Applications of a High-Throughput Screen Detecting CRISPR-Cas Spacer Acquisition

Christopher Cannon, MA

School of Life Sciences

University of Nottingham

PhD Thesis - November 2024

Abstract

The arms race between prokaryotes and their foreign genetic elements has driven the evolution of a diverse and enigmatic array of immune systems. CRISPR-Cas systems uniquely provide adaptive immunity against these foreign elements. Immunity is achieved through the acquisition of DNA fragments derived from invaders' genomes, catalysed by the Cas1-Cas2 integrase complex. These fragments are stored as spacers within CRISPR arrays and are transcribed to specifically direct the effector machinery against their complementary sequences.

Mutations in the Cas1 protein of the *E. coli* Type I-E CRISPR-Cas system have been identified that exhibit an increased rate of spacer acquisition. It has not been practical to perform systematic large-scale screening of mutagenic libraries using previous assays. I used a papillation reporter assay system to identify novel hyperactive mutations which exhibit up to five-fold increases in rates of spacer acquisition. These mutants also induce an elevated SOS response, suggesting increased integrase activity has the potential to confer a negative fitness cost to the host cell.

The screening of metagenomic libraries to identify anti-CRISPRs targeting the integrase machinery identified homologs of genes ancillary to the process of spacer acquisition as potential inhibitors. This suggests a mechanism whereby genetic elements may inhibit CRISPR systems without directly targeting the *cas* genes.

Finally, I constructed and tested several versions of a genetic circuit for the high-throughput detection of spacer acquisition, which may be universally applied to CRISPR systems.

Contents

Figures

Tables

Abbreviations

a.a.	Amino acid(s)
Abi	Abortive Infection
AbR	Antibiotic Resistance
Aca	Anti-CRISPR Associated
Acr	Anti-CRISPR
Amp	Ampicillin
AmpR	Ampicillin resistance
Ara	Arabinose
Bp	Base pair(s)
Cas	CRISPR Associated
CBASS	Cyclic-Oligonucleotide-Based Anti-Phage Signalling Systems
chi	Crossover hot-spot instigator
Cm	Chloramphenicol
CmR	Chloramphenicol resistance
Cracr	Cas/CreT-repressing RNA anti-CRISPRs
CreT	Cascade-repressed toxin
CRISPR	Clustered and Regularly Interspaced short Palindromic Repeats
crRNA	CRISPR RNA
CymR	Cysteine metabolism repressor
DNA	Deoxyribonucleic acid
DSB	Double stranded break
dsDNA	Double-stranded DNA
frt	flippase recognition target
FLP	Flippase Recombinase
g	Gram

	Gravitational Force
GFP	Green Fluorescent Protein
IHF	Integration Host Factor
Kan	Kanamycin
KanR	Kanamycin resistance
kb	Kilobase
kDa	Kilodalton
L	Litre
LB	Lysogeny Broth
M	Molar
mRNA	Messenger RNA
NA	Nalidixic Acid
NEB	New England Biolabs
O.D.	Optical Density
ORF	Open Reading Frame
Ori	Origin of replication
OriT	Origin of Transfer
PAM	Protospacer Adjacent Motif
PCR	Polymerase chain reaction
pre-crRNA	Precursor-CRISPR RNA
Racr	RNA-based anti-CRISPRs
RBS	Ribosome Binding Site
RM	Restriction Modification
RNA	Ribonucleic acid
RNAP	RNA polymerase
RPM	Revolutions per minute
RT	Room Temperature
ssDNA	Single-stranded DNA
Spect	Spectinomycin

SpectR	Spectinomycin Resistance
Tet	Tetracycline
TetR	Tetracycline Resistance
VanR	Vanillic acid Repressor
WT	Wild type
X-Gal	5-bromo-4-chloro-3-indolyl-β-D- galactopyranoside

Chapter 1

Introduction

Prokaryotes are hosts for an immensely diverse array of bacteriophages and other mobile genetic elements (MGEs) [1]. These viruses are speculated to have emerged billions of years ago, preceding even the appearance of the first prototypical cells [2]. Therefore, prokaryotes and their viruses may have been engaged in an arms race since the dawn of cellular life. Following billions of years of selective pressure, prokaryotes and their viruses have evolved many enigmatic immune systems and evasion strategies [3][4].

Whilst often conferring considerable fitness disadvantages to their hosts, the introduction of exogenous genetic material by these MGEs can also provide prokaryotes access to a broader gene pool. This may be sampled via horizontal gene transfer (HGT) to facilitate the introduction of advantageous genetic material into their genome. As such, MGE-host interactions can provide fitness advantages to the host cell, such as allowing for the expansion of the host's niche through the introduction of antibiotic resistance genes (ARGs) or pathogenicity islands [5][6]. Prokaryotes must maintain a balance as they attempt to defend themselves from those MGEs that confer harm whilst maintaining access to the fitness advantages of the wider gene pool.

1.1 Diversity of Prokaryotic Defences Systems

Prokaryotes have evolved an immensely diverse arsenal of immune systems that attempt to tightly regulate and control the genetic content of the individual cell and broader population. Restriction-Modification (RM) systems were the first to be described in the 1950s [7][8]. Over the ensuing decades, several further systems were identified, including Abortive Infection (Abi) [9] and Toxin-Antitoxin (TA) systems [10][11]. CRISPR-Cas systems were identified as anti-phage immune systems in 2007 [12].

Following the incremental identification of novel prokaryotic immune systems, the identification that these systems cluster into 'defence islands' within the genome

facilitated a rapid expansion of known systems [13]. This clustering may facilitate functional cooperation and joint horizontal transfer between hosts. This behaviour enabled the identification of further systems through the investigation of genes enriched locally to other defence systems [14][15][16]. Guilt-by-association studies have expanded the pool of known prokaryotic immune systems to >150 [Figure 1.1][3].

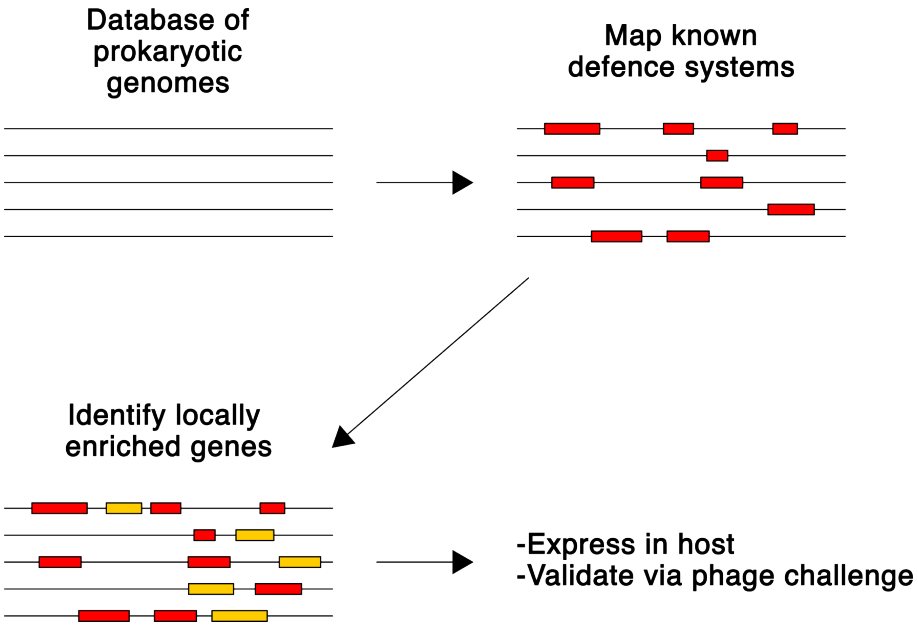

Figure 1.1: Pipeline for identifying candidate defence systems via guilt-by-association.
Genome databases are mined to identify genes physically enriched proximal to known defence systems such as RM, Abi, Argonaute and CRISPR-Cas (red). Predicted systems (yellow) can then be cloned into bacteria and experimentally validated via phage challenge.

Mechanisms of prokaryotic defence from MGEs can be broadly divided into three categories depending upon the precision and manner with which the respective system acts: Mechanisms that act exogenous to the cell to prevent initial infection or invasion. For a review, see Labrie *et al.* [17]. Innate mechanisms within the cell recognise and respond to predetermined non-self and damage associated motifs or signals [18]. Finally, the CRISPR-Cas system can be uniquely categorised as separate due to its ability to generate a specific, adaptive immune response against a broad range of invasive MGEs.

1.2 Restriction Modification

Restriction modification systems are a highly abundant prokaryotic immune system found in ≈90% of genomes [19]. Over 5,000 RMs have been identified, with individual genomes found to encode as many as 24 systems [19][20]. RM systems cleave invading DNA recognised as non-self [Figure 1.2].

RMs typically possess two opposing but synergistic enzymatic activities: the modification of host DNA and restriction endonuclease (REase) activity against unmodified DNA. Modification of host DNA is typically carried out by a methyltransferase (MTase), which methylates specific residues within a recognition site [21]. The absence of DNA modification at a recognition site causes invading DNA to be recognised as 'non-self', resulting in cleavage by a host REase [21]. DNA modification is necessary to prevent self-targeting by a REase against the host chromosome following the recognition of short target sequences [21].

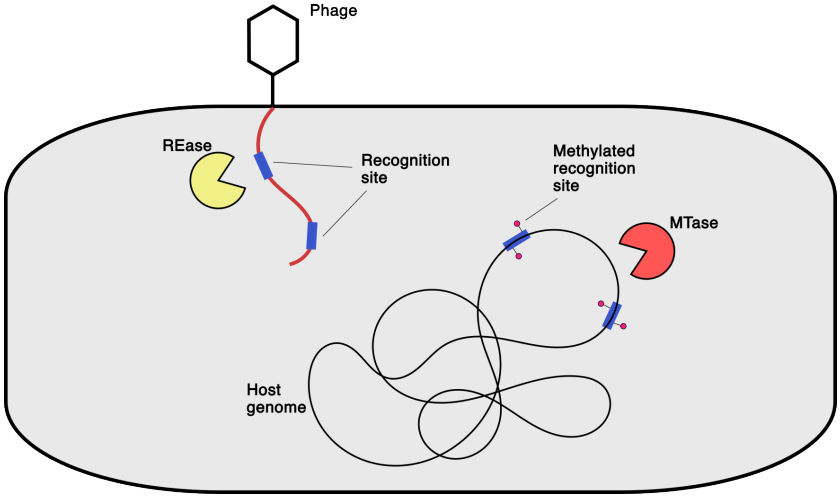

Figure 1.2: Overview of the role of restriction-modification systems in host defence. The host genome is modified to inhibit targeting by restriction endonucleases. Foreign genetic material within the cell lacking such modification is subjected to degradation by REase activity.

RMs can provide significant protection against phage infection [22]. Subsequently, bacteriophages have evolved anti-restriction proteins such as Phage T3 and T7 encoded *ocr*, an inhibitor of the EcoK and EcoB RM systems [23]. Ocr sterically

blocks DNA recognition sites, inhibiting surveillance by RM complexes [24]. Alternatively, phages may acquire host modification to escape modification [23]. In turn, *Streptococcus coelicolor* encodes *pglX*, which modifies phage DNA such that it becomes susceptible to its next host [25]. An evolutionary arms race exists between prokaryotes encoding RM systems and the parasitic MGEs they target [22].

1.3 Abortive Infection

Abortive infection (Abi) is a collective term for phage resistance strategies which involves the cell committing suicide or inducing dormancy before an infecting phage can complete a replication cycle. The traditionally limited host range of phages, combined with the typically isogenic nature of bacterial populations, means this is a kin selection strategy [26]. Abi systems are typically active at later stages of the infection cycle when initial defences have failed to halt infection [27].

Abi systems are typically composed of two modules: a phage infection sensor and a cell death trigger [Figure 1.3]. A diverse range of signals of phage infection can trigger Abi, including phage structural proteins [28], phage DNA transcription [29], or intermediates of phage genome replication [30]. Following the identification of a phage infection, a cell-killing module is activated. Mechanisms of cell-killing are similarly diverse, including depolarisation of the inner membrane [31], non-specific DNA degradation [32], or transfer RNA (tRNA) degradation [33]. Examples of phage signals and subsequent methods of cell death are inexhaustive.

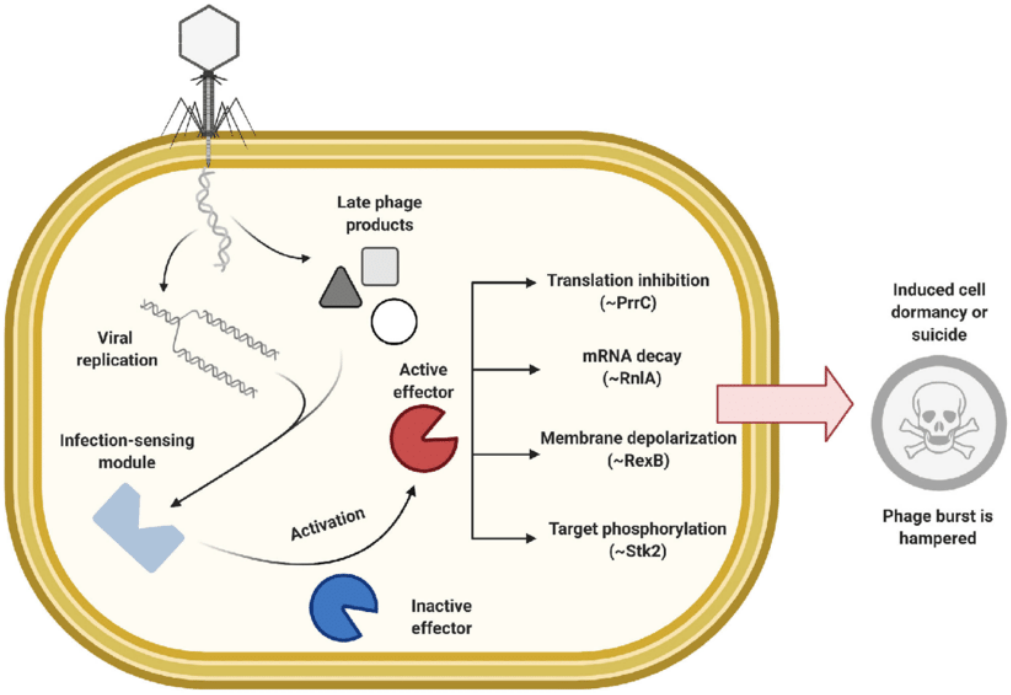

Figure 1.3: Overview of the role of Abi systems in host defence.
A sensor module identifies a signal of phage infection and triggers an effector module to induce cell death or dormancy, inhibiting further phage replication and propagation. Incoming signals and outing effector activities are diverse. Taken from (Isaev *et al.*, 2021).

The mechanism of the *E. coli* Rex Abi system was one of the first to be elucidated. A lambda prophage encodes the genes *rexA* and *rexB* [34]. RexA, the sensor module, detects protein-DNA complexes formed during phage replication or recombination, activating RexB [30]. RexB is the cell-killing module, forming an ion channel in the inner membrane [35]. The resultant loss of membrane potential results in a drop in cellular ATP concentration, aborting phage infection and potentially causing cell death [35]. *E. coli*'s Rex Abi system is inhibited by T4 phage, which encodes *RIIA* and *RIIB*, their mechanism of inhibition is unknown [36].

1.3.1 Toxin-Antitoxin (TA) systems

TA systems are typically composed of a pair of genes. The first gene encodes a toxin targeted against an essential cellular process, whilst the second gene encodes an antitoxin, which confers immunity to this toxin [37]. As the toxin is more

stable than the antitoxin, the antitoxin must be constantly produced. Cellular stress inhibiting antitoxin production prevents toxin suppression [38]. TA systems contribute to many cellular processes, including plasmid addiction [39], antibiotic resistance [40], and immunity against MGEs [41].

ToxIN is an example of a TA system which confers immunity against phages [41]. The *toxN* gene encodes an endoribonuclease, presumed to target both phage and cellular RNA non-specifically. ToxN is bound to and inhibited by the non-coding RNA ToxI [42]. Homologs of *toxN* systems are widespread across gram-positive and gram-negative bacteria [41]. Phages may inhibit the system by expressing their own non-coding RNA mimics of *toxI* [43].

1.3.2 CBASS

Cyclic Oligonucleotide-Based Antiphage Signalling Systems (CBASS) refer to a large family of recently described Abi systems [44]. CBASS is characterised by the use of a diverse array of cyclically linked nucleotides as secondary messengers to convey signalling from a sensor module to a cell-killing module [45]. Unlike other Abi systems, these modules do not directly interact.

Vibrio cholera El Tor encodes a CBASS system which provides protection against a large number of phages [44]. The system produces cyclic GMP-AMP (cGAMP) in response to a phage trigger [46]. This binds to and activates the phospholipase CapV, which breaks down the inner membrane to induce lysis [44]. T4 phage encodes *acb1*, which facilitates the evasion of CBASS through the hydrolysis of cyclic secondary messengers [47].

1.3.3 CRISPR-associated Abi

CRISPR-Cas systems typically target and cleave invading DNA [48]. However, Type III systems, which bind RNA, have been shown to induce Abi [49][29][50]. Actively transcribed RNA is recognised by a cognate CRISPR RNA (crRNA) in

complex with Csm3 or Cmr4. Alongside inducing DNA and RNA cleavage, recognition induces the synthesis of the secondary messenger molecule cyclic oligoadenylate (cOA) by Cas10. cOA binds to and triggers Csm6, a non-specific RNase which induces cell dormancy or death [29][50].

1.4 Argonaute

Argonaute proteins were first identified in *Arabidopsis thaliana* [51]. However, before they adopted a role in complex eukaryotic regulatory pathways, they initially evolved as prokaryotic host defence systems [52][53].

Eukaryotic argonaute proteins are involved in a range of cellular processes, including the regulation of post-translational gene expression and the formation of heterochromatin [54][55]. The Dicer endoribonuclease produces 20-25 bp RNA guides, targeting eukaryotic argonaute proteins (eAgo) against cognate mRNA sequences [56][54]. eAgos may then sequester these sequences, or cleave them via PIWI endoribonuclease domains [57]. Phylogenetic reconstructions of eAgos suggest a functional RNAi pathway was present in the last eukaryotic common ancestor (LECA), where it likely functioned as an immune system against MGEs [52].

pAgos were identified from within prokaryotic defence islands [13]. Unlike eAgos, pAgos have undergone extensive horizontal gene transfer and exhibit much greater diversity than eAgos [53]. The phylogeny of pAgos can be divided into two trees, dependent on domain architecture and operon organisation, these are short Agos and long Agos. Unlike eAgos targeting RNA, they can also target single- and double-stranded DNA [53][58]. Short Agos, and some clades within long Agos, contain catalytically inactive Ago proteins. However, these frequently associate with various putative nucleases, suggesting modularity in the organisation of pAgo-derived defence [53]. A collection of genes of unknown function have similarly been shown to associate with pAgos, suggesting modularity may extend beyond

nuclease recruitment to the recruitment of ancillary functions [53].

The *Geobacter sulfurreducens* Sir2-APAZ-pAgo (SPARSA) system exemplifies the recruitment of atypical argonaute functions in pAgos [59]. The pAgo binds to a guide RNA, which, upon binding cognate viral DNA, activates the Sir2 domain NAD+ hydrolase (NADase). This induces an Abi response as NAD degradation inhibits cellular functions, inhibiting viral replication and potentially leading to cell death.

1.5 Wadjet

Wadjet, a recently identified prokaryotic defence system, exemplifies the effectiveness of guilt-by-association studies in identifying such systems [15]. Wadjet is a bacterial structural maintenance of chromosomes (SMC) complex which confers defence against plasmid [15]. Uniquely amongst prokaryotic immune systems, it monitors topology to target DNA for cleavage [60].

The Wadjet complex conducts DNA loop extrusion to determine topology [60]. Extrusion is conducted by JetABC, homologs of the MukBEF SMC family involved in chromosome condensation and segregation [15]. Recognition of a closed, circular plasmid activates the JetD nuclease, inducing DNA cleavage [61]. Plasmids of sizes 50–100 kbp are cleaved, preventing their successful transformation into prokaryotic cells [15][62].

1.6 CRISPR-Cas

All prokaryotic immune systems discussed previously follow the same broad format, in which a sensor domain recognises a narrow and specific phage- or damage-associated motif. This signal is then transduced to an effector domain, which mounts an immune response. CRISPR-Cas systems stand apart from these; they are able to generate a specific response against a breadth of potential signals

through an adaptive immune response.

1.6.1 Discovery of CRISPR-Cas Adaptive Immunity

The identification of an array of 29 bp repeats interspersed with 32 bp spacers, found adjacent to the *iap* gene in *E. coli*, was the first step towards the discovery of CRISPR-Cas [63]. Subsequent studies identified similar structures in *Shigella dysenteriae* and *Salmonella typhimurium* [64]. These discoveries preceded the broad identification of such loci across prokaryotes [65]. It was only following the advancement and proliferation of widespread sequencing that MGEs were identified as the principal source of these spacers and *cas* genes were identified as associating with these loci [66][67].

Bioinformatic analysis of the *cas* genes identified nuclease domains within Cas3, alongside identifying Cas4 as a RecB family exonuclease [68]. This suggested CRISPR-Cas was involved in DNA repair. These findings were unified with the identification of spacers as deriving from MGEs, and CRISPR-Cas systems were then hypothesised to form a bacterial immune system against the MGEs from which they derived spacers [69]. This was demonstrated experimentally in *Streptococcus thermophilus*, in which the acquisition of spacers into a CRISPR array containing phage DNA conferred resistance to said phage [12].

Insights into the mechanism underpinning CRISPR-Cas induced phage defence commenced with the co-purification of an effector complex with a mature CRISPR RNA (crRNA) [70]. The mature crRNA was composed of the transcribed repeat and flanking sequences; the 5' contained the final 8 bp of the repeat, whilst the 3' contained a repeat sequence. Maturation of pre-crRNA was conducted by the Cas proteins, this was shown to be essential for phage defence [70] Analysis of targets of CRISPR Cas identified DNA as the target [71]. Together, this demonstrated that CRISPR-Cas was an adaptive immune system which mounted a specific immune response against DNA, fundamentally differing from the RNAi style system to which

it was originally compared [69]

1.7 Overview of the CRISPR-Cas Immune Response

The CRISPR-Cas immune response comprises three stages: adaptation, crRNA processing and interference [Figure 1.4]. Below, I outline each stage with a focus on the *E. coli* Type I-E system.

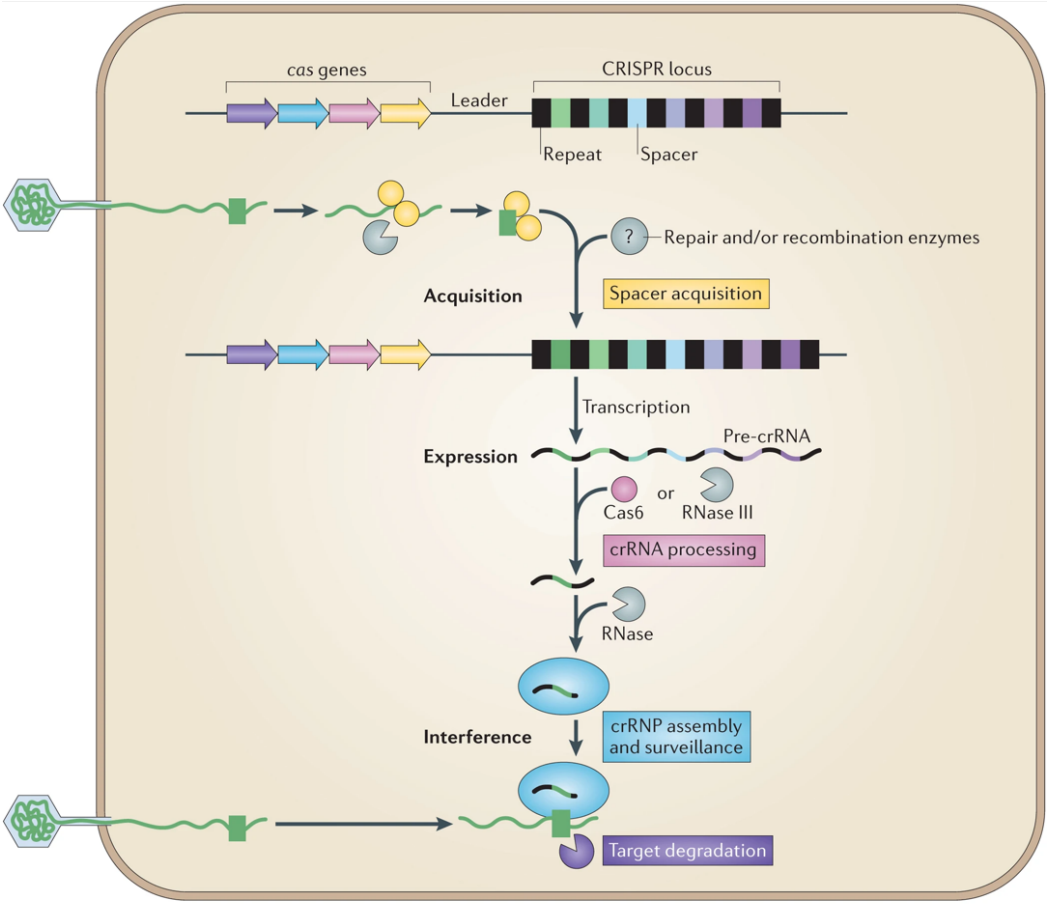

Figure 1.4: Overview of the CRISPR-Cas immune system.
The CRISPR array contains a set of repeats (black) separated by spacers (multi-coloured). The CRISPR array is further characterised by the presence of a leader sequence involved in spacer integration and crRNA expression. Furthermore, they are often found associated with *cas* genes (multi-coloured). Adaptation involves the acquisition of a novel spacer sequence into the CRISPR array. The CRISPR array is transcribed, producing a long-pre-crRNA, which is processed into a mature crRNA. This nucleates the formation of the interference complex (blue). This complex scans the cell for cognate sequences against which nuclease activity can be targeted. Depending on the system, nuclease activity is either intrinsic to the complex or recruited following cognate sequence identification (purple). Taken from (van der Oost *et al.*, 2014).

1.7.1 Adaptation

The first stage of the CRISPR-Cas immune response is adaptation. Adaptation is the process whereby a segment of DNA is integrated into the CRISPR array by the adaptation machinery. In the *E. coli* Type I-E systems this is conducted by Cas1-Cas2 in a hetero-hexameric integrase complex [72]. This complex possesses a $(Cas1_2\text{-}Cas2)_2$ stoichiometry, in which a pair of Cas1 dimers bind opposing sides of a central Cas2 dimer [72]. Whilst this complex is sufficient for spacer acquisition in *E. coli* Type I-E, many systems recruit further Cas genes to conduct spacer acquisition [73][48]. Adaptation may be naïve or primed. Naïve adaptation is the acquisition of spacers from an element novel to the CRISPR array. In contrast, primed acquisition involves spacer acquisition being strongly promoted by the presence of a spacer in the CRISPR array targeting a sequence [74]. Despite naïve and primed acquisition having different methods of prespacer generation, they share a pathway for prespacer processing and integration [75].

1.7.1.1 Prespacer binding and trimming

A pair of asymmetric Cas1 homodimers (Cas1a-b and Cas1a'-b') bind opposingly upon a symmetric Cas2 homodimer, forming a 'wings-up butterfly' configuration as the central Cas2 dimer is flanked by respective Cas1 dimers [Figure 1.5] [76]. Cas1b and Cas1b' deviate from a symmetric protein-protein interface with the Cas2 dimer due to the C-terminus of Cas1a, but broadly confer symmetry in their binding [72]. Cas1a and Cas1a' contact their respective Cas1 monomers but do not contact Cas2.

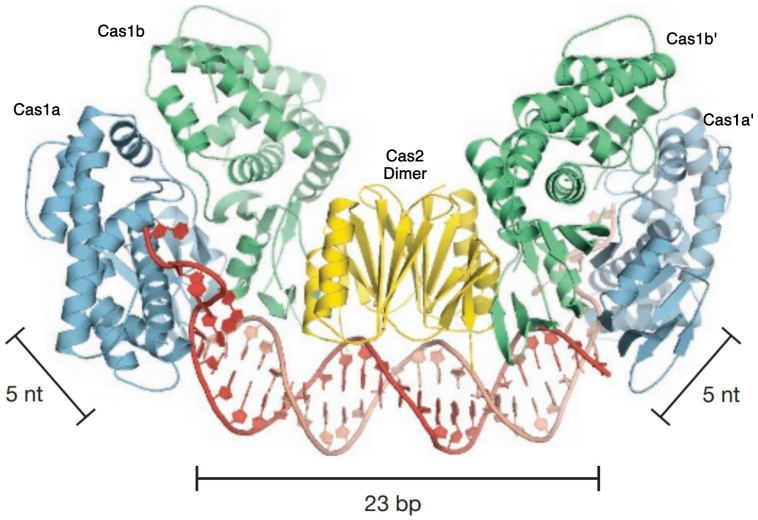

Figure 1.5: Adaptation complex bound to protospacer.
Crystal structure of the Cas1-Cas2 adaptation complex bound to a protospacer. The central 23 bp duplex of the protospacer (red) lies flat across the bottom of the complex, with 3' 5 bp overhangs protruding away in interactions with the C terminal domain of Cas1a/a' (light blue). Adapted from (Nuñez *et al*., 2015).

Conformational changes associated with the binding of a protospacer are associated with the adoption of a 'wings-down butterfly' configuration, as each Cas2 monomer undergoes reciprocal clockwise/anticlockwise rotations relative to one another [76]. The resultant flattened binding interface is thought to stabilise interaction with the incoming protospacer. Furthermore, rearrangements result in the formation of an optimal catalytic pocket in Cas1a to permit site-specific cleavage. Finally, a pair of Tyr22 residues within Cas1a and 1a' form brackets at either end of the 23 bp duplex, acting as a ruler to determine the length of the incoming spacer [76].

The optimal protospacer within the *E. coli* Type I-E system contains a central 23 bp dsDNA duplex flanked by 5 bp 3' overhangs [Figure] [76]. As DNA duplex size is capped by the Cas1a-1a' tyrosine bracket, 5 bp 3' overhangs must protrude away from the duplex. These overhangs are flipped away from the duplex to be threaded through the Cas1a and 1a' C-terminal tails [76]. Prespacer overhangs larger than this are cleaved to size in a site-specific manner.

Protospacer substrate in the form of dsDNA can be captured by the integrase complex [76]. Cas3 and RecBCD degradation products are ssDNA, necessitating reannealing to form substrates. It is unclear if this is spontaneous, or aided by Cas1-2 or another host factor.

1.7.1.2 Spacer integration

The Cas1-Cas2 integrase complex catalyses a pair of trans-esterification reactions during the full site integration of a spacer into a CRISPR array [Figure 1.6] [77]. As the two active sites of the integrase complex cannot access DNA at the same time, half-site integration of the captured spacer occurs sequentially [78]. The first half-site integration occurs between the leader-distal end of the first repeat and the PAM end of the spacer [77]. The second half-site integration then occurs between the leader proximal end of the repeat and the non-PAM end of the spacer. The 3' -OH on either side of the incoming spacer conducts nucleophilic attacks on their respective 5' borders of the repeat [77]. Cas1-Cas2 binding then causes strong deformation and unwinding of the repeat DNA [79]. Host DNA repair enzymes then fill the ssDNA present in the repeats flanking the integrated protospacer [80].

The leader sequence is essential for the targeted integration of spacers at the CRISPR array [73][81]. Specific recognition of this locus is necessary to confer polarisation of spacers present in the CRISPR array, such that recently acquired spacers generate the most immunogenic response [82]. In the *E. coli* Type I-E system, essential recognition sites are contained within the terminal 60 bp of the leader [73]. Cas1 binds specifically to positions -1 to -4 of the leader to confer firstly polarisation of the leader array, but also appropriate PAM orientation of the ensuing spacer [83].

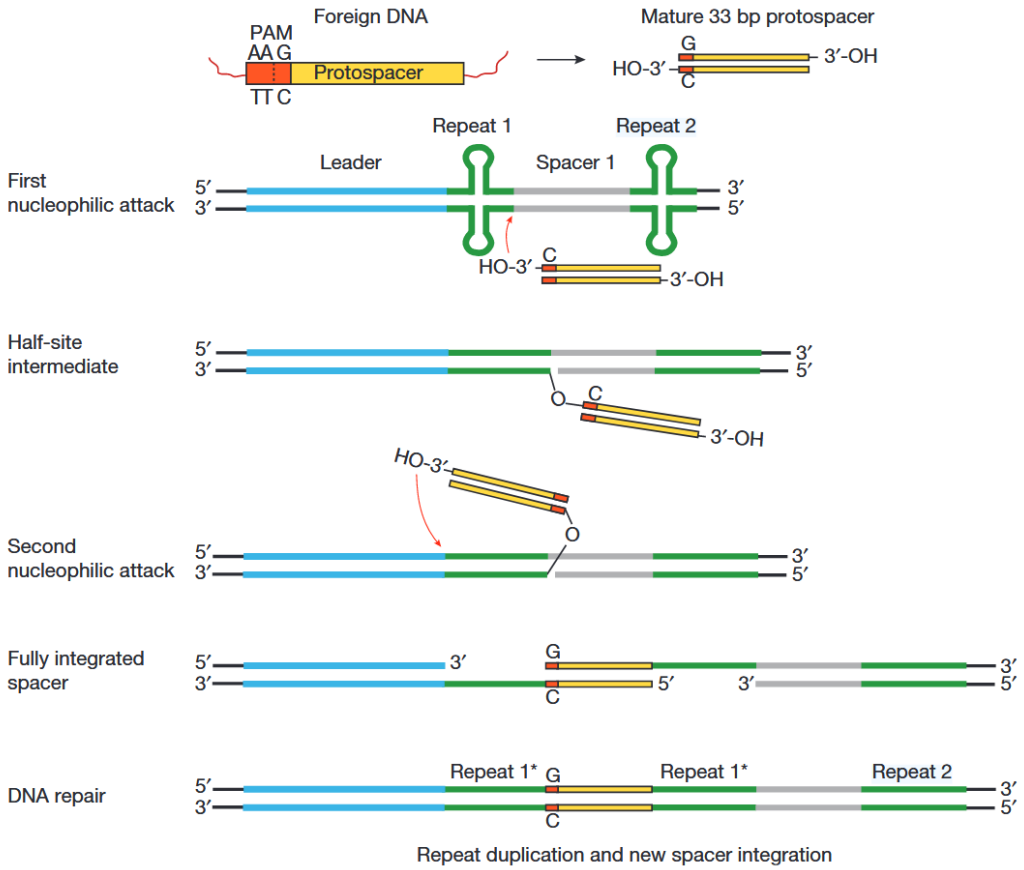

Figure 1.6: Mechanism of spacer integration into the CRISPR array.
Spacer integration commences with the nucleophilic attack of the first repeat's leader-distal 5' boundary by the PAM distal 3' -OH of the incoming protospacer, resulting in the formation of a half-site intermediate. Subsequent nucleophilic attack by the protospacer's opposing 3' -OH on the repeat's leader-proximal 5' boundary results in full site integration. Cas1-Cas2 binding and bending of the repeat DNA stimulates unwinding. The resultant ssDNA is repaired by host factors. Taken from (Nunez *et al.*, 2015).

Integrated host factor (IHF), composed of the monomers IHFα and IHFβ, is essential for *in vivo* integration in Type I-E and I-F systems [84]. IHF binds from -9 to -35 within the *E. coli* leader, inducing a 180 degree bend into the leader sequence DNA [85][79]. Hydrogen bonding between IHF and the leader stabilises the formation of this bend. DNA bending facilitates essential interaction between -50 to -55 bp of the leader and Cas1 [79].

The absence of IHF inhibits the integration of novel spacers into linear but not supercoiled DNA *in vitro* [85]. However, in the absence of IHF, supercoiled DNA

acquires spacers adjacent to every repeat in the CRISPR array. IHF is therefore essential to the polarisation of integration into the CRISPR array, with its supplementation restoring 80% polarisation [85].

Despite being essential for polarising spacer acquisition into CRISPR arrays in Type I-E and I-F CRISPR systems, IHF is only present in gram-negative bacteria [86]. An alternative mechanism must be used by bacteria that do not possess IHF. Polarisation in Type II systems is intrinsic to Cas1, which recognises the 3' leader anchoring sequence (LAS) of the leader sequence to direct orientation of integration [82]. *Enterococcus faecium*'s Type II-A Cas1-Cas2 integration complex scans all repeats within the CRISPR array, but only upon recognition of the LAS does it conduct integration [87]. This process is imperfect, as ectopic spacers have been identified integrating at repeats within the CRISPR array distal to the leader in *Strep. pyogenes* and *Strep. thermophilus* Type II-A CRISPR systems [81].

1.7.2 crRNA Processing

Interference by CRISPR systems is guided by mature crRNAs, processed from the products of CRISPR array transcription [Figure 1.7] [70]. These long precursor crRNAs (pre-crRNAs) contain stem-loops derived from the palindromes present in the repeats. Cas6 processes pre-crRNAs in Type I-E and I-F systems to produce mature crRNAs [88]. Cas6 specifically binds to and cleaves stem-loop structures in pre-crRNA, resulting in the mature crRNA containing a spacer and bases from the flanking sequences [88]. Following cleavage, Cas6 remains bound to the crRNA to nucleate the formation of the Cascade surveillance complex.

Cascade formation commences with the assembly of additional subunits along the mature crRNA [Figure 1.8] [89]. crRNA non-specifically interacts with six Cas7 subunits, which form the complex's backbone. Cas5 binds to 5' crRNA, forming a cap opposing Cas6. Cas11 binds centrally to stabilise crRNA-DNA interactions. Finally, Cas8 binds to Cas6 at the terminus.

