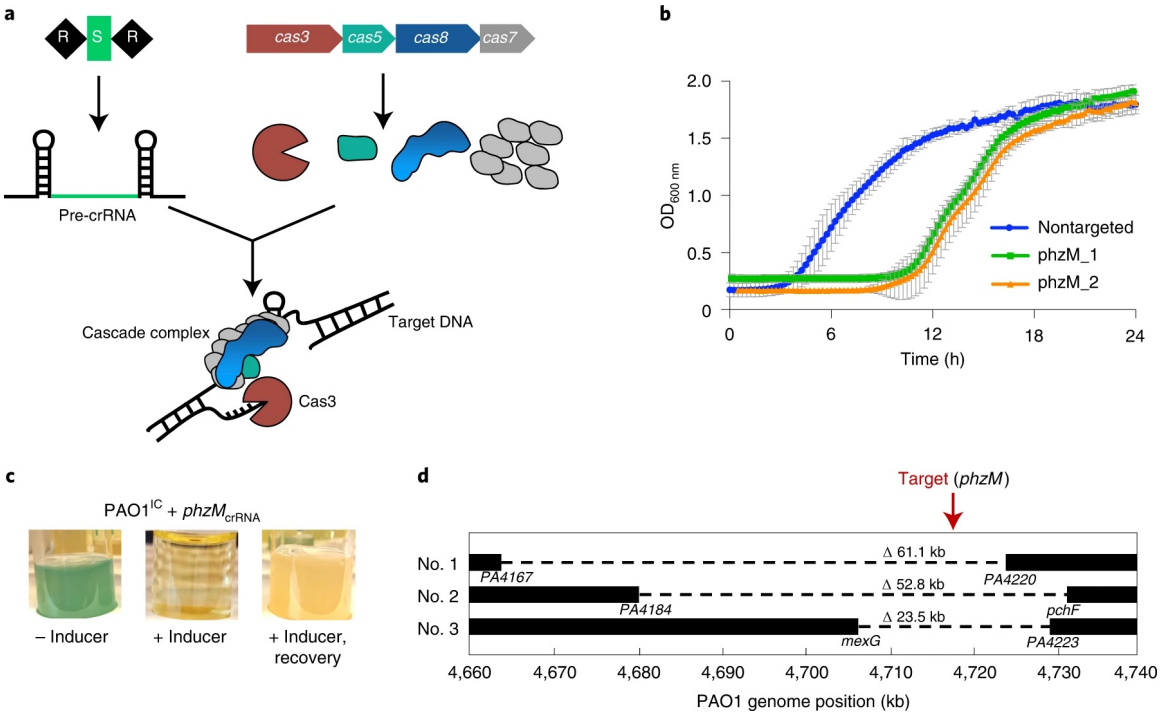


Fig 2.1. Type I-C CRISPR-mediated self-targeting leads to genomic deletions. (A) A schematic of the Type I-C cas gene operon and CRISPR array. The surveillance complex is made up of Cas proteins (Cas5, Cas8, Cas7) and one crRNA, which recruits Cas3 on target DNA recognition. Cas3 then degrades DNA through its dual helicase-nuclease activity. (B) Growth curves of two PAO1^{IC} strains expressing different crRNAs targeting *phzM* (green and orange) compared to a nontargeting strain (blue). Values are the mean of eight biological replicates each, error bars indicate s.d. values. (C) Cultures resulting from *phzM* targeting, in the absence of inducer (-ind), presence (+ind) and after recovery. (D) WGS of three PAO1^{IC} self-targeted survivor strains. Bars indicate boundaries of deletions with open reading frame (ORF) indicated below; red arrow indicates genomic position of targeted sequences.

blue-green survivors maintained it (Fig. 2.6B). Three of these deletion strains were sequenced, revealing deletions of 23.5, 52.8 and 60.1kb, and each one was bidirectional relative to the crRNA target site (Fig. 2.1D). This demonstrated the ability of Type I-C Cascade–Cas3 to induce large genomic deletions surrounding a programmed target site.To determine the in vivo processivity of the Cas3 enzyme, we identified 16 extended nonessential (XNES) regions >100kb in length (Supplementary Table 1) that lack a known essential gene [91]. Targeting XNES1 and XNES2 (along with additional targeting of *phzM*, which is found in XNES15) with two crRNAs each, led to deletions in 20–40% of the surviving colonies (Fig. 2.2A). To understand how cells lacking large deletions had survived self-targeting, three possibilities were considered: (1) cas gene loss-of-function mutations, (2) PAM or protospacer mutations or (3) mutations in the crRNA-expressing plasmid. Three survivors lacking target deletions from each of the six self-targeting crRNAs were assayed. All had functional cas genes when the self-targeting crRNA was replaced with a D3 phage-targeting crRNA (Fig. 2.7A), leading to a reduction in phage replication. Additionally, target sequencing revealed no point mutations. We did, however, observe spacer loss from the crRNA-expressing plasmid, via recombination between the direct repeats (Fig. 2.7B). An additional 17 survivors that lacked target deletions also had plasmids that were missing the spacer (Fig. 2.7C, D). Spacer excision was successfully prevented by modifying the second repeat, introducing six mutated nucleotides in the stem and three in the loop (Fig. 2.2B) and thus disrupting homology between the two direct repeats. A phage-targeting crRNA with this new design targeted phage as well or better than the same crRNA with unmodified repeats (Fig. 2.8A). Targeting of the same six genomic sites with modified repeat crRNAs resulted in consistent growth delays (Fig. 2.2C) and in a robust increase in editing efficiencies to 94–100% (Fig. 2.2A). Of 216 surviving colonies assayed with deletions generated by the six different crRNAs, 211 (98%) had large deletions (that is, >1kb), while the remaining 5 had inactive CRISPR–Cas systems when tested with the phage-targeting crRNA (Fig. 2.8B).When targeting poorly characterized genomes, essential genes may be unknowingly targeted leading to confounding editing outcomes. To assess the phenotype of such an event, we intentionally targeted an essential gene, *rplQ*. Two different modified repeat crRNAs targeting *rplQ* led to a severely extended lag time compared to nonessential gene targeting. Only 8 out of 36 *rplQ*-targeting biological replicates grew after 24h (Fig. 2.9A). Subsequent analysis of these eight

survivor cultures with phage-targeting assays revealed nonfunctional *cas* genes (Fig. 2.9B). No spacer excision events were detected in this experiment, confirming the robustness of the crRNA-engineering and of the deletion method, as the outcome of essential gene versus nonessential gene targeting is noticeably distinct.

2.3.2 Cas3 generates larger deletions than Cas9 and is recombinogenic

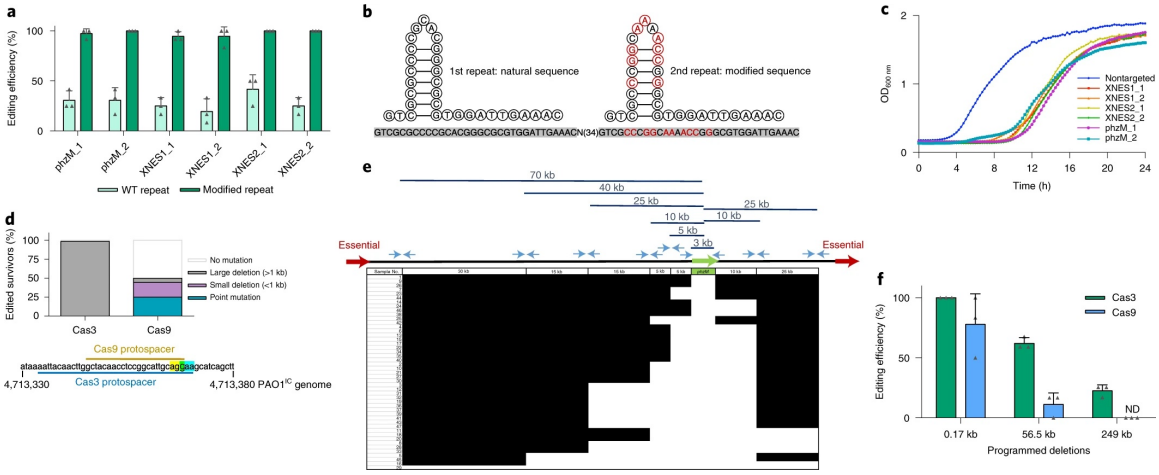


Fig 2.2. Optimization and characterization of Cascade–Cas3-directed genomic editing. (A) Percentage of survivors with a genomic deletion at the location targeted. Six different crRNA constructs with either wild-type (WT) repeat sequences (light green) or with the second repeat being modified (dark green). Values are means of three biological replicates each, where 12 individual surviving colonies were assayed per replicate, error bars show s.d. values. (B) Sequence and structure of natural and modified repeat sequences. Specifically engineered modified nucleotides shown in red; repeat sequences highlighted in gray with an arbitrary intervening spacer sequence. (C) Growth curves of PAO1^{IC} strains expressing distinct self-targeting crRNAs flanked by modified repeats. Nontargeting crRNA-expressing control is marked in blue. Values depicted are averages of four biological replicates each. (D) Gene-editing outcomes for distinct survivor cells targeted with either a Type II-A SpyCas9 system or a Type I-C Cas3 system (n=72 individual colonies isolated from distinct cultures of self-targeting cells for each system). (E) Deletion size distribution of 47 independently generated *phzM* deletion strains determined using tiling PCR. Black segments indicate the presence of a PCR product for a given sample, while white segments indicate the absence, thus giving an estimate of the size for the generated deletions. (F) Percentage of analyzed survivors with the specific deletion size present (0.17, 56.5 or 249 kb) generated with homologous repair templates, in tandem with the Cascade–Cas3 system (green) or the SpyCas9 system (blue). Values are means of three biological replicates each, where 12 individual surviving colonies were assayed per replicate and have been normalized to the total percentage of edited samples for each enzyme shown in Fig. 2d (98.6% for Cas3, 50% for Cas9); error bars show s.d. values; ND, not detected.

To determine whether large deletions are a direct consequence of the processive Cas3 enzyme, as

opposed to selecting for rare preexisting large genomic deletions [92], we compared Cas3-mediated self-targeting outcomes to those resulting from targeting using *Streptococcus pyogene*s Cas9, which lacks a helicase domain, expressed in an isogenic strain ($PAO1^{IIA}$). If the large deletions we observed in *P. aeruginosa* are preexisting in the population, we would expect these to be selected for at comparable frequencies regardless of the nuclease used to target the genome. A Cas9 single-guide RNA (sgRNA) was used with a spacer that overlapped with one of the Cas3 crRNAs used to target *phzM* (Fig. 2.2D, Fig. 2.10). Sequence analysis of these surviving cells revealed that deletions larger than 1kb were a rare occurrence in the presence of Cas9 (5.6% assayed survivor cells, n=72) compared to 98.6% with Cas3 (Fig. 2.2D). Whole-genome sequencing (WGS) of two large deletion survivors selected by Cas9 showed lesions of 5 and 23kb around the target site. The more common modes of survival after Cas9 targeting were small deletions between 0.1 and 0.5kb in length (25% of all survivors), or 1–3 basepair protospacer/PAM deletions/mutations (19.4%), with the remaining 50% of survivors unedited at their target sites. In sum, the apparent shift of deletions toward smaller size resulting from targeting with SpyCas9, compared to a nearly completely distinct set of outcomes when using Cas3, implicates Cas3's enzymatic activity as the cause of large deletions.To achieve a more granular measurement of the deletion sizes generated by Cas3-mediated editing, we examined 47 individual *phzM*-targeting events. Tiling PCR was used to determine the presence of flanking segments at various intervals spanning a total of 95kb surrounding *phzM* (Fig. 2.2E). Of the 47 independent deletion outcomes examined with this method, 44 had deletions of at least 5kb and 22 had deletions of at least 35 kb in size, with 1 strain having a deletion larger than 95kb. The average deletion was larger than 26.6 kb and smaller than 48.2kb, as based on the resolution of the tiling experiment. This comprehensive assessment confirms both the variability and size of Cas3-induced deletion outcomes.The processive ssDNAse activity of nuclease-helicase Cas3 led us to hypothesize that it may promote recombination. To test this, we provided a double-stranded DNA repair template with 500 bp of the upstream and downstream regions flanking a desired deletion to enable homology-directed repair (HDR) during targeting. We chose 0.17 and 56.5 kb deletions around *phzM*, to model a gene and prophage deletion, respectively, and a large 249 kb deletion within XNES8 for the programmed deletions (see Supplementary Information). The specific editing efficiencies were significantly higher with Cas3 than with Cas9 (Fig. 2.2F). The 249 kb

deletion was incorporated in 22% of the Cas3-generated survivors, compared to 0% using Cas9 (χ^2(1, N=72=9, $P = 2.7 \times 10^{-3}$). The 56.5 kb deletion was present in 61 versus 11% (χ^2(1, N=72)=25, P=5.73×10^{-7}), and the 0.17 kb deletion in 100 versus 78% of survivors, when targeting with Cas3 or Cas9, respectively (χ^2(1, N=72)=31.68, P=1.82×10^{-8}). During Cas3 targeting, strains that did not incorporate the HDR-programmed deletion had other deletions of random size. Most of the strains that survived Cas9-sgRNA induction without incorporating the HDR template had no change at the target site (84.7% and 80.6% for the 56.5 and 249 kb deletions, respectively), similar to data reported above (Fig. 2.2E). We presume that mutation or loss of Cas9 occurs more frequently than loss of the Cas3-based system under this experimental set-up. To account for this, normalizing the editing efficiency by roughly twofold (derived from the frequency of unedited clones in Fig. 2.2D) revealed that whether one considers the absolute percentage of colonies with the desired edits or the normalized value, Cascade–Cas3 targeting outperforms Cas9 for generating large specific deletions (Fig. 2.2F).