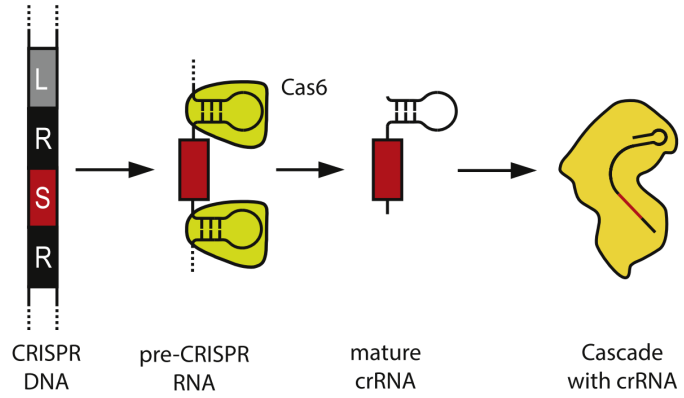

Figure 1.7: Overview of crRNA maturation by Cas6.
The CRISPR array is transcribed, producing a long-pre-crRNA. Cas6 binds to and cleaves the stem-loop structures to produce a mature crRNA. Remaining bound, Cascade complex formation is nucleated from the Cas6-mature crRNA complex. Cascade may then conduct surveillance within the cell. Adapted from (Rath *et al.*, 2015).

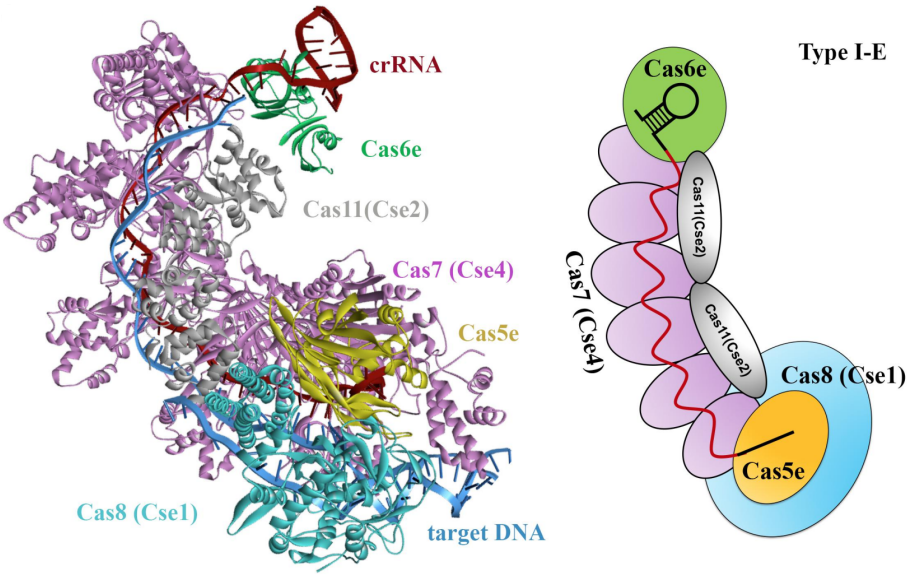

Figure 1.8: Architecture of the *E. coli* Type I-E Cascade interference complex.
The crystal structure of the Cascade surveillance complex is shown in complex with a target DNA sequence (left). Recognition of an optimal PAM in the sequence by Cas8 has facilitated the unwinding of target DNA. This would stimulate Cas3 recruitment *in vivo*. A model of Cascade complex architecture illustrates the seahorse shape adopted by the complex (right), following complex nucleation around the Cas6-mature crRNA complex. Taken from (Zheng *et al.*, 2020).

1.7.3 Interference

Cascade surveillance commences with recognition of a PAM in the target sequence, triggering the unwinding of target DNA for further querying [Figure 1.9] [90]. Establishing PAM complementarity before querying a target sequence for full complementarity minimises the number of sequences that are queried, increasing the efficiency of Cascade surveillance. PAM recognition is performed by Cas8, with targeted interactions between Cas8 and the PAM minor groove stalling Cascade to initiate target DNA unwinding [91]. Non-specific interactions between lysine-rich loops of Cas7 promote DNA interaction [91].

Cas8-PAM interactions cause localised DNA bending, causing proximal unwinding [92]. Unwound DNA may be queried by Cascade's crRNA to generate a crRNA-DNA heteroduplex. R-loop formation then occurs, commencing from PAM-proximal to PAM-distal regions of the query sequence. Complementary sequence binding triggers conformational changes in Cascade. Cas11 shifts to interact with Cas8, widening the DNA binding channel and stabilising the R-loop [93].

Conformational changes in Cas8's C-terminus upon R loop formation facilitate Cas3 recruitment [94]. Without these conformational changes, Cas3 is sterically hindered from binding. Cas3 can then conduct targeted nucleotide degradation of the cognate sequence [95]. Cas3 contains a C-terminal 3'-5' helicase domain and an N-terminal HD nuclease domain [96]. The helicase domain processively unwinds target DNA, generating ssDNA to be cleaved by the nuclease.

Target degradation commences with the nicking of the non-target strand. This nicked dsDNA provides a substrate for Cas3's helicase activity. Cas3 then translocates along the non-target strand, intermittently generating nicks in the sequence [97]. As Cas3 processes along the non-target strand, the target strand is extruded as a ssDNA loop. Strand bias during degradation causes bias in the source of spacers during primed acquisition [98].

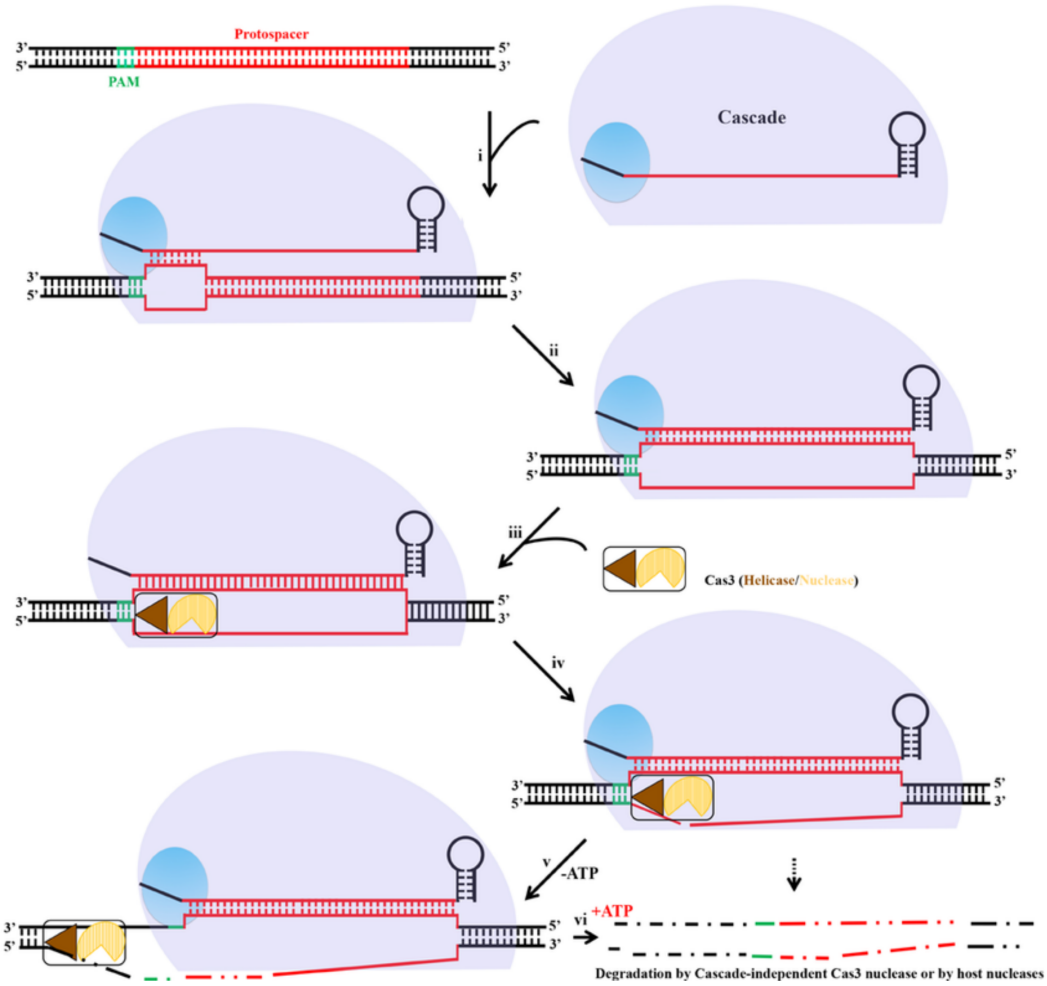

Figure 1.9: Overview of the mechanism of interference by Cascade in *E. coli* Type I-E systems.
Cascade, in complex with a mature crRNA, interrogates a DNA sequence. Cas8 (blue) interacts with the PAM (green), facilitating Cascade binding. R-loop formation nucleates from the PAM, eventually displacing the non-target strand across the length of the crRNA spacers. Target binding induces conformational changes, which result in Cas3 recruitment. Cas3 then conducts cleavage of the displaced strand and unwinds the dsDNA in an ATP-dependent manner to facilitate complete cleavage. Taken from (Zheng *et al.*, 2020).

1.8 Spacer Selection

1.8.1 Self vs Non-self Discrimination

CRISPR-Cas systems can cause autoimmunity when interference is targeted against the host genome. A range of mechanisms are employed to mitigate incidences of autoimmunity.

1.8.1.1 Protospacer Adjacent Motif (PAM)

Spacers within the CRISPR array are potential targets of the crRNAs that they themselves encode. Recognition of a PAM sequence is therefore necessary to induce cleavage [99]. PAMs are short DNA sequences that flank a protospacer but are not integrated into the CRISPR locus. They range from 2-8 nucleotides in length [100]. PAMs ensure the effector arm only cleaves foreign DNA, they are present in both type I and II systems [101].

In the *E. coli* Type I-E CRISPR system, the PAM 'AAG' is specifically recognised by the Cas1a C-terminal domain during spacer acquisition [102]. The integration of spacers containing a consensus PAM is biased [102]. Following PAM recognition within a protospacer, the sequence is cleaved such that the cytosine of the PAM-complementary sequence (5' CTT 3') becomes the terminal nucleotide of the incoming spacer [76]. Following crRNA expression and maturation, surveillance by Cascade proceeds in a PAM-first model, with PAM complementarity necessary to initiate the querying of a target sequence [103]. In type I systems, PAM recognition is necessary to induce endonuclease activity from Cas3. Cas3 will still be recruited in the absence of a PAM, however, it will behave as a molecular motor and processively move along DNA to promote spacer uptake distal to the site of recruitment, facilitating primed adaptation [92]. In type II systems, PAM recognition is necessary to conduct primed adaptation [104].

Cas1-Cas2 of Type I-E and I-F systems conduct PAM selection independently of other host factors [48]. However, Type I systems generally use Cas4 to cleave protospacers in a PAM-dependent manner [48]. In Type II systems, Cas9 plays a key role in PAM-specific protospacer recruitment, with PAM recognition by Cas9 during adaptation and interference mitigating autoimmunity [105].

Hosts have been shown to avoid target sequences in their genomes to mitigate autoimmunity [106]. In this respect, longer PAM sequences may reduce this se-

lective pressure whilst requiring more stringency during spacer acquisition. MGEs similarly exhibit a preference away from the presence of host PAM sequences in their genomes [107].

1.8.1.2 RecBCD and chi sites

RecBCD is a helicase-nuclease complex involved in double-strand break (DSB) repair and homologous recombination. It has also been implicated in the creation of substrates for spacer acquisition [Figure 1.10A] [108][109]. RecBCD binds to exposed linear ends of dsDNA, exhibiting helicase activity as it unwinds DNA at a rate of 1 kb/s [110]. RecB exhibits nuclease activity against both strands [108].

RecC scans for *chi* site sequences (5'-GCTGGTGG-3') during translocation [111]. *Chi* site recognition induces a conformational change in RecBCD, inhibiting continued degradation of the DNA strand. The complex then nicks the 3' end strand near the *chi* site, recruiting RecA to promote homologous recombination [112]. *Chi* sites act as a regulatory checkpoint for RecBCD and are enriched on the genome relative to invasive MGEs [112]. Host DNA is preferentially protected whilst foreign DNA is degraded. This promotes the generation of protospacers from foreign DNA over host DNA.

The exposed linear ends of dsDNA that RecBCD targets are often introduced during plasmid transfer and phage infection but are comparably rare in the host genome, further biasing towards acquisition from MGEs [Figure 1.10D]. Furthermore, the processing of free DNA ends generated during restriction nuclease activity against MGEs means RecBCD facilitates synergy between CRISPR systems and other host defences. CRISPR Cas systems are enriched alongside certain restriction-modification systems [113].

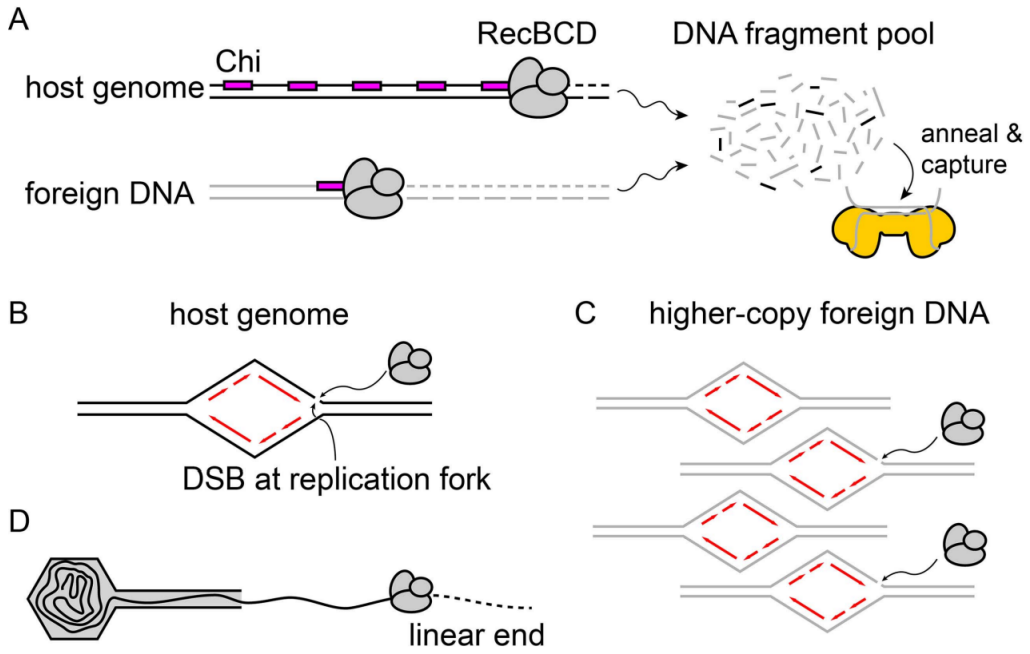

Figure 1.10: Creation of substrate for naïve acquisition by RecBCD.
(A) RecBCD generates substrate for adaptation by Cas1-Cas2. Enrichment of *chi* sites within the host genome relative to foreign DNA biases towards the preferential acquisition of spacers from foreign DNA. (B, C) RecBCD activity is initiated at double-strand breaks, found at replication forks in both host and foreign DNA. The high copy number of foreign DNA creates more replication forks from which RecBCD may generate spacers. (D) Linear DNA, typically found following injection of phage genomes, is vulnerable to degradation by RecBCD. Taken from (Xue and Sashital, 2019).

1.8.1.3 Substrate bias

Some CRISPR systems exhibit bias towards acquisition from actively replicating DNA, leading to a bias towards non-self in the event of MGE invasion [114][115]. A preference for free DNA ends has been demonstrated in *E. coli* Type I-E systems, these are present at stalled replication forks[114]. As high copy number MGEs will contain a greater number of replication forks, this biases towards the acquisition of spacers from non-host DNA [Figure 1.10B, C][114]. Preference for the use of free DNA ends as a source of spacers has been reported in Type II-A, I-G, and III-B systems [116][115]. This bias is not always the case, with no bias detected in an *S. thermophilus* Type II-A system [117]. Spacer acquisition hotspots have also been identified from highly expressed genes, suggesting transcription may make DNA

physically vulnerable to the Cas machinery or may induce double-stranded breaks [116][118].

1.8.2 Primed Adaptation

Primed adaptation is the acquisition of novel spacers from regions proximal to pre-existing crRNA targets, requiring both adaptation and interference machinery [102]. Point mutations within the PAM or crRNA target may facilitate MGE evasion of the CRISPR response. Primed adaptation facilitates the targeting of such mutants. This is the dominant form of adaptation in Type I-E systems, with novel spacers acquired at x1000 the rate of naïve acquisition [118]. Up to 13 mutations may be present in a sequence targeted by primed acquisition, and it may induce acquisition of novel spacers as far as 100 kb from the target sequence [74][119].

Primed acquisition in *E. coli* Type I-E systems exhibit specificity towards the non-target strand, derived from its unidirectional degradation by Cas3 [120]. In comparison, the Cas2/Cas3 fusion in Type I-F systems degrades both strands, resulting in no bias [121]. Type I-E primed spacers exhibit a high fidelity for the 'AAG' PAM. This may derive from Cas3's preferential cleavage of T-rich sequences, generating ssDNA fragments rich in 5'-NTT-3' sites from which protospacers with the complementary 5'-AAG-3' PAM can derive [122].

1.9 Regulation of CRISPR-Cas

Without invasive genetic elements present, CRISPR-Cas systems confer a fitness cost primarily through autoimmunity. Autoimmunity arises from the acquisition of spacers targeted against the host genome [123]. Regulation of expression such that CRISPR-Cas is only active during periods of high risk of MGE infection mitigates this fitness cost [124].

The regulon of histone-like nucleoid-structuring protein (H-NS), a DNA-binding

transcriptional repressor, silences roughly 5% of the *E. coli* genome [125]. H-NS binding induces DNA condensation into looped DNA-protein-DNA structures, sequestering promoters from the transcriptional machinery [126]. Binding to AT-rich sequences, H-NS interacts with the leader and the P_{Cas3} promoter to suppress the expression of crRNAs and *cas* genes in *E. coli* [127]. AT-rich sequences in invading elements may sequester H-NS from these sites to facilitate CRISPR-Cas targeting [128]. LeuO antagonises H-NS repression, inhibiting cooperative polymerisation of H-NS at the CRISPR locus. LeuO binds upstream of Cas8 and can induce expression when overexpressed [129].

The P_{Cas8} promoter is inhibited by the binding of catabolite repressor protein (CRP) [130]. The absence of glucose stimulates cAMP production by adenylate cyclase, inducing CRP repression [131]. CRP shares a binding site with LeuO, competitively inhibiting Cas gene expression at low glucose concentrations.

Finally, the two-component regulatory system BaeSR upregulates *cas* gene expression [132]. BaeS, a histidine kinase, phosphorylates BaeR in response to envelope stress [133]. Phosphorylated BaeR may then upregulate the *cas* genes. The induction of envelope stress by phage injection may induce CRISPR-Cas expression.

1.10 Classification of CRISPR Systems

CRISPR-Cas systems have undergone immense diversification. However, major unifying themes have underpinned this evolution, facilitating their organisation into groups based on phylogeny and function. The hierarchical structure of CRISPR-Cas classification commences with their division into classes [48]. Class 1 systems are characterised by multi-protein interference complexes, which must assemble to conduct target recognition and cleavage. In comparison, Class 2 systems are characterised by a single multidomain protein that integrates these activities.

Each class is further divided into types and subtypes [Figure 1.11] [48][134]. While types may be determined by the *cas* gene complement of a system, subtypes are delineated based on gene organisation and protein sequence. Five types were initially proposed [48]. However, a Type VI has since been categorised, and a further Type VII system has been proposed [134][135]. Both Type VI and VII are RNA-targeting systems.

Expansions in known system variants, alongside the identification of rarer or atypical systems, have been driven by advances in high-throughput sequencing, growing sequence depth, and diversity in databases [136][135]. This has been compounded by integrating community resources such as CRISPRCasFinder into workflows [137]. Together, these have aided the identification of CRISPR systems encoding potentially novel functions [135].

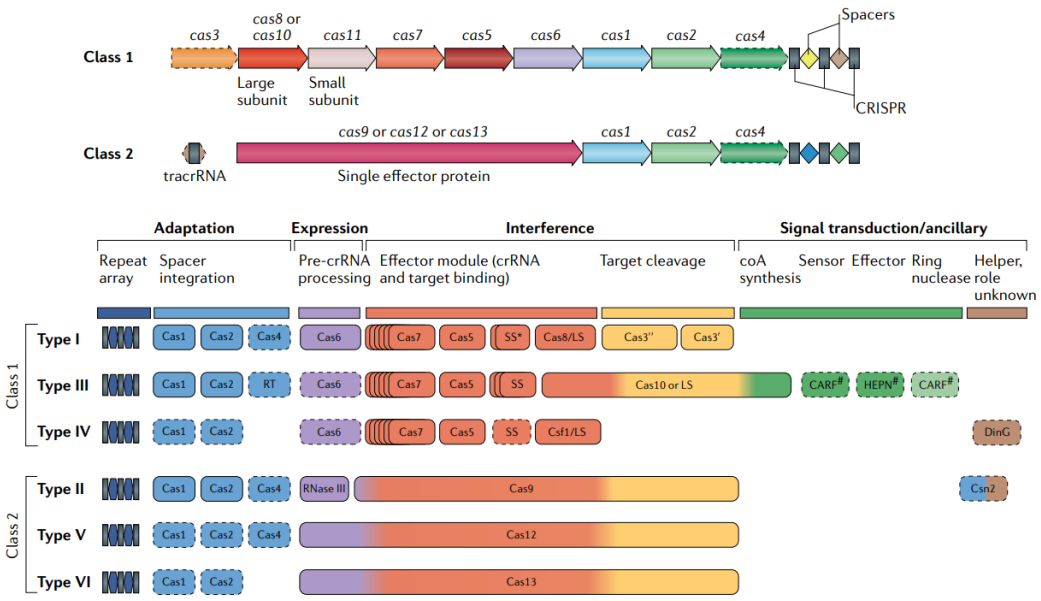

Figure 1.11: Modular organisation of CRISPR-Cas Classes and Types.
CRISPR-Cas systems are divided into Class 1 and 2 systems depending on the structure of their interference machinery. Class 1 system effector complexes possess multiple subunits, while Class 2 systems encode a single, multidomain protein for interference. Schematics of the typical *cas* gene complement of the six core types of CRISPR-Cas systems are outlined, there is variation within subtypes. Dotted lines indicate *cas* genes that may be absent in their respective CRISPR-Cas type subtypes. Taken from (Makarova *et al.*, 2020).

1.10.1 Class I CRISPR Systems

Class I CRISPR systems comprise Types I, III, IV, and VII. These are found ubiquitously in prokaryotes. Class I systems are characterised by their use of multi-protein interference complexes [136].

1.10.1.1 Type I CRISPR Systems

Type I systems are amongst the most common and well-studied CRISPR systems. The hallmark of Type I systems is the presence of the Cas3 gene, frequently found with an HD nuclease domain fusion to facilitate DNA cleavage [96]. These systems also contain the CRISPR-associated complex for antiviral defence (Cascade), a multi-protein effector complex responsible for crRNA binding, target scanning, and Cas3 recruitment. Cascade is comprised of repeat-associated mysterious proteins (RAMPs), which use an RNA recognition motif (RRM) to interact with crRNA [70].

Cascade assembled into a seahorse-shaped architecture [89]. Cascade is composed of five Cas proteins in different stoichiometry; these are Cas5, Cas6, Cas7, Cas8, and Cas11. Cas6 selectively binds and cleaves long crRNA transcripts at the 3' end to produce mature crRNA [88]. Multiple Cas7 proteins oligomerise from Cas6 along the crRNA to form the backbone of Cascade. Cas5 occupies the 5' handle of the crRNA, while Cas8 and Cas11 bind within the Cas7 backbone. Two copies of the Cas11 small subunit interact with Cas7 to stabilise the crRNA and target DNA, while Cas8 caps Cas5 at the terminus of the complex. It is Cas8 which conducts PAM recognition and Cas3 recruitment [89].

Type I CRISPR systems are highly diverse, comprising nine subtypes (I-A to I-E, I-F1 to I-F3) [136]. Type I systems typically use Cas1, Cas2, and Cas4 for adaptation. However, Cas4 is absent from the monophyletic clade of subtypes I-E and I-F, with Cas2 present in fusion with Cas3 in I-F [48]. Several subtypes (I-B and I-F) have been identified within transposons whilst lacking Cas3. This

suggests alternate functions have emerged from within Type I systems, such as the RNA-guided transposition of MGEs [138][139].

1.10.1.2 Type III CRISPR Systems

Type III systems are proposed to be ancestral to all Class 1 systems [140]. They are diverse and widespread, exhibiting unique functionalities such as producing secondary messengers to trigger auxiliary defences [29]. Type III system interference is triggered by transcription from a target sequence. Effector complexes bind at transcription bubbles on mRNA, activating the Cas10 nuclease to cleave both cognate RNA and foreign DNA in close proximity [141][142].

Effector complexes of Type III systems are divided into two families. The Csm and Cmr effector complexes have similar structures to that of Cascade [143]. Cas5 and Cas7 comprise the effector complex backbone, while Cas10 occupies the position of Cas8. However, Cas6 is rare in type III systems. CrRNA processing is therefore usually conducted by ancillary CRISPR systems [143].

Type III systems do not use PAM sequences. Instead, autoimmunity at the CRISPR locus is prevented through a 'tag-anti-tag' mechanism in which complementarity in the crRNA with repeats inhibits interference [144]. However, a PAM-like sequence within cognate RNA is required to induce DNA targeting [145]. Interference is concomitant with the activation of palm domains within Cas10. These are cyclases that produce cyclic oligonucleotide secondary messengers that induce ancillary effector activation [29].

1.10.1.3 Type IV CRISPR Systems

Type IV systems are unique amongst CRISPR systems for their strong association with MGEs and their role in inter-plasmid conflict [146]. Principally found on plasmids, Type IV systems are divided into five subtypes (IV-A to IV-E). IV-A systems are associated with DinG, a helicase shown to possess 3'-5' exonuclease activity

[147]. In comparison, IV-B lacks an associated helicase or CRISPR array, suggesting it possesses an alternative role to host defence [148]. Type IV-C represents an ancestral out-group identified in archaea that shares similarities with Type III systems, suggesting an origin for Type IV systems from within this group. Type IV-D systems possess RecD family helicases. The pervasiveness of helicases in Type IV systems, alongside the frequent absence of nucleases, suggest they may rely on host-encoded nucleases to cleave targets, rendering them vulnerable to DNA degradation through unwinding of target DNA.

Type IV systems typically lack Cas1-Cas2 adaptation modules, but their presence correlates with Type I systems. This suggests they may rely on other systems, such as Type I, to acquire spacers, which they can use as guides [146].

1.10.1.4 Type VII CRISPR Systems

Type VII systems are the most recently identified type of CRISPR system [135]. Like other Class 1 systems, Cas5 and Cas7 assemble in complex with a crRNA processed by Cas6 [149]. The unique effector complex Cas14, which contains a β-CASP nuclease, is then recruited. Tetrameric Cas14 conducts 5'-3' RNA exonuclease cleavage. Uniquely, Type VII spacers principally target transposons. This suggests these systems are akin to Type IV systems as they may specifically modulate the horizontal transfer of a subset of MGEs.

1.10.2 Class 2 CRISPR Systems

Class 2 CRISPR-Cas systems encompass Types II, V, and VI. They are uniquely characterised by their use of a single multidomain effector protein [150]. Broadly, these target and cleave cognate nucleotide sequences of invading MGEs. The single protein nature of effector domains has resulted in their prolific use as genome editing tools, notably Cas9 [151].

Cas9, the effector protein of Type II systems, contains two nuclease domains that

conduct DNA cleavage. Cas9 is guided by a crRNA, which undergoes maturation through interaction with RNase III in cooperation with Cas9 [152]. Following binding to a cognate target, an HNH nuclease domain cleaves the complementary strand while a RuvC-like nuclease cleaves the non-complementary strand [153].

In comparison, Type V systems use the single effector protein Cas12. Smaller than Cas9, Cas12 encodes a single RuvC-like nuclease domain, which performs cleavage on the target and non-target strand sequentially [154]. Cleaved DNA is characterised by staggered cuts, compared to the blunt ends produced by Cas9 [155]. Specific Cas12 variants conduct indiscriminate ssDNA cleavage following cognate strand recognition [155].

Finally, Type VI systems encode the effector protein Cas13. Cas13 contains two higher eukaryotes and prokaryotes nucleotide-binding (HEPN) nuclease domains. Uniquely amongst Class 2 systems, Cas13 exclusively cleaves RNA [156]. This system does not use PAM sequences, instead relying on complementarity to RNA to derive specificity of response [157]. Following cognate RNA binding, Cas13 indiscriminately degrades RNA [156].

Broadly, Class 2 systems differ from Class 1 in how they process crRNA. Class 1 systems contain a dedicated Cas protein for pre-crRNA maturation. In contrast, Class 2 systems either rely on host nucleases (such as RNase III in Type II and V) or encode that functionality into their single effector protein (such as Cas12 in Type VI) [158][159].

1.11 Ancillary CRISPR Functions

1.11.1 Regulation by CRISPR-Cas

The potential for CRISPR-Cas system activity to induce toxic autoimmunity necessitates regulation. One such form is the autoregulation of the Cas gene operon by

interference machinery. A long trans-activating crRNA (tracr-L RNA) may be bound by Cas9 instead of a standard crRNA [160]. This crRNA possesses partial homology to the promoter upstream of *cas9*. Partial complementarity facilitates binding without nuclease activation, inhibiting expression but not inducing autoimmunity. Cas9 is essential to spacer acquisition and interference in Type II-A systems; autoregulation tightly represses all aspects of CRISPR immunity.

Regulation through partial spacer complementarity may extend beyond autoregulation. In *Pseudomonas aeruginosa* UCBPP-PA14, Phage DMS3 infection inhibits biofilm formation and swarming motility [161]. Partial complementarity of a spacer to phage DMS3 is proposed to induce low-rate nicking of DNA by Cas3. Subsequent induction of the SOS response is suggested to prevent biofilm formation [162]. Such roles in regulating biofilm formation and virulence have been reported in other bacteria [163][164].

1.11.2 Toxin-Antitoxin systems (CreR/CreT)

Regulation of endogenous loci by CRISPR-Cas systems provides a mechanism to induce abortive infection following inactivation of interference machinery. Cascade-repressed toxin (CreT) sequesters a rare arginine tRNA to inhibit growth [165]. This locus is repressed by Cascade bound to CreA. CreA is an antitoxin RNA that resembles a crRNA. Therefore, through partial complementarity, it targets cascade to occupy and repress but not cleave *creT*. Thus, this system may serve as an anti-anti-CRISPR, as inhibiting the interference machinery by an anti-CRISPR would induce toxin expression and growth arrest. Furthermore, this TA-pair serves as an addiction module for its cognate CRISPR system, preventing its loss from the host.

1.12 Inhibition of Prokaryotic Defence Systems

As their hosts have adapted and evolved a diverse array of immune systems, MGEs have themselves evolved an arsenals of countermeasures that can disrupt, subvert and evade the immune defences of their respective hosts [4].

1.12.1 Anti-RM

The first inhibitor of a prokaryotic defence system identified was Ocr, encoded by the phages T7 and T3, targeted against the *E. coli* EcoK and EcoB restriction-modification systems [166]. Ocr is a small, highly negatively charged protein resembling dsDNA, exhibiting mimicry of the target of an RM system [167]. However, this protein also potently inhibits BREX, an RM-like system [168][169]. Ocr exemplifies how defence system inhibitors may exhibit broad-spectrum inhibition.

Several further anti-restriction proteins have been identified [170][171][172]. However, an absence of biotechnology applications combined with difficulty of screening restrained the discovery of further RM inhibitors.

1.12.2 Anti-RecBCD

The RecBCD complex degrades DNA at double-stranded breaks. As such, it contributes to host defence through the degradation of phage DNA either following CRISPR-Cas cleavage or through the detection of linearised dsDNA present at certain stages of the phage life cycle [108]. Gam is a highly compact DNA mimic expressed by Phage λ, which binds to RecBCD and inhibits the degradation of phage DNA [173]. Gam is used in lambda red mediated recombineering to inactivate RecBCD and facilitate recombination into the *E. coli* genome [174].

Inhibition of RecBCD by Gam has been shown to trigger the induction of abortive infection via the retron Ec48 [175]. Retrons are tripartite systems composed of a reverse transcriptase, non-coding RNA, and an effector complex [176]. They have

been shown to guard first-line immune systems, the inactivation of which triggers an immune response [175]. Subsequently, phage-encoded retron inhibitors have been identified, which prevent the induction of abortive infection [177].

1.12.3 Anti-SOS

MGEs have been shown to inhibit the broader SOS immune response [178][179][180]. SOS inhibitors have been shown to target RecA, inhibiting the formation of RecA nucleofilaments along single-stranded DNA (such as those present during invasion by an MGE). These trigger LexA-dependent induction of the SOS response [179]. Inactivation of the SOS response can help the propagation of an invading element by preventing the arrest of the cell cycle [181] and activation of lytic prophages [182]. Therefore, alongside the inhibition of specific bacterial immune defences, MGEs may target broader systems within the cell to maintain an environment more conducive to their replication.

1.12.4 Expansion of Known Defence System Inhibitors

The massive expansion of identified prokaryotic defence systems [3][14][15][16] has occurred concomitant with a significant increase in known defence system inhibitors [4]. Whilst the identification of a greater number of systems has provided more targets against which inhibitors may be identified [47][183][184], changes in experimental approaches have driven a rapid increase in the number of known defence system inhibitors [Figure 1.12].

Just as defence systems have been found to cluster into defence islands [13], their inhibitors similarly cluster into anti-defence islands [185]. Subsequently, guilt-by-association has facilitated the identification of defence system inhibitors as it has the systems themselves [186]. Furthermore, these anti-defence islands accumulate in specific regions of MGE genomes, such as those expressed early during infection [185]. Therefore, behaviour-dependent clustering of MGE-encoded anti-

defence proteins facilitates their homology-independent identification. Identified proteins may then be tested and experimentally verified [14].

Functional metagenomic screens have facilitated the identification of additional defence system inhibitors [187]. The sequence-independent aspect of this approach enables the identification of novel inhibitors, which may seed further guilt-by-association studies. Functional metagenomics has been particularly useful in identifying novel anti-CRISPRs [187][188]. The targeting of interference against an antibiotic resistance gene during the absence of selection, followed by a period of selection in the absence of interference, facilitates the selection of novel inhibitors, as opposed to their identification through comparably lower throughput screening efforts.

Finally, direct interrogation of phages has also facilitated the identification of inhibitors. Comparative genomics of closely related phages resistant and sensitive to a CRISPR-Cas system facilitated the identification of the first anti-CRISPR (Acr) [189]. However, screening methods may extend beyond measuring survival via plaquing assay. CBASS and Pycsar system inhibitors were identified from the cyclic nucleotide hydrolase activity present in post-infection bacterial lysates, suggesting respective phages contained genes which degraded the secondary messenger molecules present in these systems [47]. Finally, the experimental evolution of phages has facilitated the identification of escape mutants possessing methods of inhibition which arose *de novo* during screening [190].

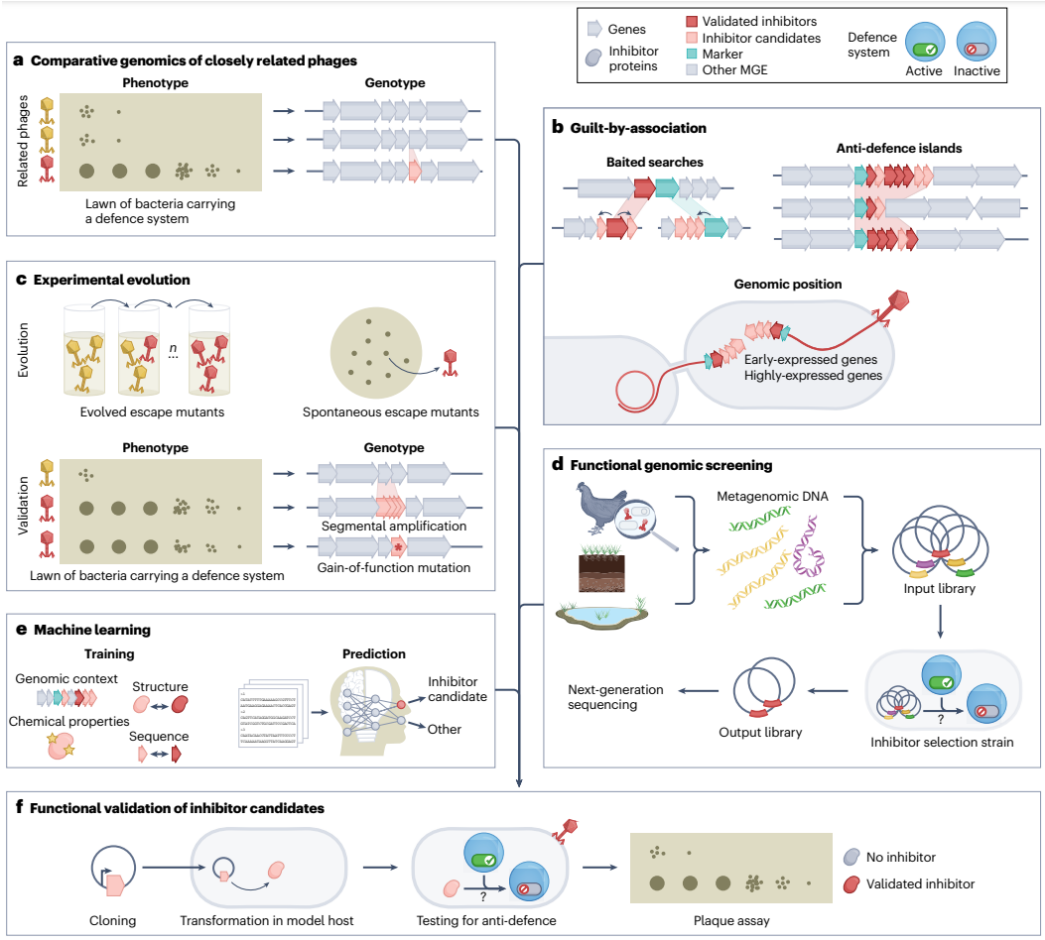

Figure 1.12: Approaches for identifying inhibitors of bacterial immune systems. Candidates for novel bacterial immune system inhibitors may be identified via: (A) Comparative genomics of closely related phages identifying specific genes which facilitate propagation in a host containing a specific immune system. (B) Guilt by association studies identifying genes enriched proximal to known inhibitors. (C) Experimental evolution identifying novel gain-of-function phage escape mutants. (D) Functional metagenomic screens identifying genes through a screening or selection process targeted against a bacterial immune system of interest. (E) Machine learning applying conserved characteristics of known inhibitors to identify unknown proteins with similar functions. (F) These approaches identify candidates that can be experimentally validated through testing against an appropriate immune system in a model host. Taken from (Mayo-Muñoz *et al.*, 2024).