2.3.3 Rapid genome minimization of *P. aeruginosa* with CRISPR–Cas3 editing

Large deletions with undefined boundaries provide an unbiased mechanism for genome streamlining, screening and functional genomics. To demonstrate the potential for Cas3, we aimed to minimize the *P. aeruginosa* genome through a series of deletions of the XNES regions (Fig. 2.3A). Six XNES regions were iteratively targeted in six parallel lineages (Fig. 2.3B), resulting in 35 independent deletions (WGS revealed no deletion at XNES2 in one of the strains). Deletion efficiency remained high (>80%) throughout each round of self-targeting (Fig. 2.11A). WGS of these six multiple deletion strains (Δ61–Δ66) revealed that no two deletions had the exact same coordinates, highlighting the stochastic nature of Cas3. The smallest isolated deletion was 7 kb and the largest 424 kb (mean 92.9 kb, median 58.2 kb). Of note, four genes (PA0123, PA1969, PA2024 and PA2156) previously identified as essential [91] were deleted in at least one of the lineages. Most deletions appeared to be resolved by flanking microhomology regions ranging from 4 to 14 bp in length (Fig. 2.11B and Supplementary Table 2), implicating alternative-end joining [93] as the dominant repair process.To minimize the genome further, one of the already reduced strains was subjected to four additional rounds of deletions at XNES regions for a total of ten genomic deletions (Δ10, Fig. 2.3B). WGS of

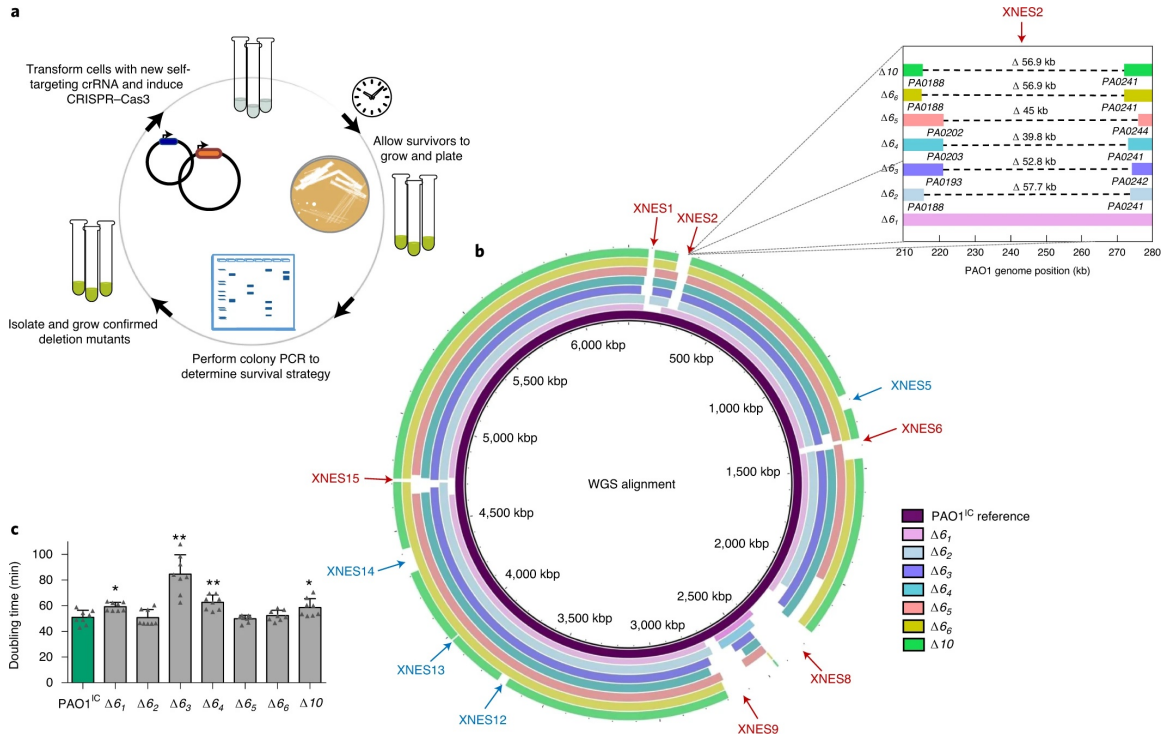


Fig 2.3. Iterative generation of multiple genomic deletions in *P. aeruginosa*. (A) Schematic overview of the iterative deletion generating process. (B) Whole-genome sequences of six PAO1^{IC} strains that have been iteratively targeted at six distinct genomic positions and one (derived from strain Δ66) with ten total deletions (Δ10) aligned to the parental P. aeruginosa PAO1^{IC} strain. The first six targeted sites are marked with red arrows and the final four are marked with blue arrows. Inset shows deletion coordinates of XNES2 region of the various strains in finer detail. (C) Calculated doubling times of the seven genome-reduced strains (strains Δ61–66 with six deletions, Δ10 with ten) compared to the parent PAO1^{IC} strain (green). Values are means of eight biological replicates, error bars represent s.d. values, *P<0.05 (Δ61 P=0.0103, Δ10 P=0.0173), **P<0.01 (Δ63 P=0.00064, Δ64 P=0.00764), paired two-sided t-test compared to PAO1^{IC}.

the Δ10 strain showed a genome reduction of 849 kb (13.6% of the genome). Generation of large deletions resulted in a growth defect in some cases, with significantly slower growth in three of the six deletion strains (Δ61, Δ63 and Δ64), with the other three growing normally (Fig. 2.3C). Strain Δ10 also displayed a slight decrease in fitness, showing a 15% increase in doubling time compared to the parent strain. Stronger growth defects were likely avoided by the selection of fast-growing colonies at each deletion round. To determine whether any deletions may be preexisting at low frequencies in unedited cells, PCR primers probed for specific deletions at XNES1, 6, 8 and 9, revealing no products (Fig. 2.11C). This again indicates that Cas3 has a direct role in generating large deletions.

2.3.4 CRISPR–Cas3 editing in distinct bacteria

To enable expression of this system in other hosts, we constructed an all-in-one vector (pCas3cRh) carrying the I-C specific crRNA with a modified repeat sequence, *cas3, cas5, cas8 and cas7*, under rhamnose induction (Fig. 2.12A). As a pilot experiment, we transformed wild-type PAO1 with a nontargeting crRNA and crRNAs targeting *phzM* and XNES2. Minimal leaky expression of the system was observed, as transformation efficiency was only slightly reduced comparing targeting and nontargeting constructs. Subsequent induction of the targeting crRNAs resulted in 95–100% of survivors being edited (Fig. 2.12B–D).

Having verified that pCas3cRh was functional, we tested this system in *E. coli* K-12 MG1655. pCas3cRh encoding crRNAs targeting *lacZ* or its vicinity (Fig. 2.4A) were used to transform cells, which were plated directly on inducing media containing X-gal and scored using blue/white screening. Depending on the crRNA used, directly targeting *lacZ* or 30 kb upstream yielded 51–90% and 82–85% LacZ (-) survivors, respectively (Fig. 2.4B). Of the 96 LacZ (-) survivors, 95 assayed by PCR showed an absence of the *lacZ* region. crRNAs targeting downstream of lacZ, however, had reduced efficiency as they approached the essential gene, *hemB*. *frmA* targeting (9.2 kb upstream of hemB) resulted in lower editing efficiencies (21–25%) while targeting *yaiS* (4.2 kb upstream of hemB) was even lower (2%). This decrease in efficiency was independent of the strand being targeted (and therefore the predicted strand for Cas3 loading and 3′–5′ translocation), confirming the importance of Cas3 bidirectional deletions. WGS of selected Δ*lacZ* cells revealed bidirectional

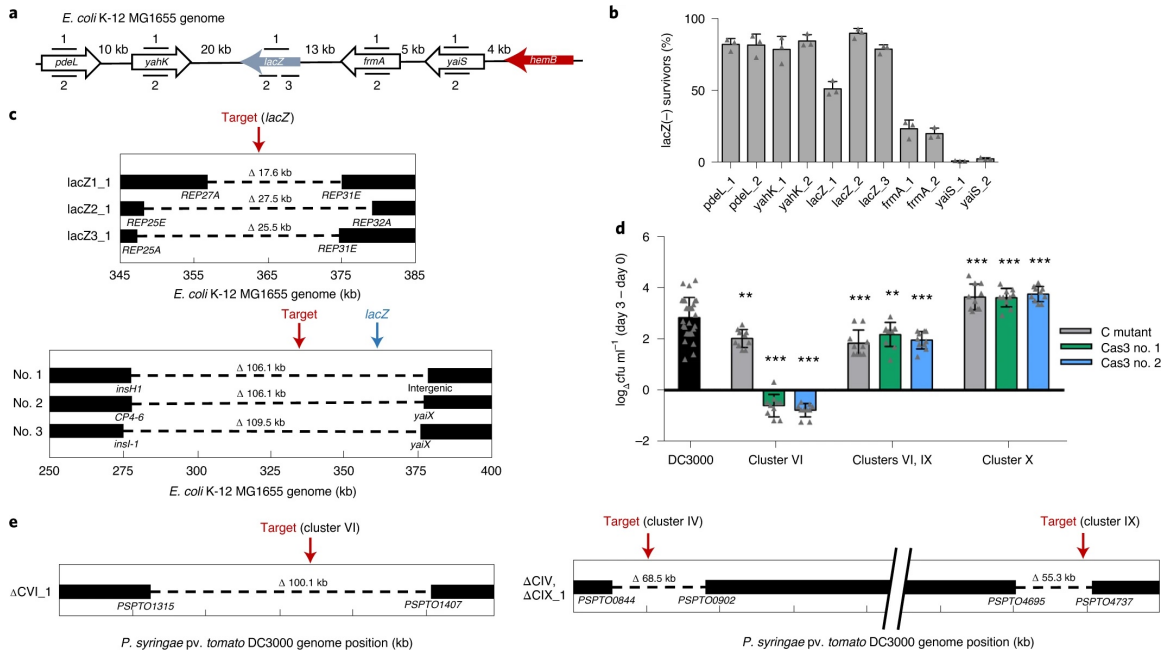
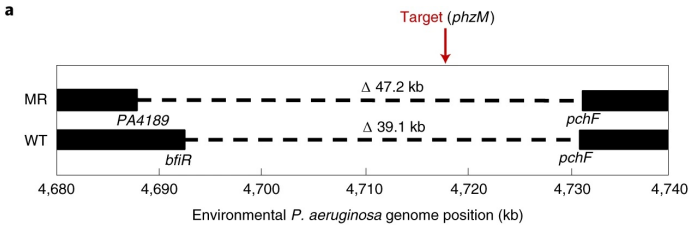
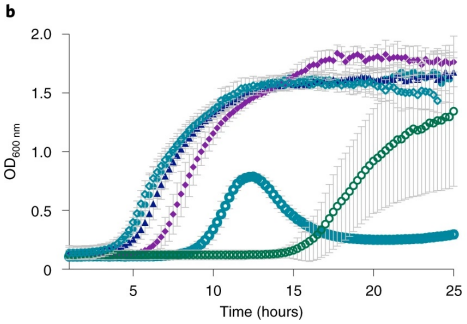


Fig 2.4. Cascade–Cas3-mediated heterologous editing in various bacteria. (A) Schematic of the crRNA-targeted sites in the *E. coli* MG1655 genome at the *lacZ* locus. (B) *lacZ* deletion efficiencies using distinct crRNAs targeting the *E. coil* K-12 MG1655 chromosome. Efficiencies calculated based on LacZ activity. Values are averagesof three biological replicates, error bars represent standard deviations. (C) WGS of *E. coli* deletion mutants targeted at *lacZ* and 30 kb upstream at *pdeL*. Bars indicate boundaries of deletions with ORF indicated below. (D) Bacterial growth of deletion mutants in *Arabidopsis thaliana*. Values are differences in colony forming units (cfu)/ml counted on day 0 of the experiment and day 3, shown on a logarithmic scale. The wild-type DC3000 strain is shown in black, while gray bars represent previously constructed polymutant control (labeled as 'C') strains of the different clusters (labeled at bottom), and green and blue bars show deletion mutants generated using Cas3 (two isolated strains for each targeted cluster, nos. 1 and 2). Values shown are means of ten biological replicates each (30 for DC3000), error bars show standard deviation values, P<0.01 (C mutant VI P= 0.00567, Cas3_1 IV, IX P=0.00829) P<0.005, (Cas3_1 VI P= 2.3310^{-15}, $Cas3\_2VIP$=9.27×10^{-18}, C mutant IV,IX P= 9.9510^{-5}, Cas3_2 IV,IX P=0.00165, C mutant X P=0.00185, Cas3_1 X P=0.000858, Cas3_2 P=0.000864), determined based on ANOVA analysis on the day 0 group of values and the day 3 group of values. (E) WGS of *P. syringae* deletion mutants. The left panel shows virulence cluster VI targeting, while the right panel shows virulence cluster IV and IX targeting with a single crRNA, as the clusters share sequence identity. Bars indicate boundaries of deletions with ORF indicated below.

deletions ranging from 17.6 to 109.5 kb encompassing the targeted region (Fig. 2.4C). Based on these findings, the nearby presence of an essential gene can substantially lower editing efficiency and must be considered when targeting a selected region. Next, we tested Cas3-mediated editing in the plant pathogen *P. syringae* pv. tomato DC3000, which does not naturally encode a CRISPR–Cas system. *P. syringae* encodes many nonessential virulence effector genes whose activities are difficult to disentangle due to their redundancy [94]. We designed crRNAs targeting four chromosomal virulence effector clusters (IV, VI, VIII and IX), or one plasmid cluster (pDC3000, ref. [95]; cluster X). Two clusters (IV and IX) shared identical sequences that could be targeted simultaneously using a single crRNA. Expression of targeting crRNAs caused a growth delay compared to nontargeted controls (Fig. 2.13A) and 67–92% of survivors had deletions (Fig. 2.13B). In planta and in vitro growth assays of three deletion mutants effectively recapitulated the phenotypes of previously described mutants [95] (Fig. 2.4D and Fig. 2.13C–H). Targeting cluster X cured the 73 kb plasmid (as observed by the absence of plasmid-specific reads in WGS) and simultaneous cluster IV and IX targeting led to dual deletions in eight out of 12 survivors, with a sequenced representative having 68.5 and 55.3 kb deletions, respectively, at the target sites (Fig. 2.4E). The effector cluster VI Cas3-derived mutant (100.1 kb deletion) had a more severe growth defect *in vitro* and *in planta* than the control mutant, likely from a fitness defect owing to the missing genetic material (Fig. 2.4D, E and Fig. 2.13C,F). In contrast to *P. aeruginosa*, IS elements were present at deletion junctions, suggesting the involvement of homologous recombination between insertion sequence (IS) elements flanking the virulence gene clusters. In such instances, we have not ruled out that the loss of these large regions was not a natural occurrence in the population, as seen in *S. thermophilus* [92]. In two out of three cases, however, the deletions did entail significant fitness costs (Fig. 2.4D), decreasing this likelihood. Using our portable streamlined system, we achieved three distinct applications in *P. syringae*: plasmid curing (similar to previous observations with a I-E system [96]), single-step deletion of large virulence regions and multiplexed targeting. Finally, we tested the feasibility of heterologous editing using the I-C system in more distantly related and clinically relevant bacteria *K. pneumoniae*. Using pCas3cRh, *K. pneumoniae* strain KPPR1 was targeted with four distinct crRNAs, with two each targeting *rfaH* and *sacX*, which are flanked by nonessential genes [97]. On induction, all four crRNAs resulted in a substantial growth delay compared to a nontargeted control, indicating





	Prophage *acr*	crRNA	*cas*/crRNA induction	*aca1*	Edited survivors
○	AcrIC1	phzM_1	+/+	+	6/10
●	AcrIC1	phzM_1	+/+	–	0/10
◆	AcrIC1	NT	+/+	–	0/10
▲	AcrIC1	phzM_1	++/++	–	NA
○	AcrIIA4	phzM_1	+/+	–	NA
◆	AcrIIA4	NT	+/+	–	NA

Fig 2.5. Cascade–Cas3-mediated gene editing in native settings. (A) Schematic of WGS of an environmental isolate of *P. aeruginosa* with an endogenous Type I-C system. Two survivors were isolated post-targeting using either wild-type (WT) direct repeats flanking the spacer or modified repeats (MR). Bars indicate boundaries of deletions with ORF indicated below. (B) Growth curves of PAO1^{IC} lysogenized by recombinant DMS3m phage expressing *acrIIA4* or *acrIC1* from the native *acr* locus. CRISPR–Cas3 activity is induced with either 0.5mM (+) or 5mM (++) IPTG and 0.1% (+) or 0.3% (++) arabinose. Edited survivors reflect number of isolated survivor colonies missing the targeted gene (*phzM*). NA means that editing was not assessed as no growth defect was seen and NT means a nontargeting crRNA was expressed. Each growth curve is the average of ten biological replicates and error bars represent s.d.

functionality (Fig. 2.14A). Individual surviving clones were isolated and 38–63% of survivors had deletions, showing the feasibility of Cas3 editing in *K. pneumoniae* as well (Fig. 2.14B,C). *rfaH* deletion mutants had a smaller colony morphology, consistent with previous work [97] (Fig. 2.14D). Overall, we have demonstrated portable Type I-C Cascade–Cas3 editing to be a generally applicable tool capable of generating large genomic deletions in four distinct species.

2.3.5 Repurposing endogenous Cascade–Cas3 systems for gene editing

Type I CRISPR–Cas3 systems are the most common CRISPR–Cas systems in nature [11]. Therefore, many bacteria have a built-in genome editing tool to be harnessed. We introduced self-targeting *phzM* crRNAs into the environmental isolate (PaLML1) from which our Type I-C system was derived. Genome targeting led to the isolation of 33.7 and 39 kb deletions (Fig. 2.5A and Fig. 2.15A). Additionally, HDR-based editing with a single construct was again efficacious, with 7/10 survivors acquiring the specific 0.17 kb deletion (Fig. 2.15B). We next evaluated the feasibility of repurposing other Type I systems, using the naturally active Type I-F systems [98] encoded by laboratory strain *P. aeruginosa* PA14, and the clinical strain *P. aeruginosa* z8. Plasmids with Type I-F specific crRNAs were expressed, targeting various genomic sites for deletion (Supplementary Table 3). HDR templates (600 bp arms on average) were included in the plasmids to generate deletions of defined coordinates ranging from 0.2 to 6.3 kb. Overall, at five different genomic target sites in strain z8 and 2 sites in PA14, we observed desired deletions in 29–100% of analyzed survivor colonies (Supplementary Table 3). Together, these experiments demonstrate the capacity for different forms of high-efficiency genome editing using a single plasmid and an endogenous CRISPR–Cas system.

Finally, one potential impediment to the implementation of any CRISPR–Cas bacterial genome editing tool is the presence of anti-CRISPR (Acr) proteins that inactivate CRISPR–Cas activity [99]. In the presence of a prophage expressing AcrIC1 (a Type I-C anti-CRISPR protein [78]) from a native *acr* promoter, targeting was completely inhibited, but not by an isogenic prophage expressing a Cas9 inhibitor AcrIIA4 [36] (Fig. 2.5B). To attempt to overcome this impediment, we expressed *aca1* (anti-CRISPR-associated gene 1), a direct negative regulator of *acr* promoters [38], from the same construct as the crRNA. Using this repression-based 'anti-anti-CRISPR' strategy, CRISPR–Cas function was enabled, allowing the isolation of edited cells despite the presence of acrIC1 (Fig. 2.5B

and Fig. 2.15C). In contrast, simply increasing *cas* gene and crRNA expression did not overcome AcrIC1-mediated inhibition as assessed by growth kinetics (Fig. 2.5B). Therefore, using anti-CRISPR repressors to combat anti-CRISPR impediments presents a viable route toward enhanced efficiency of CRISPR–Cas editing and necessitates continued discovery and characterization of anti-CRISPR proteins and their cognate repressors.

2.4 Discussion

By repurposing a minimal Cascade–Cas3 system (referred to as PaeCas3c), large deletions of random or programmed sizes can be obtained with high efficiencies. Using only a single crRNA with modified repeat sequences, we isolated deletions with variable sizes, one as large as 424 kb, without requiring the insertion of a selectable marker. Notably, the mean (92.9 kb) and median (58.2 kb) deletion sizes are roughly in the range of the average size of *Pseudomonas* bacteriophages (35–100 kb for 92% of sequenced genomes [100]), suggesting that the Cascade–Cas3 machinery can efficiently degrade entire phage genomes. Few studies have directly measured the processivity of the Cas3 enzyme in vivo. Additionally, the I-C system appears to produce bidirectional deletions contrary to unidirectional deletions observed with Type I-E [57, 58, 72]. Cascade–Cas3 presents a genome editing tool useful for the targeted removal of large elements for genome streamlining. As a long-term goal of microbial gene editing has been genome minimization [101, 102], we used our optimized Cascade–Cas3 system to generate ten iterative deletions, achieving greater than 13% genome reduction of the targeted strain. This spanned only 30 days while maintaining editing efficiency, a great improvement over previous genome reduction methods [102]. Some basic microbial applications of Cas3 include studying chromosome biology (for example, replichore asymmetry [103]), pathogenesis, the impact of the mobilome and a better understanding the essential building blocks for life.

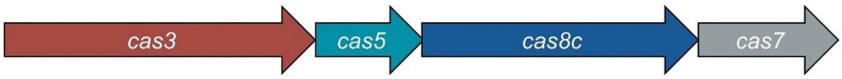
An important outcome of this work is the high efficiency of recombination observed at cut sites when comparing Cas3 and Cas9 directly. The potential for Cas3 to be recombinogenic through the generation of exposed ssDNA may be advantageous for both programmed knockouts and knockins. Although knockins were not systematically explored here, a preliminary attempt to affix

a chromosomal mCherry tag to *cas3*, using the Type I-C system for selection, was successful. The direct comparison presented here between Cas3 (large deletions) and Cas9 (small deletions), coupled with the high variability of deletions not observed to be preexisting also demonstrates the causality of Cas3 in the deletion outcomes.

Our study revealed benefits and challenges of applying Cascade–Cas3. While the Type I-C system was functional in heterologous hosts, it remains unclear whether the approach will be limited by differences in DNA repair mechanisms. Indeed, in *E. coli* and *P. syringae*, larger regions of homology, such as 34-bp long repetitive extragenic palindromic sequences were observed [104], indicating the role of RecA-mediated homologous recombination in the repair process. Meanwhile in *P. aeruginosa*, the borders of the deletions showed either small (4–14 bp) microhomology or no noticeable sequence homology. The former implies a role for alternative-end joining [93], which has also been observed in *P. atrosepticum* [60], while the latter implicates nonhomologous end joining [105] in the repair process. Downstream studies are required to dissect the roles of each mechanism in the deletion generation process.

CRISPR–Cas3 is an especially promising tool for use in eukaryotic cells as it would facilitate the interrogation of large segments of noncoding DNA, much of which has unknown function. Additionally, it was recently shown that Cas9-generated 'gene knockouts' (that is, small indels causing out-of-frame mutations) frequently encode pseudo-messenger RNAs that may produce protein products, necessitating methods for full gene removal [106, 107]. Type I-E CRISPR–Cas systems were recently shown to generate large (up to 100–200 kb) deletions in human cells [72–74], demonstrating the potential wide applicability of Cas3. Overall, the intrinsic properties of Cas3 make it a promising tool to fill a void in current gene-editing capabilities. Using Cas3 to make large genomic deletions will facilitate the manipulation of repetitive and noncoding regions, having a broad impact on genetics research by providing a tool to probe genomes en masse.

A



	Cas3 (%)	Cas5 (%)	Cas8 (%)	Cas7 (%)
*Bacillus halodurans**	Query cover.: 98 ID: 32	Query cover.: 93 ID: 31	Query cover.: 96 ID: 24	Query cover.: 94 ID: 35
Bacillus cereus	Query cover.: 91 ID: 31	Query cover.: 93 ID: 29	Query cover.: 19 ID: 29	Query cover.: 94 ID: 33
Lactobacillus fermentum	Query cover.: 97 ID: 27	Query cover.: 93 ID: 28	Query cover.: 78 ID: 22	Query cover.: 87 ID: 31
Mycobacterium canettii	Query cover.: 93 ID: 36	Query cover.: 93 ID: 41	Query cover.: 99 ID: 28	Query cover.: 95 ID: 45
Streptococcus pyogenes	Query cover.: 92 ID: 28	Query cover.: 91 ID: 29	Query cover.: 23 ID: 31	Query cover.: 84 ID: 34

B

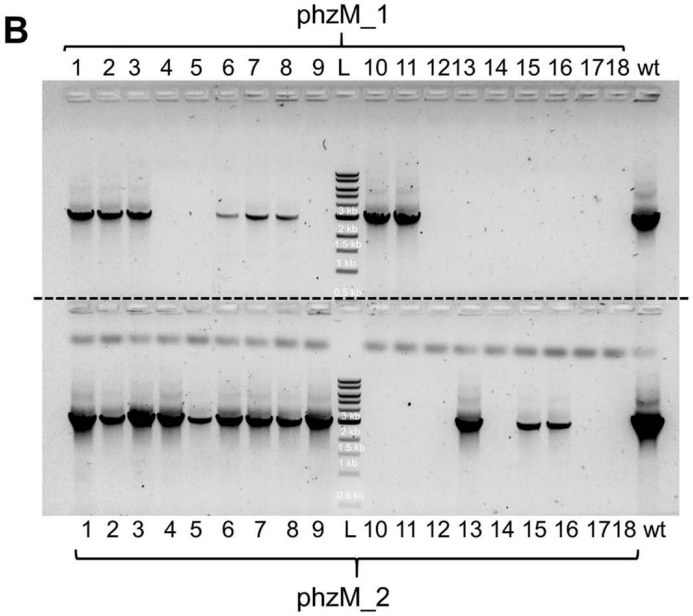


Fig 2.6. Type I-C CRISPR targeting leads to genomic deletions. (A) Comparison of Type I-C CRISPR system from *P. aeruginosa* used in the study, to various other previously identified I-C systems from a range of different bacteria. Values show query coverage and percent identity (ID) percentages comparing the four genes of the *P. aeruginosa* system to each of the other four. * Denotes the reference Type I-C CRISPR system referred to in Ref. 1. (B) PCR amplification of a 3kb genomic fragment flanking the *phzM* gene targeted using two different crRNAs, phzM_1 and phzM_2. Colony PCRs were performed on 18 biological replicates of self-targeted strains for each crRNA. The PAO1^{IC} parental strain is used as a positive control (wt). L indicates a 1 kb DNA ladder.