1.12.5 Inhibition of CRISPR-Cas

Whilst significant advances have been made in the identification of inhibitors of a diverse array of prokaryotic immune systems [4], the majority of identified inhibitors target CRISPR-Cas systems [191]. Alongside expanding our understanding of the prokaryote-MGE arms race, Acrs have several applications. Acrs may act as a

selectable marker during the editing of phage genomes; they have facilitated the fluorescent tagging of proteins within jumbo phage genomes, helping illuminate mechanisms of genome packaging by these phages [192]. Furthermore, they can help modulate the activity of CRISPR interference machinery as a biotechnology tool [193]. Finally, rational engineering of phages to contain defence system inhibitors tailored against pathogenic bacteria may facilitate their practical use as a treatment option in the absence of viable antibiotic alternatives [194][195].

1.12.5.1 ACRs target interference

Acrs targeting interference have been the primary means identified whereby MGEs may evade CRISPR-Cas defences [4]. Over 100 families of such Acrs have been identified, employing diverse mechanisms to achieve inhibition [191]. Broadly, these can be divided into either enzymatic or non-enzymatic mechanisms.

There are several methods whereby non-enzymatic inhibition of interference may be achieved. The most common method involves occluding functional domains; for instance, AcrIIC1 occludes the HNH nuclease domain of Cas9 to inhibit DNA cleavage [196]. Alternatively, an Acr may multimerise components of interference; AcrID1 forms a homodimer with each monomer binding a Cas10d monomer such that they are sequestered from interaction partners and unable to partake in an immune response [197]. Furthermore, an Acr may co-opt interference components through mimicry of targets. AcrIF2 mimics the negative charge of the dsDNA backbone, competitively inhibiting DNA binding at the Cas7-Cas8 interface [198]. Finally, a non-enzymatic Acr may induce degradation of immune components; this is exemplified by AcrIIA1 inducing Cas9 degradation following binding at the HNH domain [199].

Alternatively, an Acr may enzymatically inhibit interference. AcrVA1 targets interference through the cleavage of crRNAs bound to Cas12a. This irreversibly inhibits Cas12a whilst the Acr may inactivate further complexes [200]. An Acr may

instead covalently modify a target to induce inactivation; AcrVA5 covalently modifies Cas12a to induce its inactivation [201]. Fewer methods of enzymatic inactivation have been identified compared to non-enzymatic inhibition, the reason for this is unclear. Discovery methods may favour the identification of non-enzymatic inhibitors, they may impart a higher fitness cost on the encoding MGE, or they may be more challenging to evolve. They may also present a regulatory challenge to the MGE, as repression of further Acr activity following inhibition of CRISPR-Cas is more difficult for enzymatic inhibitors than non-enzymatic inhibitors, which exhibit single-turnover mechanics [202].

A final group of Acrs identified recently are RNA-based anti-CRISPRs (Racrs) [203]. The prophage encoded RacrIF1 has been shown to encode a CRISPR repeat mimic, such that it is bound by Cas6f and Cas7f, but the absence of a 5' handle to the RNA sequence prevents the binding of Cas5f and Cas8f, preventing assembly of the interference complex. These inhibitors have been shown to target a diverse array of CRISPR systems [203]. The identification of RNA-based inhibitory mechanisms continued with the identification of viral Cas/CreT-repressing RNA anti-CRISPRs (Cracrs), which guide CRISPR interference machinery to repress expression from Cas operons and CRISPR-regulated toxin genes [204]. CracrIF1 was demonstrated to repress the expression of Cas genes, while CracrIF2 was shown to inhibit a CRISPR-repressed toxin gene, which itself conferred anti-anti-CRISPR protection. CracrIF2, therefore, can be considered an anti-anti-anti-CRISPR.

1.12.5.2 Inhibition of CRISPR adaptation

The anti-CRISPR AcrVA5 is the only Acr demonstrated to inhibit spacer acquisition directly. It inhibited adaptation in the *Treponema denticola* Type II-A system [205]. AcrVA5 was initially discovered in a screen for novel inhibitors of the Cas12a interference machinery [206]. The presence of self-targeting spacers was used to predict Acrs, which were then experimentally validated through the inhibition of the

cleavage of fluorescent reporter genes. The mechanism of inhibition was shown to be the steric hindrance of dsDNA binding by the Cas12a protein [200]. This inhibition derived from acetyltransferase activity modifying a lysine residue in the PAM recognition domain [201]. This acetylation was reversible, with CobB deacetylase removing acetyl groups to return Cas12a activity [207]. AcrVA5 is suggested to have evolved from a bacterial acetyltransferase, which MGEs have co-opted for the modulation of host defences [208].

AcrVA5's anti-adaptation behaviour was identified in a screen of known Acrs via *in vitro* protospacer integration assay [205]. Integration by *Treponema denticola* Type II-A Cas1-Cas2 was inhibited by AcrVA5 [Figure 1.13]. This inhibition was determined to derive from strong binding between AcrVA5 and Cas2, with AcrVA5 found to acetylate Lys55 on Cas2. This was found to inhibit Cas2's non-specific endonuclease activity in not just *Treponema denticola* Type II-A system but also in *Moraxella bovoculi* Type V-A, suggesting it may exhibit broad-spectrum inhibition of adaptation.

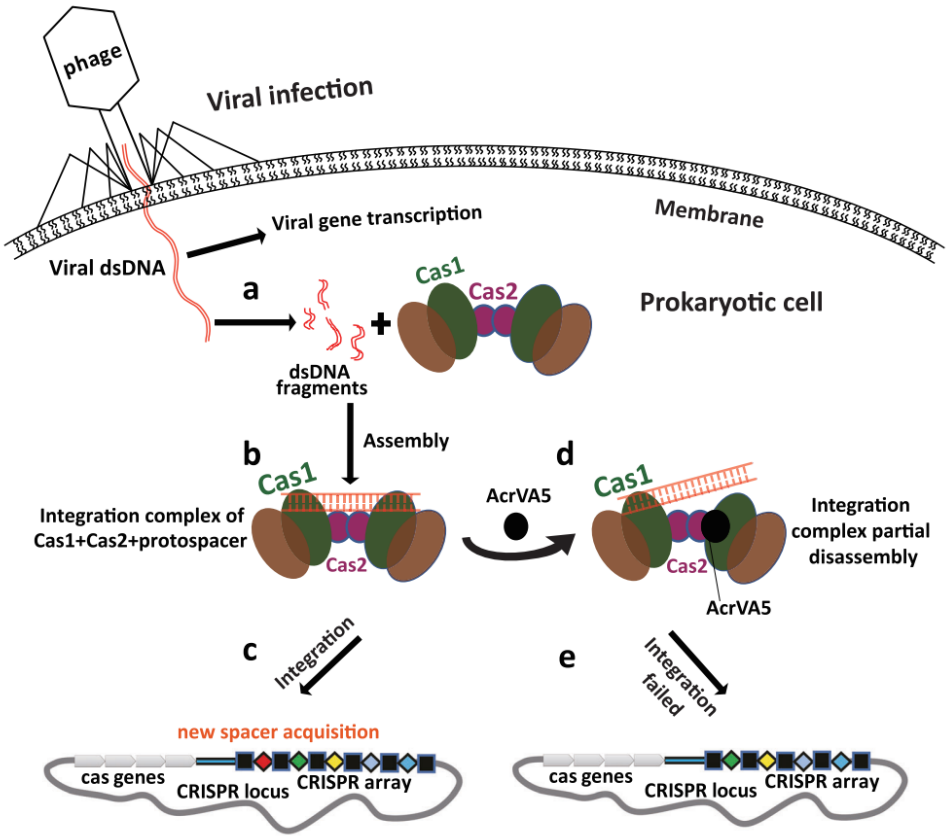

Figure 1.13: Mechanism of adaptation inhibition by AcrVA5.
(A) Double-stranded DNA derived from invading phage is available as a substrate for Cas1-Cas2. (B) Cas1-Cas2 binds to protospacers to form a functional integration complex. (C) Invading phage DNA is integrated into the CRISPR array, facilitating a targeted immune response against this phage. (D) Phage-encoded AcrVA5 triggers partial disassembly of the integration complex. (E) Disassembled integration complexes cannot integrate spacers into the CRISPR array, inhibiting an adaptive immune response against invading phage DNA. Taken from (Bi *et al.*, 2024).

1.13 Scope of Thesis

The work described in this thesis aims to leverage advances in assay sensitivity and throughput to investigate adaptation in *E. coli* Type I-E CRISPR-Cas systems. This includes the identification of Cas1-Cas2 mutants possessing hyperactive phenotypes, alongside a functional metagenomic screen to identify Acrs targeting adaptation. Additionally, I applied deep sequencing to determine patterns and trends in biases in the source of spacers acquired during adaptation. Finally, I

describe work to develop a high-throughput reporter for the detection of adaptation in a manner that can be applied to CRISPR systems universally.

An overview of this work is detailed below:

Chapter 3 describes patterns and trends identified in the distribution of spacers acquired in a spacer acquisition reporter strain under different culturing conditions. Firstly, I refute prior evidence of an absence of bias in the source of spacers acquired in this strain. Next, I investigate the role of growth media on biases in the sources of spacers. I then analyse spacer distribution, alongside PAM preference, in different regions of the plasmid to determine if this explains the biases seen. Deep sequencing is then conducted to develop a broader dataset of spacers to strengthen conclusions. Finally, this dataset is used to interrogate the genome and plasmid for novel hotspots of spacer acquisition.

Chapter 4 describes the screening, identification, and characterisation of Cas1 mutants possessing a hyperactive phenotype. First, it outlines the creation of a mutagenised Cas1-Cas2 library. Next, it describes the application of a papillation assay to screen for spacer acquisition activity, alongside the identification of a suite of hyperactive mutations in Cas1. These mutations are verified and quantified using a YFP-based reporter for spacer acquisition. Assessment of SOS response induction by these mutants identifies that they are associated with an increased response. Subsequent deep sequencing of Cas1 I28K, the mutant with the highest spacer acquisition activity, identifies a reduction in fidelity to the consensus PAM. Codon mutagenesis and screening of Cas1 I28 and D29 libraries identify further hyperactive mutants, for which SOS induction is similarly measured. Finally, mutations are combined, showing antagonism in the enhancement of spacer acquisition but synergism in the enhancement of the SOS response.

Chapter 5 described a functional metagenomic screen to identify anti-CRISPRs targeted against adaptation. First, it outlines the development of a strain capable

of screening a set of BAC libraries containing metagenomic DNA. I then conduct a pair of functional metagenomic screens using a papillation assay. These screens identify a large number of initial hits that undergo successive verification steps, ruling them out as Acrs targeted against adaptation. Sequencing several hits identifies the enrichment of homologs to ancillary genes involved in spacer acquisition. Finally, assays demonstrate that the broad-spectrum anti-CRISPR AcrVA5 shows no inhibition of adaptation in the *E. coli* Type I-E CRISPR-Cas system.

Chapter 6 outlines various efforts to create a stop codon-based reporter system capable of detecting spacer acquisition in all CRISPR systems. First, it outlines the creation and troubleshooting of stop codon-based reporters for adaptation using the TetR and VanR repressors. Next, it outlines the process of identifying a repressor capable of tolerating the addition of a CRISPR array encoded as an N-terminal peptide. Finally, it describes the construction, testing, and troubleshooting of a reporter system using the CymR repressor.

Finally, the data presented is discussed within the context of previously published work. I discuss possible explanations for biases identified in the spacer selection between the plasmid and genome. Furthermore, I discuss potential mechanisms behind the identified hyperactive residues, and how these relate to hyperactive residues of Cas1-Cas2 identified in other screens. Additionally, I critique and provide alternative approaches for a functional metagenomic screen to identify Acrs targeting adaptation, alongside the construction of a reporter for the universal detection of spacer acquisition. Finally, I discuss future approaches to these areas derived from the results within this thesis.

Chapter 2

Materials and Methods

2.1 Reagents

PCR and enzyme materials were purchased from New England Biolabs (NEB). Oligonucleotides and primers were purchased from Sigma Aldrich. DNA Purification Kits (QIAquick Gel Extraction Kit, QIAquick PCR Purification Kit, QIAGEN Plasmid Mini Kit and QIAGEN Plasmid Maxi Kit) were purchased from Qiagen™. All constructed plasmids were verified by restriction digestion and sanger sequencing by Source Bioscience.

2.2 Strains used and created

Strain	Relevant Genotype	Source and Description
BW25113	F- DE(araD-araB)567 lacZ4787(del)::rrnB-3 LAM- rph-1 DE(rhaD-rhaB)568	Parental strain of the Keio collection. [209]
BW25113 hsdR :: FRT-Kan-FRT	F- DE(araD-araB)567 lacZ4787(del)::rrnB-3 LAM- rph-1 DE(rhaD-rhaB)568 hsdR :: FRT-Kan-FRT	Keio Collection *hsdR* knockout strain. [209]
DH10β	F– mcrA Δ(mrr-hsdRMSmcrBC) Φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara-leu)7697 galU galK λ– rpsL(StrR) nupG	Thermofisher. Cloning strain.
EPI3000	F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ- rpsL (StrR) nupG trfA dhfr	Host strain of the libraries AK21 and AK22. [210] [211]

Strain	Relevant Genotype	Source and Description
ER1793	F- fhuA2 △(lacZ)r1 :: Frt-CmR-Frt glnV44 (= amber sup) e14-(McrA-) trp-31 his-1 rpsL104 xyl-7 mtl-2 metB1 △(mcrC-mrr)114::IS10	Derived from NEB. E.coli K Strain by Dr Jack Braithewaite. Frt-CmR-Frt replaces LacZ.
JB023	MG1655 △lacZ △CRI locus and Cas genes △CRII locus, T7 RNAP TetR	Dr. Jack Braithewaite. JB028 strain lacking a papillation reporter. [212]
JB028	MG1655 △lacZ △CRII locus araBAD::T7 RNAP tetA argE:: J23100 Type I-E II(A) Reporter	Dr. Jack Braithwaite. Type I-E II(A) CRISPR papillation reporter strain. [212]
JB028 hsdR :: FRT-Kan-FRT	MG1655 △lacZ △CRII locus araBAD::T7 RNAP tetA argE:: J23100 Type I-E II(A) Reporter hsdR :: FRT-KanR-FRT	This Study. JB028 with hsdR:: FRT-KanR-FRT introduced via P1 transduction from the Keio Collection.
JB028 △hsdR	MG1655 △lacZ △CRII locus araBAD::T7 RNAP tetA argE:: J23100 Type I-E II(A) Reporter dhsdR	This Study. JB028 with hsdR:: FRT-Kan-FRT introduced via P1 transduction from the Keio Collection. KanR was removed by FLP-frt recombination.
MG1655	K-12 F- λ- ilvG- rfb-50 rph-1	Wildtype E. coli K Strain

Strain	Relevant Genotype	Source and Description
MLS989	MG1655 araB::T7pol tetA, ΔaraA, ΔCas3- CRISPR-I, ΔCRISPR-II, GalK::J23101 CRISPR-II link SYFP2 opt	Type I-E II(A) YFP CRISPR reporter. [213]
MLS990	MG1655 araB::T7pol tetA, ΔaraA, ΔCas3- CRISPR-I, ΔCRISPR-II, GalK::J23101 CRISPR-II +1nt link SYFP2 opt	Type I-E II(A) YFP CRISPR reporter positive control. [213]
NEB5α	fhuA2 a(argF-lacZ)U169 phoA glnV44 a80a(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	NEB. Cloning strain.
RC5231	F- DE(araD-araB)567 Δ(lacZ)r1 :: Frt-CmR-Frt LAM- rph-1 DE(rhaD-rhaB)568 hsdR514	This Study. BW25113 with LacZ replaced by Frt-CmR-Frt from ER1793 introduced via P1 transduction.
RC5281	F- DE(araD-araB)567 Δ(lacZ)r1 LAM- rph-1 DE(rhaD-rhaB)568 hsdR514	This Study. RC5231 with Frt-CmR-Frt removed via FLP-frt recombination. Polar deletion of lacZ remains.
RC5282	F- DE(araD-araB)567 Δ(lacZ)r1 LAM- rph-1 DE(rhaD-rhaB)568 hsdR514 ArgE::TII Reporter-KanR	This Study. RC5281 with ArgE :: Type I-E II(A) papillation reporter introduced via P1 transduction.

Table 2.1: Strain List

2.3 Plasmids and Oligonucleotides

2.3.1 Plasmid List

Name	Description
pCP20	For FLP-frt mediated recombination. Yeast FLP recombinase expression vector with the pL lambda promoter. Temperature-sensitive SC101 *ori*. AmpR and CmR.
pRC1620	pBAD Cas1-Cas2 expression vector. ColEI *ori*. AmpR. Arabinose inducible Cas1-Cas2 expression vector.
pRC1656	pBAD Cas1-Cas2 expression vector. ColEI *ori*. KanR.
pRC1657	pBAD Cas1-Cas2 expression vector. ColEI *ori*. CmR.
pRC2802	$pMar_{Ara}$ Cas1-Cas2 expression vector. P15A *ori*. KanR. NdeI cut site present at ATG start codon of Cas1. Arabinose inducible Cas1-Cas2 expression vector derived from the marionette collection. [214]
pRC2804	pRC1656 Cas1 M17V
pRC2805	pRC1656 Cas1 M17T
pRC2807	pRC1656 Cas1 D26G
pRC2808	pRC1656 Cas1 I28K
pRC2809	pRC1656 Cas1 D29N
pRC2811	pRC1656 Cas1 Q90R
pRC2812	pRC1656 Cas1 Y101H
pRC2790	SOS reporter plasmid. GFP is under the control of the sulA promoter. P15A *ori*. CmR. Purchased from addgene. [215]
pRC3116	pRC1656 Cas1 I28R
pRC3117	pRC1656 Cas1 I28N
pRC3118	pRC1656 Cas1 I28S
pRC3119	pRC1656 Cas1 D29H

Name	Description
pRC3120	pRC1656 Cas1 D29K
pRC3121	pRC1656 Cas1 D29A
pRC2880	pRC1656 Cas1 M177T I28K
pRC2881	pRC1656 Cas1 D26G I28K
pRC2882	pRC1656 Cas1 I28K Q90R
AK Libraries	See [210] and [211]
pOX38 LacZ+	pOX38 containing conjugation machinery and LacZ expressed from the constitutive P_{J23100} promoter. F-*ori*.
pRC3136	P_{van} AcrVA5. P15A *ori*. KanR. The Acr AcrVA5 cloned into the vanillic acid inducible marionette expression vector.
pRC1680	P_{tet} Type I-E II(A) universal reporter. F-*ori*. KanR and CmR. Tet repressor-based universal reporter construct.
pRC1681	P_{tet} Type I-E II(A) universal reporter positive control. F-*ori*. KanR and CmR. Tet repressor-based universal reporter construct positive control.
pRC1682	P_{tet} Type I-E I(T) universal reporter.F-*ori*. KanR and CmR. Tet repressor-based universal reporter construct.
pRC1683	P_{tet} Type I-E I(T) universal reporter positive control.F-*ori*. KanR and CmR. Tet repressor-based universal reporter construct positive control.
pRC3108	P_{van} Type I-E II(A) universal reporter. Vanillic acid inducible expression vector. KanR. F-*ori*. Vanillic acid inducible repressor-based universal reporter construct expressing *lacZ*.
pRC3104	P_{van} YFP. Vanillic acid inducible expression vector. F-*ori*. KanR. Vanillic acid inducible repressor-based universal reporter construct expressing *YFP*.

Name	Description
pRC3105	P_{van} lacZ. Vanillic acid inducible *lacZ* expression vector. F-*ori*. KanR.
pRC3122	pMarPhlF with YFP. CRISPR array-PhlF repressor fusion *YFP* expression vector. P15A-*ori*. KanR.
pRC3123	pMarCymR with YFP. CRISPR array-CymR repressor fusion *YFP* expression vector. P15A-*ori*. KanR.
pRC3124	pMarVanR with YFP. CRISPR array-VanR repressor fusion *YFP* expression vector. P15A-*ori*. KanR.
pRC3125	pMarTetR with YFP. CRISPR array-TetR repressor fusion *YFP* expression vector. P15A-*ori*. KanR.
pRC3126	pMarBetI with YFP. CRISPR array-BetI repressor fusion *YFP* expression vector. P15A-*ori*. KanR.
pRC3127	pMarTtgR with YFP. CRISPR array-TtgR repressor fusion *YFP* expression vector. P15A-*ori*. KanR.
pRC3128	pMarPcaU with YFP. CRISPR array-PcaU repressor fusion *YFP* expression vector. P15A-*ori*. KanR.
pRC2763	pMarPhlF with YFP. PhlF repressor *YFP* expression vector. P15A-*ori*. KanR.
pRC2764	pMarCymR with YFP. CymR repressor *YFP* expression vector. P15A-*ori*. KanR.
pRC2766	pMarVanR with YFP. VanR repressor *YFP* expression vector. P15A-*ori*. KanR.
pRC2868	pMarTetR with YFP. TetR repressor *YFP* expression vector. P15A-*ori*. KanR.
pRC2770	pMarBetI with YFP. BetI repressor *YFP* expression vector. P15A-*ori*. KanR.
pRC2771	pMarTtgR with YFP. TtgR repressor *YFP* expression vector. P15A-*ori*. KanR.

Name	Description
pRC2772	pMarPcaU with YFP. PcaU repressor *YFP* expression vector. P15A-*ori*. KanR.
pRC2751	pMarPhlF with YFP. PhlF repressor *YFP* expression vector. No repressor present. P15A-*ori*. KanR.
pRC2752	pMarCymR with YFP. CymR repressor *YFP* expression vector. No repressor present. P15A-*ori*. KanR.
pRC2754	pMarVanR with YFP. VanR repressor *YFP* expression vector. No repressor present. P15A-*ori*. KanR.
pRC2756	pMarTetR with YFP. TetR repressor *YFP* expression vector. No repressor present. P15A-*ori*. KanR.
pRC2758	pMarBetI with YFP. BetI repressor *YFP* expression vector. No repressor present. P15A-*ori*. KanR.
pRC2759	pMarTtgR with YFP. TtgR repressor *YFP* expression vector. No repressor present. P15A-*ori*. KanR.
pRC2760	pMarPcaU with YFP. PcaU repressor *YFP* expression vector. No repressor present. P15A-*ori*. KanR.
pRC3129	pMarCymR with YFP. CymR repressor *YFP* expression vector. No repressor present. P15A-*ori*. KanR.
pRC3130	pMarCymR with YFP. CymR repressor *YFP* expression vector. P15A-*ori*. KanR.
pRC3131	pMarCymR Type I-E II(A) universal reporter. KanR F-ori.
pRC3133	pMarCymR Type I-E II(A) universal reporter. KanR F-ori. Two stop codons in spacer. Positive Control
pRC3135	pMarCymR type Type I-E II(A) universal reporter. KanR F-ori. Frameshift. Positive Control.

Table 2.2: Plasmid List

2.3.2 Oligonucleotide List

Name	Sequence	Description
Sp Aq F	GGCTAGCAGGAGGAAT TCACC	Standard SPIN assay primers to amplify the CRISPR locus of the LacZ reporter.
Sp Aq R	CGCATCGTAACCGTGC ATCTG	
Deep Seq Fwd 1	GACGGATATCAATACG GACC	SPIN assay primers to amplify the CRISPR locus of the LacZ reporter for deep sequencing. Produces a 4 kb product.
Deep Seq Rev 1	CTGCTGGTGTTTTGCT TCC	
Deep Seq Fwd 3	GGTTCTCTTCCAGTTG TTCG	SPIN assay primers to amplify the CRISPR locus of the LacZ reporter for deep sequencing. Produces a 2 kb product.
Deep Seq Rev 3	GTAACAACCCGTCGGA TTCTC	
Mar.Ara.Vec.Fwd	GGCTCGGTACCAAATT CCAG	Amplification of pRC2802 backbone to be used to generate epPCR Cas1-Cas2 mutant library.
Mar.Ara.Vec.Rev	CAGGCCATATGGTATT TC-CCCTCTTTCTCTAG	

Name	Sequence	Description
Mar C12 Mut Fwd	CTCTACTGTTTCTCCA TACCCGAG	Amplification of Cas1-Cas2 to be used to generate ep-PCR Cas1-Cas2 mutant library.
Mar C12 Mut Rev	GATGTACTGACAAGCC TCGC	
M17T F	CGATCTTTCTGCAATA TGGGCAG	Introduction of M17T mutation into Cas1.
M17T R	TGGAGACGCGATCTTT GAG	
M17V F	GTGATCTTTCTGCAAT ATGGGCAG	Introduction of M17V mutation into Cas1.
M17V R	GGAGACGCGATCTTTG AGT	
D26G F	GTGTAATAGATGGCGC GTTTGTAC	Introduction of D26G mutation into Cas1.
D26G R	CGATCTGCCCATATTG CAGA	
D29N F	AATGGCGCGTTTGTAC TTATCG	Introduction of D29N mutation into Cas1.
D29N R	TATTACATCGATCTGC CCATATTGC	
I28K F	AAGATGGCGCGTTTGT ACTTATC	Introduction of I28K mutation into Cas1.
I28K R	TTACATCGATCTGCCC ATATTGC	
Y101H F	CTCAGGCAAAACTTGC TCTGG	Introduction of Y101H mutation into Cas1.

Name	Sequence	Description
Y101H R	GAGCAGCTTATCTGAA CGC	
Q90R F	GGCCTGGAGGTGCG	Introduction of Q90R mutation into Cas1.
Q90R R	GACCAGAAGCATAAAC ACGAAC	
YFP Reporter F	GCTAGCTACTAGAGAA AGAGG	SPIN primers to amplify the YFP reporter locus in MLS989.
YFP Reporter R	AATTCAACTAAGATCG GTACTAC	
pVan acrVA5 F	AATGTAGCCGCCGCTC AGTTCAATTTTCATCT AGTATTTCCCCTCTTT CTC	PCR amplification of the pMarVan marionette expression vector backbone for the introduction of AcrVA5.
pVan acrVA5 R	GTGGATGGCCGCCTGA TGCGCTGGAGCTAACT CGGTACCAAATTCCAG	
2766 SC F	AAAACAGCATATCCAC TCAGTTCCACATTTGA AGATCATCTTATTAAG GGGTCTGACGC	PCR amplification of pMar marionette collection plasmids to introduce an F-*ori*.
2766 SC R	GGGCAGTGAAAGGAAG GCCCATGAGGCCCAGA ACGGCTTTGCCGCG	
2766 SC Seq F	CTACGATAATGGGAGA TTTTCC	PCR amplification of an F-*ori* to be introduce into pMar marionette plasmids.

Name	Sequence	Description
2766 SC Seq R	ACAACATTTAAGGAAA CAGCTATG	
LacZ F	ATGACCATGATTACGG ATTCACTGGC	PCR amplification of LacZ to be introduced into pMar marionette plasmids.
LacZ R	TTATTTTTGACACCAG AC-CAACTGGTAATGGT AG	
pRC3131 BB F	GGGGAACACTCTAAAC ATAAC	PCR amplification of the pCym Universal Reporter to introduce stop codons for positive controls into the CRISPR arrays first spacer.
pRC3131 BB R	GGGATAAACCGGCAAA AAC	
pRC3131 Double STOP	GCGATTGCCCGGTTTT TGCCGGTTTATCCCTA ATAACGCGGGGAACAC TCTAAACATAACCTAT TATTA	For the insertion of a double stop codon into pRC3131 via Gibson assembly.
MLS990 L/R F	ATGACCATGATTACGC CAAGCTTG	PCR amplification of the CRISPR array of MLS990 for introduction into pMar marionette plasmids.
MLS990 L/R R	GCTTCCAGAACCAGAA CCTGAC	

Name	Sequence	Description
2763 F	GGTTCAGGGTCAGGTT CTGGTTCTGGAAGCGC ACGTACCCCGAGC	PCR amplification of pRC2763 to introduce the MLS990 CRISPR array N-terminal to the PhlF repressor.
2763 R	CGGTACCAAGCTTGGC GTAATCATGGTCATAT TCACCCCCATGAATTG ACTCTC	
2764 F	GGTTCAGGGTCAGGTT CTGGTTCTGGAAGCAG CCCGAAACGTCGTACC	PCR amplification of pRC2764 to introduce the MLS990 CRISPR array N-terminal to the CymR repressor.
2764 R	CGGTACCAAGCTTGGC GTAATCATGGTCATAT TCACCACCCTGAATTG ACTCTC	
2766 F	GGTTCAGGGTCAGGTT CTGGTTCTGGAAGCGA CATGCCTCGTATTAAA CCGGG	PCR amplification of pRC2766 to introduce the MLS990 CRISPR array N-terminal to the VanR repressor.
2766 R	CGGTACCAAGCTTGGC GTAATCATGGTCATTA TTCACCCCCCTGAATT GACTCTC	

Name	Sequence	Description
2768 F	GGTTCAGGGTCAGGTT CTGGTTCTGGAAGCTC CAGATTAGATAAAAGT AAAGTGATTAACAGCG CATTAG	PCR amplification of pRC2768 to introduce the MLS990 CRISPR array N-terminal to the TetR repressor.
2768 R	CGGTACCAAGCTTGGC GTAATCATGGTCATAT TCACCACCCTGAATTG ACTCTC	
2770 F	GGTTCAGGGTCAGGTT CTGGTTCTGGAAGCCC GAAACTGGGTATGCAG AGC	PCR amplification of pRC2770 to introduce the MLS990 CRISPR array N-terminal to the BetI repressor.
2770 R	CGGTACCAAGCTTGGC GTAATCATGGTCATAT TCACCCCCGTGAATTG ACTC	
2771 F	GGTTCAGGGTCAGGTT CTGGTTCTGGAAGCGT TCGTCGTACCAAAGAA GAGGC	PCR amplification of pRC2771 to introduce the MLS990 CRISPR array N-terminal to the TtgR repressor.
2771 R	CGGTACCAAGCTTGGC GTAATCATGGTCATAT TCACCTCCCTGAATTG ACTCTC	

Name	Sequence	Description
2772 F	GGTTCAGGGTCAGGTT CTGGTTCTGGAAGCTG GTCGAACATGGATGAC AAGAAAG	PCR amplification of pRC2772 to introduce the MLS990 CRISPR array N-terminal to the PcaU repressor.
2772 R	CGGTACCAAGCTTGGC GTAATCATGGTCATAT TCACCTCCCTGAATTG ACTCTC	

Table 2.3: Oligonucleotide List

2.4 Solution Compositions

2.4.1 Media

Media	**Composition**
Lysogeny Broth (LB)	10 g/L Yeast Extract, 5 g/L NaCl, 5 g/L Tryptone, 2 mL/L 1 M NaOH. (Agar added to 15 g/L for plates)
SOC	20 g/L Tryptone, 5 g/L Yeast Extract, 0.5 g/L NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose
Minimal Media	1 X M9 salts, 0.1% carbon source, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$. (Agar added to 15 g/L for plates). 0.1 g/L L-arginine supplemented when a strain has ArgE disrupted.
10x M9 Salts	60 g/L Na_2HPO_4, 30 g/L KH_2PO_4, 5 g/L NaCl, 10 g/L NH_4Cl
Terrific Broth (TB)	24 g/L Yeast Extract, 20 g/L Tryptone, 4 mL/L glycerol, 0.017 M KH_2PO_4, 0.072 M K_2HPO_4

Table 2.4: Media Recipes

2.4.2 Antibiotics

Antibiotics	**Single-copy resistance gene working concentration (μg/ml)**	**Multi-copy resistance gene working concentration (μg/ml)**
Tetracycline	12.5	25
Chloramphenicol	5	35
Kanamycin	30	50
Ampicillin	50	100

Table 2.5: Antibiotic Concentrations

Sub-inhibitory concentrations of antibiotics were used as follows:

Antibiotic	**Concentration Range μg/ml**
Nalidixic Acid	0-0.8

Table 2.6: Sub-inhibitory Antibiotic Concentrations

2.5 Cloning and DNA Manipulation

2.5.1 Polymerase Chain Reaction

PCRs for cloning steps were performed with Q5^{TM} polymerase, according to manufacturer instructions. Culture-based PCRs were performed using Taq DNA polymerase, according to manufacturer instructions. 5 ng of purified DNA was used in cloning PCRs. 1 μL of prepared bacterial suspension was used in culture-based PCRs. Primer melting temperatures were determined using the NEB Tm Calculator. PCR amplification from bacterial suspensions included an initial lysis step of heating to 95 $^{\circ}$C for 10 minutes. Products were visualised by gel electrophoresis.

2.5.2 Error Prone Polymerase Chain Reaction

Standard Taq Polymerase PCR reactions were supplemented with excess dGTP and $MnSO_4$ at μM concentrations. Specific values are given in respective result sections.

2.5.3 Restriction Digestion

Restriction digests were performed in accordance with manufacturer instructions.

2.5.4 Phosphorylation

Phosphorylation reactions were performed using T4 Polynucleotide Kinase (T4 PNK) in accordance with manufacturer instructions.

2.5.5 Dephosphorylation

Dephosphorylation reactions were performed using Calf Intestinal Phosphatase (CIP) in accordance with manufacturer instructions.

2.5.6 Ligation

Ligation reactions were performed using T4 DNA Ligase. Per reaction, 200 enzyme units were used, alongside 1x CutSmart buffer supplemented with 1 mM ATP or 1x T DNA Ligase buffer. Ligations were typically performed using 50 ng of linearised vector, with insert DNA at a 3:1 molar excess. Reactions were incubated at room temperature for 15 minutes or at 16 °C overnight, for cohesive or blunt ends respectively. Following incubation, reactions were transformed into appropriate competent cells.

2.5.7 Gibson Assembly

Gibson assembly was performed as an alternative to traditional ligation-dependent cloning. DNA fragments were designed with 20 bp of homology at junctions. These sections of homology were introduced to the 5' end of PCR primers. Amplified PCRs were purified and introduced to the Gibson assembly reaction mix, never exceeding 20% final reaction volume. A home-made Gibson assembly mix gave a final concentration of 100 mM Tris HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM NAD+, 1 mM of each dNTP, 5 % w/v PEG 8000, 0.005 U/µL of T5 Exonuclease, 0.03 U/µL Phusion Polymerase and 5.3 U/µL Taq DNA ligase. DNA was present at a concentration of 0.02-0.5 $_p$moles per fragment. Final reaction volumes of 20 μL were incubated at 50 $^\circ$C for 2 hours. Following incubation, reactions were transformed into appropriate competent cells.

2.5.8 Plasmid Purification

Plasmid purification was conducted using Qiagen miniprep or maxiprep kits in accordance with the manufacturer's instructions. To increase DNA yields, low copy number plasmids were grown for purification with Terrific Broth.

2.5.9 Phenol Chloroform Extraction

A mixture of sample and phenol:chloroform: isoamyl alcohol (25:24:1) present at equal volumes was prepared. This was vigorously vortexed to produce a milky white consistency. The sample was then centrifuged for 5 minutes at 16,000 x *g*. The upper aqueous phase was then aspirated off for subsequent ethanol precipitation.

2.5.10 Ethanol Precipitation

DNA samples were purified and concentrated by ethanol precipitation. One volume of each sample was mixed with 20 μg glycogen, 0.5 X volume 7.5M NH_4OAc and

3 X volume 100% ethanol. Samples were incubated at -80 °C for 1 hour before being centrifuged at 16,000 x *g* for 30 minutes at 4 °C. The supernatant was discarded, and the pellet was washed with 150 μL 70% ethanol. The sample would be centrifuged at 16,000 x *g* for 2 minutes at 4 °C and the supernatant removed. This step was repeated to remove all supernatant. Precipitated DNA samples were resuspended into an appropriate volume of TE buffer (10 mM Tris-HCl, 1 mM EDTA pH 8.0).

2.6 Gel Electrophoresis

2.6.1 Agarose Gel Electrophoresis

Isolation and analysis of DNA fragments was conducted through their separation by molecular weight via agarose gel electrophoresis. This was performed using a BioRad gel electrophoresis system. Gels were comprised of 0.5X Tris-Borate-EDTA (TBE) containing 1-1.8% agarose. Agarose concentration was dependent on the molecular weight of DNA fragments. Electrophoresis of standard cloning gels was conducted at s voltage of 5-8 V/cm in the presence of 0.5 μg/mL Ethidium Bromide (EtBr). Analytical gels would be electrophoresed at 1.5 V/cm overnight. These would be post-stained in 200 ml 0.5X TBE with EtBr present at 1 μg/mL. Gels were imaged under UV light.

2.6.2 Agarose Gel Extraction

Gel extraction procedures were performed in accordance with the manufacturer's instructions.