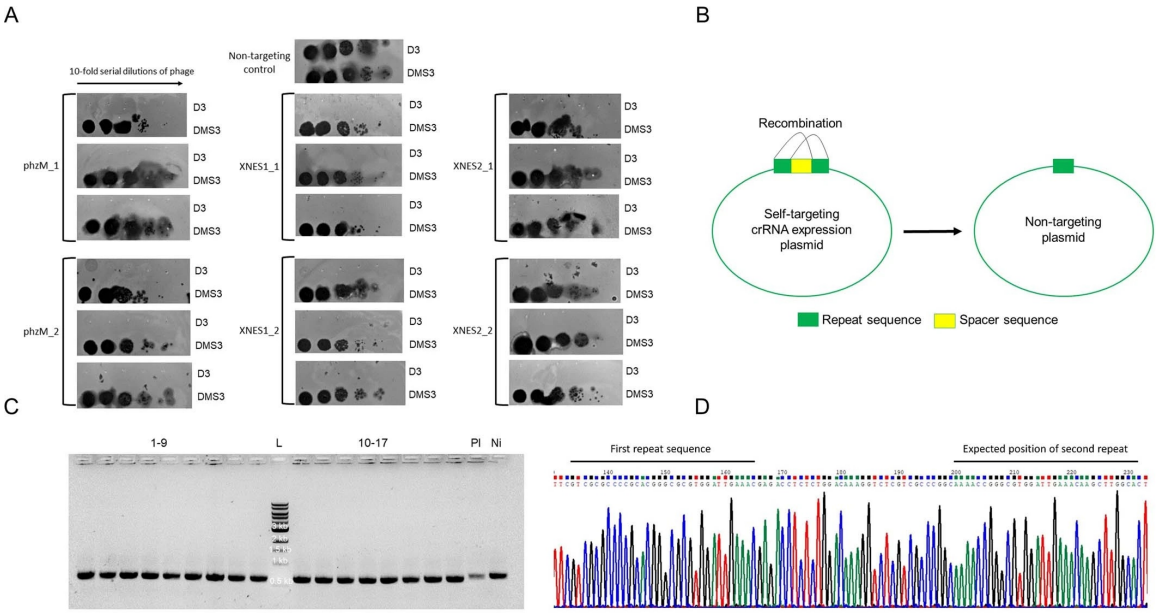


Fig 2.7. Excision of plasmid-encoded spacer sequences. (A) Phage targeting assays with survivors that had no discernable deletion of the crRNA-targeted genomic site. Strains were transformed with a D3 phage-targeting crRNA to assay for IC CRISPR-Cas3 activity. Three unique survivors were isolated from six self-targeting assays for a total of 18 survivors. Control is a non-targeting crRNA. (B) Schematic of spacer excision events where the two direct repeats recombine, resulting the loss of the targeting spacer. (C) PCR amplification of the crRNA sequence from plasmids isolated from 17 non-deletion self-targeted survivors (selected from 3 biological replicates of 12 analyzed colonies). Pl indicates the original plasmid as the PCR template, Ni indicates a sample where the crRNA was not induced, L indicates a 1kb DNA ladder. (D) Sample chromatogram of a sequenced plasmid with the spacer flipped out. Only one 32bp repeat sequence remains in the plasmid, the 34bp spacer sequence and other 32bp repeat are missing.

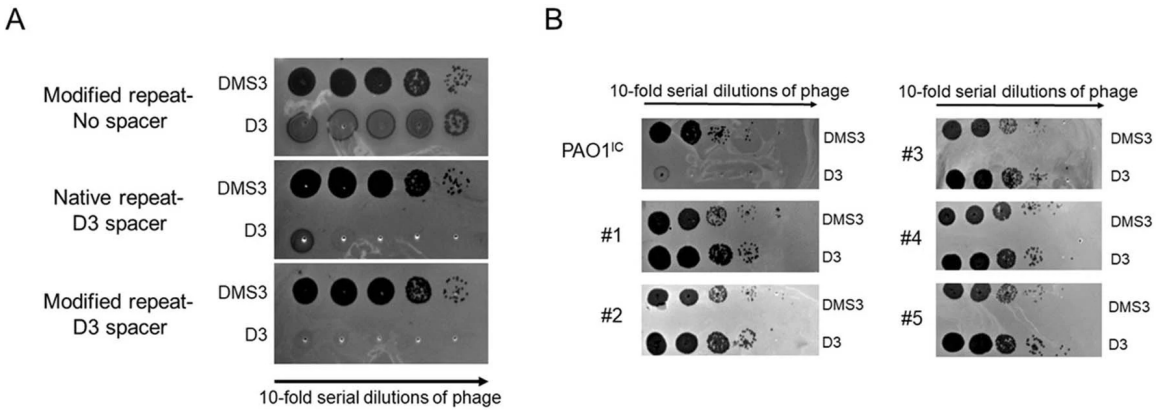


Fig 2.8. Phage-targeting assays to confirm CRISPR-Cas functionality. (A) Phage-targeting assay showing the activity of the modified repeat crRNA constructs. Ten-fold serial dilutions of DMS3 phage and D3 phage were spotted on lawns of PAO1^{IC} expressing either empty vector (top), a crRNA targeting D3 with WT direct repeats (middle), or a crRNA targeting D3 with modified repeats (bottom). (B) Phage targeting assay of five non-deletion self-targeting survivors expressing a D3 phage targeting crRNA. Unsuccessful targeting of phage indicates a non-functional CRISPR-Cas system in these strains. The parental PAO1^{IC} strain with a functional CRISPR-Cas system was used as a control.

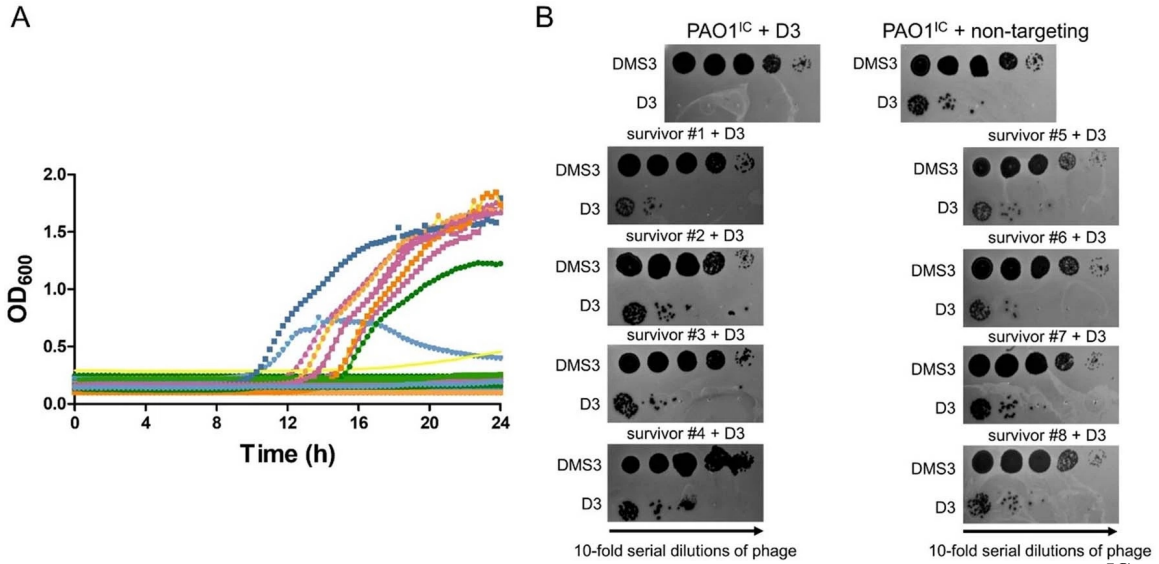


Fig 2.9. Genomic targeting of essential gene *rplQ*. (A) Growth curves of 36 PAO1^{IC} biological replicates targeting the essential gene, *rplQ*, using the MR crRNA plasmid. (B) Phage targeting assays with eight isolated *rplQ*-targeted survivors to assay for I-C CRISPR-Cas activity. Serial dilutions of DMS3 phage and D3 phage were spotted on lawns of PAO1^{IC} expressing a crRNA targeting phage D3. The parent PAO1^{IC} strain expressing a D3 targeting crRNA (top left) was used as a positive control, while PAO1^{IC} expressing a non-targeting crRNA was used as a negative control.

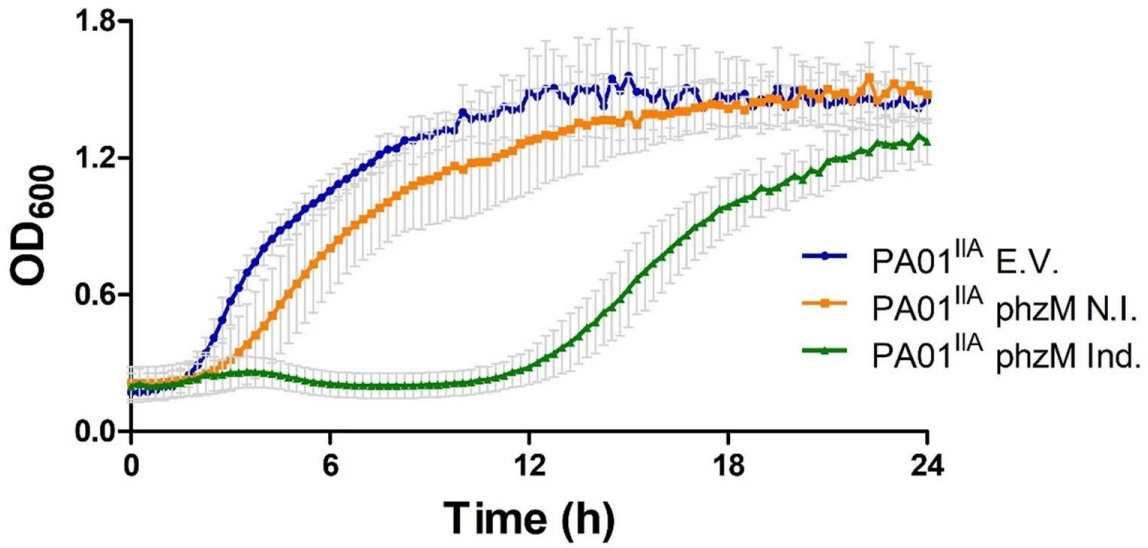


Fig 2.10. Genomic targeting using a Type II-A CRISPR-Cas system. Growth of self-targeting strains of PAO1^{IIA} expressing a self-targeting gRNA targeting the genome at *phzM* (Ind.). An empty vector (E.V.) and a non-induced *phzM* targeting strain (N.I.) were used as controls. Mean OD values measured at 600nm are shown for 8 biological replicates each, error bars indicate SD values.

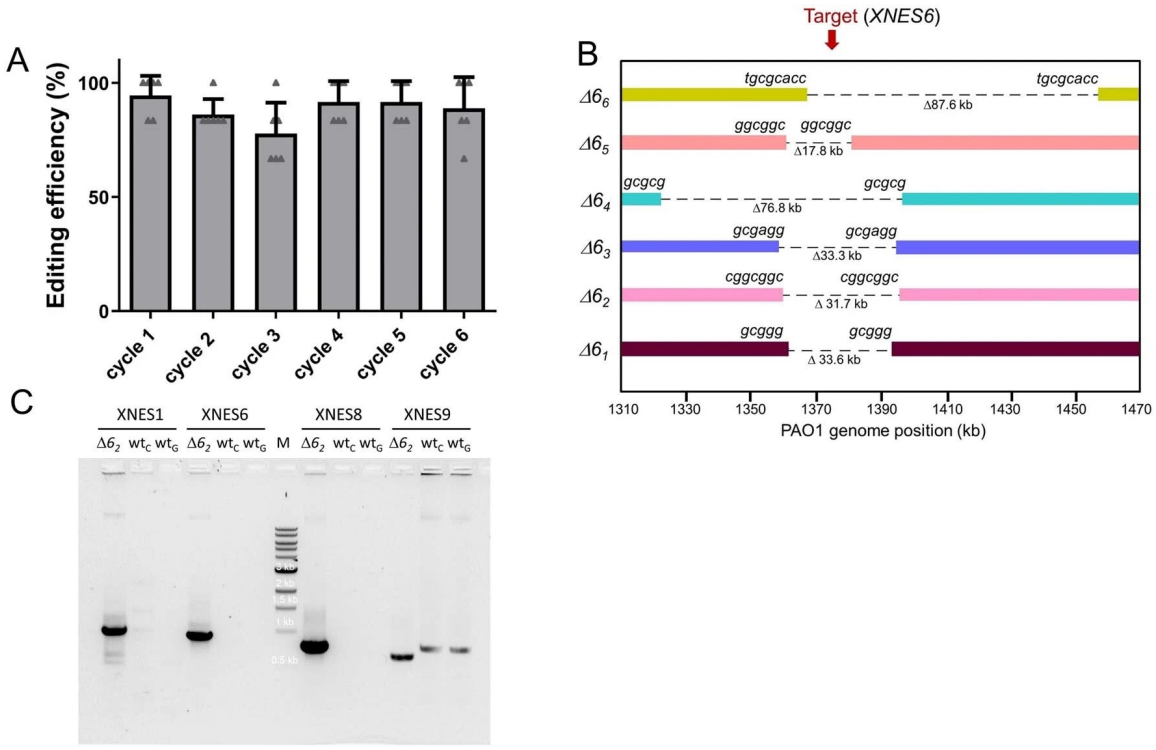


Fig 2.11. Genomic deletions and junction sites. (A) Deletion efficiencies observed over six cycles of iterative self-targeting. Six genomic targets were targeted in six different orders. Six survivors were analyzed using site-specific PCR after each cycle, for a total of 36 analyzed colonies (6*6) after each cycle, error bars represent standard deviations. (B) Deletion junctions at XNES6 target site of the 6 PAO1^{IC} strains with 6 iterative targeting events each. Sequences of each specific microhomology for the junctions are shown for each strain above the bars representing the given genomes at both ends, deletion sizes are shown below dashed lines for each strain. (C) PCR analysis using a representative set of primers amplifying various large deletion junctions (at XNES1, 6, 8, and 9 regions) of the whole-genome sequenced Δ62 strain. Δ62 served as a positive control template, while wtC represents untargeted PAO1^{IC} cells scraped from a lawn of colonies from a single overnight culture grown on plates serving as templates, and wtG represents isolated genomic DNA from a different 1.5ml overnight culture of untargeted PAO1^{IC} used as templates. Bands appearing for the XNES9 deletion junction for the PAO1^{IC} samples were aspecific and when sequenced, did not match any genomic region of the PAO1^{IC} genome. L indicates a 1kb DNA ladder.