2.6.3 Running Buffers

Running Buffer	Composition
10X TBE	1 M Trizma Base, 1 M Boric Acid, 0.02 M EDTA pH 8.0

Table 2.7: Running Buffer Compositions

2.7 Competent Cells

2.7.1 Making Chemically Competent Cells

Glycerol stocks of strains to be made chemically competent were streaked out onto LB-agar supplemented with appropriate antibiotics. Plates were grown overnight at 37 °C. A single colony was picked and inoculated into 5 mL LB, which was incubated at 37 °C overnight. This culture was then diluted 1:100 in fresh LB broth and grown at 37 °C, shaking at 165 rpm until an $O.D_{600}$ of 0.4-0.5 was reached. The flask was then cooled in a shaking ice slurry for 20 minutes. Cells were to be kept cold for the remainder of the protocol. All centrifugation steps are assumed to be at 4 °C. Cells were pelleted for 10 minutes at 6000 x *g*. The supernatant was discarded, and the pellet was resuspended in 1/3 of the original culture volume in TFB1 (100 mM RbCl, 30 mM CH_3CO_2K, 50 mM $MnCl_2$ and 15% Glycerol, pH 5.8). The resuspension was incubated on ice for an hour. A subsequent round of centrifugation was conducted. The supernatant was discarded, and the pellet was resuspended in 1/12 the original culture volume of TFB2 (10 mM MOPS pH 7.0, 75 mM NaCl, KCl, 15% glycerol). Resuspensions were aliquoted into 200 μL units and flash frozen on dry ice for long-term storage at -80 °C.

2.7.2 Chemically Competent Cell Transformation Protocol

Chemically competent cell aliquots were thawed on ice for 30 minutes. 5 ng of plasmid was added to the aliquot, and they were incubated together on ice for

a further 30 minutes. Cells were then heat shocked at 42 °C for 1 minute and returned to the ice for 5 minutes. Aliquots were then supplemented with 900 μL LB broth pre-warmed to 37 °C. Cultures were incubated at 37 °C on a spinning wheel for 1 hour. Cultures were then evenly spread on LN agar plates supplemented with appropriate antibiotics. Plates were incubated overnight.

2.7.3 Making Electrocompetent Cells

Glycerol stocks of strains to be made chemically competent were streaked out onto LB-agar supplemented with appropriate antibiotics. Plates were grown overnight at 37 °C. A single colony was picked and inoculated into 5 mL LB, which was incubated at 37 °C overnight. This culture was then diluted 1:100 in fresh LB broth and grown at 37 °C, shaking at 165 rpm until an $O.D_{600}$ of 0.4-0.5 was reached. The flask was then cooled in a shaking ice slurry for 20 minutes. Cells were to be kept cold for the remainder of the protocol. All centrifugation steps are assumed to be at 4 °C. 100 mL of culture was pelleted for 10 minutes at 6000 x *g*. The supernatant was discarded, and the cell pellet was resuspended in 50 mL dH_20 and once again pelleted for 10 minutes at 6000 x *g*. This step was repeated twice. The supernatant was discarded, and the cell pellet was resuspended in 2 mL 10% glycerol. 50 μL aliquots were then flash frozen on dry ice for long-term storage at -80 °C.

2.7.4 Electrocompetent Cell Transformation Protocol

Electrocompetent cell aliquots were thawed on ice for 10 minutes. 1-5 μL DNA was added to the aliquot. Cells were introduced into an ice-cold 1mm electroporation cuvette. Cells were electroporated at 1.80 kV, 200 Ω, and 25 μF. 900 μL of SOC broth, pre-warmed to 37 °C, was added to the cuvette and pipetted up and down fully twice to mix the two solutions fully. The solution was transferred to a 12 cm culture tube (Greiner) and incubated at 37 °C for 1 hour on a spinning wheel. Dilutions were then performed, and the culture was spread evenly onto LB-agar plates

supplemented with appropriate antibiotics. Plates were incubated overnight.

2.8 Conjugation

A donor strain (possessing a conjugative plasmid of interest) and a recipient strain (to receive the plasmid of interest) were grown overnight in 5 mL LB at 37 °C on a spinning wheel in the presence of appropriate antibiotics. Overnight cultures were pelleted via centrifugation for 1 minute at 6000 x *g*. Cells were resuspended in 1 mL LB broth after the supernatant was discarded. This wash step was repeated a further three times. 100 μL donor and recipient culture were added to filter paper placed on an LB agar plate. This plate was incubated at 37 °C for 4 hours. Following incubation, the filter paper was placed into a 50 mL falcon tube containing 1 mL LB broth. This was then vortexed vigorously for 30 seconds. The solution was then looped onto LB agar plates supplemented with appropriate antibiotics. The plate was then incubated overnight at 37 °C. Following incubation, successful transconjugant colonies could be picked.

2.9 Genetic Manipulation by P1 Transduction

2.9.1 Preparation of P1 Bacteriophage Lysate

Lysates of P1 bacteriophage were prepared from overnight cultures of *E. coli* carrying genetic markers of interest. An *E. coli* culture was diluted 1:40 in LB broth supplemented with 5 mM $CaCl_2$. Cultures were grown until they reached an $O.D_{600}$ of 0.7-0.8. 0.6% agarose was prepared in transformation tubes at 42 °C. 100 μL culture was introduced into transformation tubes, alongside a gradient of P1 phage stocks from 0-200 μL. This was poured onto P1 plates. These are LB-agar plates supplemented with 5 mM $CaCl_2$ and 0.13% glucose. Plates were incubated upright in the incubator overnight at 37 °C.

The two plates which produced the best lysis were scraped into a single 50 mL

falcon. 1 mL MC buffer [100 mM $MgSO_4.7H_2O$ and 5 mM $CaCl_2$] and 500 μL chloroform were added to the falcon. After thorough vortexing for 30 seconds, this was centrifuged at 7000 x *g* for 10 minutes. The supernatant was poured into a 15 mL falcon tube and topped up with a further 500 μ L chloroform. The pellet was discarded. The P1 lysate was then stored at 4 °C until use.

2.9.2 Genetic Manipulation by P1 Transduction

A recipient strain was inoculated into 5 mL LB broth. This was incubated overnight at 37 °C. The overnight culture was pelleted and resuspended in 1 mL MC buffer. The supernatant was discarded. 100 μL resuspension solution was added to separate tubes within a 37 °C water bath. A phage titration of prepared P1 lysate was added to these tubes. Following a 30 minute incubation, 300 μL 1 M sodium citrate was added to halt the infection process. Cultures were placed on ice briefly before being added to transformation tubes containing 4 mL 42 °C 0.6% agar. The mixture was poured evenly over an LB agar plate supplemented with appropriate antibiotics and 5 mM sodium citrate. Plates were incubated upright overnight in an incubator at 37 °C. Single colonies were picked and subcultured three consecutive times on LB agar plates supplemented with 5 mM sodium citrate and appropriate antibiotics to remove P1 phage. P1 transduction was then verified via colony PCR and/or Sanger sequencing.

2.9.3 FLP-frt Recombination

P1 transduction was carried out using the Keio collection as a donor strain for genetic knockouts in *E. coli* [214]. The FRT-KanR-FRT cassettes were removed via a FLP recombinase encoded on pCP20 [216]. pCP20 contains the FLP recombinase under a temperature sensitive promoter P_λ, with a temperature sensitive pSC101 *ori*. pCP20 was transformed into a strain of interest. Transformations were spread onto LB agar plates containing appropriate antibiotics and incubated overnight at 30 °C. Single colonies were picked and inoculated into 5 mL LB broth,

grown overnight at 42 °C on a spinning wheel, growing without antibiotic selection. Incubation facilitates FLP recombinase expression alongside plasmid curing. Following overnight growth, the culture was titrated and spread evenly over LB-agar plates. Following overnight incubation at 37 °C, colonies were patched onto LB-agar plates, plates containing ampicillin, and antibiotics containing the antibiotic marker of interest. Colonies that grew only on LB-agar plates were assumed to have been cured of pCP20 and the resistance cassette. This can then be verified through diagnostic colony PCR of the site of interest.

2.10 Spacer Acquisition Assays

2.10.1 Papillation Assay

Papillation assays were performed by transforming a Cas1-Cas2 expression vector plasmid into a reporter strain. Transformations were serially diluted onto papillation plates. Papillation plates are LB agar plates supplemented with 40 μg/mL X-Gal, 0.1% Lactose, 0.02% L-arabinose and relevant antibiotics unless otherwise stated. Plates were incubated at 37 °C for 5 days. Relative frequencies of spacer acquisition could then be determined via visual inspection, and representative colony images were then taken.

2.10.2 Spacer Integration (SPIN) Assay

SPIN assays were performed by transforming a Cas1-Cas2 expression vector into a strain of interest. Transformations were diluted to give single colonies following overnight incubation. Colonies were inoculated into 5 mL overnight cultures supplemented with relevant antibiotics and grown overnight at 37 °C on a spinning wheel. A 1:1000 dilution was performed into a 5 mL culture supplemented with appropriate antibiotics and 0.2 % L-arabinose (unless other conditions are specified). 1 mL of the overnight culture was pelleted by centrifugation and frozen for later analysis. The fresh culture is then grown as before, before being subject to

two more 1:1000 dilutions on subsequent days. Each day, 1 mL culture is taken for later analysis.

After three days of growth, pellets were resuspended in 100 μL LB broth. Taq colony PCRs of the CRISPR loci of interest were performed with 1 μL of the resuspension as templates. PCR reactions were then analysed via agarose gel electrophoresis using 1.8 % agarose gels stained with EtBr.

2.10.3 YFP Reporter Assay (Flow Cytometry)

This is a modified version of the method described by Amlinger *et al.* [213]. MLS989 was transformed with a pBAD Cas1-Cas2 expression vector. An individual colony was picked and inoculated into 5 mL LB broth supplemented with appropriate antibiotics. This was grown overnight at 37 °C on a spinning wheel. A 1:1000 dilution was performed into 5 mL LB broth supplemented with appropriate antibiotics and 0.2% L-arabinose. This was grown up as before. Over the following two days, two more 1:1000 dilutions were made. The day 3 culture was diluted 1:100 in LB broth. 200 μL was inoculated into a well of a 96-well plate. Fluorescence was measured using a Beckman Coulter CytoFLEX S. Cultures of MLS989 and MLS990 were also measured to set initial gating conditions. Samples were run at 60 μL/min until 20000 cells had been sorted. Data was analysed using the CytoExpert program. Data was plotted using PRISM Graphpad 10, and statistical comparisons were drawn using two-way t-tests. Datasets can be found in the Appendix.

2.10.4 YFP Reporter Assay (Plate Based)

This is a modified version of the method described by Amlinger *et al.* [213]. Day 3 cultures were prepared as previously described. Cultures were titrated and diluted to produce ≈200 colonies per plate. Cultures were plated onto LB agar plates containing appropriate antibiotics. Plates were incubated overnight at 37 °C. Fol-

lowing growth, plates were imaged for fluorescence using the Cy3 filter set on an Amersham typhoon plate reader. Visible light images of plates were also taken to facilitate comparison.

2.11 Culture Fluorescence Assays

2.11.1 Measurement of SOS Response Induction by GFP Fluorescence

Strains of interest were transformed with pRC2790 and plated onto LB agar plates containing appropriate antibiotics. Following overnight growth, individual colonies were picked and inoculated into LB broth supplemented with appropriate antibiotics and grown at °C on a spinning wheel overnight. Following overnight growth, cells were put on ice, and $O.D_{600}$ was determined. Cells were kept on ice for the remainder of the protocol. Cells were diluted to an $O.D_{600}$ of 1. 200 μL culture was inoculated into a well on a 96 well plate. LB-only controls were also introduced into separate wells. GFP fluorescence was measured using the Cy2 filter set on an Amersham typhoon plate reader. FIJI was used to determine the average image intensity across the well. Data was plotted using PRISM Graphpad 10, and statistical comparisons were drawn using two-way t-tests. Datasets can be found in the Appendix.

2.11.2 Measurement of YFP Fluorescence

Cells were either streaked or transformed and plated onto LB agar plates supplemented with appropriate antibiotics. Following overnight growth, individual colonies were picked and inoculated into LB broth supplemented with appropriate antibiotics and grown at °C on a spinning wheel overnight. Following overnight growth, cells were put on ice, and $O.D_{600}$ was determined. Cells were kept on ice for the remainder of the protocol. Cells were diluted to an $O.D_{600}$ of 1. 200 μL culture was

inoculated into a well on a 96 well plate. LB-only controls were also introduced into separate wells. GFP fluorescence was measured using the Cy3 filter set on an Amersham typhoon plate reader. FIJI was used to determine the average image intensity across the well. Data was plotted using PRISM Graphpad 10, and statistical comparisons were drawn using two-way t-tests. Datasets can be found in the Appendix.

2.12 Sequencing of Spacers

2.12.1 Solid Culture Derived Spacer Protocol

2.12.1.1 Low-throughput sequencing

A papillation assay was conducted as previously described with 0.002% L-arabinose. A plate was scraped, and colonies were resuspended in 2 mL M9 salts. Resuspensions were vortexed vigorously for 20 seconds. 150 μL Taq colony PCRs were performed using the primers described. 8 μL PCR was used to verify the product of interest was generated via agarose gel electrophoresis. The remaining PCR product was purified via a PCR clean-up kit. The product was sent for sequencing using the Genewiz PlasmidEZ service.

2.12.1.2 High-throughput sequencing

A papillation assay was conducted as previously described with 0.002% L-arabinose. Ten plates were scraped, and colonies were resuspended in 20 mL M9 salts. Resuspensions were vortexed vigorously for 20 seconds. 150 μL Q5 colony PCRs were performed using the primers described. 8 μL PCR was used to verify the product of interest was generated via agarose gel electrophoresis. The remaining PCR product was purified via a PCR clean-up kit. The product was sent for sequencing using the University of Nottingham Deep Sequencing nanopore service.

2.12.2 Liquid Culture Derived Spacer Protocol

2.12.2.1 Low-throughput sequencing

A strain of interest was transformed with a pBAD Cas1-Cas2 expression vector and plated on LB agar containing appropriate antibiotics. Following overnight incubation at 37 °C, an individual colony was inoculated into 5 mL LB broth supplemented with appropriate antibiotics. Following overnight growth at 37 °C, a 1:1000 dilution was made into 5 mL LB broth supplemented with appropriate antibiotics and 0.2% L-arabinose. Over the two successive days, two further 1:1000 dilutions were made. The day 3 solution was titrated and plated evenly on M9 minimal media plates supplemented with 0.1% lactose. The plate with the most individual colonies (544) was scraped and resuspended in 2 mL M9 salts. Resuspensions were vortexed vigorously for 20 seconds. 150 μL Taq colony PCRs were performed using the primers described. 8 μL PCR was used to verify the product of interest was generated via agarose gel electrophoresis. The remaining PCR product was purified via a PCR clean-up kit. The product was sent for sequencing using the Genewiz PlasmidEZ service.

2.12.2.2 High-throughput sequencing

A strain of interest was transformed with a pBAD Cas1-Cas2 expression vector and plated on LB agar containing appropriate antibiotics. Following overnight incubation at 37 °C, an individual colony was inoculated into 5 mL LB broth supplemented with appropriate antibiotics. Following overnight growth at 37 °C, a 1:1000 dilution was made into 5 mL LB broth supplemented with appropriate antibiotics and 0.2% L-arabinose. Over the two successive days, two further 1:1000 dilutions were made. 50 μL Taq colony PCRs were performed using the primers described. The PCR was run down an agarose gel, and the parental and +1 acquisition event bands were separated via agarose gel electrophoresis. The +1 band was excised and extracted from the gel using a gel extraction kit. 15 ng of the product was

amplified in a 150 Q5 PCR reaction. 8 μL PCR was used to verify the product of interest was generated via agarose gel electrophoresis. The remaining PCR product was purified via a PCR clean-up kit. The product was sent for sequencing using the University of Nottingham Deep Sequencing nanopore service.

2.12.3 Data Extraction from Sequencing

Data was returned in FASTQ format. A sequence comprised of the terminal 10 bp of the leader and the 28 bp repeat sequence was queried against the FASTQ file contents. Degeneracy of up to three mismatches was permitted in the search. Accounting for strand directionality, the 33bp downstream of the query was extracted into a library of spacers. Duplicate spacers were purged from the library, but counts were taken and assigned to each spacer. BLAST libraries were prepared for the genome and plasmid contents of the source culture. Spacer libraries were BLAST searched against these libraries. The directionality and location of mapped spacers were noted and used to assign PAM motifs to each spacer. Python scripts are available at DOI: 10.5281/zenodo.14911635.

Chapter 3

Biases in Genome- and Plasmid-derived Spacer Distributions

3.1 Introduction

3.1.1 Methods of Detecting Spacer Acquisition

Several techniques have been developed to visualise spacer acquisition [217]. The most widely used is the Spacer Integration (SPIN) assay, in which acquisition is observed by PCR across the CRISPR locus [Figure 3.1] [218]. In the *E. coli* Type I-E system adaptation expands the CRISPR array by 61 bp, from the acquisition of a novel 33 bp spacer and the duplication of the 28 bp repeat.

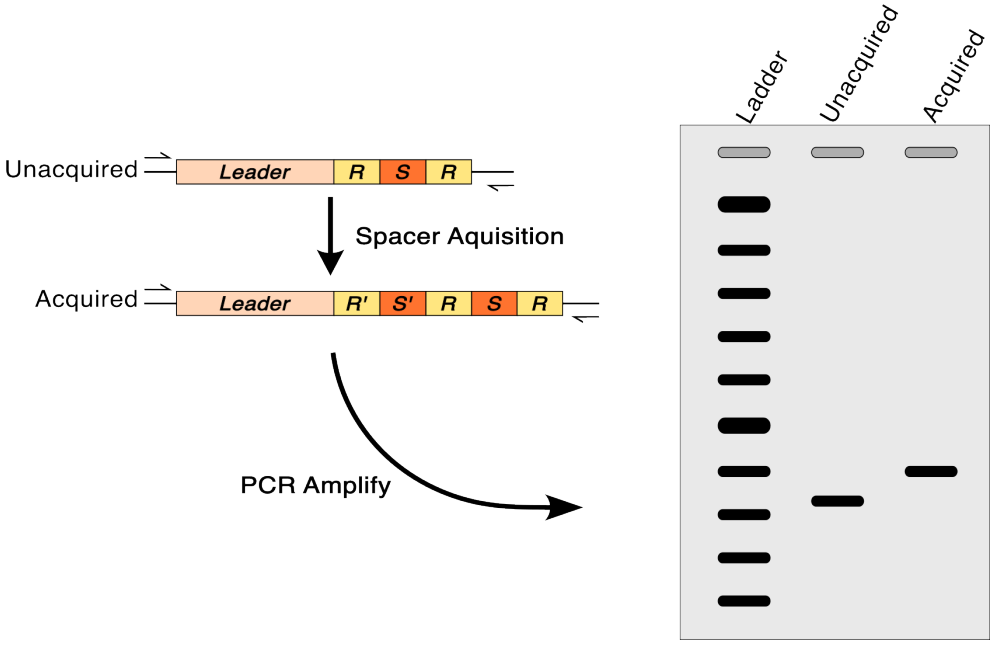

Figure 3.1: Overview of the SPIN assay.
Expansion of the CRISPR array through spacer acquisition can be visualised via PCR. Flanking the leader sequence (brown), S and R refer to the original spacers (orange) and repeats (yellow), respectively, while S' and R' refer to novel spacers and repeats. PCRs can be visualised via agarose gel electrophoresis.

Whilst the SPIN assay can detect the expansion of CRISPR loci within a population, its reliance on PCR amplification presents several drawbacks [217]. Firstly, it lacks sensitivity, with expanded loci required to comprise as much as 5% of a population before they become detectable. Secondly, the smaller (and typically much more abundant) unexpanded locus is more efficiently amplified. These prevent the use of the SPIN assay to analyse rates of spacer acquisition quantitatively.

Several alternative assays using reporter genes have been developed to circumvent the SPIN assay's limitations. These assays use a promoter to drive transcription and translation from a start codon through a minimal array into a reporter gene. In the unexpanded state, stop codons within the CRISPR array's leader sequence prevent the reporter gene's expression. However, expansion of the CRISPR array shifts the reading frame such that the reporter gene is expressed fused to an N-terminal peptide encoded by the array itself [Figure 4.6]. To date, a chloramphenicol resistance protein (CmR) and yellow fluorescent protein (YFP) have been used as reporter genes [219] [213].

These reporter assays can be more than 1000-fold more sensitive than SPIN assays, and are able to detect spacer acquisition at the individual cell level [219] [213]. However, the workflows involved with these systems are slow and labour-intensive, making them poorly suited to high-throughput analysis.

3.1.2 A Papillation-Based Reporter System

To overcome the limitations of the CmR- and YFP-based reporter systems, a papillation assay was developed for the high-throughput study of spacer acquisition in the *E. coli* Type I-E CRISPR system [Figure 4.6] [212]. A constitutive promoter drives transcription through the Type II (A) CRISPR array into a *lacZ* reporter. Stop codons within the leader prevent translation of LacZ when the CRISPR array is unexpanded. Expansion of the CRISPR array by 61 bp moves the stop codons out of frame and places *lacZ* in frame.

Growth of *E. coli* on LB-agar is carbon limited [220]. Cells expressing LacZ have a growth advantage in the presence of lactose and form papillae, or microcolonies, on the main colony [221]. Supplementation of X-gal into the media gives papillae a blue phenotype [222].

Papillae correspond to individual spacer acquisition events. This has been verified through the isolation and PCR amplification of the reporter locus CRISPR arrays

found within papillae [212]. Visual inspection of the number of papillae on the surface of a colony is a semi-quantitative measure of spacer acquisition.

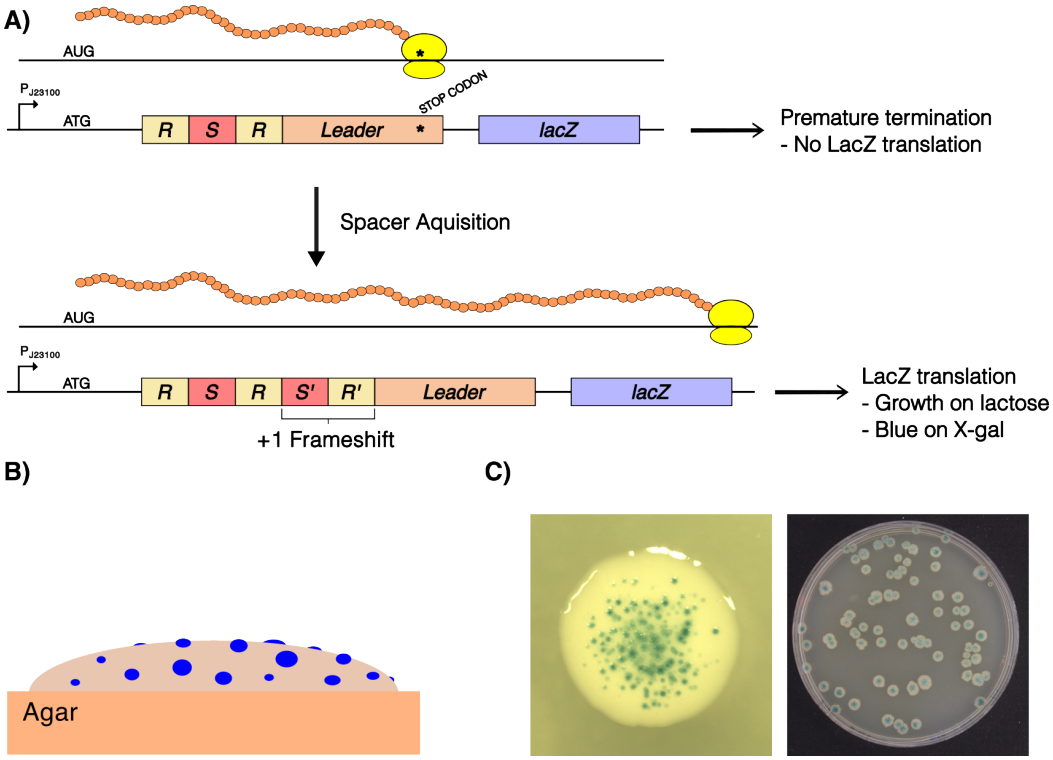

Figure 3.2: A papillation assay to detect spacer acquisition.
(A) A promoter drives expression from an ATG start codon through a minimal CRISPR array into *lacZ*. Initially, the *lacZ* gene is not expressed due to a stop codon within the leader. Acquisition of the 61 bp spacer-repeat unit expands the CRISPR array, moving the stop codons out of the *lacZ* reading frame. The *lacZ* gene is now expressed. The ribosome is represented by a pair of yellow circles. The translated polypeptide chain is represented by the string of orange circles. The leader sequence is denoted in brown. Spacers are in red, and repeats are in yellow.
(B) A diagram of a single colony of the LacZ-based papillation reporter strain. Expression of LacZ results in the formation of blue papillae when grown on LB agar plates supplemented with X-gal and lactose. Papillae appear as small raised blue bumps on the surface of the colony.
(C) An image of a single colony of the LacZ-based reporter strain following incubation at 37°C for 5 days in the presence of Cas1-Cas2 expression. Adjacent is a full plate of papillated colonies.

Whilst every papilla will derive from a spacer acquisition event, not every spacer acquisition event will yield a papilla. Several factors dictate the relationship between the rate of adaptation and papillation. Firstly, 5% of spacer acquisition events are not exactly 61 bp in size and will therefore not correct the *lacZ* reading frame [219].

Secondly, spacer acquisition will add 12 new codons into the lacZ reading frame. There is a 3/64 probability that each of them will be a stop codon. However, only 10 of the 12 codons can encode a stop signal because they are constrained by the first base of the repeat and the conserved G at the start of most spacers. The probability that acquisition of a 61 bp spacer-repeat will lead to LacZ expression is therefore:

$(61/64)^{10} \times 0.95 = 0.59$

The true probability is likely lower than this because the spacer-repeat segment could interfere with *lacZ* expression in other ways, such as destabilising the fusion protein.

3.1.3 Disparities in Spacer Selection Biases

Prior work investigating the nature of spacers acquired by the papillation reporter identified all ten sequenced spacers as deriving from the host chromosome [212]. Since genomic DNA is present in the cell at about a 35:1 ratio compared to the plasmid present in this work, we would expect all ten spacers to derive from the genome in 75.0% of instances if spacer selection followed a binomial distribution [212] [223]. Therefore, these results suggest an absence of bias between the genome and plasmid in spacer selection. However, a bias towards the acquisition of plasmid-derived spacers has been reported in the literature [114] [213]. For example, 88% of spacers acquired by a YFP-based reporter derived from the plasmid [213].

Spacer selection from the host chromosome is thought to be biased against by the enrichment of *chi* sites on the genome, disfavouring the production of spacer substrate by RecBCD [114]. *Chi* sites prevent extensive degradation at double-strand breaks (DSB) by RecBCD. Mobile DNA elements, which are not enriched for *chi* sites, suffer more extensive degradation by RecBCD at DSBs, generating more spacer substrate [114].

The established bias against chromosomal spacer acquisition and its hypothesised underpinning mechanisms, identified in prior work, are at odds with the fact that the first ten papillae sequenced by a previous PhD student in the lab all harboured chromosomal spacers [212]. I characterised spacer acquisition in the papillation reporter to identify the biases present.

3.2 Results

3.2.1 Genome vs Plasmid Derived Bias

A 100-1000 fold bias towards the acquisition of spacers from the plasmid was reported in a prior study assessing rates of acquisition of plasmid- and genome-derived spacers during growth in liquid culture [114]. As prior work with the papillation reporter strain, JB028, was conducted on solid LB-agar media, this raised the question of whether acquisition in the CRISPR-Cas system behaved differently depending on culture conditions [224]. To address this, I benchmarked the papillation assay by sequencing spacers acquired by the JB028 reporter strain during growth on solid media versus in liquid culture.

First, I benchmarked spacer acquisition during growth in liquid culture. JB028 was transformed with a pBAD Cas1-Cas2 expression vector and grown over three days in LB supplemented with 0.2% L-arabinose inducer. A daily 1:1000 dilution was performed. A titration of the day three culture was plated onto M9 minimal media supplemented with 1% lactose as a carbon source. Therefore, only those cells that

had undergone adaptation could express *lacZ* and form colonies [Figure 3.3A]. Plates were incubated at 37 °C for 3 days. The plate with the highest number of individual colonies (554) was scraped, and the bacteria were resuspended in 2 ml M9 salts. Two PCRs were performed to amplify the reporter CRISPR loci present in the colony resuspension, producing products either 2 kb or 4 kb in length. The CRISPR array was close to the end of the 2 kb product and in the middle of the 4 kb product [Figure 3.3B]. PCR products were sequenced using a relatively shallow commercial nanopore sequencing service.

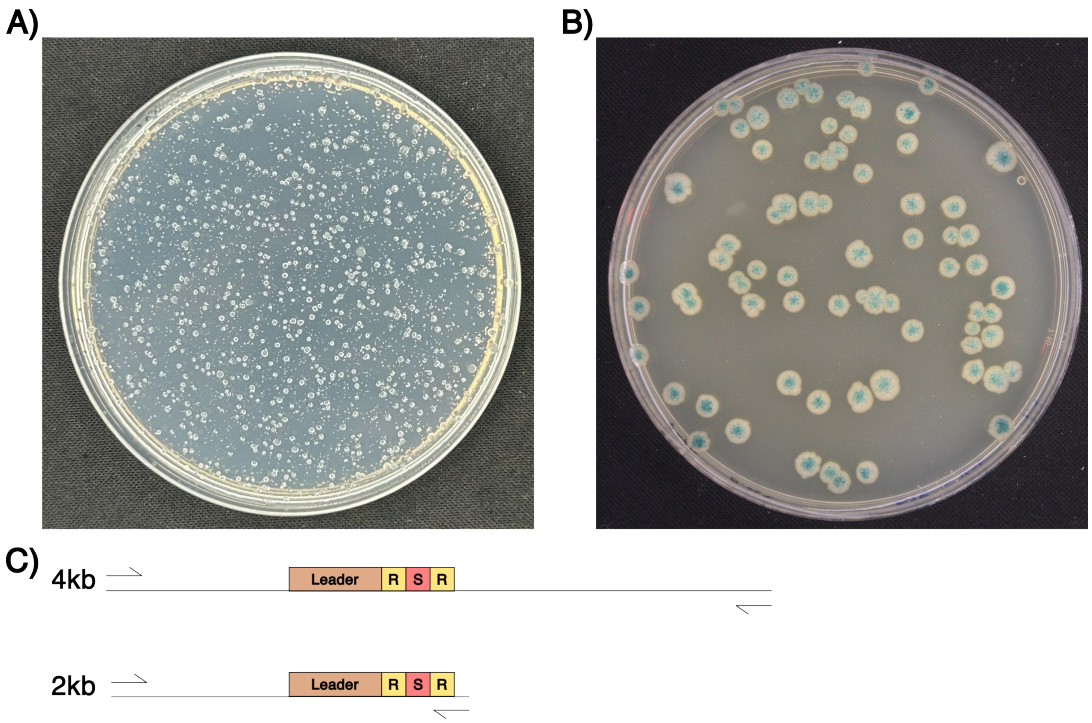

Figure 3.3: Exemplar plates and PCR strategy for the sequencing of spacers.
(A) JB028 was transformed with a pBAD Cas1-Cas2 plasmid. L-arabinose was present at 0.2%. Cultures were grown overnight, followed by a daily 1:1000 dilution for three days. The culture was titrated and plated onto M9 minimal media supplemented with 1% lactose. Plates were incubated at 37°C for three days. The plate which produced the largest number of individual colonies is shown. A total of 554 colonies are present.
(B) A papillation assay was conducted in JB028 transformed with a pBAD Cas1-Cas2 expression vector. L-arabinose was present at 0.002%. Plates were incubated at 37 °C for five days. An image of the whole plate following five days of incubation is shown.
(C) Overview of the orientation and location of reporter CRISPR arrays in the 2 kb and 4 kb products of the two PCRs, conducted to identify optimised conditions for the nanopore sequencing of spacers.

Analysis of nanopore sequencing results identified a library of novel spacers acquired into the JB028 CRISPR reporter locus. Spacer libraries were BLAST searched against the JB028 genome and pBAD Cas1-Cas2 plasmid. If a spacer mapped to a site present on both genome and plasmid, such as the P_{Ara} promoter, Cas1, or Cas2, it was assumed to have derived from the plasmid. Duplicate spacers were removed, as per (Levy *et al.*, 2015) [114]. Nanopore sequencing of the 2 kb PCR product yielded more mappable spacers than the 4 kb product [Table 3.1].

PCR Product	Spacers Mapping to Genome	Spacers Mapping to Plasmid	Total Spacers
4 kb	9	13	22
2 kb	37	76	113
Cumulative	46	89	135

Table 3.1: Distribution of genome- and plasmid-derived spacers for PCR reactions following spacer acquisition in liquid culture.

To assess spacer acquisition rates during growth on solid media, a papillation assay was performed in JB028 transformed with a pBAD Cas1-Cas2 expression vector [Figure 3.3B]. L-arabinose inducer was present at 0.002%. A plate was incubated at 37 °C for five days. The plate was resuspended in 2 ml M9 salts, and a PCR of the reporter CRISPR locus was performed using the primer pair which generated a 2 kb product. The PCR product was sequenced and analysed as previously described.

Media	Spacers Mapping to Genome	Spacers Mapping to plasmid	Plasmid Bias (-fold)	Total Spacers
Liquid Culture	46	89	23.0	135
Solid Culture	134	75	12.4	209

Table 3.2: Source of spacers following acquisition in different media types.

Solid media culture exhibited a reduced bias for the acquisition of spacers from the plasmid compared to liquid culture [Table 3.2]. However, the rate of plasmid-derived spacers acquired during growth on LB-agar was still 12.4-fold greater than that expected in the absence of bias (35.9% vs 2.9%). It remained unclear how the absence of a bias was identified in prior work with JB028 [212].

3.2.2 Characterisation of Plasmid-Derived Spacers

Spacer acquisition has been shown to exhibit a bias towards regions proximal to origins of replication [114]. I analysed how the spacers in Table 3.2 were distributed around the plasmid relative to the ColEI origin of replication [Figure 3.4]. Spacers acquired during growth on liquid and solid media were enriched proximal to the unidirectional *ori*. This bias was more pronounced for the spacers acquired in liquid culture, with the first 1000 bp downstream of the origin providing a source of spacers at five-fold the rate of the plasmid's remaining 4277 bp (44 spacers per 1000 bp vs 9 spacers per 1000 bp).

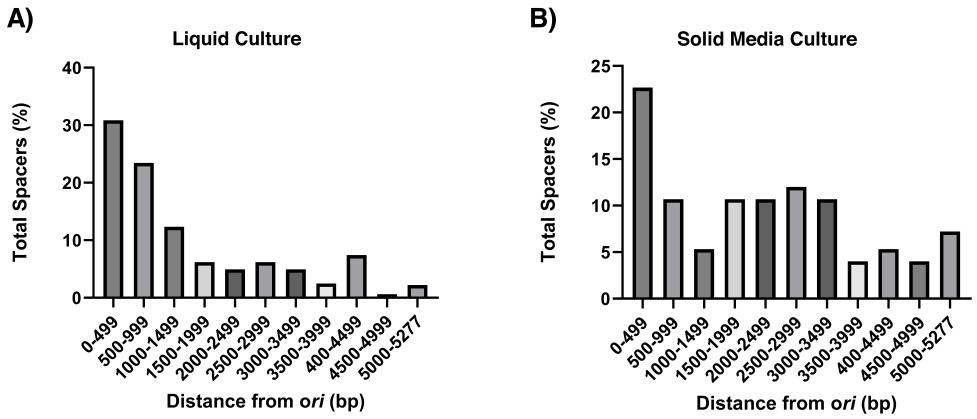

Figure 3.4: Percentage frequency of spacer acquisition from different regions of the plasmid relative to their distance from the *ori*.
(A) The spacers mapping to the plasmid in Table 3.2 were binned depending on their location proximal to the *ori*. The percentage spacer distributions around the plasmid for spacers acquired during growth in liquid culture are shown.
(B) Spacers were binned as described in (A). The percentage spacer distributions around the plasmid for spacers acquired on solid media are shown.
Each bin comprised a 500 bp region of the plasmid. Spacers derived from 5000-5277 have been weighted to reflect the reduced size of this bin.
Data values are found in Supplementary Table S2.

The enriched selection of spacers near the *ori* has previously been attributed to dosage effects during DNA replication [114]. However, this bias has only been reported for the genome [114], with analysis of naive [73] or primed [102] selection from plasmid DNA not describing this bias. It was unclear whether the bias towards the selection of spacers proximal to the *ori* could be fully explained by DNA dosage effects [Figure 3.4].

I performed DNA motif analysis on the fidelity of Protospacer Adjacent Motifs (PAMs) associated with spacers mapping to different plasmid regions [Figure 3.5A]. The small sample size of spacers in the 500 bp bins distal from the origin in Figure 3.4 necessitated their clustering into a single larger group. Origin-proximal PAMs were those described as deriving from spacers mapping to the first 500 bp from the *ori*, whilst origin-distal PAMs were those deriving from spacers mapped to the furthest 2500 bp from the unidirectional *ori*. Each group contained 19-28 spacers. The plasmid and genome groups contained the PAMs for all the respective spacers reported in Table 3.2.

The E. coli Type I-E consensus PAM is AAG [73] [Figure 3.5B]. The consensus for adenine at bases 1 and 2 is much weaker than the consensus for guanine at the third base. Spacers acquired during growth in solid media exhibited a stronger preference for the consensus PAM than those acquired during growth in liquid culture. Under both growth conditions, spacers acquired from the plasmid in sequences proximal to the *ori* exhibited a reduced fidelity to the consensus PAM when compared to those acquired from sequences distal to the *ori*. This suggests that an increased frequency of spacer acquisition near the origin of replication may derive from a reduced PAM preference during spacer selection from these sequences. The low sample size of spacers analysed necessitated deeper sequencing to corroborate these findings.

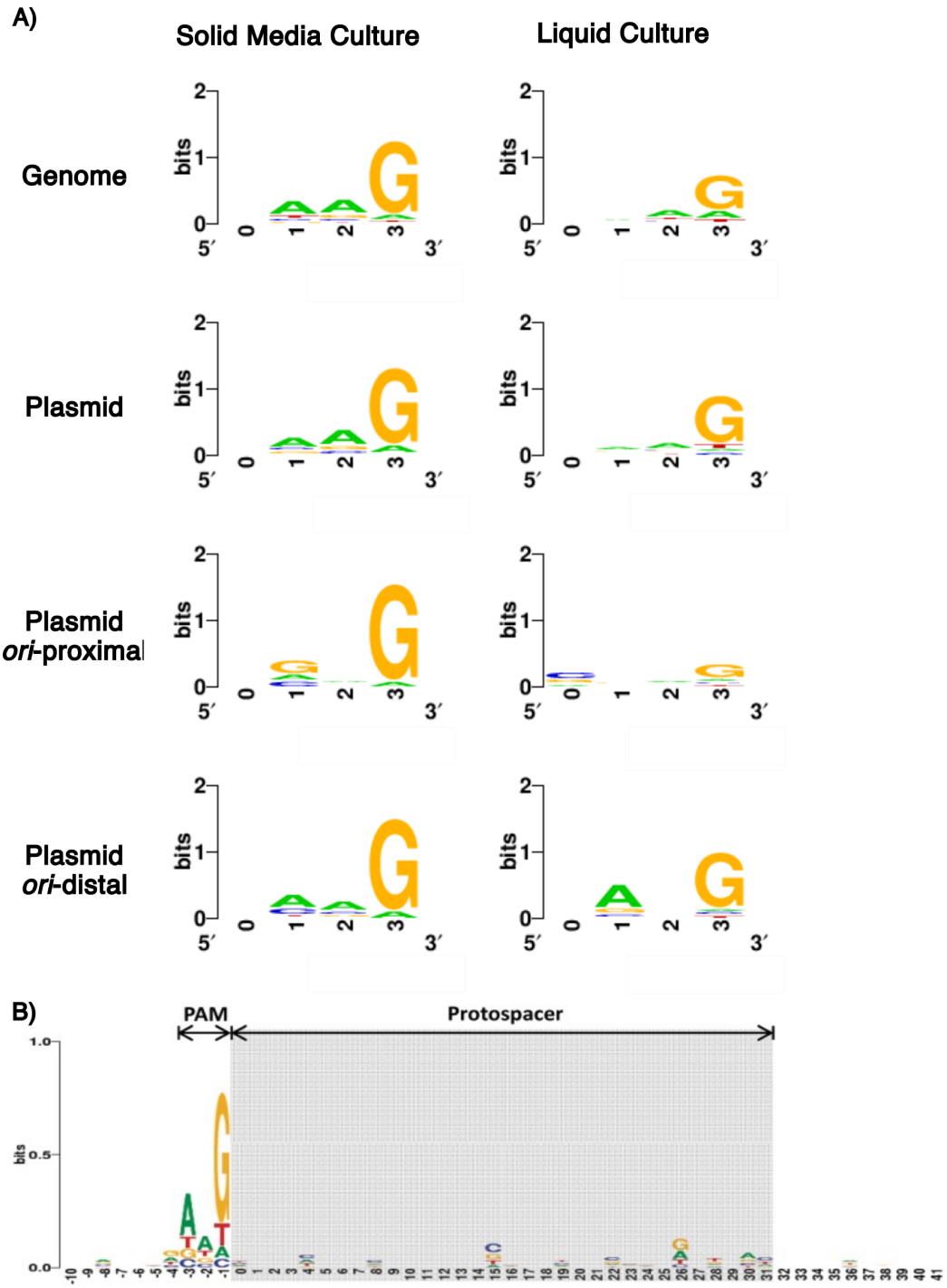

Figure 3.5: DNA motif analysis of PAM motifs of various spacer populations.
(A) DNA motif analysis was performed for the PAM sequences of spacers acquired during liquid and solid media growth. Analysis was conducted for all spacers mapped to JB028 (Genome), the pBAD Cas1-Cas2 expression vector (Plasmid), the 500 bp downstream of the origin (Plasmid *ori*-proximal) and the 2500 bp furthest from the origin (Plasmid *ori*-distal). An initial base at site 0 was included as a control, for which there should be no strong consensus. DNA motif analysis was conducted using WebLogo [225].
(B) Results of the initial DNA consensus analysis carried out to determine the PAM in the *E. coli* Type I-E CRISPR-Cas system. This work preceded recognition of the terminal PAM 'G' as the first base of the 33bp spacer, hence why this is at position -1 and not 0. Taken from [73].

3.2.3 High-throughput Plasmid Spacer Sequencing

Deeper sequencing of spacers acquired at the CRISPR reporter locus of JB028 was conducted. A liquid culture with Cas1-Cas2 expression induced over three days was prepared as previously described. A papillation assay was conducted, as previously described. Following incubation, the plate was resuspended and pooled in a total volume of 2 mL M9 salts. CRISPR reporter loci were then PCR amplified and sequenced via a deeper commercial nanopore service. Spacer libraries were generated and mapped as previously described. The distribution of mapped spacers is described in Table 3.3.

Source of Spacers	Spacers Mapping to Genome	Spacers Mapping to Plasmid	Plasmid Bias (-fold)	Total Spacers
Liquid Culture (Unique)	2811	1671	12.9	4482
Solid Culture (Unique)	26817	1709	2.1	28526

Table 3.3: Distribution of genome- and plasmid-derived unique spacers from the deep sequencing of a liquid culture and papillation assays.

Deeper sequencing revealed a reduced bias towards the acquisition of spacers from the genome in both liquid and solid media cultures compared to shallower prior sequencing [Tables 3.2 and 3.3]. This was an artefact derived from the removal of duplicate spacers during the mapping process. Duplicate spacers were removed to mitigate biases in spacer counts introduced by the papillation assay, PCR, and sequencing. Spacer counts with duplicate included, and respective biases are described in Table 3.4.

Source of Spacers	Spacers Mapping to Genome	Spacers Mapping to Plasmid	Plasmid Bias (-fold)	Total Spacers
Liquid Culture (With Duplicates)	38580	69481	22.5	108061
Solid Culture (With Duplicates)	132754	15983	3.8	148737

Table 3.4: Distribution of genome- and plasmid-derived duplicate spacers from the deep sequencing of a liquid culture and papillation assays.

Shallow and deep sequencing of liquid cultures exhibited comparable biases towards the acquisition of spacers from the plasmid when duplicate spacers were included in spacer counts [Tables 3.2 and 3.4]. In comparison, spacers identified in the deep sequencing of solid media cultures exhibited a reduced bias towards acquisition from the plasmid, in contrast to the results of shallower sequencing [Tables 3.2 and 3.4]. This reduced bias likely derives from the over-representation of spacers encoding genomically derived RBS and start codon sequences. Eight of the ten most frequently identified spacers from the deep sequencing of papillation assay CRISPR loci contained these sequences [Supplementary Figure S3]. These spacers may increase *lacZ* expression, conferring a growth advantage over other papillae in the papillation assay and leading to their over-representation in the dataset.

I analysed the distribution of the spacers described in Table 3.4 around the plasmid relative to the ColEI origin of replication [Figure 3.6]. Duplicate spacers were included in this analysis. A strong bias towards acquisition proximal to the *ori* was seen, concordant with the results of shallower sequencing [Figure 3.4]. Again, this bias was more pronounced for spacers acquired during growth in liquid culture.

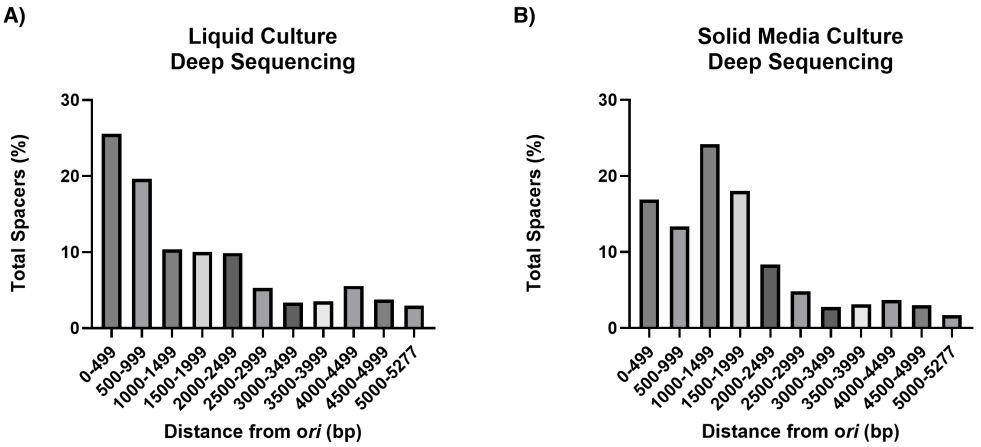

Figure 3.6: Percentage frequency of spacer acquisition from different regions of the plasmid relative to their distance from the *ori* - Deep sequencing.
(A) The spacers mapping to the plasmid in Table 3.4 were binned depending on their location proximal to the *ori*. The percentage spacer distributions around the plasmid for spacers acquired during growth in liquid culture are shown. Duplicate spacers were included.
(B) Spacers were binned as described in (A). The percentage spacer distributions around the plasmid for spacers acquired on solid media are shown.
Each bin comprised a 500 bp region of the plasmid. Spacers derived from 5000-5277 have been weighted to reflect the reduced size of this bin.
Data values are in Supplementary Table S2.

Consensus analysis was performed on the PAMs of spacers identified from the deep sequencing of the CRISPR reporter locus [Figure 3.7]. Spacers, including their duplicates, were clustered as described in Figure 3.5. The disparity in PAM fidelity between regions distal and proximal to the unidirectional *ori*, identified in Figure 3.5, was not present in Figure 3.7. This prior observation was likely an artefact from the small datasets with which PAM DNA motif analysis was initially conducted.

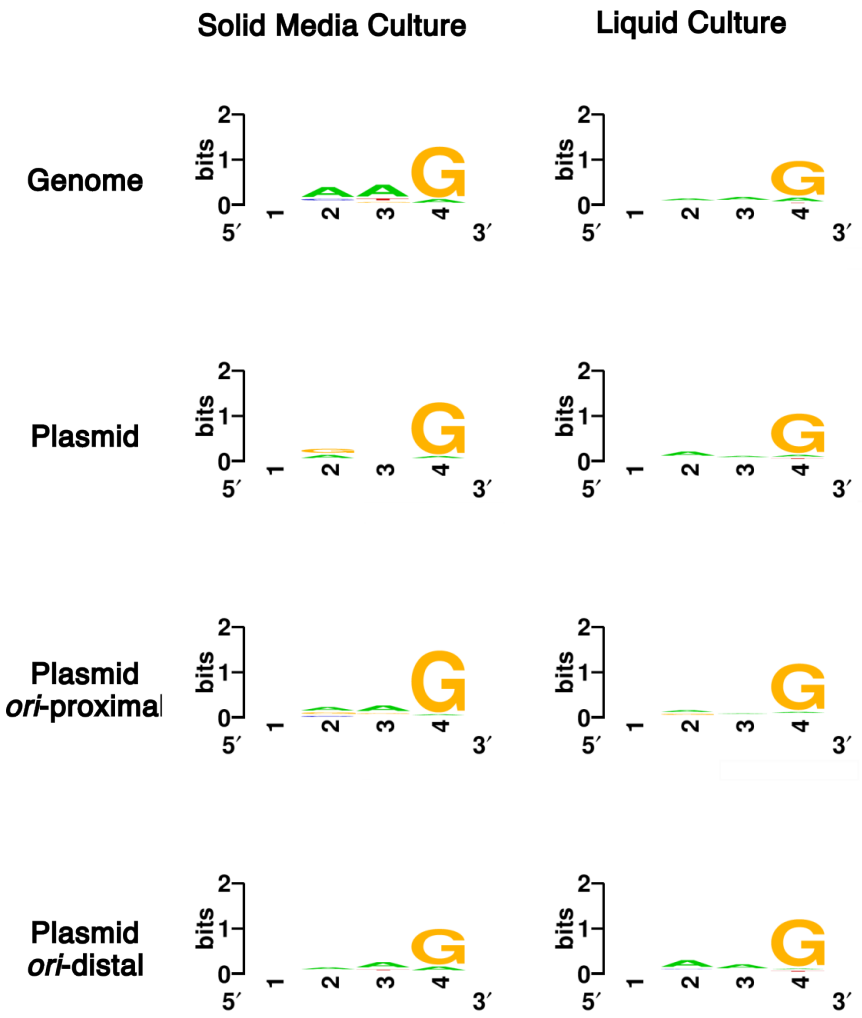

Figure 3.7: DNA motif analysis of PAM motifs of deep sequenced spacer populations.
DNA motif analysis was performed for the PAM sequences of spacers acquired during liquid and solid media growth. Analysis was conducted for all spacers mapped to JB028 (Genome), the pBAD Cas1-Cas2 expression vector (Plasmid), the 500 bp downstream of the origin (Plasmid ori-proximal) and the 2500 bp furthest from the origin (Plasmid ori-distal). DNA motif analysis was conducted using WebLogo [225].

3.2.4 Identification of Hotspots for Spacer Acquisition

Hotspots for spacer acquisition in the genome have been identified near the *ori*, at the terminus and at CRISPR arrays [114]. Spacers were mapped to the chromosome and plasmid to validate these locations as hotspots and identify any further sites [Figures 3.8 and 3.9].

Hotspots on the chromosome were identified from the mapping of spacers derived during growth on solid media and in liquid cultures [Figure 3.8A and 3.9A]. These hotspots were identified around the origin, at the terminus, and at the two CRISPR arrays present on the genome. The sequencing of spacers acquired at the *argE*:: CRISPR reporter locus identified the Type I (T) CRISPR array as a hotspot. The remaining hotspots derive from 16S and 23S rRNA genes alongside cryptic prophage insertion sequences. The conserved nature of these sites around the genome increases the likelihood of a spacer mapping, therefore over-representing them in the final sample.

Plasmids exhibited a unidirectional bias in spacer distribution starting from the ColEI *ori*. No *chi* sites are present on the plasmid to act as potential hotspots of acquisition. No clear individual hotspots were identified in the plasmid [Figure 3.8B and 3.9B].

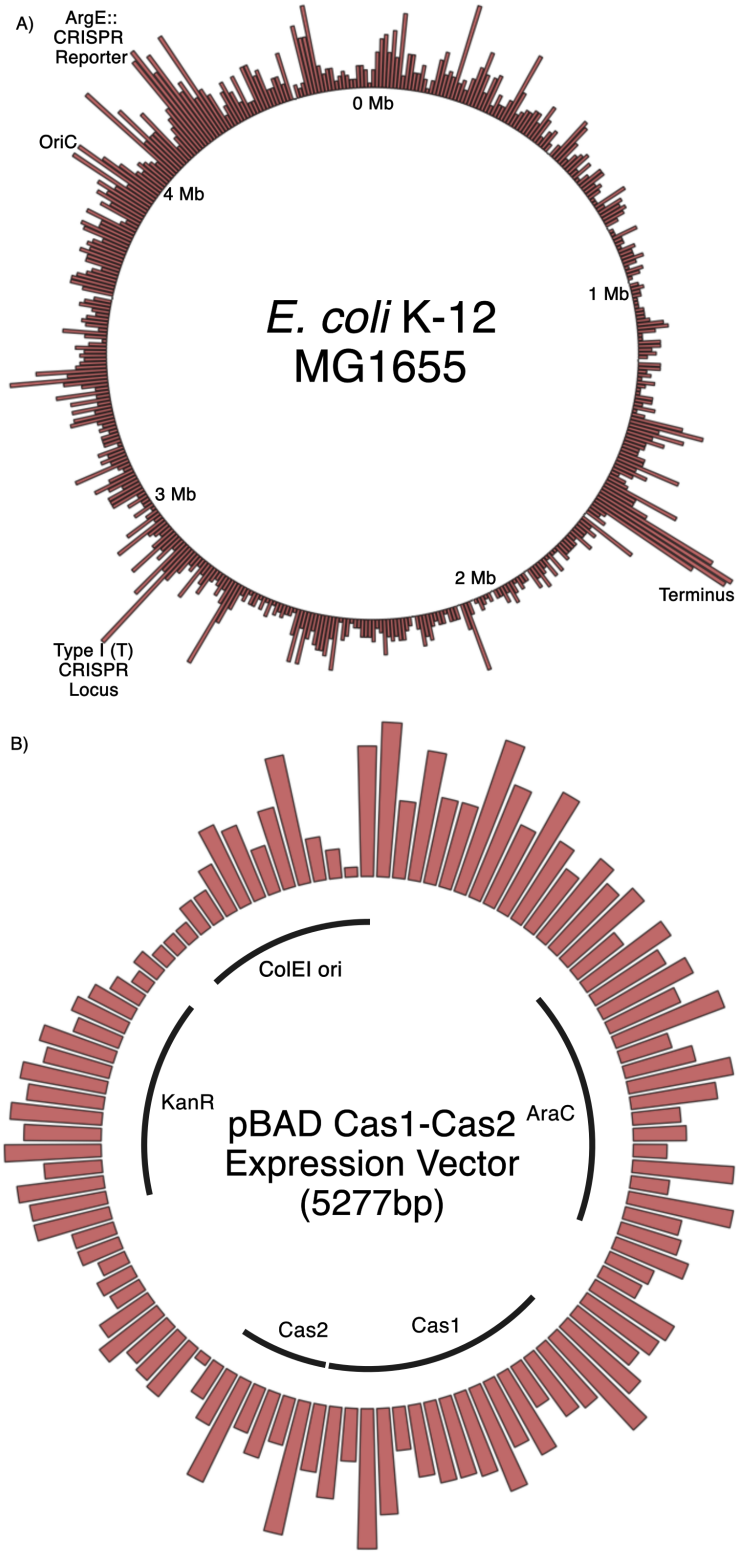

Figure 3.8: Hotspot analysis of spacers mapping to the chromosome and plasmid, acquired during growth in liquid culture.
(A) Spacers were mapped to the chromosome. Spacers were binned into 10 kb regions. The chromosome commences from the origin of transfer (*oriT*) at 0°. Hotspots are annotated.
(B) Spacers were mapped to the plasmid. Spacers were binned into 50 bp regions. The plasmid is annotated with its contents.

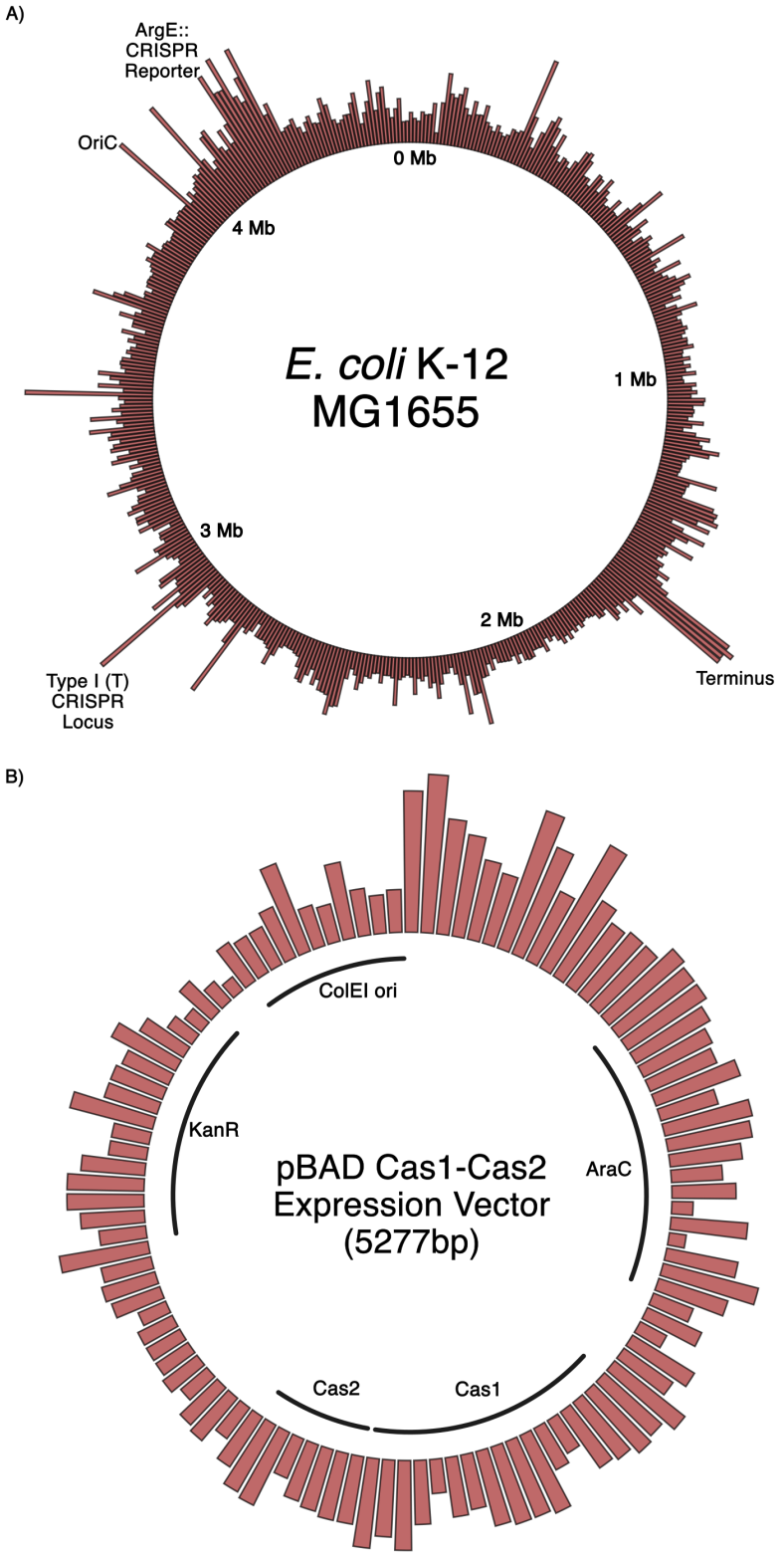

Figure 3.9: Hotspot analysis of spacers mapping to the chromosome and plasmid, acquired during growth on solid media.
(A) Spacers were mapped to the chromosome. Spacers were binned into 10 kb regions. The chromosome commences from the origin of transfer (*oriT*) at 0°. Hotspots are annotated.
(B) Spacers were mapped to the plasmid. Spacers were binned into 50 bp regions. The plasmid is annotated with its contents.

3.3 Chapter Summary

- The papillation assay shows significant bias towards the acquisition of plasmid-derived spacers. This is concordant with the literature and contrary to prior results with the strain.
- Bias towards the acquisition of spacers from the plasmid is more pronounced during growth in liquid culture than on solid media.
- Spacer acquisition is enriched proximal to the plasmid ColEI origin of replication.
- The increased rate of spacer acquisition proximal to the *ori* is not due to a reduction in the fidelity of PAM selection in this region.
- Spacer acquisition from the genome is enriched proximal to *oriC*, active CRISPR arrays, and the terminus.

Chapter 4

Discovery and Characterisation of Cas1-Cas2 Integrase Mutants with a Hyperactive Spacer Acquisition Phenotype

4.1 Introduction

4.1.1 The Detection of Hyperactive Cas1-Cas2 Mutants

The process of integrating novel spacers into the CRISPR array is mechanistically similar to transposition [77]. Hyperactive variants of transposases have frequently been identified via mutagenesis [226][227][228]. Such mutants are a product of the pleiotropic nature of transposases, as damage to the host provides selective pressure against greater amplification rates. Cas1-Cas2 exhibits a similar pleiotropy, as expression activates the SOS response and may induce non-specific integrase activity against the genome [229] [81]. This suggests acquisition rates of Cas1-Cas2 may be reduced to mitigate autoimmunity. Recently, Cas1-Cas2 mutants with hyperactive spacer acquisition phenotypes have been identified [230] [218]. Systematic mutagenesis of the *Streptococcus pyogenes* Cas1-Cas2-Csn2 operon identified a suite of mutants [230], whilst a phage-based enrichment protocol applied to *E. coli* Cas1-Cas2 identified hyperactive mutants possessing up to seven-fold increased rates of spacer acquisition [218]. I use a papillation assay to screen a mutagenised Cas1-Cas2 library and identify further hyperactive mutants.

4.2 Results

4.2.1 Assessment of Papillation Rates at Different Cas1-Cas2 Expression Levels

A low rate of papillation is necessary to identify hyperactive mutants. In prior work [Figure 3.3B], a ColEI-based pBAD Cas1-Cas2 expression vector was used to produce a high rate of papillation. I performed a papillation assay in the chromosomal reporter strain JB028, using this plasmid and a P15A-based $pMar_{Ara}$ Cas1-Cas2 expression vector [Figure 4.1]. The inducer was titrated to identify the optimal conditions for the identification of hyperactive phenotypes.

The $pMar_{Ara}$ promoter was evolved from the pBAD promoter to reduce leaky expression and increase titraitability [214]. These attributes are reflected in their rates of papillation across a range of arabinose concentrations [Figure 4.1]. $pMar_{Ara}$, at an inducer concentration of 0.002%, was selected to provide Cas1-Cas2 expression in a screen for Cas1 mutants with a hyperactive phenotype.

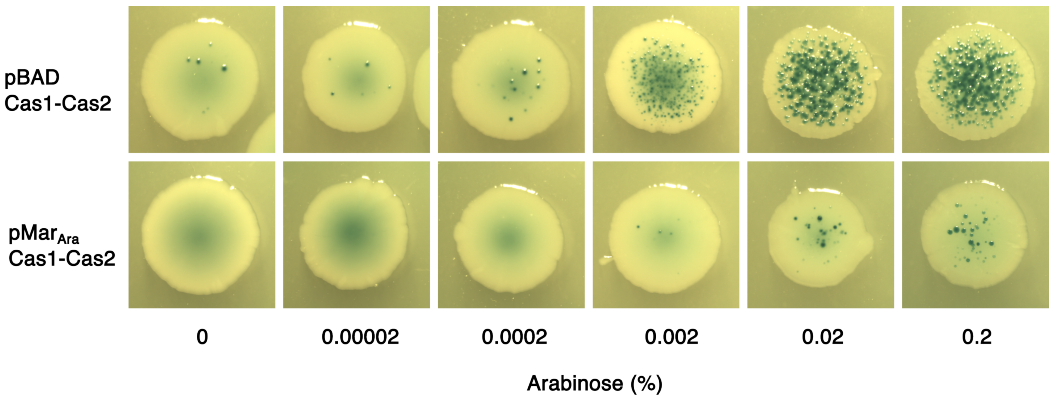

Figure 4.1: Comparison of papillation rates of different Cas1-Cas2 expression vectors.
Papillation assays were performed in JB028, with Cas1-Cas2 expressed from the pBAD and $pMar_{Ara}$ expression vector. L-arabinose inducer was titrated from 0-0.2%. Following incubation at 37 °C for five days, representative colonies are shown.

4.2.2 Cas1-Cas2 Mutant Library Construction

I performed error-prone PCR (epPCR) of Cas1-Cas2 across a range of mutagen conditions to identify those that would generate a mutation rate of one mutation across the 1.2 kb *Cas1-Cas2* operon. A range of concentrations of dGTP and Mn^{2+} were tested [Table 4.1]. This desired mutation rate was achieved with 10 μM dGTP and 20 μM Mn^{2+}.

I performed restriction digestion reactions upon the epPCR products and ligated them into the $pMar_{Ara}$ plasmid. This was transformed into electrocompetent NEB 5α cells and plated onto LB-agar supplemented with kanamycin. Colonies were pooled, and the plasmids were harvested via maxiprep to generate a library of approximately 1.3 x10^5 Cas1-Cas2 mutants. This is equivalent to 9.6 fold coverage of the Cas1-Cas2 operon. Sanger sequencing identified a mutational bias of this library as follows: T-C (45%), A-G (27%), G-C (9%), G-A (9%) and T-A (9%).

$MnSO_4$ (μM)	dGTP (μM)	Mutational frequency of Cas1-Cas2 (mutations per 1.2 kb)
0	0	0
100	40	4.1
200	0	11.6
0	80	2.0
200	80	10.2
0	0	0
20	10	0.9
10	5	0.8

Table 4.1: Mutation frequency with different epPCR conditions.
EpPCR was performed at various concentrations of $MnSO_4$ and dGTP. dGTP concentrations listed are in addition to the standard concentration. Mutagenesis rates were determined by averaging the frequency of mutations across five sequences per condition.

4.2.3 Cas1-Cas2 Mutant Library Screening

The mutagenised Cas1-Cas2 library was assayed for the presence of hyperactive spacer acquisition mutants. Papillation assays were performed in JB028 transformed with the mutagenised Cas1-Cas2 library [Figure 4.2]. A total of 27,000 colonies were screened, giving a 2.1-fold coverage of the library. Papillation plates were incubated for five days before hyperactive colonies were identified by visual inspection.

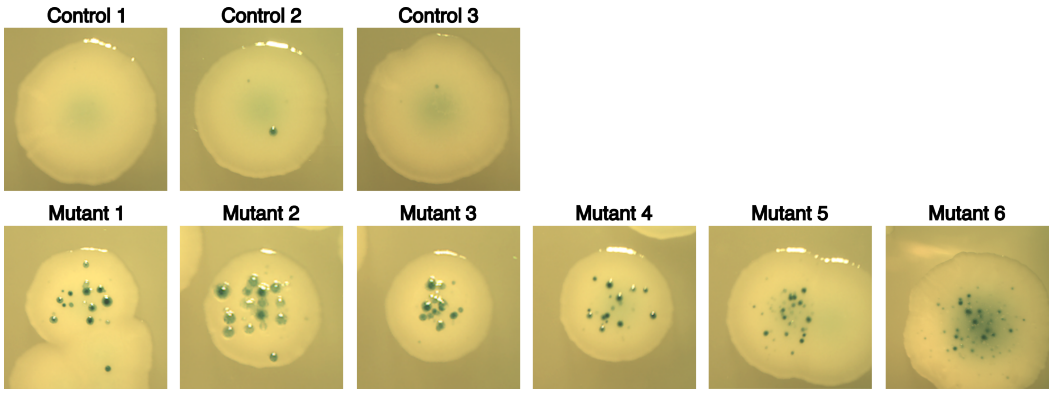

Figure 4.2: Papillation assays identified hyperactive Cas1-Cas2 mutants.
Papillation assays were performed in JB028 to screen a mutagenised Cas1-Cas2 library. Control colonies were transformed with the pMar$_{Ara}$Cas1-Cas2 plasmid to provide a rate of papillation for wildtype Cas1-Cas2. L-arabinose was present at a concentration of 0.002%. Hyperactive colonies were identified through visual inspection. Following incubation at 37 °C for 5 days, day five candidates and representative control colonies are shown.

To verify hyperactive phenotypes, a second round of papillation assays were performed using plasmid DNA isolated from hyperactive candidates [Figure 4.3]. A total of six candidates maintained a hyperactive phenotype in the validation screen [Figure 4.3]. Images were taken daily to track the development of papillae. On day two, papillae were visible on mutants 2 and 3. By day three, papillae were visible on all mutants at a frequency greater than wildtype. This disparity remained present at day five. Papillation rates in the verification assay were much greater than the initial screen due to a significantly reduced colony density per plate.

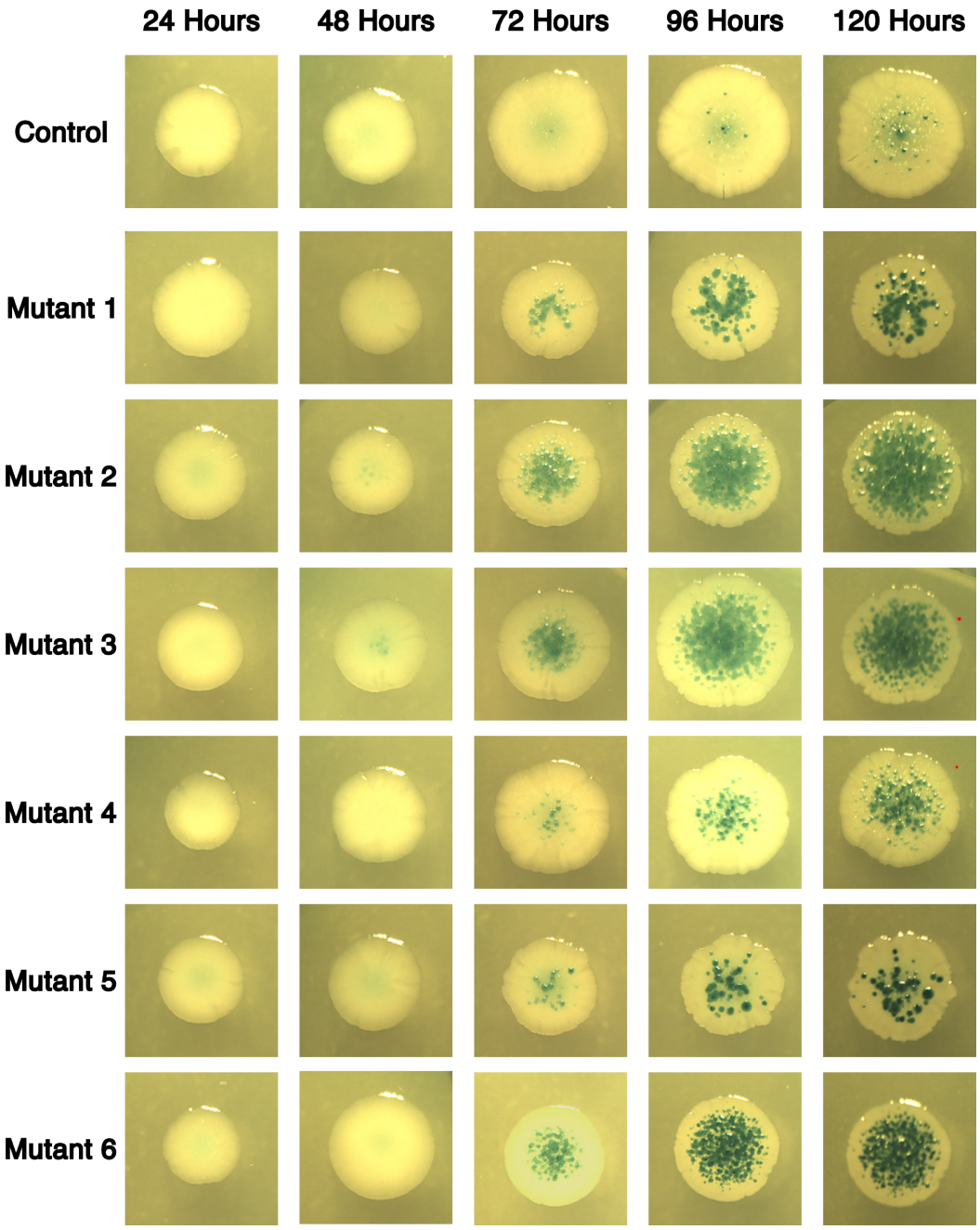

Figure 4.3: Validation papillation assay of initial hyperactive Cas1-Cas2 candidates. Papillation assays were conducted in JB028 transformed with plasmid from candidates identified in Figure 4.2. Control colonies were transformed with the pMar$_{Ara}$Cas1-Cas2 plasmid. L-arabinose inducer was present at 0.002%. A representative colony is shown for each condition. Plates were incubated at 37 °C for five days. Images were taken daily for five days.

To discover the mutation causing the hyperactive phenotypes, the Cas1-Cas2 loci of mutants 1-6 were sequenced. The nucleotide substitutions and their respective amino acid substitutions are listed in Table 4.2.

Mutant	DNA sequence mutations	Amino acid substitutions
1	T50C, T648C*, T657C*	M17T
2	A49G	M17V
3	A77G	D26G
4	G85A, T816C*	D29N
5	T83A, T301C	I28K, Y101H
6	A269G	Q90R

Table 4.2: The DNA and amino acid substitutions in mutants 1-6.
Base positions are with respect to the start of the Cas1 coding sequence. No mutations were found in *cas2*.
*Indicates synonymous mutations.

4.2.4 Quantification of Rates of Spacer Acquisition in Hyperactive Cas1 Mutants

SPIN assays were performed to verify that increases in papillation were concomitant with normal integration activity at the CRISPR locus. Initial work was performed using wildtype Cas1-Cas2 to discover conditions that produce a visible +1 band in a SPIN assay, corresponding to the acquisition of a novel spacer [Figure S1]. pBAD Cas1-Cas2 produced detectable acquisition in a SPIN assay at 0.2% L-arabinose, whilst $pMar_{Ara}$ Cas1-Cas2 did not produce detectable acquisition. Therefore, all subsequent work was conducted with pBAD expression vectors unless otherwise stated.

The non-synonymous mutations listed in Table 4.2 were cloned into the pBAD expression vector. SPIN assays were performed [Figure 4.4]. The acquisition behaviour at the CRISPR locus was normal for all mutations compared to wildtype.

Wildtype and mutant Cas1-Cas2 exhibited significant variability in the strength of PCR bands associated with the expansion of the CRISPR locus. This cannot be used to quantify acquisition rates as PCR bias, selection against LacZ expression, and toxicity derived from strong Cas1-Cas2 expression may affect the strength of bands present in a SPIN assay.