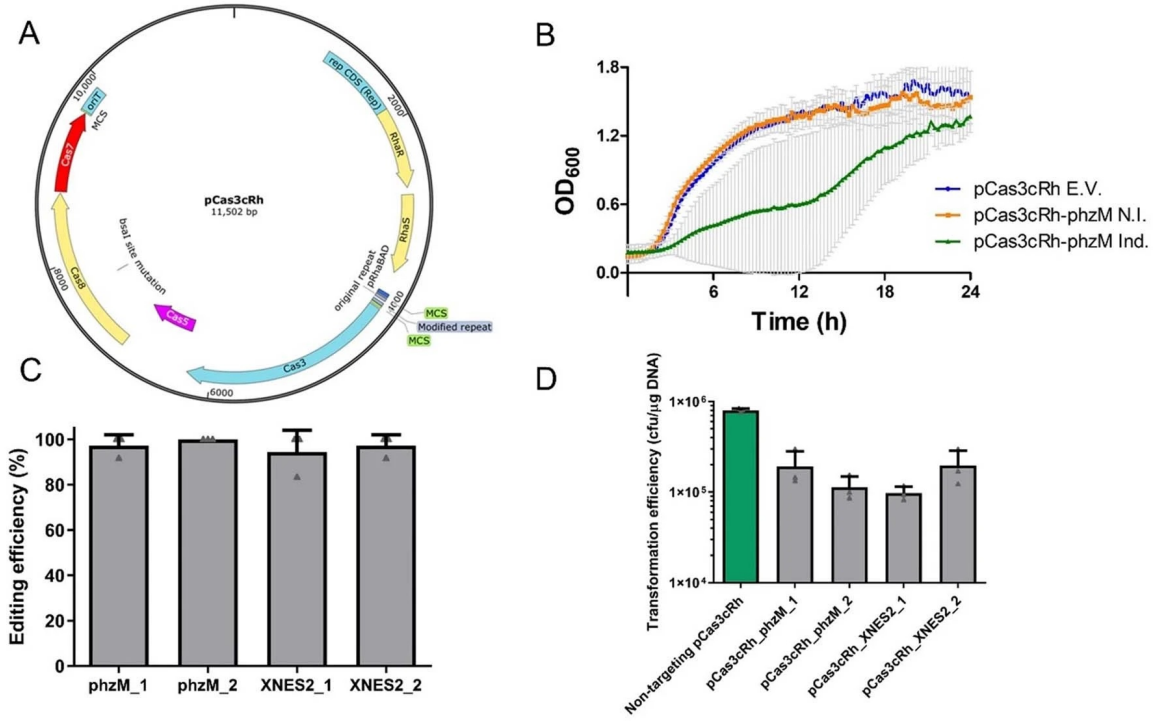


Fig 2.12. Genomic targeting of PAO1^{IC} with all-in-one vector pCas3ch. (A) Map of the I-C CRISPR-Cas all-in-one plasmid pCas3cRh carrying I-C crRNA and genes *cas3, cas5, cas8*, and *cas7* under the control of the rhamnose-inducible rhaSR-PrhaBAD system. (B) Growth curve of PAO1^{IC} transformed with the pCas3cRh vector expressing a self-targeting crRNA targeting phzM (Ind.). An empty vector (E.V.) and a non-induced *phzM* targeting strain (N.I.) were used as controls. Mean OD values measured at 600nm are shown for six biological replicates each. (C) Deletion efficiencies for WT PAO1^{IC} using the all-in-one vector pCas3cRh carrying all necessary components of the I-C CRISPR-Cas system. Values are averages of three replicates where 12 individual colonies were analyzed using site-specific PCR. Error bars show standard deviations. (D) Transformation efficiencies with self-targeting pCas3cRh vectors expressing crRNAs for *phzM* or XNES 2 compared to a non-targeting control (green bar) in PAO1^{IC}. Values are means of 3 replicates each, error bars represent SD values.

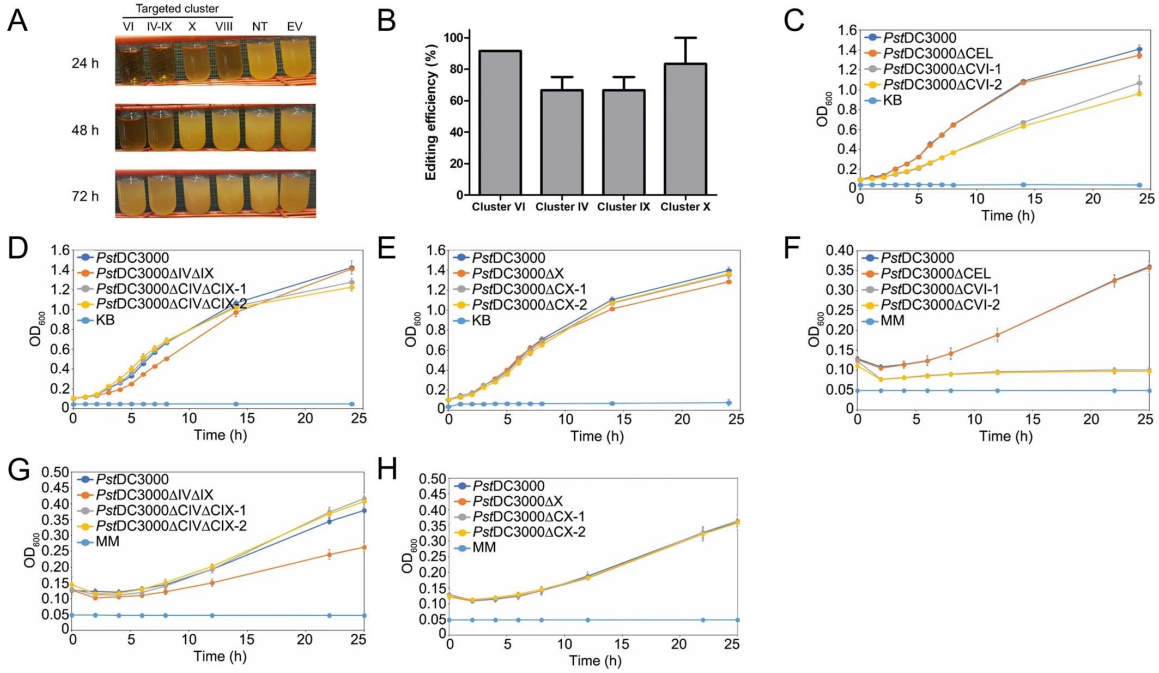


Fig 2.13. Genomic targeting of *Pseudomonas syringae* and growth phenotypes of deletion strains. (A) Growth of *P. syringae* DC3000 strains expressing the I-C system and distinct crRNAs. Constructs VI, IV-IX, and VIII target *P. syringae* DC3000 non-essential chromosomal genes, non-targeting crRNA (NT), empty vector (EV). (B) Percentage of survivors with targeted deletions in clusters of non-essential virulence effector genes in *P. syringae* pv. tomato DC3000. Values are averages of three biological replicates where 12 individual colonies were analyzed using site-specific PCR for each, error bars show standard deviations. (C) In vitro growth of cluster VI deletion strains in King's medium B (KB). ΔCEL is the previously published polymutant, while ΔCVI-1 and ΔVI-2 are Cas3-generated mutants. Values shown are the means of 4 biological replicates each, error bars represent standard deviations. (D) In vitro growth of cluster IV, cluster IX deletion strains in KB. ΔCEL is the previously published polymutant, while ΔCIVΔCIX-1 and ΔCIVΔCIX-2 are Cas3-generated mutants. Values shown are the means of 4 biological replicates each, error bars represent standard deviations. (E) In vitro growth of cluster X deletion strains in KB. ΔCEL is the previously published polymutant, while ΔCX-1 and ΔCX-2 are Cas3-generated mutants. Values shown are the means of 4 biological replicates each, error bars represent standard deviations. (F) In vitro growth of cluster VI deletion strains in apoplast mimicking minimal media (MM). ΔCEL is the previously published polymutant, while ΔCVI-1 and ΔCVI-2 are Cas3-generated mutants. Values shown are the means of 4 biological replicates each, error bars represent standard deviations. (G) In vitro growth of cluster IV, cluster IX deletion strains in MM. ΔCEL is the previously published polymutant, while ΔCIVΔCIX-1 and ΔCIVΔCIX-2 are Cas3-generated mutants. Values shown are the means of 4 biological replicates each, error bars represent standard deviations. (H) In vitro growth of cluster X deletion strains in MM. ΔCEL is the previously published polymutant, while ΔCX-1 and ΔCX-2 are Cas3-generated mutants. Values shown are the means of 4 biological replicates each, error bars represent standard deviations.

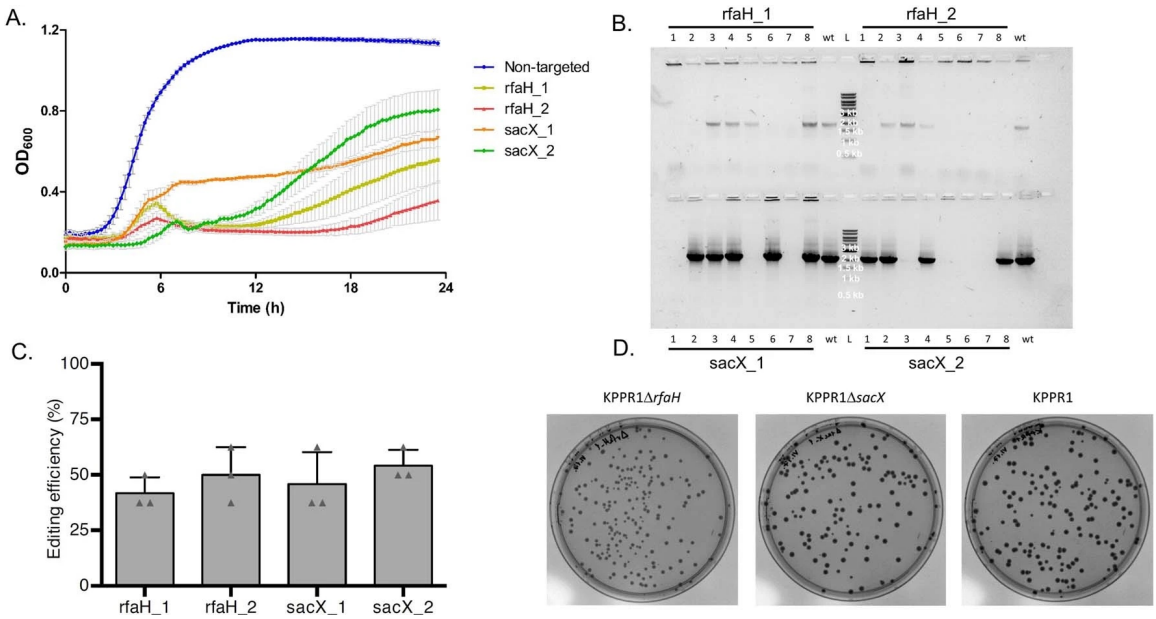


Fig 2.14. CRISPR-Cas3 editing in *Klebsiella pneumoniae*. (A) Growth curves of *K. pneumoniae* strains expressing distinct crRNAs targeting *rfaH* and *sacX* (2 each). Non-targeting crRNA expressing control is marked in blue. Values depicted are averages of 8 biological replicates each. (B) Representative gel electrophoresis of PCR fragments amplified from 8 total surviving colonies each from the 4 crRNA targeting constructs (representing 1 biological replicate of 3 total). Primer pairs amplified regions flanking the targeted position at *rfaH* and *sacX*. Wild-type KPPR1 (wt) colonies were used as controls, L represents 1kb DNA marker ladder. (C) Percentage of survivors with targeted deletions at the targeted genomic positions. Values are averages of three biological replicates where 8 individual colonies were analyzed using site-specific PCR for each, error bars show standard deviations. (D) Colony morphologies of deletion candidate strains of *rfaH* and *sacX* compared to wild-type *K. pneumoniae* KPPR1.

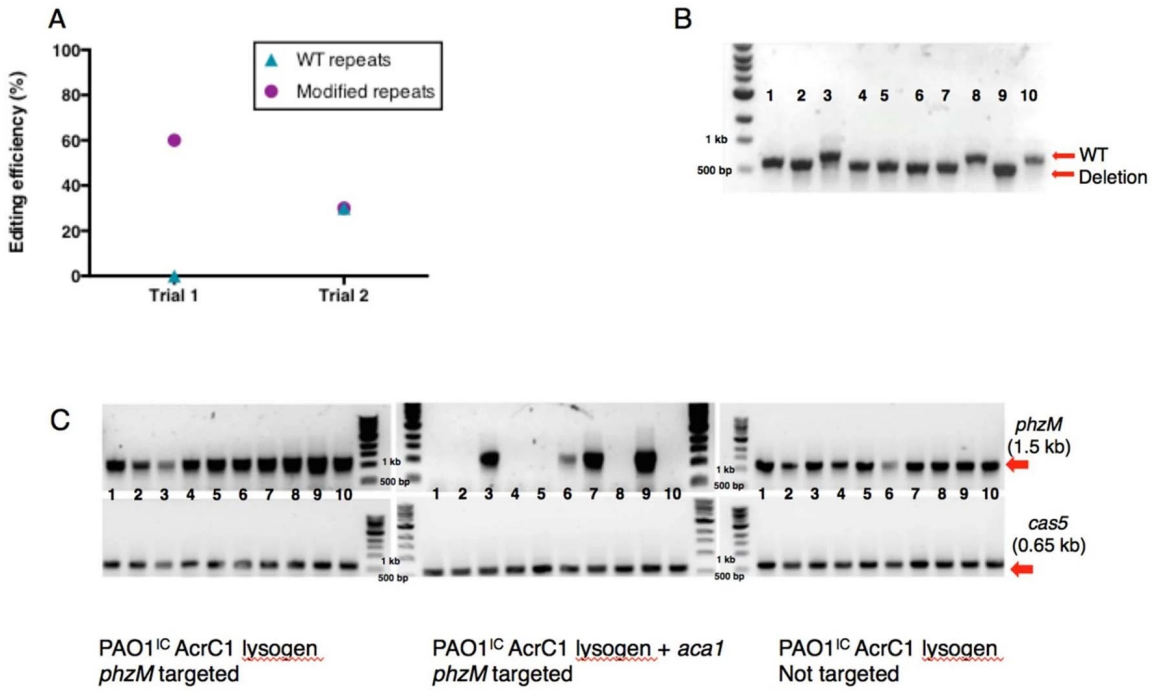


Fig 2.15. Genomic editing in native host of Type I-C CRISPR-Cas system and effect of I-C specific anti-CRISPR protein on the process. (A) Editing efficiencies for the *Pseudomonas aeruginosa* environmental isolate naturally expressing the Type I-C *cas* genes, transformed with a plasmid targeting *phzM* with WT repeats or modified repeats. Each data point represents the fraction of isolates with the deletion out of ten isolates assayed. (B) Genotyping results for the *Pseudomonas aeruginosa* environmental isolate using the 0.17kb HDR template. 10 biological replicates were assayed. Larger band corresponds to the WT sequence, smaller band corresponds to a genome reduced by 0.17kb. (C) Genotyping results of PAO1^{IC} AcrC1 lysogens after self-targeting induction in the presence or absence of *aca1* and a non-targeted control. Ten biological replicates per strain were assayed. gDNA was extracted from each replicate and PCR analysis for the *phzM* gene (targeted gene, top row of gels) or *cas5* gene (non-targeted gene, bottom row) was conducted. Only cells that co-expressed *aca1* with the crRNA showed loss of the *phzM* band, indicating genome editing. All replicates had a *cas5* band, indicating successful gDNA extraction and target specificity for the *phzM* locus.

Bibliography

1. Kira S. Makarova, Yuri I. Wolf, Jaime Iranzo, Sergey A. Shmakov, Omer S. Alkhnbashi, Stan J. J. Brouns, Emmanuelle Charpentier, David Cheng, Daniel H. Haft, Philippe Horvath, Sylvain Moineau, Francisco J. M. Mojica, David Scott, Shiraz A. Shah, Virginijus Siksnys, Michael P. Terns, Česlovas Venclovas, Malcolm F. White, Alexander F. Yakunin, Winston Yan, Feng Zhang, Roger A. Garrett, Rolf Backofen, John van der Oost, Rodolphe Barrangou, and Eugene V. Koonin. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology*, 18(2):67–83, February 2020.

2. Kimberley D. Seed, David W. Lazinski, Stephen B. Calderwood, and Andrew Camilli. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature*, 494(7438):489–491, February 2013.

3. Basem Al-Shayeb, Rohan Sachdeva, Lin-Xing Chen, Fred Ward, Patrick Munk, Audra Devoto, Cindy J. Castelle, Matthew R. Olm, Keith Bouma-Gregson, Yuki Amano, Christine He, Raphaël Méheust, Brandon Brooks, Alex Thomas, Adi Lavy, Paula Matheus-Carnevali, Christine Sun, Daniela S. A. Goltsman, Mikayla A. Borton, Allison Sharrar, Alexander L. Jaffe, Tara C. Nelson, Rose Kantor, Ray Keren, Katherine R. Lane, Ibrahim F. Farag, Shufei Lei, Kari Finstad, Ronald Amundson, Karthik Anantharaman, Jinglie Zhou, Alexander J. Probst, Mary E. Power, Susannah G. Tringe, Wen-Jun Li, Kelly Wrighton, Sue Harrison, Michael Morowitz, David A. Relman, Jennifer A. Doudna, Anne-Catherine Lehours, Lesley Warren, Jamie H. D. Cate, Joanne M. Santini, and Jillian F. Banfield. Clades of huge phages from across Earth's ecosystems. *Nature*, 578(7795):425–431, February 2020.

4. Alan R. Davidson, Wang-Ting Lu, Sabrina Y. Stanley, Jingrui Wang, Marios Mejdani, Chantel N. Trost, Brian T. Hicks, Jooyoung Lee, and Erik J. Sontheimer. Anti-CRISPRs: Protein Inhibitors of CRISPR-Cas Systems. *Annual Review of Biochemistry*, 89(1):309–332, June 2020.

5. K. C. Cady, J. Bondy-Denomy, G. E. Heussler, A. R. Davidson, and G. A. O'Toole. The CRISPR/Cas Adaptive Immune System of Pseudomonas aeruginosa Mediates Resistance to Naturally Occurring and Engineered Phages. *Journal of Bacteriology*, 194(21):5728–5738, November 2012.

6. Joe Bondy-Denomy, April Pawluk, Karen L. Maxwell, and Alan R. Davidson. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*, 493(7432):429–432, January 2013.

7. April Pawluk, Joseph Bondy-Denomy, Vivian H. W. Cheung, Karen L. Maxwell, and Alan R. Davidson. A New Group of Phage Anti-CRISPR Genes Inhibits the Type I-E CRISPR-Cas System of Pseudomonas aeruginosa. *mBio*, 5(2):e00896–14, April 2014.

8. April Pawluk, Megha Shah, Marios Mejdani, Charles Calmettes, Trevor F. Moraes, Alan R. Davidson, and Karen L. Maxwell. Disabling a Type I-E CRISPR-Cas Nuclease with a Bacteriophage-Encoded Anti-CRISPR Protein. *mBio*, 8(6):mBio.01751–17, e01751–17, December 2017.

9. Valerie M. Crowley, Adam Catching, Hannah N. Taylor, Adair L. Borges, Josie Metcalf, Joseph Bondy-Denomy, and Ryan N. Jackson. A Type IV-A CRISPR-Cas System in *Pseudomonas aeruginosa* Mediates RNA-Guided Plasmid Interference *In Vivo*. *The CRISPR Journal*, 2(6):434–440, December 2019.

10. Alex van Belkum, Leah B. Soriaga, Matthew C. LaFave, Srividya Akella, Jean-Baptiste Veyrieras, E. Magda Barbu, Dee Shortridge, Bernadette Blanc, Gregory Hannum, Gilles Zambardi, Kristofer Miller, Mark C. Enright, Nathalie Mugnier, Daniel Brami, Stéphane Schicklin, Martina Felderman, Ariel S. Schwartz, Toby H. Richardson, Todd C. Peterson,

Bolyn Hubby, and Kyle C. Cady. Phylogenetic Distribution of CRISPR-Cas Systems in Antibiotic-Resistant Pseudomonas aeruginosa. *mBio*, 6(6):e01796–15, November 2015.

11. Kira S. Makarova, Yuri I. Wolf, Omer S. Alkhnbashi, Fabrizio Costa, Shiraz A. Shah, Sita J. Saunders, Rodolphe Barrangou, Stan J. J. Brouns, Emmanuelle Charpentier, Daniel H. Haft, Philippe Horvath, Sylvain Moineau, Francisco J. M. Mojica, Rebecca M. Terns, Michael P. Terns, Malcolm F. White, Alexander F. Yakunin, Roger A. Garrett, John van der Oost, Rolf Backofen, and Eugene V. Koonin. An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology*, 13(11):722–736, November 2015.

12. Chitong Rao, Denny Chin, and Alexander W Ensminger. Priming in a permissive type I-C CRISPR–Cas system reveals distinct dynamics of spacer acquisition and loss. page 14.

13. Paola Soto-Perez, Jordan E. Bisanz, Joel D. Berry, Kathy N. Lam, Joseph Bondy-Denomy, and Peter J. Turnbaugh. CRISPR-Cas System of a Prevalent Human Gut Bacterium Reveals Hyper-targeting against Phages in a Human Virome Catalog. *Cell Host & Microbe*, 26(3):325–335.e5, September 2019.

14. Megan L. Hochstrasser, David W. Taylor, Jack E. Kornfeld, Eva Nogales, and Jennifer A. Doudna. DNA Targeting by a Minimal CRISPR RNA-Guided Cascade. *Molecular Cell*, 63(5):840–851, September 2016.

15. Hayun Lee, Yukti Dhingra, and Dipali G Sashital. The Cas4-Cas1-Cas2 complex mediates precise prespacer processing during CRISPR adaptation. *eLife*, 8:e44248, April 2019.

16. Ekaterina Semenova, Maxim Nagornykh, Mikhail Pyatnitskiy, Irena I. Artamonova, and Konstantin Severinov. Analysis of CRISPR system function in plant pathogen *Xanthomonas oryzae*. *FEMS Microbiology Letters*, 296(1):110–116, July 2009.

17. Bálint Csörgő, Lina M. León, Ilea J. Chau-Ly, Alejandro Vasquez-Rifo, Joel D. Berry, Caroline Mahendra, Emily D. Crawford, Jennifer D. Lewis, and Joseph Bondy-Denomy. A compact Cascade–Cas3 system for targeted genome engineering. *Nature Methods*, 17(12):1183–1190, December 2020.

18. Ki Hyun Nam, Charles Haitjema, Xueqi Liu, Fran Ding, Hongwei Wang, Matthew P. DeLisa, and Ailong Ke. Cas5d Protein Processes Pre-crRNA and Assembles into a Cascade-like Interference Complex in Subtype I-C/Dvulg CRISPR-Cas System. *Structure*, 20(9):1574–1584, September 2012.

19. Yoon Koo, Donghyun Ka, Eun-Jin Kim, Nayoung Suh, and Euiyoung Bae. Conservation and Variability in the Structure and Function of the Cas5d Endoribonuclease in the CRISPR-Mediated Microbial Immune System. *Journal of Molecular Biology*, 425(20):3799–3810, October 2013.

20. E. L. Garside, M. J. Schellenberg, E. M. Gesner, J. B. Bonanno, J. M. Sauder, S. K. Burley, S. C. Almo, G. Mehta, and A. M. MacMillan. Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases. *RNA*, 18(11):2020–2028, November 2012.

21. Erianna M. Basgall, Samantha C. Goetting, Megan E. Goeckel, Rachael M. Giersch, Emily Roggenkamp, Madison N. Schrock, Megan Halloran, and Gregory C. Finnigan. Gene drive inhibition by the anti-CRISPR proteins AcrIIA2 and AcrIIA4 in Saccharomyces cerevisiae. *Microbiology*, 164(4):464–474, April 2018.

22. Saikat Chowdhury, Joshua Carter, MaryClare F. Rollins, Sarah M. Golden, Ryan N. Jackson, Connor Hoffmann, Lyn'Al Nosaka, Joseph Bondy-Denomy, Karen L. Maxwell, Alan R. Davidson, Elizabeth R. Fischer, Gabriel C. Lander, and Blake Wiedenheft. Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex. *Cell*, 169(1):47–57.e11, March 2017.

23. Tai Wei Guo, Alberto Bartesaghi, Hui Yang, Veronica Falconieri, Prashant Rao, Alan Merk, Edward T. Eng, Ashleigh M. Raczkowski, Tara Fox, Lesley A. Earl, Dinshaw J. Patel, and Sriram Subramaniam. Cryo-EM Structures Reveal Mechanism and Inhibition of DNA Targeting by a CRISPR-Cas Surveillance Complex. *Cell*, 171(2):414–426.e12, October 2017.

24. Joseph Bondy-Denomy, Bianca Garcia, Scott Strum, Mingjian Du, MaryClare F. Rollins, Yurima Hidalgo-Reyes, Blake Wiedenheft, Karen L. Maxwell, and Alan R. Davidson. Multiple

mechanisms for CRISPR–Cas inhibition by anti-CRISPR proteins. *Nature*, 526(7571):136–139, October 2015.