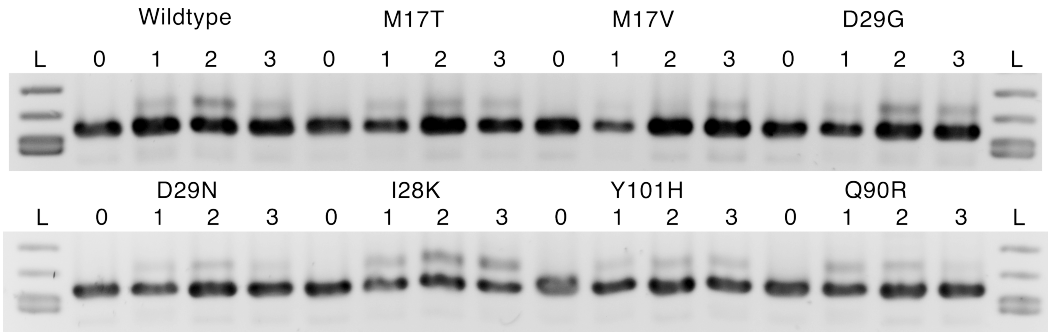

Figure 4.4: SPIN assays comparing CRISPR loci expansion by wildtype and mutant Cas1-Cas2.
SPIN assays were performed in JB028 to compare spacer acquisition at the CRISPR loci of wildtype and verified hyperactive mutant Cas1-Cas2 expressed from a pBAD plasmid. Expression was induced at 0.2% L-arabinose. A dilution of 1:1000 was performed for each culture each day over three days. Colony PCRs were performed on cultures daily. The bands at 548 bp derive from unexpanded loci, and bands at 609 bp derive from expanded loci. PCRs were visualised via agarose gel electrophoresis. L refers to an NEB 100 bp Ladder.

A YFP reporter assay was instead used to quantify the rate of spacer acquisition of hyperactive residues [213]. The YFP assay is similar to the papillation assay [Figure 4.5]. A constitutive promoter drives transcription and translation from a start codon, through a reporter CRISPR array, and into a *YFP* reporter gene. Stop codons prevent *YFP* expression. The expansion of the CRISPR array during adaptation shifts the reading frame to remove stop codons, resulting in *YFP* expression. The rate of spacer acquisition in a culture can be determined via flow cytometry.

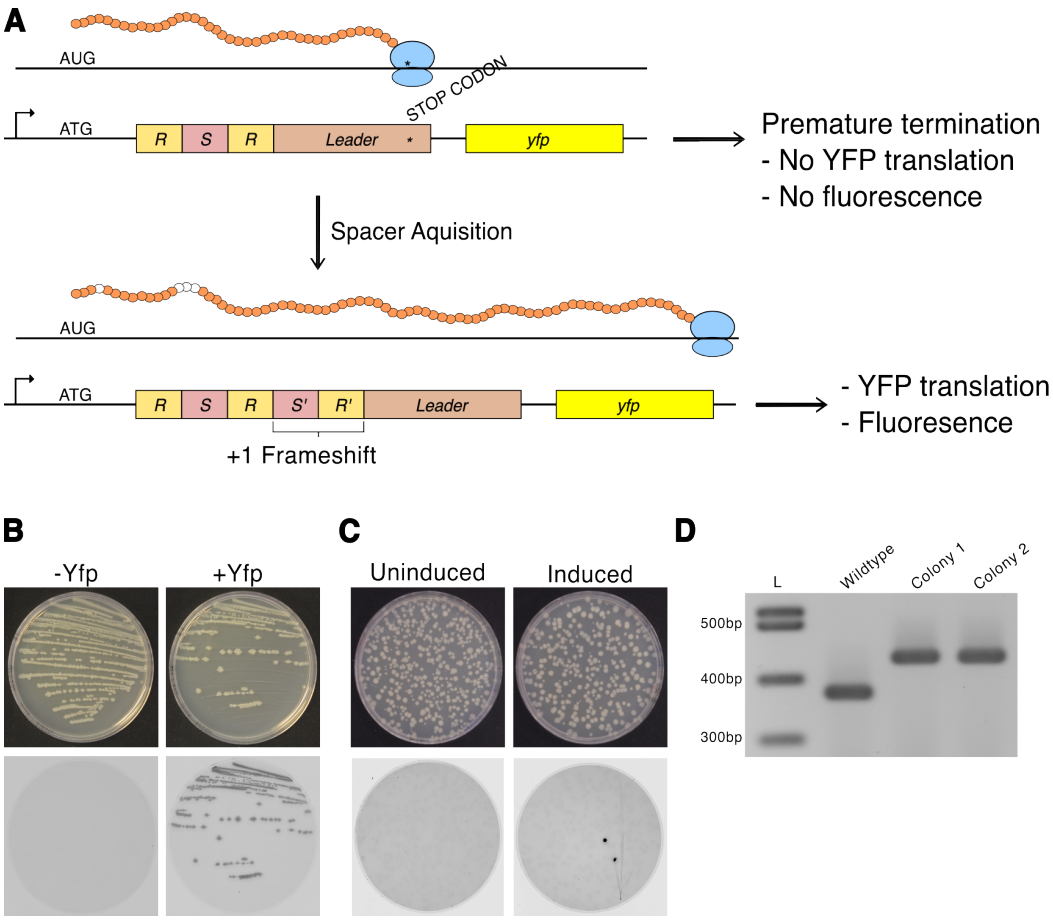

Figure 4.5: Overview of the YFP reporter assay for spacer acquisition.
(A) A constitutive promoter drives transcription and translation from a start codon through a CRISPR locus into a *YFP* reporter gene. The *YFP* gene is not expressed due to stop codons within the leader sequence and being outside the correct reading frame. The insertion of a 61 bp spacer-repeat unit into the CRISPR array causes a frameshift, which removes the stop codons and induces *YFP* expression. The ribosome is represented by a pair of blue circles. The translated polypeptide chain is represented by a string of small orange circles. *YFP* is in bright yellow, the leader sequence is in brown, repeats are in pastel yellow, and spacers are in red.
(B) On the left, a streak of the YFP reporter strain, MLS989, is imaged under visible light and with the Cy2 filter set of an Amersham Typhoon. On the right is a plate of the positive control strain, MLS990, which possesses a single base insertion to produce a frameshift mimicking spacer acquisition.
(C) Imaged as in part (B). On the left, MLS989, containing a pBAD Cas1-Cas2 expression vector, was grown without an inducer. On the right, the culture was grown with 0.2% L-arabinose. Cultures were grown overnight at 37 °C before a 1:1000 dilution was performed daily over the course of three days. Cultures were diluted and plated on LB-agar containing appropriate antibiotics. Following overnight growth at 37 °C, plates were imaged as in (C).
(D) MLS989, alongside the two colonies with enhanced fluorescence from the induced culture plate in (C), have been subjected to the PCR amplification of their CRISPR loci. PCRs were visualised via agarose gel electrophoresis. Bands at 369 bp derive from an unexpanded locus. Bands at 430 bp derive from an expanded locus. L refers to the NEB 100 bp ladder.

To test the YFP reporter, I performed an assay with MLS989 transformed wildtype and inactive Cas1-Cas2 (possessing the inactivating D221A mutation) at 0.2% L-arabinose [Supplemental Figure S2]. D221A abolishes the catalytic activity of the Cas1 nuclease but does not inhibit its binding activity with the CRISPR leader [73]. Flow cytometry determined that a significantly greater proportion of the wild-type Cas1-Cas2 population was fluorescent in comparison to that of the inactive D221A mutant. The level of acquisition by D221A was insignificant. This confirmed the effectiveness of the YFP reporter assay in quantifying rates of spacer acquisition.

The rates of spacer acquisition by the candidate hyperactive Cas1 mutants were determined using the YFP reporter assay [Figure 4.6]. The Cas1-Cas2 mutants were expressed at 0.2% L-arabinose for three days, with cultures diluted 1:1000 daily before YFP+ populations were determined via flow cytometry. All assessed mutants acquired spacers at a rate significantly higher than wildtype. The I28K mutant had the highest rate of adaptation, at roughly five-fold greater than wildtype. This identified this as the key mutation from candidate 5, which also possessed the Y101H mutation. However, Y101H also exhibited a smaller significant increase in acquisition rate compared to wildtype.

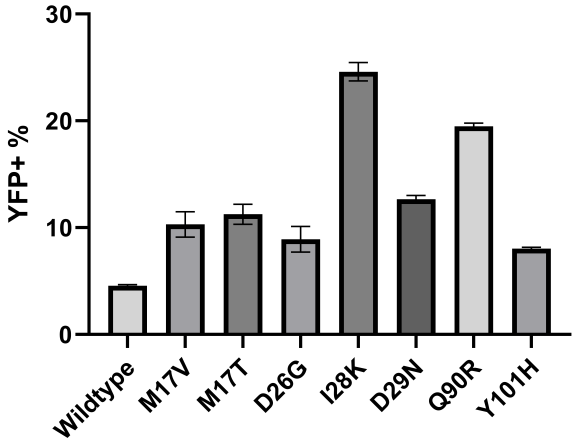

Figure 4.6: YFP reporter assay quantifying adaptation in hyperactive Cas1 mutants.
The MLS989 YFP reporter strain was transformed with wildtype Cas1-Cas2 or the hyperactive mutants outlined in Table 4.2. Cultures were grown in the presence of 0.2% L-arabinose inducer. A dilution of 1:1000 was performed each day over three days. The proportion of YFP+ cells was determined via flow cytometry. Five replicates were performed for each condition, with 20,000 cells analysed per replicate. Mean values are plotted, and error bars are SEM. Data values are in Supplementary Table S5.

4.2.5 Mutant Cas1-Cas2 Proteins Exhibit Enhanced Induction of the SOS Response

The SOS response is a broad genetic programme which is upregulated in response to DNA damage [231]. LexA is the master regulator of the SOS response [232] [233] [234]. Free RecA protein binds to and polymerises along single-stranded DNA (ssDNA) to form nucleoprotein filaments [235]. Sites within these filaments induce autocatalytic self-cleavage of LexA, releasing repression of the SOS regulon [236]. SulA is an archetypal gene of the SOS regulon [237]. It binds free FtsZ monomers to prevent their polymerisation, inhibiting cell division [238] [239]. SulA expression, therefore, triggers a checkpoint, inhibiting cell division until SOS induction has subsided [240].

A reporter system has been developed to measure the induction of the SOS response [215]. The *GFP* gene is placed under the control of the *sulA* promoter on

a reporter plasmid [Figure 4.7A]. When SulA is expressed from its native locus on the chromosome, the reporter *GFP* gene is also expressed. Fluorescence can be measured to determine the level of induction of the SOS response.

To test the SOS reporter system, I treated MG1655 with nalidixic acid. This is a DNA gyrase inhibitor which induces the SOS response through the introduction of double-stranded breaks [241]. Cells were grown across a range of nalidixic acid concentrations, and GFP fluorescence was measured [Figure 4.7b]. As nalidixic acid concentration increased, GFP fluorescence increased. The reporter, therefore, provides a titratable system to measure the induction of the SOS response.

The integrase activity of Cas1-Cas2 has been shown to induce the SOS response in *E. coli* [229]. To determine the effect that Cas1 mutants have on the induction of the SOS response, I performed SOS induction assays. Overnight cultures were prepared of MG1655 transformed with the P_{sulA}*GFP* reporter plasmid and either wildtype or mutant pBAD Cas1-Cas2 plasmids, induced at 0.2% L-arabinose [Figure 4.7C]. All mutants exhibited a significantly higher induction of the SOS response relative to the wildtype.

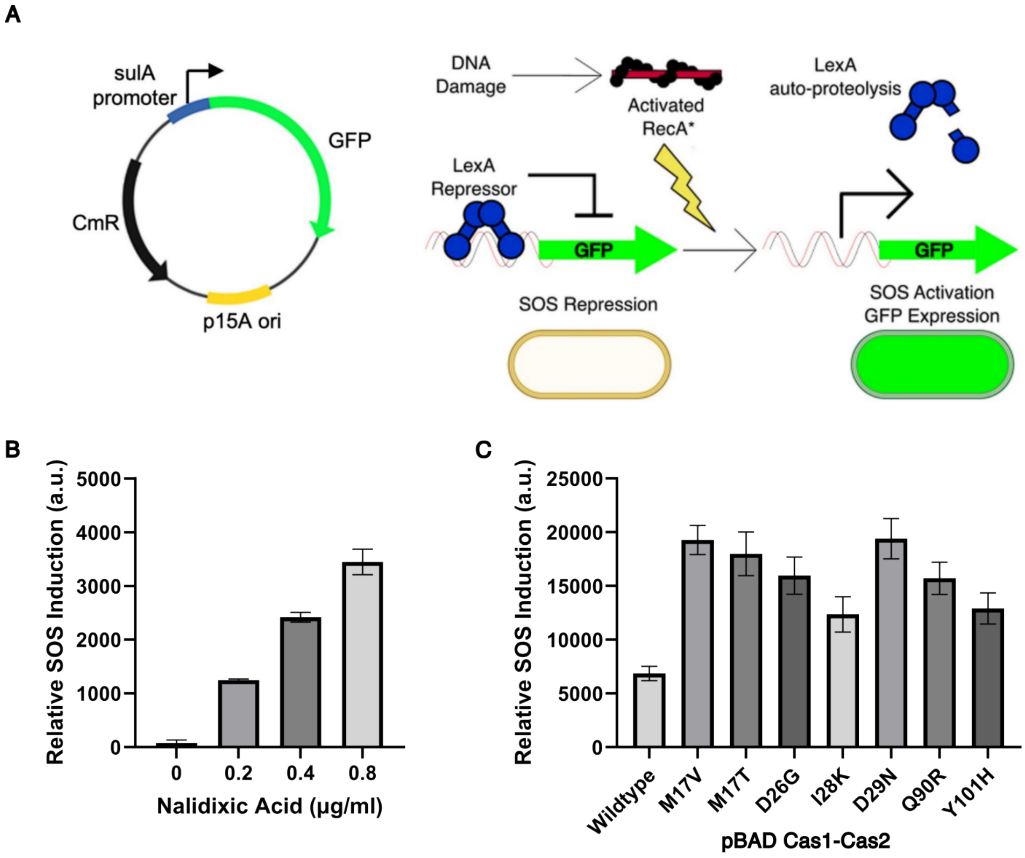

Figure 4.7: Hyperactive Cas1 mutants induce the SOS response to a higher level than wildtype.
(A) A diagram depicting the mechanism of a plasmid-based genetic circuit for the detection of induction of the SOS response. This reporter system has *GFP* under the control of the *sulA* promoter. DNA damage induces the activation of RecA, inducing the formation of RecA-ssDNA nucleoprotein filaments. These filaments induce the auto-proteolysis of LexA dimers, releasing repression of the *sulA* promoter. When the SOS response is induced following DNA damage, the *GFP* reporter gene is expressed. This generates detectable fluorescence.
(B) Overnight cultures of MG1655 containing the P_{sulA}*GFP* reporter plasmid were incubated overnight in the presence of 0-0.8 μg of nalidixic acid. Overnight cultures were diluted to an O.D 600 of 1. Fluorescence was measured using an Amersham Typhoon with a Cy2 filter set. Three replicates were performed per condition. Mean values are plotted, and error bars are SEM. Data can be found in Supplementary Table S10.
(C) Overnight cultures of MG1655 containing the P_{sulA}*GFP* reporter plasmid alongside wildtype or mutant pBAD Cas1-Cas2, respectively, were grown up overnight. Cultures were grown in the presence of 0.2% L-arabinose inducer. Overnight cultures were diluted to an O.D 600 of 1. Fluorescence was measured using an Amersham Typhoon with a Cy2 filter set. Three replicates were performed per condition. Mean values are plotted, and error bars are SEM. Data can be found in Supplementary Table S10.

4.2.6 Cas1 I28K Exhibits Reduced PAM Fidelity in Protospacer Selection

As the Cas1 I28K mutant had the highest rate of adaptation, I decided to investigate any behavioural difference in protospacer selection compared to wildtype. I performed papillation assays in JB028 containing a wildtype Cas1-Cas2 or I28K expression vector, induced at 0.002% L-arabinose. Colonies from ten plates per condition were each resuspended in 1 mL 1x M9 salts and pooled. These resuspensions were used as templates for PCR amplification across the reporter CRISPR array. The PCR product was sent for nanopore sequencing. A spacer library was generated from the sequencing data and a BLAST search was performed, mapping spacers against the JB028 genome and pBAD Cas1-Cas2 expression plasmid. Results are described in Tables 4.3 and 4.4.

Spacer Source	Total Spacers	Total Consensus PAM	% Consensus PAM
All	28526	8068	28.28
Genome	26817	7929	29.57
Plasmid	1709	139	8.13

Table 4.3: Overview of unique spacers acquired by a wildtype Cas1-Cas2 in a pBAD expression vector in JB028.

Spacer Source	Total Spacers	Total Consensus PAM	% Consensus PAM
All	70155	11851	16.89
Genome	66451	11705	17.61
Plasmid	3704	146	3.94

Table 4.4: Overview of unique spacers acquired by Cas1-Cas2 containing the Cas1 I28K mutant in a pBAD expression vector in JB028.

Spacers acquired from Cas1 I28K exhibited a reduced fidelity to the consensus PAM compared to wildtype. This reduced fidelity was present for both plasmid-

and genome-derived spacers. The plasmid possesses 152 potential AAG PAM sites and is therefore approaching saturation at 139 and 146 respective consensus spacers. In comparison, the E. coli genome possesses 127,081 AAG PAM sites and is therefore not approaching saturation in either condition [242].

4.2.7 Mutational Analysis of Hyperactive Residues of Cas1

Random mutagenesis of Cas1-Cas2 via epPCR successfully generated mutants with a hyperactive phenotype. However, epPCR is associated with a range of biases in mutational profiles, which diminish the coverage of sequence space present in libraries [243]. This is exemplified by the transition: transversion bias seen in the Cas1-Cas2 library screened (82% transitions). Site-directed mutagenesis was therefore performed to identify further hyperactive substitutions of residues already identified.

Saturating mutagenesis was performed through the construction of codon libraries for the residues I28 and D29. A library with x10-fold coverage was generated for each residue of interest. I28 was selected as this mutant acquired spacers at the greatest rate. D29 was chosen because of its proximity to I28. They are located in a beta-turn at the interface between Cas1, Cas2 and an incoming protospacer. Both of their R-groups extend towards the negatively charged phosphodiester backbone of the 5' end of the protospacer [Figure 4.8] [76]. The change in R-groups of the I28K and D29N mutants is characterised by a shift away from a negative charge. I hypothesised similar mutations would generate similar phenotypes.

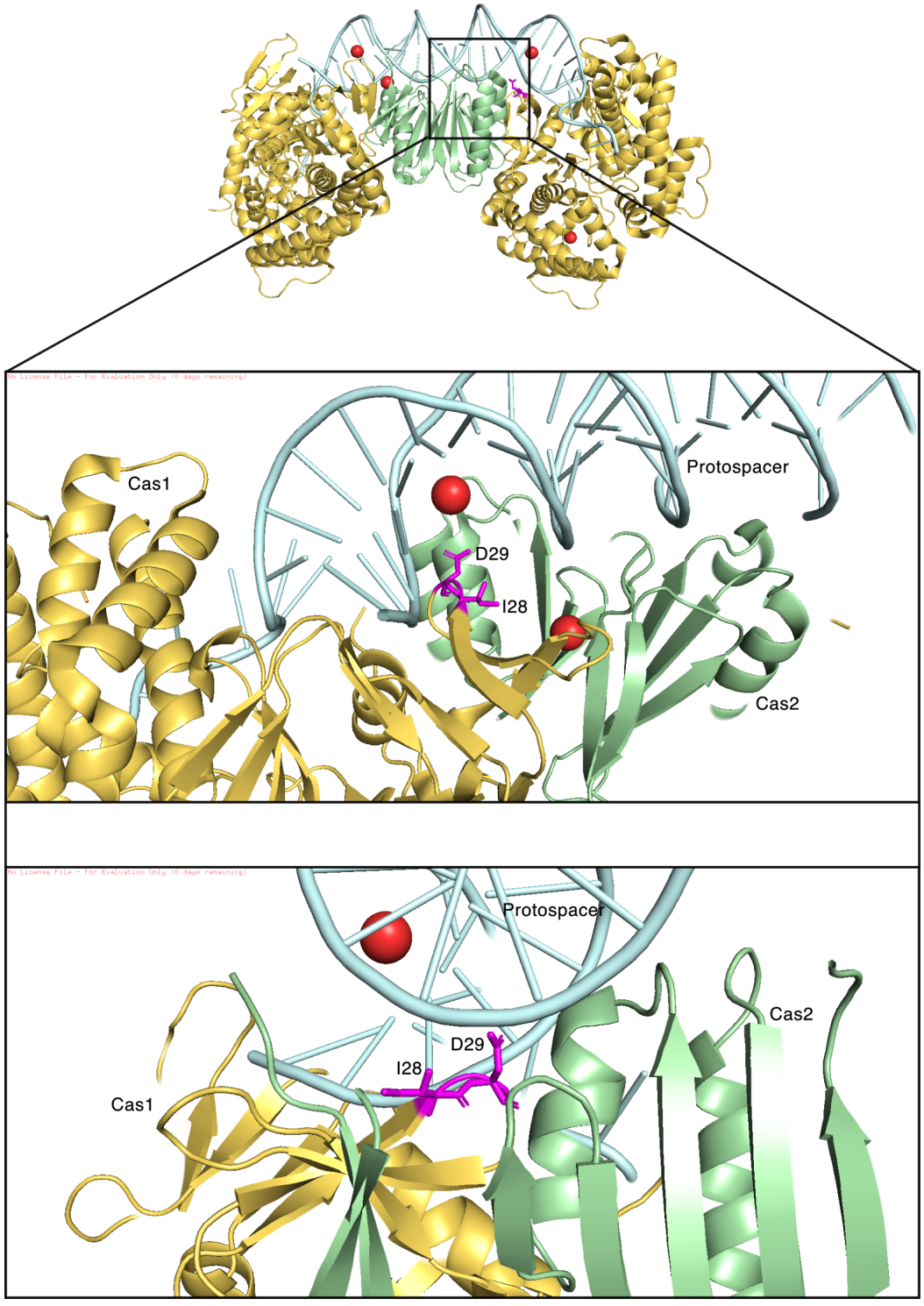

Figure 4.8: The crystal structure of *E. coli* wildtype Cas1-Cas2 bound to protospacer DNA, determined via X-ray diffraction.
The protospacer is in cyan, Cas1 is in yellow, Cas2 is in green, Mg^{2+} ions are in red, and the residues of interest I28 and D29 are highlighted in pink. The R-groups of these residues can be seen facing towards the protospacer.
Figure derived from PDB: 5DS4 [244].

The two saturating mutagenic libraries were screened by papillation assays [Figure 4.9]. Cas1-Cas2 was induced with 0.002% L-arabinose to facilitate the visual identification of hyperactive colonies [Figure 4.1]. Colonies with hyperactive phenotypes were isolated and their plasmids were sequenced. The nucleotide sequence of hyperactive mutants and their associated amino acid substitutions are presented in Table 4.5.

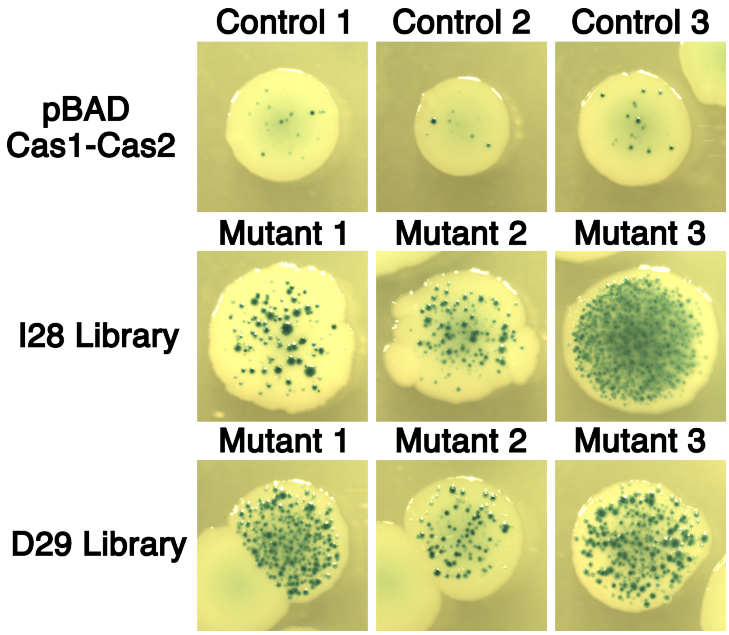

Figure 4.9: Hyperactive candidates from the saturating mutagenesis of I28 and D29.
Papillation assays were performed in JB028. Control colonies were transformed with the wildtype Cas1-Cas2, whilst mutants had been transformed with respective I28 and D29 libraries. L-arabinose was present at 0.002%. Representative day five colonies are shown.

Library	Mutant	Amino acid substitutions
I28	1	I28S
	2	I28N
	3	I28R
D29	1	D29H
	2	D29K
	3	D29A

Table 4.5: The amino acid mutational profiles for novel hyperactive Cas1 I28 and D29 mutants.

YFP reporter assays were conducted to verify the hyperactive phenotypes of each mutation [Figure 4.10A]. All mutants deriving from I28 and D29 saturation mutagenesis returned hyperactive phenotypes when rates of spacer acquisition were tested. I28K retained the highest identified rate of spacer acquisition of any mutant.

SOS induction assays were conducted to identify the extent of SOS response activation by the hyperactive mutants identified in the I28 and D29 library screens [Figure 4.10B]. Most mutants generated a significantly greater induction of the SOS response than wildtype Cas1-Cas2. However, this was not seen for all mutants with I28S, I28N and D29A generating no greater SOS induction than wildtype.

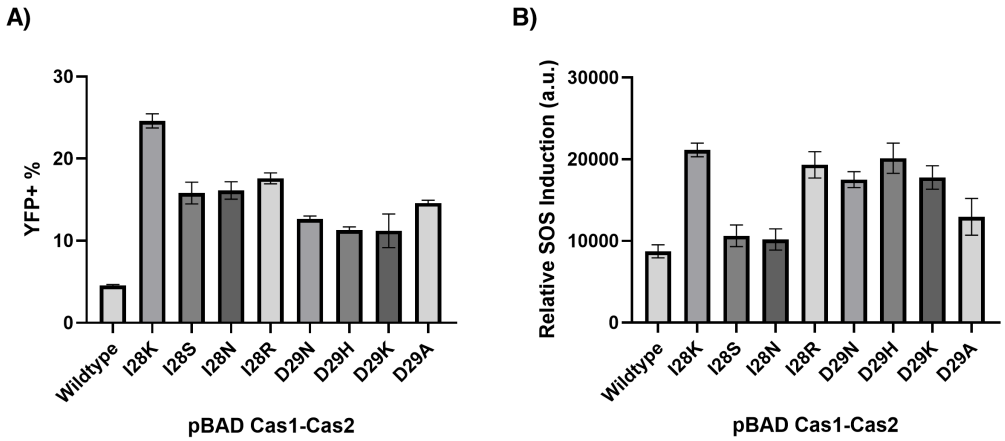

Figure 4.10: Rates of spacer acquisition and SOS induction for I28 and D29 mutants.
(A) YFP reporter assays were performed in MLS989 transformed with wildtype Cas1-Cas2 or I28 and D29 mutants. Cultures were grown in the presence of 0.2% L-arabinose inducer. A dilution of 1:1000 was performed each day over three days. The proportion of YFP+ cells was determined via flow cytometry. Five replicates were performed for each condition, with 20,000 cells analysed per replicate. Mean values are plotted, and error bars are SEM. Data values are in Supplementary Table S5.
(B) SOS induction assays were performed on overnight cultures of MG1655 containing the P_{sulA}*GFP* reporter plasmid alongside wildtype or I28/D29 mutant pBAD Cas1-Cas2. Cultures were grown in the presence of 0.2% L-arabinose inducer. Overnight cultures were diluted to an O.D 600 of 1. Fluorescence was measured using an Amersham Typhoon with a Cy2 filter set. Six replicates were performed per condition. Mean values are plotted, and error bars are SEM. Data can be found in Supplementary Table S10.

4.2.8 Combination of Hyperactive Cas1 Mutants

Hyperactive transposase mutations have previously shown that the combination of multiple mutations into the same reading frame results in the synergistic enhancement of transposition efficiency. This is exemplified by a hyperactive Tn5 transposase derived from the combination of several mutations which individually confer hyperactivity through separate mechanisms of action [245]. Hyperactive spacer acquisition mutants in *S. pyogenes* Cas1 were shown to be synergistic [230]. I therefore combined several hyperactive mutants identified in *E. coli* Cas1 to determine if their phenotypes were synergistic.

As the I28K mutant has the highest rate of spacer acquisition, this was individually combined with the M17T, D26G, and Q90R mutants. These mutants were selected because they are outside of the beta-turn from which the hyperactive mutant codon substitutions of I28 and D29 derive. They may, therefore, induce hyperactivity through separate and potentially additive mechanisms.

To quantify the rate of adaptation, YFP reporter assays were conducted with Cas1 double mutants [Figure 4.11A]. Of the three double mutants, only Cas1 M17T I28K acquired spacers at a rate significantly greater than wildtype (P=0.0002). Cas1 I28K Q90R conducted adaptation at a rate approaching statistical significance (P=0.0725), while Cas1 D26G I28K acquired at a rate comparable to wildtype (P=0.8622). In comparison, all respective single mutants acquired spacers at a significantly greater rate than wildtype. This suggests that the combination of hyperactive Cas1 mutants is antagonistic to rates of spacer acquisition.

SOS induction assays were also performed with Cas1 double mutants [Figure 4.11B]. All double mutants induced the SOS response significantly more than wildtype. In comparison to their respective single mutants, the M17T I28K and Q90R I28K double mutants exhibited an increased induction. The D26G I28K had an increased induction relative to its respective single mutants, although the in-

creased variance of induction between its replicates meant this was not significant. These results suggest that SOS induction by hyperactive Cas1 mutants is synergistic.

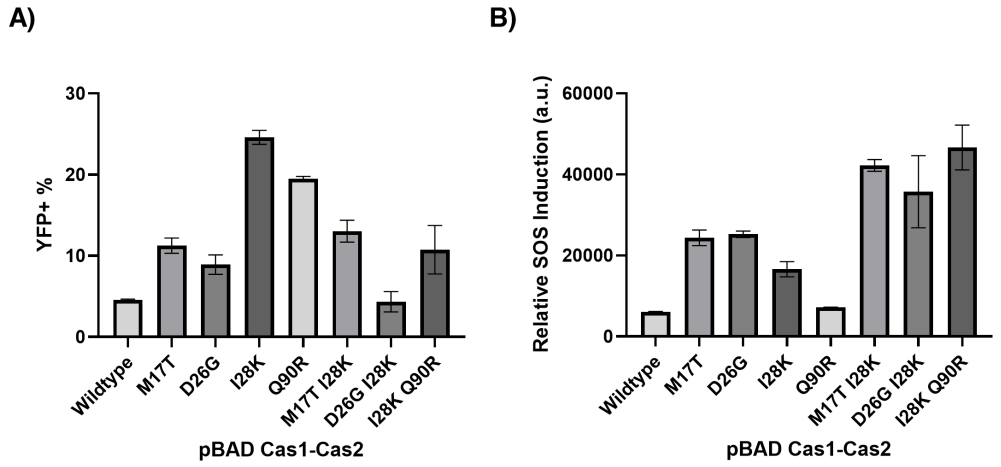

Figure 4.11: Spacer acquisition and SOS induction by single and double Cas1 mutants.
(A) YFP reporter assays were performed in MLS989 transformed with wildtype Cas1-Cas2 or single and double Cas1 mutants. Cultures were grown in the presence of 0.2% L-arabinose inducer. A dilution of 1:1000 was performed each day over three days. The proportion of YFP+ cells was determined via flow cytometry. Five replicates were performed for each condition, with 20,000 cells analysed per replicate. Mean values are plotted, and error bars are SEM. Data values are in Supplementary Table S5.
(B) SOS induction assays were performed on overnight cultures of MG1655 containing the P_{sulA}*GFP* reporter plasmid alongside wildtype or single/double Cas1 mutants in pBAD Cas1-Cas2. Cultures were grown in the presence of 0.2% L-arabinose inducer. Overnight cultures were diluted to an O.D 600 of 1. Fluorescence was measured using an Amersham Typhoon with a Cy2 filter set. Three replicates were performed per condition. Mean values are plotted, and error bars are SEM. Data can be found in Supplementary Table S10.

4.3 Chapter Summary

- High-throughput screening of a mutagenised Cas1-Cas2 library facilitated the identification of Cas1 mutants with hyperactive spacer acquisition phenotypes.
- Cas1 mutants with hyperactive spacer acquisition phenotypes are associated with enhanced rates of induction of the SOS response through the *sulA* promoter.
- Saturating codon mutagenesis of key residues facilitated the identification of further hyperactive mutants.
- The combination of hyperactive mutants within Cas1 is antagonistic to spacer acquisition but synergistic to the further induction of the SOS response.

Chapter 5

Screening for Anti-CRISPRs Targeting Adaptation: Optimisation and Results

5.1 Introduction

5.1.1 Screening for Anti-CRISPRs Targeting Adaptation

There has been a recent expansion in the identification of MGE-encoded proteins which inhibit prokaryotic defence systems [183] [47] [246]. CRISPR systems are the predominant target of these novel effectors. To date, over 100 anti-CRISPRs (Acrs) and over a dozen Acr-associated (Aca) proteins have been identified [191].

Very few Acrs are known to target spacer acquisition. AcrVA5 is the only Acr known to target the Cas1-Cas2 integrase, sterically hindering complex formation in the *Treponema denticola* Type II-A system. It was also shown to have activity against the *Moraxella bovoculi* Type I-C system [206][205]. However, there is further evidence that phages target the process of spacer acquisition. A *Campylobacter jejuni* phage encodes a Cas4-like protein which biases spacers to derive solely from the host genome [247]. Furthermore, a *Streptococcus thermophilus* phage encodes a truncated Cas9 that specifically inhibits adaptation [248]. Finally, prophages have been shown to integrate into a host's CRISPR array and Cas1 gene [249]. Together, these suggest that selective pressure exists to induce the inhibition of CRISPR adaptation by MGEs.

Screens for the identification of Acrs have focused on the identification of anti-interference proteins due to their applications in biotechnology [186] [250] [251]. Moreover, anti-interference Acrs can be selected, while anti-adaptation proteins must be screened. A high throughput papillation-based reporter for spacer acquisition can be used to rapidly screen $\geq$1 Gbp metagenomic DNA to identify novel Acrs targeting adaptation [Figure 4.6] [Figure 5.1].

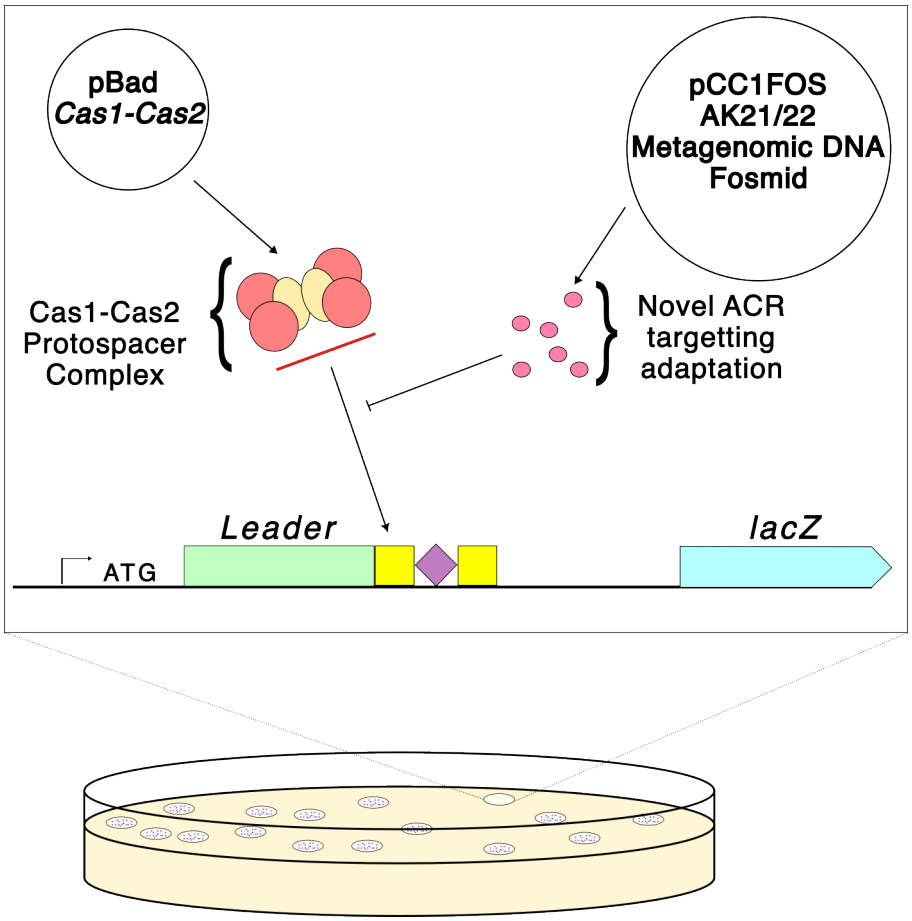

Figure 5.1: A screening approach to identify Acrs targeting adaptation.
Papillation assays will be conducted in a strain containing a papillation-based spacer acquisition reporter, a Cas1-Cas2 expression vector, and a fosmid library carrying metagenomic DNA. Fosmids carrying Acrs that can inhibit adaptation will produce hypopapillating phenotypes in their respective colonies. These colonies will be isolated for further verification and investigation.

5.1.2 The AK Libraries

The Alaska (AK) libraries are a set of metagenomic libraries derived from environmental soil samples [Table 5.1] [252] [210] [211]. They have been used to identify β-lactamases and quorum sensor inhibitors [252] [211]. A papillation assay can screen the AK libraries for anti-CRISPRs that inhibit adaptation. The collection includes several fosmid libraries, cloning vectors possessing the F-plasmid *ori* and the bacteriophage λ *cos* site, enabling the efficient cloning and stable maintenance of large DNA inserts. The large insert size and depth of the AK21/22 libraries will facilitate the rapid screening of a large volume of metagenomic DNA.