25. Lucas B. Harrington, Kevin W. Doxzen, Enbo Ma, Jun-Jie Liu, Gavin J. Knott, Alireza Edraki, Bianca Garcia, Nadia Amrani, Janice S. Chen, Joshua C. Cofsky, Philip J. Kranzusch, Erik J. Sontheimer, Alan R. Davidson, Karen L. Maxwell, and Jennifer A. Doudna. A Broad-Spectrum Inhibitor of CRISPR-Cas9. *Cell*, 170(6):1224–1233.e15, September 2017.

26. Januka S. Athukoralage, Christophe Rouillon, Shirley Graham, Sabine Grüschow, and Malcolm F. White. Ring nucleases deactivate type III CRISPR ribonucleases by degrading cyclic oligoadenylate. *Nature*, 562(7726):277–280, October 2018.

27. Nicole D. Marino, Jenny Y. Zhang, Adair L. Borges, Alexander A. Sousa, Lina M. Leon, Benjamin J. Rauch, Russell T. Walton, Joel D. Berry, J. Keith Joung, Benjamin P. Kleinstiver, and Joseph Bondy-Denomy. Discovery of widespread type I and type V CRISPR-Cas inhibitors. *Science*, 362(6411):240, October 2018.

28. Adair L. Borges, Jenny Y. Zhang, MaryClare F. Rollins, Beatriz A. Osuna, Blake Wiedenheft, and Joseph Bondy-Denomy. Bacteriophage Cooperation Suppresses CRISPR-Cas3 and Cas9 Immunity. *Cell*, 174(4):917–925.e10, August 2018.

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

30. Ambarish Biswas, Joshua N. Gagnon, Stan J.J. Brouns, Peter C. Fineran, and Chris M. Brown. CRISPRTarget: Bioinformatic prediction and analysis of crRNA targets. *RNA Biology*, 10(5):817–827, May 2013.

31. Michelle Qiu Carter, Jianshun Chen, and Stephen Lory. The Pseudomonas aeruginosa Pathogenicity Island PAPI-1 Is Transferred via a Novel Type IV Pilus. *Journal of Bacteriology*, 192(13):3249–3258, July 2010.

32. Jens Klockgether, Dieco Würdemann, Oleg Reva, Lutz Wiehlmann, and Burkhard Tümmler. Diversity of the Abundant pKLC102/PAGI-2 Family of Genomic Islands in Pseudomonas aeruginosa. *Journal of Bacteriology*, 189(6):2443–2459, March 2007.

33. Jens Klockgether, Oleg Reva, Karen Larbig, and Burkhard Tümmler. Sequence Analysis of the Mobile Genome Island pKLC102 of Pseudomonas aeruginosa C. *Journal of Bacteriology*, 186(2):518–534, January 2004.

34. X. Qiu, A. U. Gurkar, and S. Lory. Interstrain transfer of the large pathogenicity island (PAPI-1) of Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences*, 103(52):19830–19835, December 2006.

35. Ryan T. Leenay, Kenneth R. Maksimchuk, Rebecca A. Slotkowski, Roma N. Agrawal, Ahmed A. Gomaa, Alexandra E. Briner, Rodolphe Barrangou, and Chase L. Beisel. Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. *Molecular Cell*, 62(1):137–147, April 2016.

36. Benjamin J. Rauch, Melanie R. Silvis, Judd F. Hultquist, Christopher S. Waters, Michael J. McGregor, Nevan J. Krogan, and Joseph Bondy-Denomy. Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. *Cell*, 168(1-2):150–158.e10, January 2017.

37. Lun Cui and David Bikard. Consequences of Cas9 cleavage in the chromosome of *Escherichia coli*. *Nucleic Acids Research*, 44(9):4243–4251, May 2016.

38. Sabrina Y. Stanley, Adair L. Borges, Kuei-Ho Chen, Danielle L. Swaney, Nevan J. Krogan, Joseph Bondy-Denomy, and Alan R. Davidson. Anti-CRISPR-Associated Proteins Are Crucial Repressors of Anti-CRISPR Transcription. *Cell*, 178(6):1452–1464.e13, September 2019.

39. Mariann Landsberger, Sylvain Gandon, Sean Meaden, Clare Rollie, Anne Chevallereau, Hélène Chabas, Angus Buckling, Edze R. Westra, and Stineke van Houte. Anti-CRISPR Phages Cooperate to Overcome CRISPR-Cas Immunity. *Cell*, 174(4):908–916.e12, August 2018.

40. Edze R. Westra, Paul B.G. van Erp, Tim Künne, Shi Pey Wong, Raymond H.J. Staals, Christel L.C. Seegers, Sander Bollen, Matthijs M. Jore, Ekaterina Semenova, Konstantin Severinov, Willem M. de Vos, Remus T. Dame, Renko de Vries, Stan J.J. Brouns, and John van der Oost. CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3. *Molecular Cell*, 46(5):595–605, June 2012.

41. Pedro H. Oliveira, Marie Touchon, Jean Cury, and Eduardo P. C. Rocha. The chromosomal organization of horizontal gene transfer in bacteria. *Nature Communications*, 8(1):841, December 2017.

42. Christopher M. Thomas and Kaare M. Nielsen. Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. *Nature Reviews Microbiology*, 3(9):711–721, September 2005.

43. Guilhem Faure, Sergey A. Shmakov, Winston X. Yan, David R. Cheng, David A. Scott, Joseph E. Peters, Kira S. Makarova, and Eugene V. Koonin. CRISPR–Cas in mobile genetic elements: counter-defence and beyond. *Nature Reviews Microbiology*, 17(8):513–525, August 2019.

44. Rafael Pinilla-Redondo, David Mayo-Muñoz, Jakob Russel, Roger A Garrett, Lennart Randau, Søren J Sørensen, and Shiraz A Shah. Type IV CRISPR–Cas systems are highly diverse and involved in competition between plasmids. *Nucleic Acids Research*, 48(4):2000–2012, February 2020.

45. Caroline Mahendra, Kathleen A. Christie, Beatriz A. Osuna, Rafael Pinilla-Redondo, Benjamin P. Kleinstiver, and Joseph Bondy-Denomy. Broad-spectrum anti-CRISPR proteins facilitate horizontal gene transfer. *Nature Microbiology*, 5(4):620–629, April 2020.

46. Clare Rollie, Anne Chevallereau, Bridget N. J. Watson, Te-yuan Chyou, Olivier Fradet, Isobel McLeod, Peter C. Fineran, Chris M. Brown, Sylvain Gandon, and Edze R. Westra. Targeting of temperate phages drives loss of type I CRISPR–Cas systems. *Nature*, 578(7793):149–153, February 2020.

47. Hao-Ching Wang, Chia-Cheng Chou, Kai-Cheng Hsu, Chi-Hua Lee, and Andrew H.-J. Wang. New paradigm of functional regulation by DNA mimic proteins: Recent updates. *IUBMB Life*, 71(5):539–548, May 2019.

48. Gareth A. Roberts, Augoustinos S. Stephanou, Nisha Kanwar, Angela Dawson, Laurie P. Cooper, Kai Chen, Margaret Nutley, Alan Cooper, Garry W. Blakely, and David T. F. Dryden. Exploring the DNA mimicry of the Ocr protein of phage T7. *Nucleic Acids Research*, 40(16):8129–8143, September 2012.

49. M.D Walkinshaw, P Taylor, S.S Sturrock, C Atanasiu, T Berge, R.M Henderson, J.M Edwardson, and D.T.F Dryden. Structure of Ocr from Bacteriophage T7, a Protein that Mimics B-Form DNA. *Molecular Cell*, 9(1):187–194, January 2002.

50. Artem Isaev, Alena Drobiazko, Nicolas Sierro, Julia Gordeeva, Ido Yosef, Udi Qimron, Nikolai V Ivanov, and Konstantin Severinov. Phage T7 DNA mimic protein Ocr is a potent inhibitor of BREX defence. *Nucleic Acids Research*, 48(10):5397–5406, June 2020.

51. Suji Hong, Donghyun Ka, Seo Jeong Yoon, Nayoung Suh, Migyeong Jeong, Jeong-Yong Suh, and Euiyoung Bae. CRISPR RNA and anti-CRISPR protein binding to the *Xanthomonas albilineans* Csy1-Csy2 heterodimer in the type I-F CRISPR-Cas system. *Journal of Biological Chemistry*, 293(8):2744–2754, February 2018.

52. R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, and P. Horvath. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science*, 315(5819):1709–1712, March 2007.

53. Josiane E. Garneau, Marie-Ève Dupuis, Manuela Villion, Dennis A. Romero, Rodolphe Barrangou, Patrick Boyaval, Christophe Fremaux, Philippe Horvath, Alfonso H. Magadán, and Sylvain Moineau. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 468(7320):67–71, November 2010.

54. Rodolphe Barrangou and Jennifer A Doudna. Applications of CRISPR technologies in research and beyond. *Nature Biotechnology*, 34(9):933–941, September 2016.

55. S. J. J. Brouns, M. M. Jore, M. Lundgren, E. R. Westra, R. J. H. Slijkhuis, A. P. L. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin, and J. van der Oost. Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science*, 321(5891):960–964, August 2008.

56. Tomas Sinkunas, Giedrius Gasiunas, Christophe Fremaux, Rodolphe Barrangou, Philippe Horvath, and Virginijus Siksnys. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system: Cas3 nuclease/helicase. *The EMBO Journal*, 30(7):1335–1342, April 2011.

57. Tomas Sinkunas, Giedrius Gasiunas, Sakharam P Waghmare, Mark J Dickman, Rodolphe Barrangou, Philippe Horvath, and Virginijus Siksnys. In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. *The EMBO Journal*, 32(3):385–394, January 2013.

58. Sabin Mulepati and Scott Bailey. In Vitro Reconstitution of an Escherichia coli RNA-guided Immune System Reveals Unidirectional, ATP-dependent Degradation of DNA Target. *Journal of Biological Chemistry*, 288(31):22184–22192, August 2013.

59. Sy Redding, Samuel H. Sternberg, Myles Marshall, Bryan Gibb, Prashant Bhat, Chantal K. Guegler, Blake Wiedenheft, Jennifer A. Doudna, and Eric C. Greene. Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. *Cell*, 163(4):854–865, November 2015.

60. Reuben B. Vercoe, James T. Chang, Ron L. Dy, Corinda Taylor, Tamzin Gristwood, James S. Clulow, Corinna Richter, Rita Przybilski, Andrew R. Pitman, and Peter C. Fineran. Cytotoxic Chromosomal Targeting by CRISPR/Cas Systems Can Reshape Bacterial Genomes and Expel or Remodel Pathogenicity Islands. *PLoS Genetics*, 9(4):e1003454, April 2013.

61. Ahmed A. Gomaa, Heidi E. Klumpe, Michelle L. Luo, Kurt Selle, Rodolphe Barrangou, and Chase L. Beisel. Programmable Removal of Bacterial Strains by Use of Genome-Targeting CRISPR-Cas Systems. *mBio*, 5(1), February 2014.

62. Ruth Kiro, Dror Shitrit, and Udi Qimron. Efficient engineering of a bacteriophage genome using the type I-E CRISPR-Cas system. *RNA Biology*, 11(1):42–44, January 2014.

63. Yingjun Li, Saifu Pan, Yan Zhang, Min Ren, Mingxia Feng, Nan Peng, Lanming Chen, Yun Xiang Liang, and Qunxin She. Harnessing Type I and Type III CRISPR-Cas systems for genome editing. *Nucleic Acids Research*, 44(4):e34–e34, February 2016.

64. Michael E. Pyne, Mark R. Bruder, Murray Moo-Young, Duane A. Chung, and C. Perry Chou. Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in Clostridium. *Scientific Reports*, 6(1):25666, May 2016.

65. Claudio Hidalgo-Cantabrana, Yong Jun Goh, Meichen Pan, Rosemary Sanozky-Dawes, and Rodolphe Barrangou. Genome editing using the endogenous type I CRISPR-Cas system in *Lactobacillus crispatus*. *Proceedings of the National Academy of Sciences*, 116(32):15774–15783, August 2019.

66. Hannah G. Hampton, Matthew B. McNeil, Thomas J. Paterson, Blair Ney, Neil R. Williamson, Richard A. Easingwood, Mihnea Bostina, George P. C. Salmond, and Peter C. Fineran. CRISPR-Cas gene-editing reveals RsmA and RsmC act through FlhDC to repress the SdhE flavinylation factor and control motility and prodigiosin production in Serratia. *Microbiology*, 162(6):1047–1058, June 2016.

67. Feiyue Cheng, Luyao Gong, Dahe Zhao, Haibo Yang, Jian Zhou, Ming Li, and Hua Xiang. Harnessing the native type I-B CRISPR-Cas for genome editing in a polyploid archaeon. *Journal of Genetics and Genomics*, 44(11):541–548, November 2017.

68. Cassandra Cañez, Kurt Selle, Yong Jun Goh, and Rodolphe Barrangou. Outcomes and characterization of chromosomal self-targeting by native CRISPR-Cas systems in Streptococcus thermophilus. *FEMS Microbiology Letters*, 366(9):fnz105, May 2019.