AK Library	Library Type	Vector	Number of Clones	Average Insert Size (Kbps)	% Inserts	Total Size (Mbps)	Reference
1	BAC	pSuperBAC	22000	10	-	222	[252]
2	BAC	pSuperBAC	4600	5	-	23	[252]
3	BAC	pCC1BAC	2500	13	-	33	[252]
4	BAC	pSuperBAC	16800	11	-	185	[252]
5	BAC	pSuperBAC	2200	33	70	46	[252], [210]
6	Fosmid	pCC1FOS	300	39	-	12	[252]
7	BAC	pCC1BAC	2300	47	60	108	[252], [210]
8	BAC	pCC1BAC	2700	9	70	24	[252],[210]
9	BAC	-	2400	20	25	12	[211]
10	Plasmid	pCF430	16300	10	-	173	[211]
11	Plasmid	pCF430	36800	5	-	184	[211]
12	Plasmid	pCF430	36600	5	-	183	[211]
13	BAC	-	8400	29	10	24	[210]
14	Plasmid	pCF430	105800	5	-	529	[211]
15	BAC	-	32100	20	60	385	[210]
16	Plasmid	pCF430	97100	6	-	592	[211]
17	BAC	-	1000	29	20	6	[210]
18	Plasmid	pCF430	34500	10	-	338	[211]
19	Fosmid	-	3300	31	98	1003	[210]
20	BAC	pCC1BAC	48100	8	80	385	[210], [211]
21	Fosmid	pCC1FOS	333000	30	100	10100	[210], [211]
22	Fosmid	-	50500	30	100	1515	[210]
23	Fosmid	-	8200	30	100	246	[210]
Total			867500	19	66	16328	[252], [210], [211]

Table 5.1: Overview of the AK Libraries

5.2 Optimisation

5.2.1 Screening with JB028

The AK21 and AK22 libraries are in the strain EPI3000, following their construction using the Epicentre CopyControl Fosmid Library Production Kit [211]. EPI3000 has a Δ*(mrr-hsdRMS-mcrBC)* restriction minus genotype, facilitating efficient transformation with fosmid clones containing unmethylated metagenomic DNA. However, the papillation reporter strain JB028 contains an intact *hsdR* locus. Therefore, it exhibits a poor transformation efficiency with plasmids derived from EPI3000 [Figure 5.2].

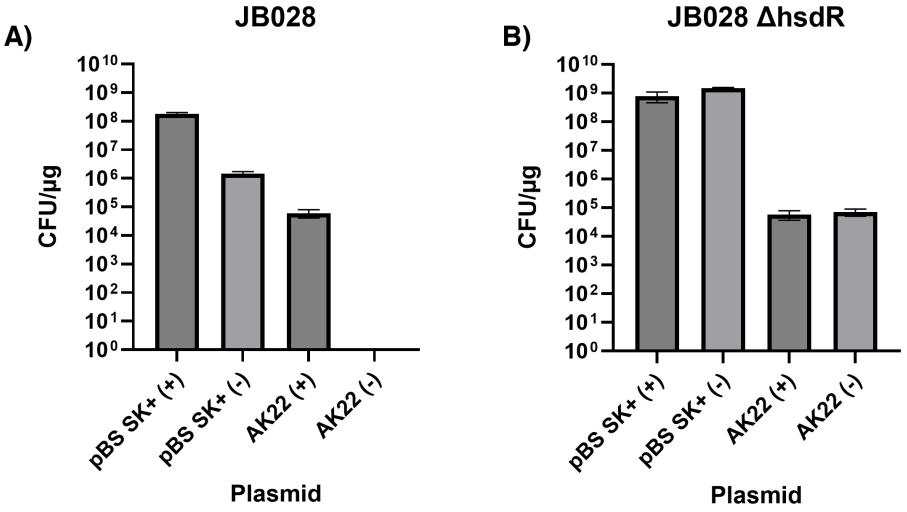

Figure 5.2: Transformation efficiencies of methylated and unmethylated AK22 and pBS SK+ plasmids in JB028 and JB028 Δ*hsdR*.
(A) JB028 was transformed with AK22 or pBS SK+ in triplicate. Plasmid was either methylated or unmethylated. Transformation efficiency was determined through titration of transformation cultures and plating on LB-agar containing appropriate antibiotics. Following overnight growth, colonies were counted to determine CFU/μg. Methylated plasmid was prepared from NEB5α, whilst unmethylated plasmid was prepared from DH10β. (+) refers to methylated plasmids, and (-) refers to unmethylated plasmids.
B) Transformation efficiencies were determined for JB028 Δ*hsdR*, as described above.
Conditions were measured in triplicate. Mean data values are plotted, and error bars are SEM. Data values are found in Supplementary Table S4.

The EcoKI restriction-modification system is encoded by the *hsdRMS* operon [253]. It is comprised of the HsdR (restriction), HsdM (methylation) and HsdS (sequence specificity) proteins [Figure 5.3]. These form two complexes with separate capabilities. The $HsdM_2S_1$ complex behaves as a methyltransferase targeted to the $AAC(N_6)GTGC$ DNA motif [254][255]. The $HsdR_2M_2S_1$ complex possesses endonuclease activity against this same sequence.

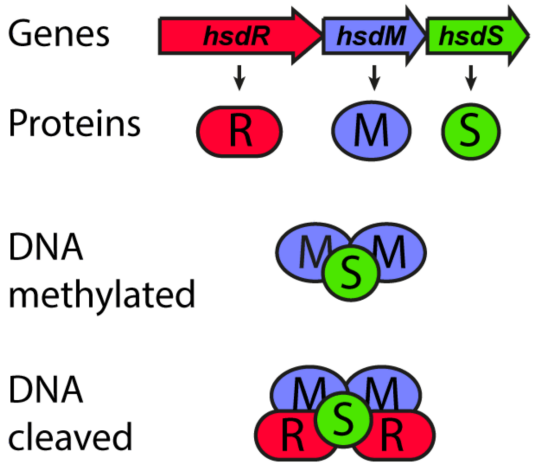

Figure 5.3: Overview of the EcoKI restriction-modification system.
The EcoKI *hsdRMS* locus encodes separate restriction (R), methyltransferase (M) and specificity (S) proteins. DNA cleavage is carried out by the pentameric $HsdR_2M_2S_1$ complex, whilst the trimeric $HsdM_2S_1$ complex acts as a methyltransferase. Figure derived from [256]

Knockout of *hsdR* in JB028 would inhibit the strain's EcoKI endonuclease activity. This would increase the transformation efficiency of the unmethylated AK21 and AK22 libraries [Figure 5.2]. Knockout of *hsdR* was performed via P1 phage transduction [Figure 5.4 and 5.5].

P1 transduction was carried out using the Keio collection [209]. This is a single-gene knockout library of the non-essential genes in *E. coli* K-12. This involves the replacement of the gene of interest with a kanamycin resistance cassette flanked by complementary *FRT* recombination sites. Following transduction, these sites facilitate the cassette's removal via FLP recombination, leaving an in-frame deletion at the target locus and a 34 residue peptide at the scar site [Figure 5.4].

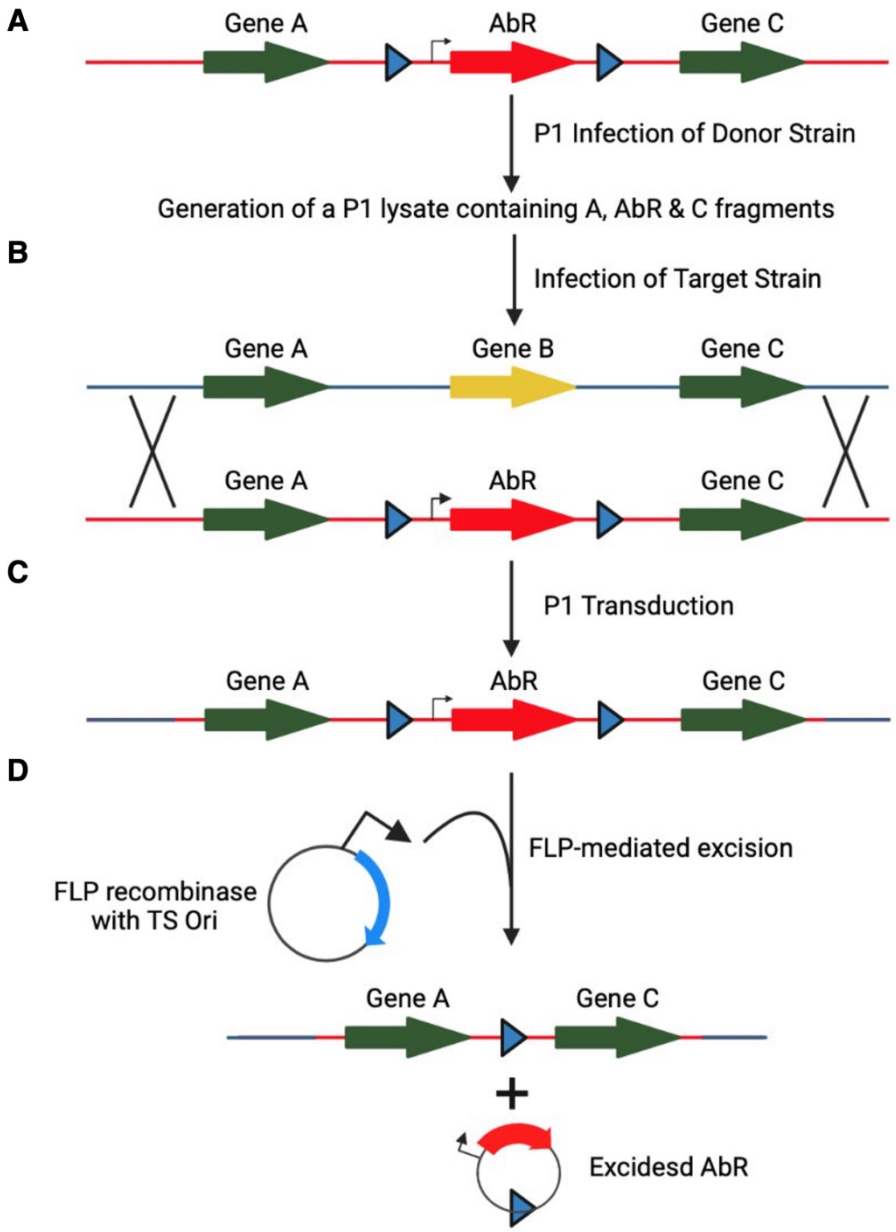

Figure 5.4: Stages of P1 phage transduction.
(A) The donor strain (red) is used to prepare a P1 lysate. This strain possesses a mutant of interest linked with a selectable marker (AbR, red arrow).
(B) The prepared P1 lysate infects the recipient strain (blue). The mutation of interest, alongside the linked marker gene, is integrated into the recipient strain's genome at low frequency via homologous recombination.
(C) The selectable marker facilitates the identification of recipient strain colonies carrying the mutation of interest.
(D) The marker can be cured from the recipient strain through FLP recombination between frt sites (blue arrow).

A P1 phage lysate was produced using the donor strain BW25113 *hsdR::FRT-KanR-FRT*. The recipient strain, JB028, was infected with this lysate containing P1 phage carrying the *hsdR::FRT-KanR-FRT* locus. Successful homologous recombination of the cassette into the JB028 genome was selected via growth on LB-agar supplemented with kanamycin. The cassette was removed via FLP recombination, producing the strain JB028 Δ*hsdR*. Knockout was confirmed via PCR [Figure 5.5A].

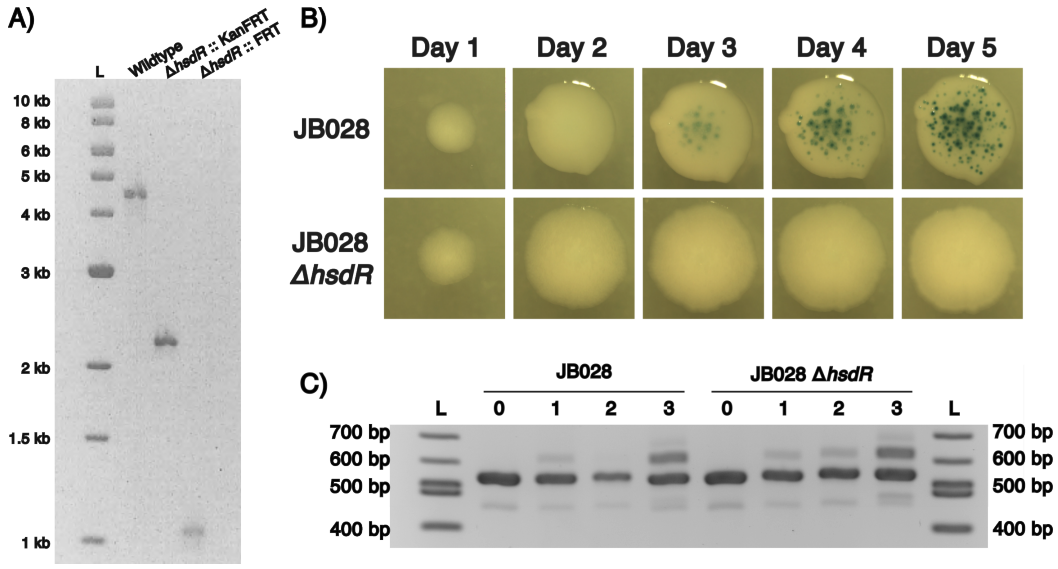

Figure 5.5: Overview of Δ*hsdR* deletion from JB028.
(A) PCR verification of the JB028 Δ*hsdR* deletion. PCRs were performed at each stage of *hsdR* knockout. PCRs were visualised via agarose gel electrophoresis. The band at 4462 bp corresponds to the wildtype *hsdR* locus. The band at 2391 bp corresponds to the Δ*hsdR*::FRT-Kanamycin-FRT locus. The band at 1169 bp corresponds to the Δ*hsdR*::FRT locus. L refers to an NEB 1 kb ladder.
(B) Papillation assays were conducted in JB028 and JB028Δ*hsdR*. Strains were transformed with a pBAD Cas1-Cas2 expression plasmid. L-arabinose was present at 0.002%. Plates were incubated at 37 °C for 5 days. An individual colony was imaged daily; these images are shown.
(C) SPIN assays were conducted in JB028 and JB028Δ*hsdR*. Strains were transformed with a pBAD Cas1-Cas2 expression plasmid. L-arabinose was present at 0.2%. Cultures were grown at 37 °C on a spinning wheel. Following overnight growth, cultures were diluted 1:1000. This was conducted daily for three days. 1 ml of culture was collected daily, and colony PCRs were performed on it. Day 0 refers to an initial culture grown without an inducer. PCRs were visualised via agarose gel electrophoresis. Bands at 548 bp correspond to the parental CRISPR locus, bands at 609 bp correspond to one spacer acquisition event, bands at 670 bp correspond to two spacer acquisition events. L refers to an NEB 100 bp ladder.

Knockout of *hsdR* in JB028 increased the transformation efficiency of unmethylated DNA into the strain [Figure 5.2]. There was no significant difference in transformation efficiency between unmethylated and methylated DNA. Therefore, JB028 $\triangle$*hsdR* could be efficiently transformed with the AK21 and AK22 libraries.

Papillation assays and SPIN assays were conducted in JB028 $\triangle$*hsdR* to verify that spacer acquisition could still occur normally at the ArgE:: CRISPR reporter locus [Figure 5.5B, C]. The strain acquired spacers in a SPIN assay at a rate comparable to wildtype but did not produce papillae in papillation assays. The complete absence of papillation in the presence of spacer acquisition suggested *lacZ* expression from the reporter was impaired. Even IHFα, reported to be an essential gene in spacer acquisition [85], produced a small number of papillae in papillation assays [229].

Leaky expression of *lacZ* at the CRISPR reporter locus in JB028 causes colonies in papillation assays to exhibit a pale blue phenotype derived from low-level X-gal metabolism [212]. The absence of this phenotype in JB028$\triangle$*hsdR* suggests it possessed a mutation in a *lacZ* [Figure 5.5B].

A pOX38 plasmid expressing *LacZ* was conjugated into JB028 and JB028$\triangle$*hsdR* [Figure 5.6]. Both strains exhibited a blue phenotype when grown on LB-agar supplemented with X-gal. This suggested JB028$\triangle$*hsdR* contained a mutation within the CRISPR reporter locus preventing expression of a functional LacZ.

A repeat of the knockout of *hsdR* from JB028 should have produced a strain capable of papillation. However, the frequent knockout of genes using the Keio collection within the lab led to demand for a papillation reporter hosted in the parental strain of the Keio collection, BW25113 [209]. I developed this reporter and used it to screen AK21 and AK22.

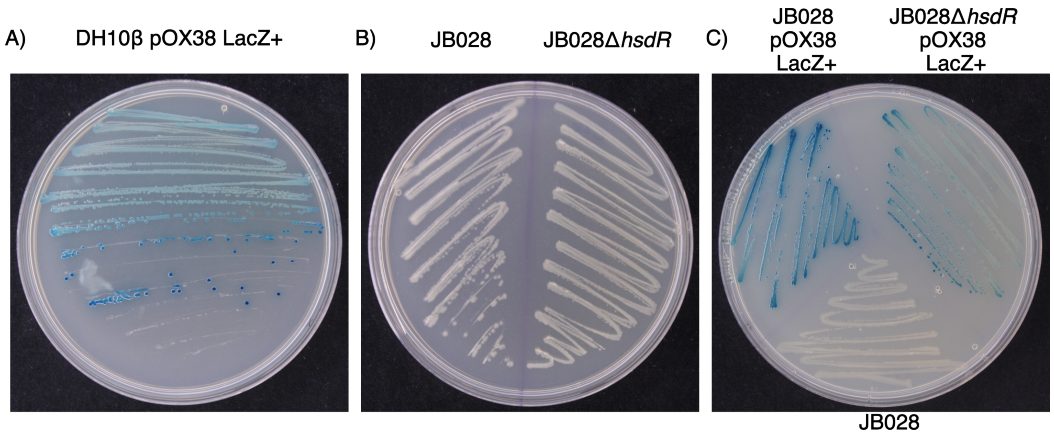

Figure 5.6: Conjugation of pOX38 LacZ+ into JB028 and JB028Δ*hsdR*.
(A) A plate of DH10β pOX38 grown on LB-agar supplemented with X-gal. This was the donor strain for the pOX38 LacZ+ plasmid. Plates were incubated overnight at 37 °C. A representative whole plate photo was taken.
(B) A plate of JB028 and JB028Δ*hsdR* grown on LB agar supplemented with tetracycline and X-gal. These were the recipient strains for the pOX38 LacZ+ plasmid. Plates were incubated overnight at 37 °C. A representative whole plate photo was taken.
(C) A plate of JB028 and JB028/JB028Δ*hsdR* containing pOX38. Overnight cultures of recipient and donor strains were incubated together on filter paper on LB-agar for 6 hours. Filter paper was resuspended in 1 ml 1x M9 salts and cells were washed. The resuspension was streaked onto LB-agar supplemented with tetracycline and X-gal. Plates were incubated overnight at 37 °C. A representative whole plate photo was taken.

5.2.2 The Papillation Assay in BW25113

BW25113 possesses a *hsdR514* genotype. This is an inactivating point mutation in *hsdR*. Therefore, the strain can be efficiently transformed with the AK21 and AK22 libraries [Supplemental Figure S3].

Two genetic manipulations were required to establish a functional papillation-based spacer acquisition reporter in BW25113. The first was the introduction of the papillation reporter genetic circuit into the *argE* locus. The second was a replacement of the Δ*lacZ4787(::rrnB-3)* mutation with a non-polar deletion of *lacZ*. The Δ*lacZ4787(::rrnB-3)* replaces the P_{lacZ} promoter and the first 1950 bp of *lacZ* with four tandem copies of the *rrnB* transcription terminator, this prevents the expression of downstream *lacY*. Both genetic manipulations were carried out via P1 transduc-

tion, creating the strain RC5282 [Figure 5.7].

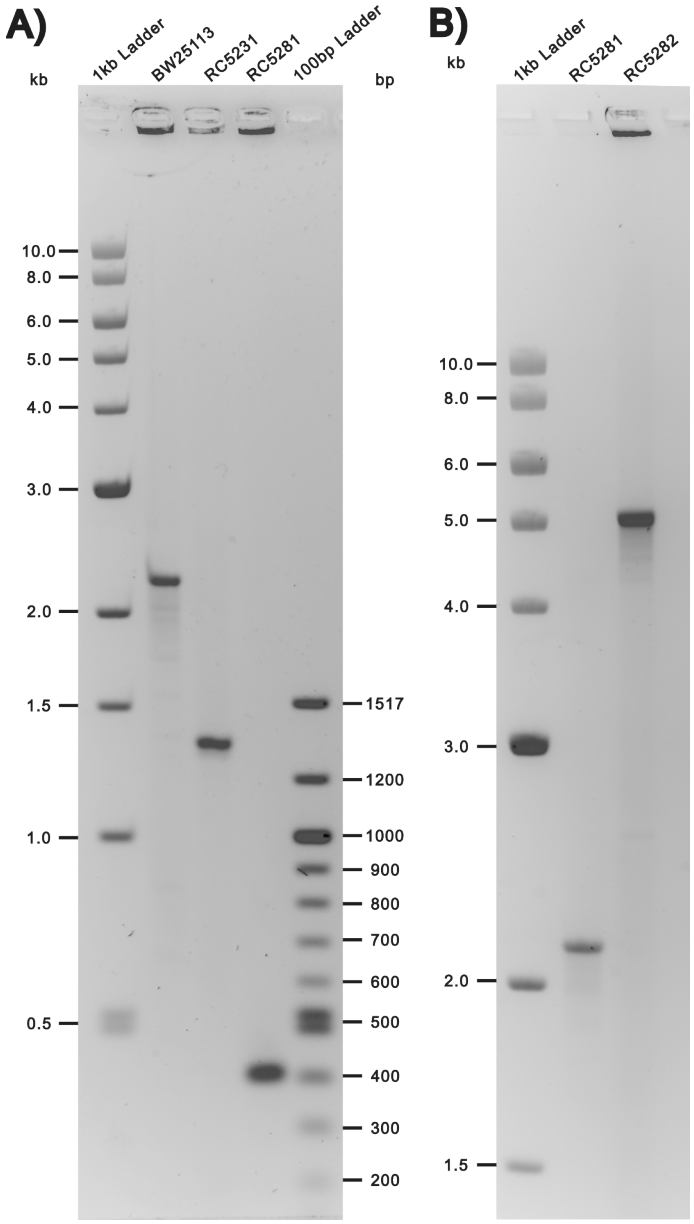

Figure 5.7: PCR analysis confirming the non-polar deletion of *lacZ* and insertion of the papillation assay genetic circuit into *argE*.
(A) The PCR confirmation of the non-polar deletion of *lacZ* in BW25113. Wildtype *lacZ* (A Lane 2, 2194 bp) was replaced with an FRT site-flanked chloramphenicol resistance marker to create RC5231 (A Lane 3, 1257 bp). FLP recombination of RC5231 removed the CmR marker to create RC5281, possessing a non-polar deletion of *lacZ* (A Lane 4, 340 bp). PCR products were visualised via agarose gel electrophoresis.
B) The PCR confirmation of the integration of a papillation reporter KanR cassette into the *argE* locus of RC5281. PCRs of the *argE* locus were conducted in respective strains. The wildtype *argE* locus of RC5281 (B Lane 2, 2128 bp) was disrupted through the introduction of the reporter cassette (B Lane 3, 5064 bp). PCR products were visualised via agarose gel electrophoresis.

SPIN and papillation assays were conducted to test the papillation reporter construct in RC5282 [Figure 5.8]. A SPIN assay showed RC5282 had comparable acquisition levels in liquid culture to that of JB028. Furthermore, in papillation assays, RC5282 produced comparable levels of papillae to JB028 across a gradient of inducer concentrations. An L-arabinose concentration of 0.002% was selected for future screening with RC5282. Individual papillae from this concentration were streaked onto LB-agar supplemented with X-gal. Following overnight growth, PCRs were conducted on individual blue colonies. Visualisation of colony PCRs demonstrated that all colonies contained expanded CRISPR reporter loci. RC5282 behaved as expected and could be used to screen the AK21 and AK22 libraries to identify inhibitors of adaptation.

Papillation assays were conducted in RC5282 transformed with the libraries AK21 and AK22 [Figure 5.9]. Papillation occurred normally in the presence of the libraries in the new reporter strain. I proceeded with library screening at scale.

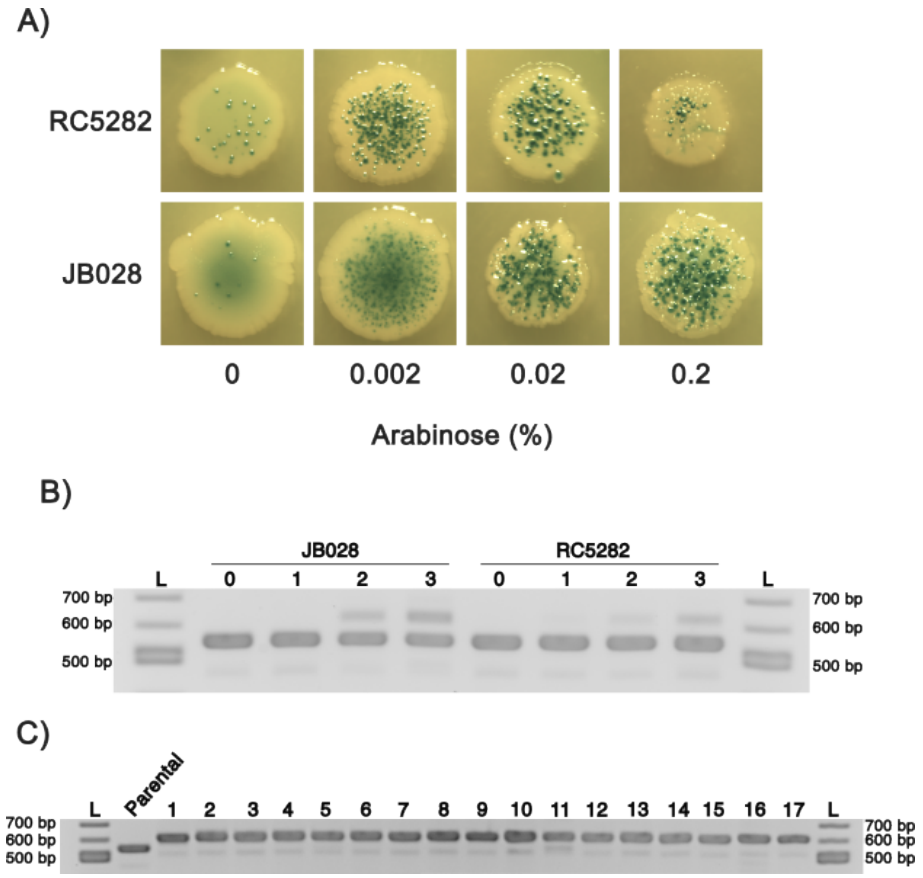

Figure 5.8: Assessment of the acquisition of spacers by RC5282.
(A) Papillation assays were conducted in JB028 and RC5282. Strains were transformed with a pBAD Cas1-Cas2 expression plasmid. L-arabinose was titrated from 0-0.2%. Plates were incubated at 37 °C for five days. Representative colonies are shown.
(B) SPIN assays were conducted with JB028 and RC5282. Strains were transformed with a pBAD Cas1-Cas2 expression plasmid. L-arabinose was present at 0.2%. Cultures were grown at 37 °C on a spinning wheel. Following overnight growth, cultures were diluted 1:1000. This was conducted daily for three days. 1 ml of culture was collected daily, and colony PCRs were performed on it. Day 0 refers to an initial culture grown without an inducer. PCRs were visualised via agarose gel electrophoresis. Bands at 548 bp correspond to the parental CRISPR locus, bands at 609 bp correspond to one spacer acquisition event, bands at 670 bp correspond to two spacer acquisition events. L refers to an NEB 100 bp ladder.
(C) Individual papillae were streaked out from the papillation assay carried out in (A) wherein RC5282 had Cas1-Cas2 expression induced at 0.002% L-arabinose. Individual papillae were subjected to PCR amplification of the CRISPR reporter locus. PCRs were visualised via agarose gel electrophoresis. Bands at 548 bp correspond to the unexpanded reporter locus. Bands at 609 bp correspond to single acquisition events.

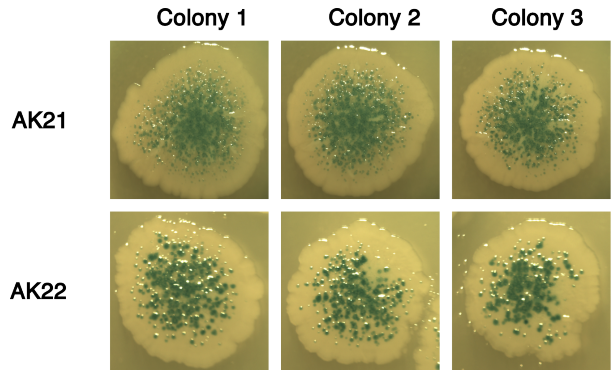

Figure 5.9: Papillation assays in RC5282 in the presence of AK21 and AK22 libraries.
RC5282 was transformed with a pBAD Cas1-Cas2 expression vector and the AK21/22 libraries. Papillation assays were conducted with L-arabinose inducer present at 0.002%. Plates were incubated at 37 °C for five days. Photographs of representative colonies are shown.

5.3 Results

5.3.1 AK22 Screen - Results

The AK22 library was screened first, with the aim of identifying anti-adaptation Acrs present in metagenomic clones producing hypoactive phenotypes in papillation assays. A secondary aim was to optimise screening methods for the subsequent interrogation of the AK21 library.

Papillation assays were conducted with RC5282 transformed with a pBAD Cas1-Cas2 expression vector and the AK22 library. A total of 311 plates were assayed. With ≈45 colonies present per plate, ≈14,000 colonies were screened. An average of 30 kbp metagenomic DNA is present in each AK22 clone. Therefore, 420 Mbp metagenomic DNA was screened, providing ≈28% coverage of the AK22 library.

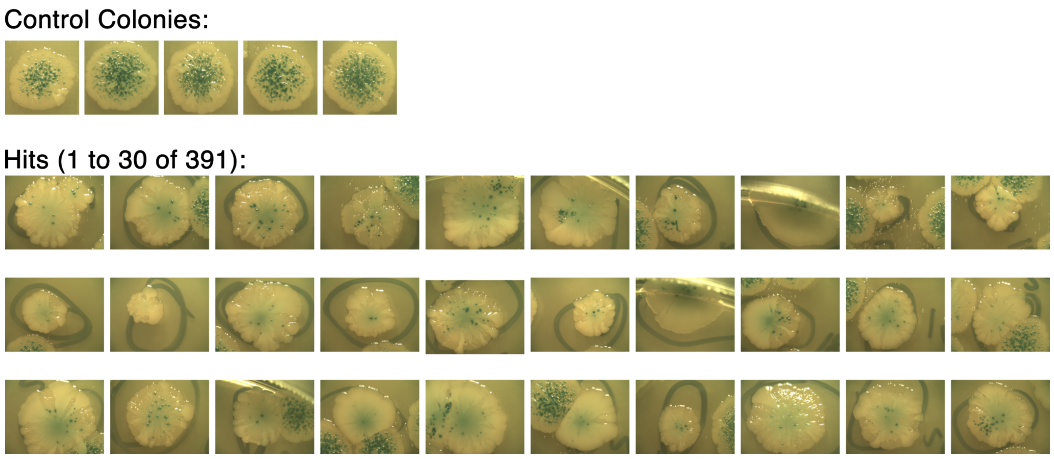

Figure 5.10: Papillation assays screening the AK22 library for Acrs targeting adaptation.
Papillation assays were conducted with RC5282 transformed with a pBAD Cas1-Cas2 expression vector and the AK22 library. L-arabinose inducer was present at 0.002%. Plates were incubated at 37 °C for five days. Images of five exemplar colonies which conducted normal papillation are shown. Images were taken of the 391 colonies identified with hypoactive phenotypes; images of thirty representative hypoactive colonies are shown.

Plates were visually inspected, and 391 colonies possessing hypoactive papillation phenotypes were identified [Figure 5.10]. Candidates were selected based on adherence to the following criteria: an estimated reduction in papillae frequency of $\geq$50%, the presence of a blue halo, the absence of an obvious growth disadvantage & healthy colony morphology, and the absence of growth into adjacent colonies. Hypoactive colonies were present at a rate of 2.8% (391/14,000).

In an initial verification round, candidates were picked and streaked onto LB-agar. Papillation assays were conducted with these streaks [Figure 5.11]. A hypoactive phenotype was maintained by 105 candidates. Therefore, 73% of colonies returned to wildtype.

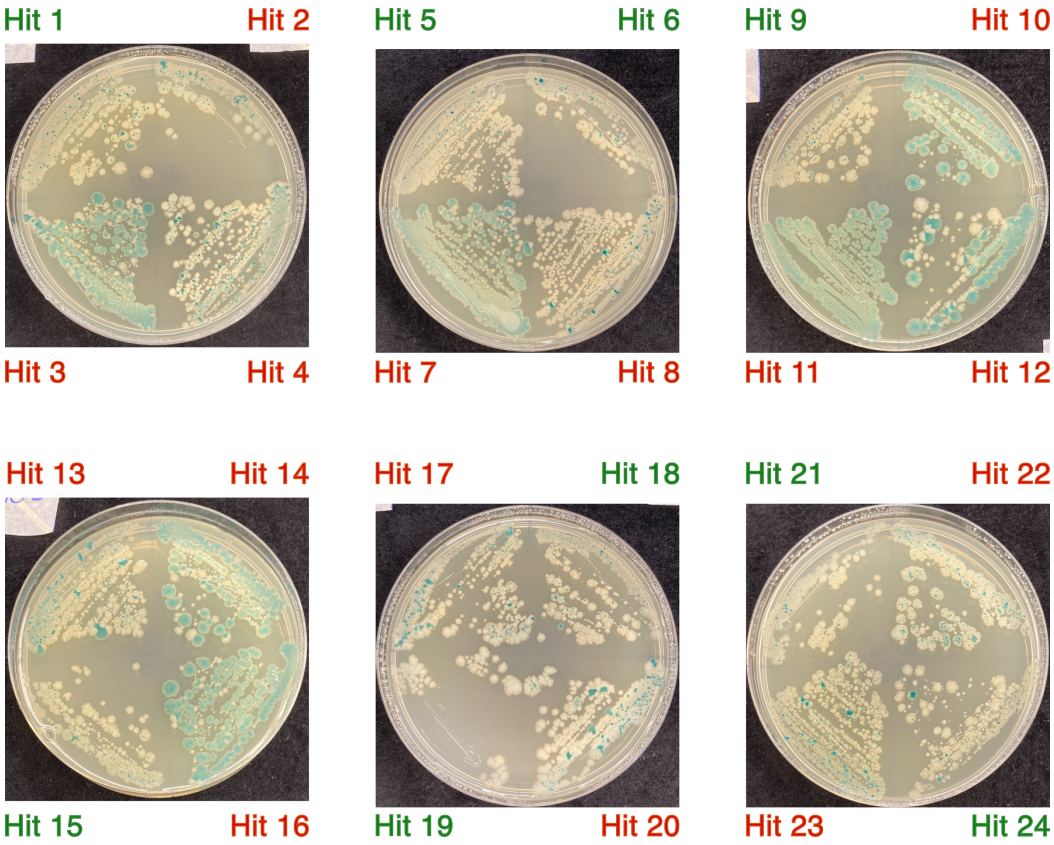

Figure 5.11: Papillation assays of AK22 library candidates streaks
Hypoactive candidates identified in Figure 5.10 were streaked onto LB-agar containing appropriate supplements for a papillation assay. L-arabinose was present at 0.002%. Plates were divided into quadrants and individual candidates were streaked onto separate quadrants. Plates were incubated at 37 °C for five days. Whole plate images were taken. Images of six exemplar plates are provided. Names of candidates maintaining a hypoactive phenotype are in green, while revertants are in red.

Plasmids were purified from the 105 candidates which retained hypoactivity in papillation assays of their streaks. JB028, containing a Cas1-Cas2 expression vector, was transformed with each plasmid individually. Papillation assays were conducted for each candidate [Figure 5.12]. Of 105 colonies tested, 3 maintained a hypoactive phenotype in JB028.

YFP reporter assays were conducted as a final verification step [Figure 5.12]. MLS989 was transformed with a pBAD Cas1-Cas2 expression vector and the 3 candidates that maintained a hypoactive phenotype in JB028 papillation assays. There was no significant decrease in the final YFP+ cell population present for any

candidate relative to a control. Therefore, all 391 initial candidates had been ruled out for further investigation at some stage of the verification procedure.

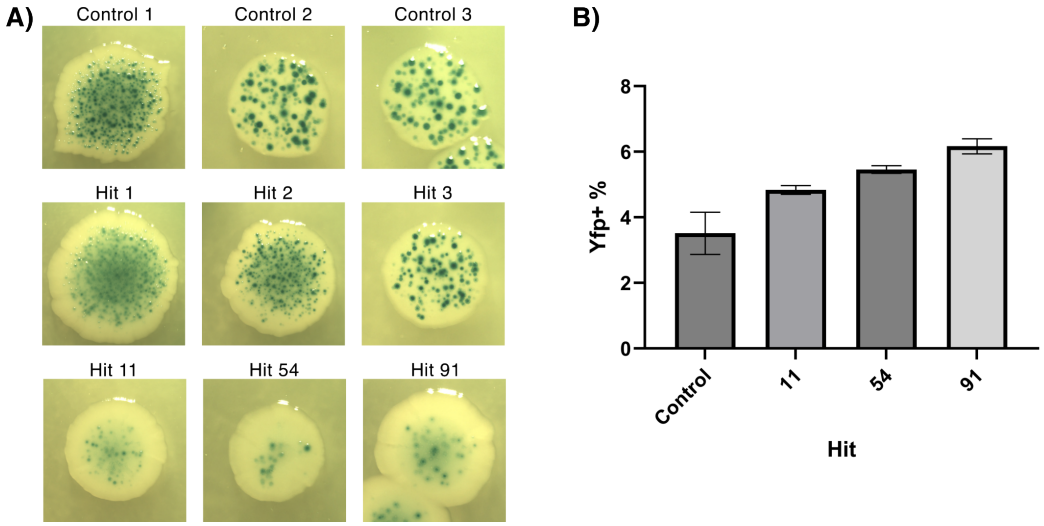

Figure 5.12: AK22 Acr screen final validation stages.
(A) Papillation assays were conducted in JB028 transformed with a pBAD Cas1-Cas2 expression vector and the candidates identified in Figure 5.11. L-arabinose inducer was present at 0.002%. Plates were incubated at 37 °C for five days. Representative images are shown. Three control colonies were derived from clones in the initial screen that did not affect papillation. Hits 1-3 are examples of revertants. Hits 11, 54 and 91 maintained a hypoactive phenotype.
(B) YFP reporter assays were conducted in MLS989 transformed with a pBAD Cas1-Cas2 expression vector and candidates 11, 54 and 91. L-arabinose inducer was present at 0.2%. A dilution of 1:1000 was performed each day over three days. On day three the proportion of YFP+ cells was determined via flow cytometry. Flow cytometry was conducted in triplicate, with 20,000 cells sorted per culture. Mean values are plotted, and error bars are SEM. Data values are found in Supplementary Table S6.