69. Zeling Xu, Ming Li, Yanran Li, Huiluo Cao, Lu Miao, Zhaochao Xu, Yusuke Higuchi, Seiji Yamasaki, Kunihiko Nishino, Patrick C.Y. Woo, Hua Xiang, and Aixin Yan. Na-

tive CRISPR-Cas-Mediated Genome Editing Enables Dissecting and Sensitizing Clinical Multidrug-Resistant P. aeruginosa. *Cell Reports*, 29(6):1707–1717.e3, November 2019.

70. Yanli Zheng, Jiamei Han, Baiyang Wang, Xiaoyun Hu, Runxia Li, Wei Shen, Xiangdong Ma, Lixin Ma, Li Yi, Shihui Yang, and Wenfang Peng. Characterization and repurposing of the endogenous Type I-F CRISPR–Cas system of Zymomonas mobilis for genome engineering. *Nucleic Acids Research*, 47(21):11461–11475, December 2019.

71. Rotem Edgar and Udi Qimron. The *Escherichia coli* CRISPR System Protects from Lysogenization, Lysogens, and Prophage Induction. *Journal of Bacteriology*, 192(23):6291–6294, December 2010.

72. Adam E. Dolan, Zhonggang Hou, Yibei Xiao, Max J. Gramelspacher, Jaewon Heo, Sara E. Howden, Peter L. Freddolino, Ailong Ke, and Yan Zhang. Introducing a Spectrum of Long-Range Genomic Deletions in Human Embryonic Stem Cells Using Type I CRISPR-Cas. *Molecular Cell*, 74(5):936–950.e5, June 2019.

73. Hiroyuki Morisaka, Kazuto Yoshimi, Yuya Okuzaki, Peter Gee, Yayoi Kunihiro, Ekasit Sonpho, Huaigeng Xu, Noriko Sasakawa, Yuki Naito, Shinichiro Nakada, Takashi Yamamoto, Shigetoshi Sano, Akitsu Hotta, Junji Takeda, and Tomoji Mashimo. CRISPR-Cas3 induces broad and unidirectional genome editing in human cells. *Nature Communications*, 10(1):5302, December 2019.

74. Peter Cameron, Mary M. Coons, Sanne E. Klompe, Alexandra M. Lied, Stephen C. Smith, Bastien Vidal, Paul D. Donohoue, Tomer Rotstein, Bryan W. Kohrs, David B. Nyer, Rachel Kennedy, Lynda M. Banh, Carolyn Williams, Mckenzi S. Toh, Matthew J. Irby, Leslie S. Edwards, Chun-Han Lin, Arthur L. G. Owen, Tim Künne, John van der Oost, Stan J. J. Brouns, Euan M. Slorach, Chris K. Fuller, Scott Gradia, Steven B. Kanner, Andrew P. May, and Samuel H. Sternberg. Harnessing type I CRISPR–Cas systems for genome engineering in human cells. *Nature Biotechnology*, 37(12):1471–1477, December 2019.

75. Adrian Pickar-Oliver, Joshua B. Black, Mae M. Lewis, Kevin J. Mutchnick, Tyler S. Klann, Kylie A. Gilcrest, Madeleine J. Sitton, Christopher E. Nelson, Alejandro Barrera, Luke C.

Bartelt, Timothy E. Reddy, Chase L. Beisel, Rodolphe Barrangou, and Charles A. Gersbach. Targeted transcriptional modulation with type I CRISPR–Cas systems in human cells. *Nature Biotechnology*, 37(12):1493–1501, December 2019.

76. Yuxi Chen, Jiaqi Liu, Shengyao Zhi, Qi Zheng, Wenbin Ma, Junjiu Huang, Yizhi Liu, Dan Liu, Puping Liang, and Zhou Songyang. Repurposing type I–F CRISPR–Cas system as a transcriptional activation tool in human cells. *Nature Communications*, 11(1):3136, December 2020.

77. Joshua K. Young, Stephen L. Gasior, Spencer Jones, Lijuan Wang, Pedro Navarro, Becca Vickroy, and Rodolphe Barrangou. The repurposing of type I-E CRISPR-Cascade for gene activation in plants. *Communications Biology*, 2(1):383, December 2019.

78. Nicole D. Marino, Jenny Y. Zhang, Adair L. Borges, Alexander A. Sousa, Lina M. Leon, Benjamin J. Rauch, Russell T. Walton, Joel D. Berry, J. Keith Joung, Benjamin P. Kleinstiver, and Joseph Bondy-Denomy. Discovery of widespread type I and type V CRISPR-Cas inhibitors. *Science*, 362(6411):240–242, October 2018.

79. Kyoung-Hee Choi, Jared B Gaynor, Kimberly G White, Carolina Lopez, Catharine M Bosio, RoxAnn R Karkhoff-Schweizer, and Herbert P Schweizer. A Tn7-based broad-range bacterial cloning and expression system. *Nature Methods*, 2(6):443–448, June 2005.

80. C K Stover, X Q Pham, A L Erwin, S D Mizoguchi, P Warrener, M J Hickey, F S L Brinkman, W O Hufnagle, D J Kowalik, M Lagrou, R L Garber, L Goltry, E Tolentino, S Westbrock-Wadman, Y Yuan, L L Brody, S N Coulter, K R Folger, A Kas, K Larbig, R Lim, K Smith, D Spencer, and M V Olson. PAO1, an opportunistic pathogen. 406:6, 2000.

81. Kyoung-Hee Choi and Herbert P Schweizer. mini-Tn7 insertion in bacteria with single attTn7 sites: example Pseudomonas aeruginosa. *Nature Protocols*, 1(1):153–161, June 2006.

82. C. R. Buell, V. Joardar, M. Lindeberg, J. Selengut, I. T. Paulsen, M. L. Gwinn, R. J. Dodson, R. T. Deboy, A. S. Durkin, J. F. Kolonay, R. Madupu, S. Daugherty, L. Brinkac,

M. J. Beanan, D. H. Haft, W. C. Nelson, T. Davidsen, N. Zafar, L. Zhou, J. Liu, Q. Yuan, H. Khouri, N. Fedorova, B. Tran, D. Russell, K. Berry, T. Utterback, S. E. Van Aken, T. V. Feldblyum, M. D'Ascenzo, W.-L. Deng, A. R. Ramos, J. R. Alfano, S. Cartinhour, A. K. Chatterjee, T. P. Delaney, S. G. Lazarowitz, G. B. Martin, D. J. Schneider, X. Tang, C. L. Bender, O. White, C. M. Fraser, and A. Collmer. The complete genome sequence of the Arabidopsis and tomato pathogen Pseudomonas syringae pv. tomato DC3000. *Proceedings of the National Academy of Sciences*, 100(18):10181–10186, September 2003.

83. F. R. Blattner. The Complete Genome Sequence of Escherichia coli K-12. *Science*, 277(5331):1453–1462, September 1997.

84. C. A. Broberg, W. Wu, J. D. Cavalcoli, V. L. Miller, and M. A. Bachman. Complete Genome Sequence of Klebsiella pneumoniae Strain ATCC 43816 KPPR1, a Rifampin-Resistant Mutant Commonly Used in Animal, Genetic, and Molecular Biology Studies. *Genome Announcements*, 2(5):e00924–14, 2/5/e00924–14, September 2014.

85. Dongru Qiu, F. Heath Damron, Takehiko Mima, Herbert P. Schweizer, and Hongwei D. Yu. P $_{\text{bad}}$ -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):7422–7426, December 2008.

86. Daniel G Gibson, Lei Young, Ray-Yuan Chuang, J Craig Venter, Clyde A Hutchison, and Hamilton O Smith. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5):343–345, May 2009.

87. Jeffrey Meisner and Joanna B. Goldberg. The Escherichia coli *rhaSR-PrhaBAD* Inducible Promoter System Allows Tightly Controlled Gene Expression over a Wide Range in Pseudomonas aeruginosa. *Applied and Environmental Microbiology*, 82(22):6715–6727, November 2016.

88. Ákos Nyerges, Bálint Csörgő, Gábor Draskovits, Bálint Kintses, Petra Szili, Györgyi Ferenc, Tamás Révész, Eszter Ari, István Nagy, Balázs Bálint, Bálint Márk Vásárhelyi, Péter Bihari,

Mónika Számel, Dávid Balogh, Henrietta Papp, Dorottya Kalapis, Balázs Papp, and Csaba Pál. Directed evolution of multiple genomic loci allows the prediction of antibiotic resistance. *Proceedings of the National Academy of Sciences*, 115(25):E5726–E5735, June 2018.

89. Nabil-Fareed Alikhan, Nicola K Petty, Nouri L Ben Zakour, and Scott A Beatson. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics*, 12(1):402, December 2011.

90. Megan L. Hochstrasser, David W. Taylor, Jack E. Kornfeld, Eva Nogales, and Jennifer A. Doudna. DNA Targeting by a Minimal CRISPR RNA-Guided Cascade. *Molecular Cell*, 63(5):840–851, September 2016.

91. Keith H. Turner, Aimee K. Wessel, Gregory C. Palmer, Justine L. Murray, and Marvin Whiteley. Essential genome of *Pseudomonas aeruginosa* in cystic fibrosis sputum. *Proceedings of the National Academy of Sciences*, 112(13):4110–4115, March 2015.

92. Kurt Selle, Todd R. Klaenhammer, and Rodolphe Barrangou. CRISPR-based screening of genomic island excision events in bacteria. *Proceedings of the National Academy of Sciences*, 112(26):8076–8081, June 2015.

93. Romain Chayot, Benjamin Montagne, Didier Mazel, and Miria Ricchetti. An end-joining repair mechanism in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 107(5):2141–2146, February 2010.

94. Magdalen Lindeberg, Sébastien Cunnac, and Alan Collmer. Pseudomonas syringae type III effector repertoires: last words in endless arguments. *Trends in Microbiology*, 20(4):199–208, April 2012.

95. Brian H. Kvitko, Duck Hwan Park, André C. Velásquez, Chia-Fong Wei, Alistair B. Russell, Gregory B. Martin, David J. Schneider, and Alan Collmer. Deletions in the Repertoire of Pseudomonas syringae pv. tomato DC3000 Type III Secretion Effector Genes Reveal Functional Overlap among Effectors. *PLoS Pathogens*, 5(4):e1000388, April 2009.

96. Brian J. Caliando and Christopher A. Voigt. Targeted DNA degradation using a CRISPR device stably carried in the host genome. *Nature Communications*, 6(1):6989, November 2015.

97. Michael A. Bachman, Paul Breen, Valerie Deornellas, Qiao Mu, Lili Zhao, Weisheng Wu, James D. Cavalcoli, and Harry L. T. Mobley. Genome-Wide Identification of Klebsiella pneumoniae Fitness Genes during Lung Infection. *mBio*, 6(3):e00775–15, June 2015.

98. K. C. Cady, J. Bondy-Denomy, G. E. Heussler, A. R. Davidson, and G. A. O'Toole. The CRISPR/Cas Adaptive Immune System of Pseudomonas aeruginosa Mediates Resistance to Naturally Occurring and Engineered Phages. *Journal of Bacteriology*, 194(21):5728–5738, November 2012.

99. Joe Bondy-Denomy, April Pawluk, Karen L. Maxwell, and Alan R. Davidson. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*, 493(7432):429–432, January 2013.

100. Anh D. Ha and Dee R. Denver. Comparative Genomic Analysis of 130 Bacteriophages Infecting Bacteria in the Genus Pseudomonas. *Frontiers in Microbiology*, 9:1456, July 2018.

101. G. Posfai. Emergent Properties of Reduced-Genome Escherichia coli. *Science*, 312(5776):1044–1046, May 2006.

102. Bálint Csörgő, Ákos Nyerges, György Pósfai, and Tamás Fehér. System-level genome editing in microbes. *Current Opinion in Microbiology*, 33:113–122, October 2016.

103. François Képès, Brian C. Jester, Thibaut Lepage, Nafiseh Rafiei, Bianca Rosu, and Ivan Junier. The layout of a bacterial genome. *FEBS Letters*, 586(15):2043–2048, July 2012.

104. Lun Cui and David Bikard. Consequences of Cas9 cleavage in the chromosome of *Escherichia coli*. *Nucleic Acids Research*, 44(9):4243–4251, May 2016.

105. Richard Bowater and Aidan J Doherty. Making Ends Meet: Repairing Breaks in Bacterial DNA by Non-Homologous End-Joining. *PLoS Genetics*, 2(2):e8, February 2006.

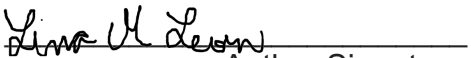
106. Rubina Tuladhar, Yunku Yeu, John Tyler Piazza, Zhen Tan, Jean Rene Clemenceau, Xiaofeng Wu, Quinn Barrett, Jeremiah Herbert, David H. Mathews, James Kim, Tae Hyun Hwang, and Lawrence Lum. CRISPR-Cas9-based mutagenesis frequently provokes on-target mRNA misregulation. *Nature Communications*, 10(1):4056, December 2019.

107. Arne H. Smits, Frederik Ziebell, Gerard Joberty, Nico Zinn, William F. Mueller, Sandra Clauder-Münster, Dirk Eberhard, Maria Fälth Savitski, Paola Grandi, Petra Jakob, Anne-Marie Michon, Hanice Sun, Karen Tessmer, Tilmann Bürckstümmer, Marcus Bantscheff, Lars M. Steinmetz, Gerard Drewes, and Wolfgang Huber. Biological plasticity rescues target activity in CRISPR knock outs. *Nature Methods*, 16(11):1087–1093, November 2019.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.



Author Signature

6/8/2021

Date