5.3.2 AK22 Screen - Discussion

The screen of the AK22 library identified a large number of initial candidates. These were all ruled out in subsequent rounds of validation. This suggests the initial screen was not sufficiently stringent. Of the ≈14000 screened, a total of 391 colonies were identified with hypoactive phenotypes. With 30 kbp of metagenomic DNA present in each colony, this suggested an Acr arose once per Mbp of metagenomic DNA.

The restreaking validation step removed 73% of colonies, while papillation assays

in JB028 removed 97%. In a subsequent screen, verification via a papillation assay in JB028 was to be prioritised. In the screen of AK21, the restreaking step would be removed, with initial candidates pooled and transformed directly into JB028 for papillation assays. This would facilitate the efficient removal of false positives.

5.3.3 AK21 Screen - Results

Following amendments to the verification methodology, the AK21 library was screened to identify metagenomic clones containing anti-adaptation Acrs. Papillation assays were conducted in RC5282 transformed with a pBAD Cas1-Cas2 expression vector and the AK21 library [Figure 5.13]. A total of 297 plates were assayed. With $\approx$137 colonies present per plate, $\approx$40,700 colonies were screened. An average of 30 kbp metagenomic DNA is present in each AK21 clone. Therefore, 1.22 Gbp metagenomic DNA was screened, providing $\approx$12% coverage of the AK21 library.

Plates were visually inspected, and 1818 colonies possessing hypoactive papillation phenotypes were identified [Figure 5.13]. Candidates were selected based on the previously described criteria. Candidates were inoculated into a pooled LB culture and grown overnight. Plasmid was then extracted, producing a library of hypoactive candidates.

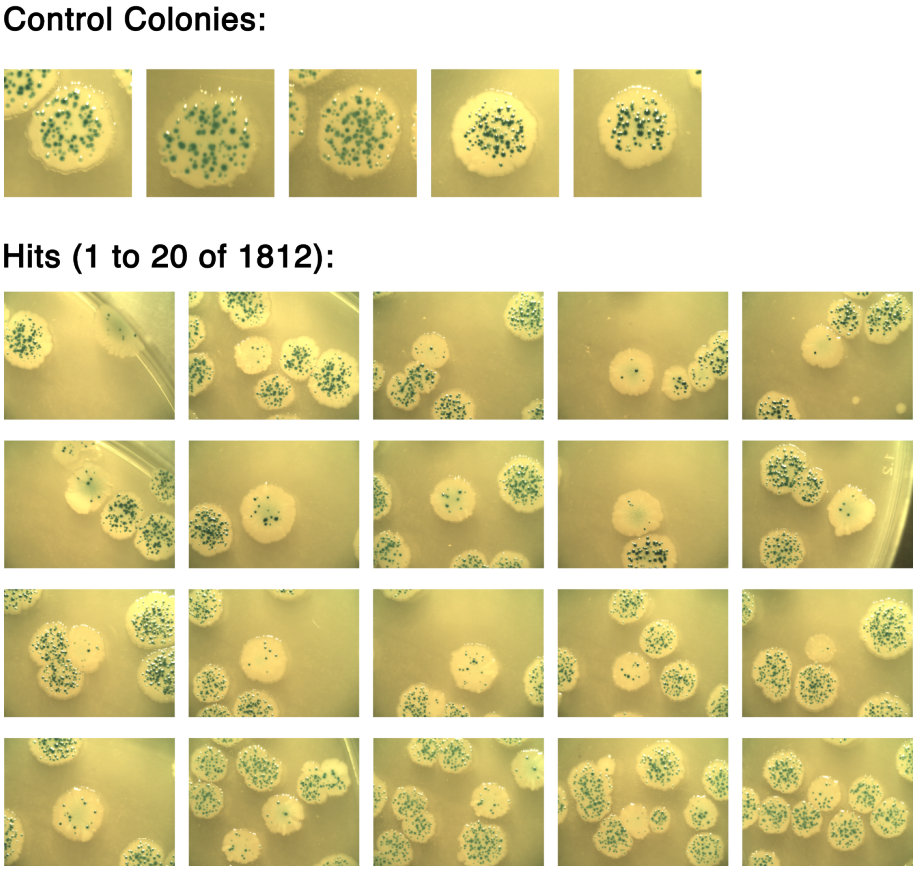

Figure 5.13: Papillation assays screening the AK21 library for Acrs targeting adaptation.
Papillation assays were conducted with RC5282 transformed with a pBAD Cas1-Cas2 expression vector and the AK21 library. L-arabinose inducer was present at 0.002%. Plates were incubated at 37 °C for five days. Images of five exemplar colonies which conducted normal papillation are shown. Images were taken of the colonies identified with hypoactive phenotypes; images of twenty representative hypoactive colonies are shown.

Papillation assays were conducted with JB028 transformed with a Cas1-Cas2 expression vector and the library of AK21 clones exhibiting a hypoactive papillation phenotype. Assays were conducted across 150 plates. With ≈35 colonies present per plate, a total of ≈5250 colonies were assayed. This gave 2.9-fold coverage of the library of hypoactive candidates. Following incubation, plates were visually inspected as previously described. A total of 46 colonies maintained a hypoactive phenotype [Figure 5.14]. Colonies were isolated, and plasmid was purified for further verification.

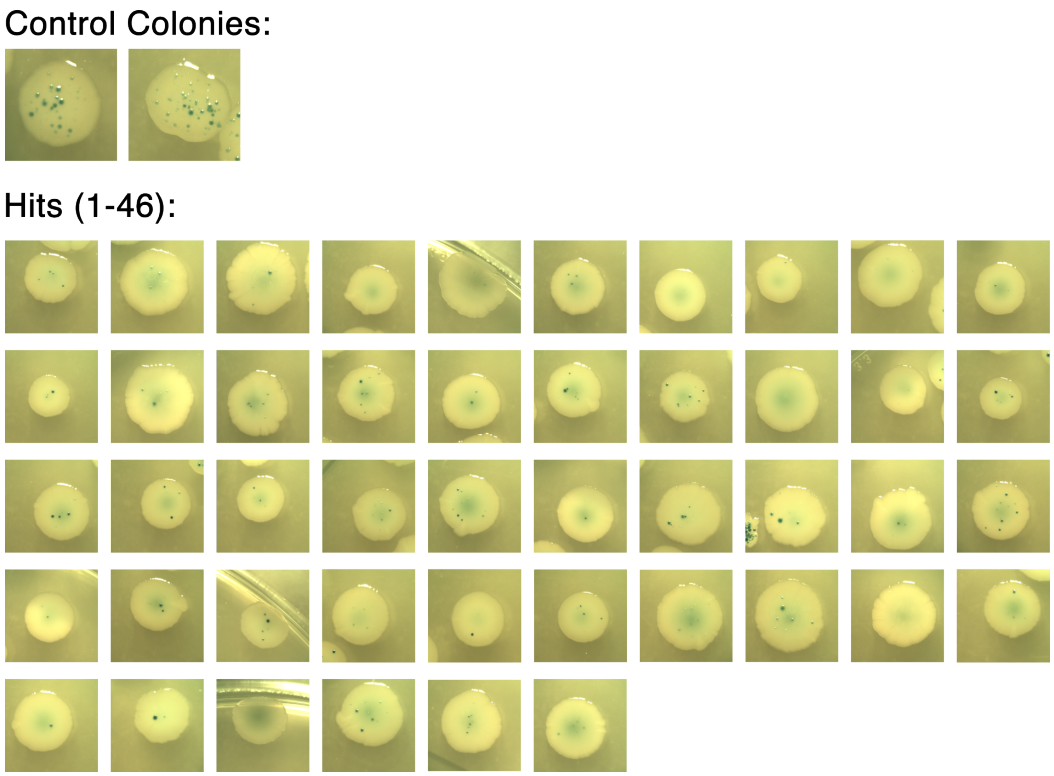

Figure 5.14: Verification papillation assay of AK21 library initial candidates . Papillation assays were conducted in JB028 transformed with Cas1-Cas2 expression vector and a library of hypoactive candidates from the AK21 library identified in Figure 5.13. L-arabinose inducer was present at 0.002%. Plates were incubated at 37 °C for five days. Images of two control colonies which conducted normal papillation are shown, these derive from clones that did not affect papillation. Photographs of the 46 colonies identified to possess hypoactive phenotypes are shown.

A series of YFP reporter assays were conducted in MLS989 transformed with a pBAD Cas1-Cas2 expression vector and plasmid from the 46 candidates [Figure 5.15]. A significant reduction in spacer acquisition was not seen for any of the candidates screened through flow cytometry. Therefore, all 1812 initially identified hypoactive colonies have been ruled out at some stage of the validation procedure.

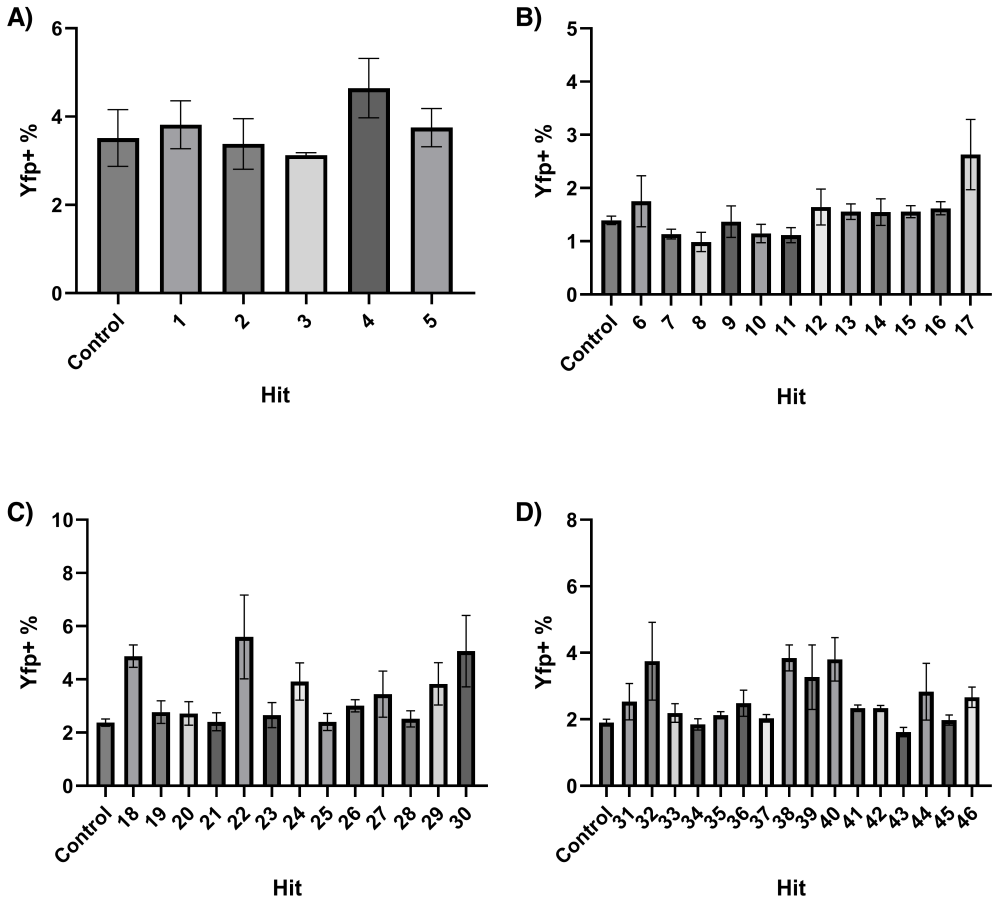

Figure 5.15: Verification YFP reporter assays of AK21 library candidates.
YFP reporter assays were performed with MLS989 transformed with a pBAD Cas1-Cas2 expression vector and plasmid from the 46 candidates identified in [Figure 5.14]. L-arabinose inducer was present at 0.2%. A dilution of 1:1000 was performed each day over three days. On day three the proportion of YFP+ cells was determined via flow cytometry. Flow cytometry was conducted in triplicate, with 20,000 cells sorted per culture. Mean values are plotted, and error bars are SEM. Data values are found in Supplementary Table S6.

5.3.4 AK21 Screen - Discussion

A screen of the AK21 library failed to identify any Acrs targeting adaptation. This is despite the screening of a significantly larger volume of metagenomic DNA and the identification of a greater number of putatively hypoactive candidates. As seen in the prior screen of the AK22 library, many initial candidates identified in RC5282 were ruled out as false positives upon testing in JB028. Future screens should be conducted in a working JB028Δ*hsdR*. This would combine the high transformation

efficiency of RC5282 with the low rate of false positives of JB028, facilitating more efficient screening.

5.3.5 Sequencing Candidates

Several candidates from the AK21 screen were sequenced to investigate why a metagenomic clone may be identified as a false positive in the papillation assay. Candidates 3, 8 and 43 were sequenced as they produced the lowest rate of spacer acquisition in each of their respective rounds of YFP flow cytometry assays. BLAST searches were performed, and a list of annotated ORFs is provided in Supplemental Table 5.3.

A gene identified in candidate 3 had 56.8% identity over 68% of *E. coli* K-12 *ftsH*. Overexpression of *ftsH* has been shown to increase the rate of spacer acquisition [229]. Expressing a truncated homolog of *ftsH* may sequester FtsH interaction partners to exhibit a dominant negative phenotype.

A gene identified in candidate 8 had 50.9% identity over 99.2% of *Prosthecochloris aestuarii uvrC*. In *E. coli*, incisions are made during nucleotide excision repair by UvrC in complex with UvrA and UvrB [257]. Unpublished data from a systematic screen of *E. coli* suggest *uvrB* knockout may inhibit spacer acquisition (Unpublished, Jack Braithewaite). Expressing a homolog of *uvrC* may sequester UvrB to exhibit a dominant negative phenotype.

A single gene with no assigned function was identified in candidate 44. I cannot speculate on why this plasmid causes hypoactivity in papillation assays.

5.3.6 AcrVA5 Does Not Inhibit Adaptation in *E. coli* Type I-E CRISPR-Cas

AcrVA5 has been identified as a potential broad-spectrum inhibitor of adaptation [205]. I conducted assays to determine if it exhibited inhibitory behaviour in the *E. coli* Type I-E system. AcrVA5 was cloned into a vanillic-acid inducible P_{van} expression vector [214]. Papillation assays and YFP reporter assays were conducted in the presence of induced AcrVA5 [Figure 5.16].

AcrVA5 induction from 5-20 μM vanillic acid caused no change in behaviour in papillation assays when compared to an uninduced control. Growth was retarded at 50 μM induction, and no growth was seen at 100 μM vanillic acid. AcrVA5 does not inhibit the papillation assay. This suggests that AcrVA5 does not inhibit spacer acquisition.

YFP reporter assays showed a very small but significant decrease in adaptation at 50 μM vanillic acid relative to an uninduced control (P=0.0238). However, this decrease disappeared when induction increased to 100 μM. This suggests that AcrVA5 likely does not inhibit spacer acquisition.

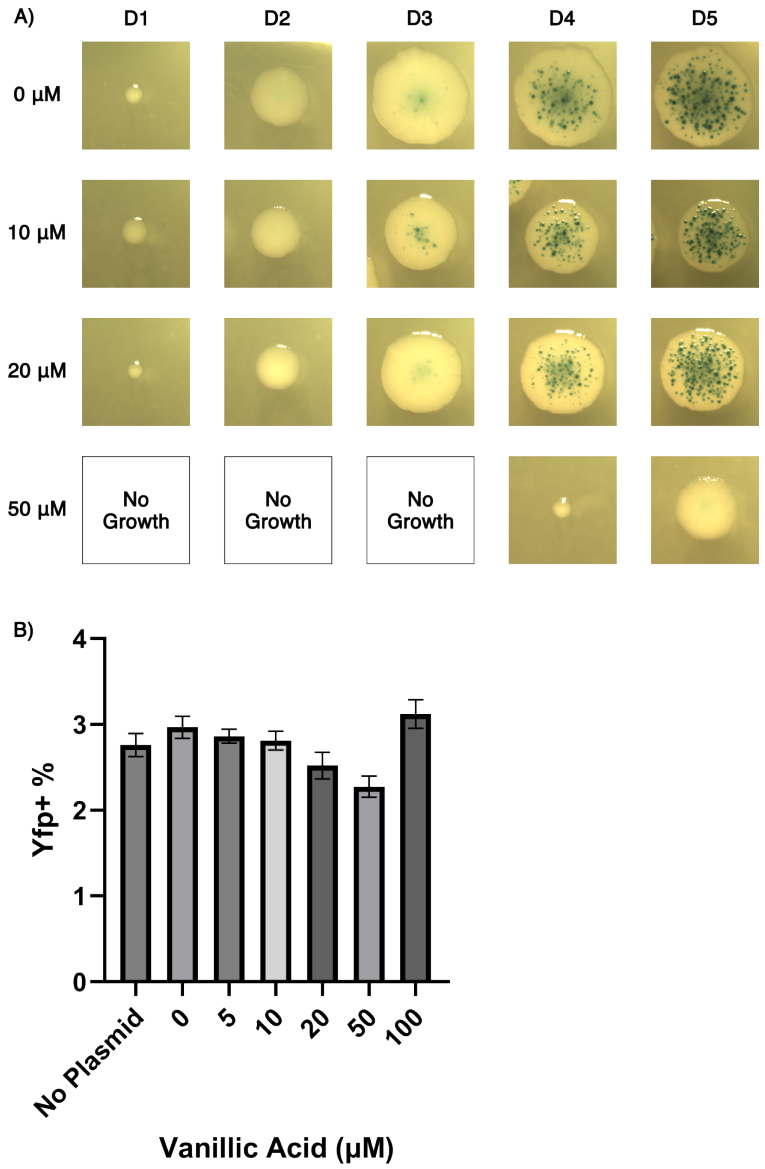

Figure 5.16: Assays determining rates of spacer acquisition in the presence of AcrVA5.
(A) Papillation assays were conducted in JB028 transformed with a pBAD Cas1-Cas2 expression vector and P_{van} AcrVA5 expression vector. Cas1-Cas2 was induced at 0.02% L-arabinose, and AcrVA5 was induced between 0-50 μM vanillic acid. Plates were incubated at 37 °C for 5 days. No growth was seen above 50μM. Representative photos were taken of an individual colony daily over 5 days (D1-D5).
(B) YFP reporter assays were conducted with JB028 transformed with a pBAD Cas1-Cas2 expression vector and P_{van} AcrVA5 expression vector. Cas1-Cas2 was induced at 0.02% L-arabinose, and AcrVA5 was induced between 0-100 μM vanillic acid. A dilution of 1:1000 was performed each day over three days. The proportion of YFP+ cells was determined via flow cytometry. Six replicates were conducted per culture, with 20,000 cells sorted per replicate. No plasmid refers to the absence of the P_{van} AcrVA5 expression vector. Mean values are plotted, and error bars are SEM. Data values are in Supplementary Table S7.

5.4 Chapter Summary

- Across two screens, 1.86 Gbp environmental metagenomic DNA was screened for inhibitors of spacer acquisition.
- Of 2,203 metagenomic clones initially identified as possessing hypoactive phenotypes, all were ruled out in successive validation stages.
- Metagenomic clones identified as hypoactive in papillation assays did not maintain their phenotype in YFP reporter assays.
- Metagenomic clones that produce hypoactive papillation assays contained homologs of ancillary spacer acquisition genes. These may have caused false positives.
- The anti-adaptation protein AcrVA5 does not inhibit spacer acquisition in the *E. coli* Type I-E CRISPR system.

Chapter 6

Development of a Repressor-Based Genetic Circuit for Universal Detection of Spacer Acquisition

6.1 Introduction

Multiple different reporter systems have been developed to study spacer acquisition [212] [219] [213] [230]. With the exception of a recently described system [230], these circuits are restricted to studying CRISPR systems in which the insertion of a spacer-repeat unit into the reporter array causes a framehsift. Efforts to circumvent this have relied on the acquisition of a prespacer containing a ribosome binding site and start codon to induce expression of a downstream reporter gene, significantly reducing the sensitivity of the assay to spacer acquisition events [230].

Frameshift-based genetic circuits can only be applied to the study of 62% of CRISPR systems [Figure 6.1]. A repressor-based reporter system which would facilitate the universal detection of spacer acquisition without a loss of assay sensitivity has been hypothesised but not built [Figure 6.3] [212]. This chapter describes my efforts to construct this reporter system in the Type I-E system of *E. coli* as a proof of principle.

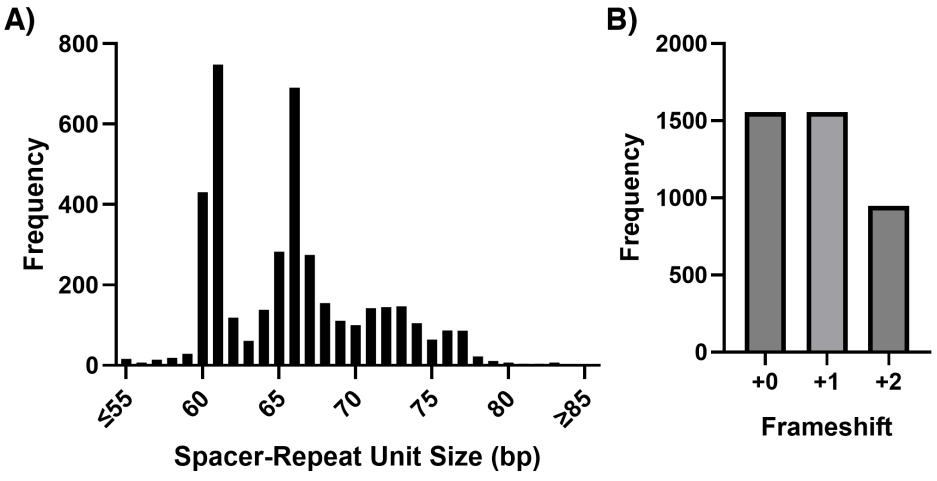

Figure 6.1: Frequency of frameshifts in CRISPR arrays.
(A) Distribution of sizes of spacer-repeat units derived from the 4061 CRISPR arrays within the CRISPRDetect Database [258].
(B) Distribution of frameshifts caused by the acquisition of the spacer-repeat units described in part (A).

6.2 Results

6.2.1 Construction and Testing of a TetR-based Universal Reporter

The tn*10 tet* locus encodes *tetA* and *tetR*, expressed from partially overlapping promoters [259]. Within the intergenic region, TetR may bind to the operator sites O_1 and O_2. The promoter P_{tetA} expresses *tetA*, however the binding of TetR to either operator represses this promoter. The promoters P_{tetR1} and P_{tetR2} express *tetR*. Binding of TetR to O_1 represses both promoters, but binding at O_2 does not repress the P_{tetR2} promoter, permitting *tetR* expression [260]. TetR exhibits a four-fold greater affinity for O_2 than O_1. This dynamic facilitates the stable repression of *tetA* and the autoregulation of *tetR* expression.

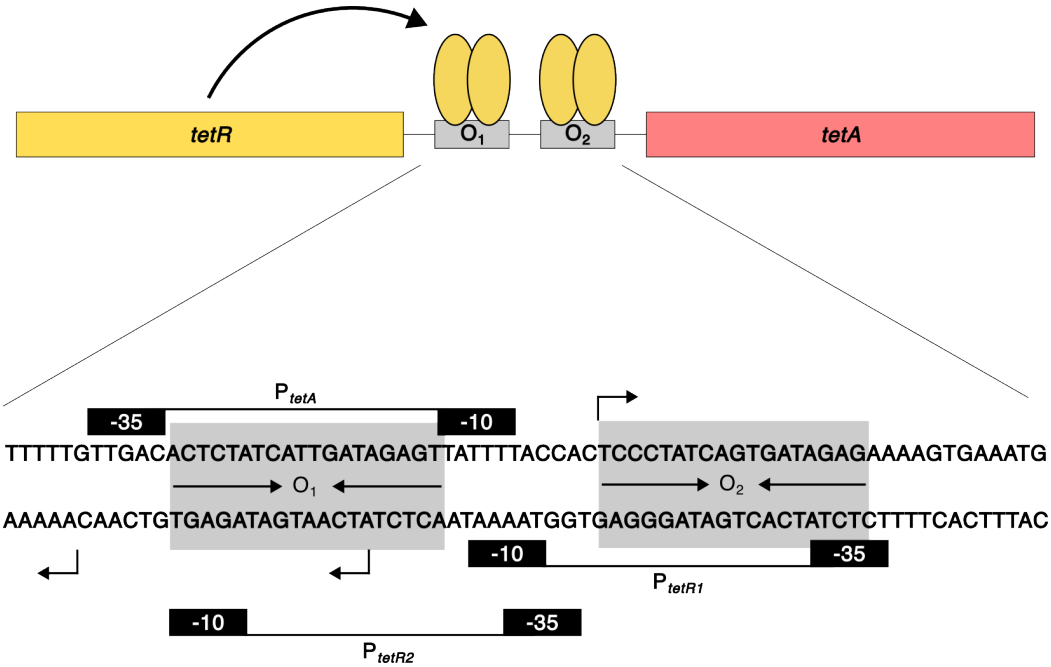

Figure 6.2: Structure of the tn*10 tet* intergenic region.
The tn*10 tet* intergenic region is made up of the inverted repeats of the O_1 and O_2 *tetO* sites (grey), alongside the pair of *tetR* promoters P_{tetR1} and P_{tetR2} and the single *tetA* promoter P_{tetA}. Transcriptional start sites associated with each promoter are indicated by their respective right-handed arrows. The binding of dimerised TetR (yellow) to each *tetO* site differentially inhibits the expression of *tetR* (yellow) and *tetA* (red).

The universal reporter system described in Figure 6.3 was constructed from the tn*10 tet* locus [Figure 6.2]. The *lacZ* gene was cloned in place of *tetA*. A CRISPR array, preceded by a start codon, was introduced upstream of *tetR*. TetR has been shown to tolerate N-terminal fusions [261]. Therefore, TetR would be expressed with the CRISPR array present as an N-terminal fusion. The O_1 site was mutated to remove the inverted repeat structure whilst maintaining promoter activity [262]. This would remove auto-repression of the CRISPR array-TetR fusion protein, inducing constitutive expression. This was anticipated to mitigate leaky expression of *lacZ*. Binding of the CRISPR array-TetR fusion to the remaining O_2 site would inhibit *lacZ* expression from the P_{tetA} promoter.

Acquiring a spacer containing a stop codon would cause premature termination of the CRISPR array-TetR fusion protein within the N-terminal peptide, leading to derepression of *lacZ*. The cell then metabolises lactose as a carbon source, providing a selective growth advantage. This would facilitate the visualisation of spacer acquisition via a papillation assay.

Reporter constructs were created for the Type I (T) and II (A) variants of the *E. coli* Type I-E CRISPR system. Positive controls for reporter systems were also made. These possessed in-frame stop codons in the TetR fusion protein, inhibiting expression.

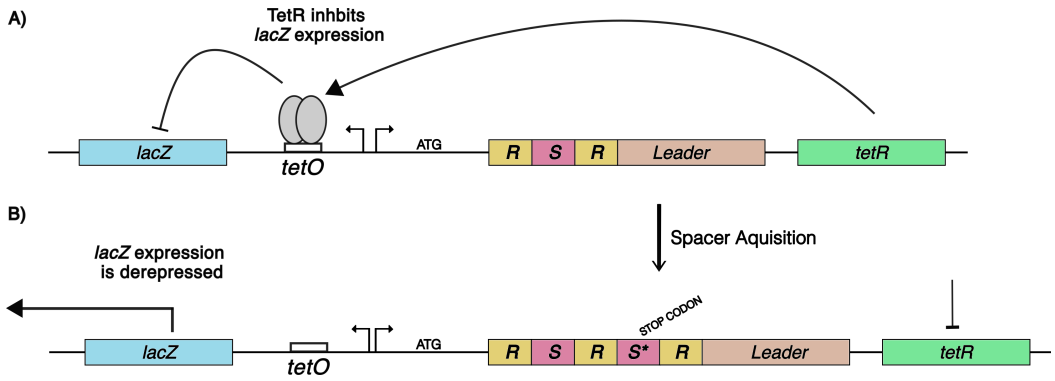

Figure 6.3: Overview of a TetR-based universal reporter for spacer acquisition [212].
(A) The expression of the tetracycline-inducible repressor (*tetR*) and *lacZ* genes are under the control of the bidirectional tet promoter (P_{tet}). The tet operator (*tetO*) site influencing the promoter of *tetR* is removed. Therefore, the reporter CRISPR locus-TetR fusion protein is constitutively expressed. This occupies the *tetO* site within the promoter of *lacZ* to repress its expression.
(B) Acquisition of a spacer containing a stop codon induces the premature termination of TetR. Therefore, repression of the *lacZ* promoter is lifted, resulting in *lacZ* expression and subsequent metabolism of lactose and X-gal.
Gene expression is depicted as a long black arrow, and gene repression is depicted as a black line ending in a black bar.

When assayed, the Type I (T) and II (A) reporter constructs did not acquire spacers. SPIN assays of both TetR-based reporters showed no detectable acquisition [Figure 6.4a]. Furthermore, growth on LB-agar supplemented with X-gal showed no phenotypic difference between each reporter and their respective positive control [Figure 6.4b]. Successive iterations of this design failed to resolve these issues (data not shown). The universal reporter, as initially proposed [212], was therefore determined to be non-functional, and an alternative approach was taken to its construction.

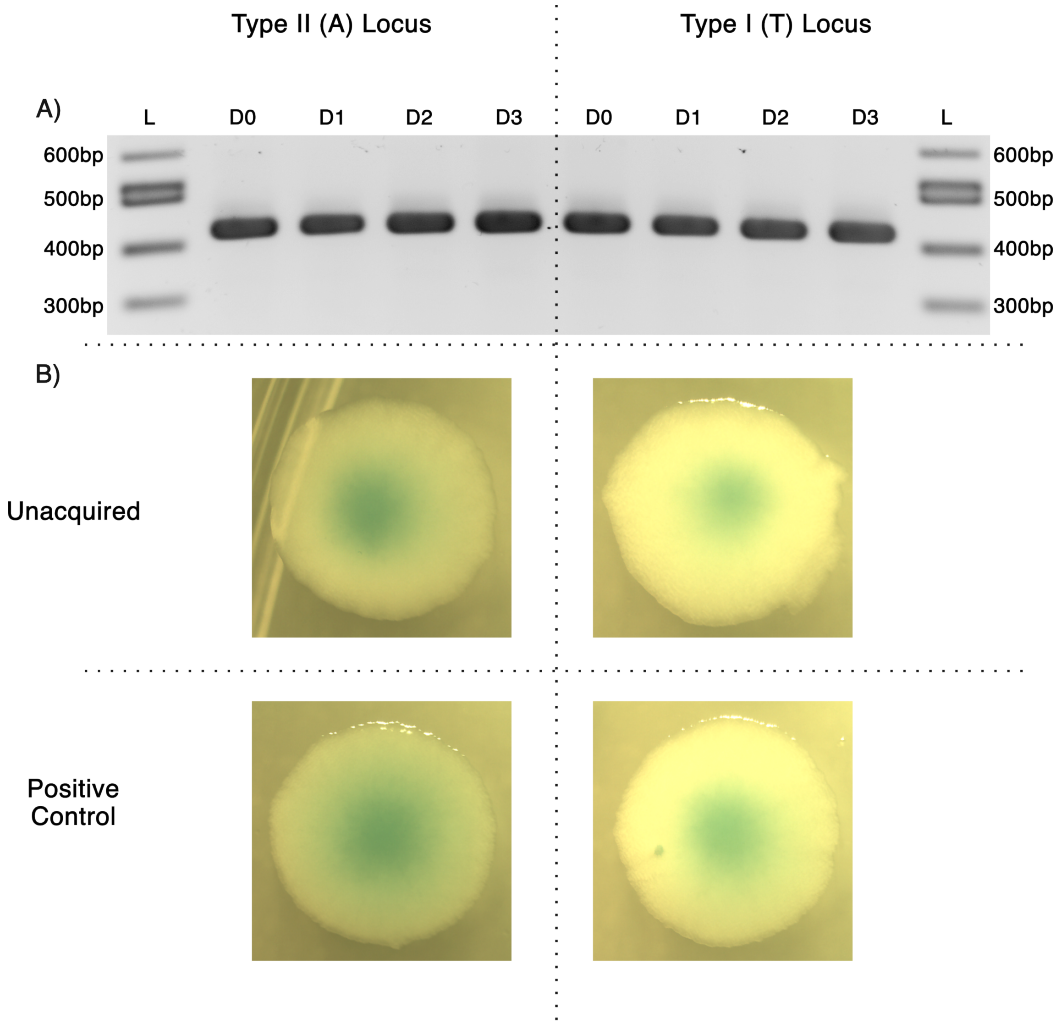

Figure 6.4: Overview of assays testing the TetR-based universal reporters.
(A) SPIN assays were conducted for both the Type I (A) and II (T) TetR-based universal reporters in BW25113. Cas1-Cas2 expression was induced from a pBAD plasmid at 0.2% L-arabinose. Cultures were grown overnight and diluted 1:1000 daily. Day 0 refers to an initial culture grown in the absence of an inducer. Colony PCRs were performed on day 0 to day 3 cultures (D0-D3). PCRs were visualised via agarose gel electrophoresis. Bands at 419 bp derive from unexpanded arrays. L refers to an NEB 100 bp ladder.
(B) The Type I (T) and II (A) TetR-based universal reporters, alongside their respective positive controls, were transformed in the Δ*lacZ* strain BW25113 and grown on LB-agar supplemented with X-gal at 37 °C for 5 days. Representative colonies are shown.

6.2.2 Construction and Testing of a VanR-based Universal Reporter

An alternative design for a universal reporter of spacer acquisition was constructed [Figure 6.5]. The vanillic acid-inducible repressor gene (*vanR*) from *Caulobacter crescentus* was expressed from a constitutive P_{J23100} promoter. A start codon and CRISPR array were introduced between *vanR* and its promoter. These would form an N-terminal fusion peptide to VanR. The reporter gene *lacZ* was expressed from the P_{van} promoter. In the parental state, functional VanR would occupy *vanO* sites within the P_{van} promoter and repress *lacZ* [214]. The acquisition of a spacer containing an in-frame stop codon would truncate the VanR protein within the N-terminal peptide, causing a loss of function. This would derepress *lacZ*, causing its expression. The cell could then metabolise lactose as a carbon source and produce a blue phenotype in the presence of X-gal.

Vanillic acid induces the derepression of P_{van} by VanR [214]. Supplementation with vanillic acid provides a positive control for this system as it derepresses *lacZ*.

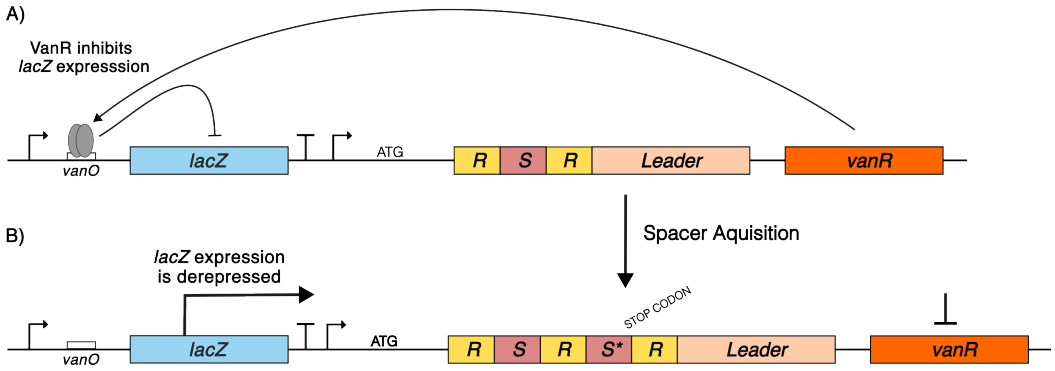

Figure 6.5: Schematic of a VanR-based universal reporter circuit.
(A) The expression of *vanR* is induced by a constitutive P_{J23100} promoter. The expression of *lacZ* is controlled by a P_{van} promoter. A CRISPR array-VanR fusion acts as a repressor of *lacZ* by binding to *vanO* sites in P_{van}.
(B) The acquisition of a spacer containing a stop codon causes premature termination of *vanR*. This induces the derepression of *lacZ*. LacZ metabolises lactose as a carbon source and produces a blue phenotype when grown in the presence of X-gal.
Gene expression is depicted as a black arrow, whilst gene repression is depicted as a black line ending in a black bar.

A VanR-based reporter was constructed using the Type II (A) CRISPR locus. A SPIN assay demonstrated the reporter acquired spacers [Figure 6.6A]. Growth on LB-agar supplemented with X-gal showed no distinguishable phenotypic difference between the reporter and its positive control supplemented with vanillic acid [Figure 6.6b].

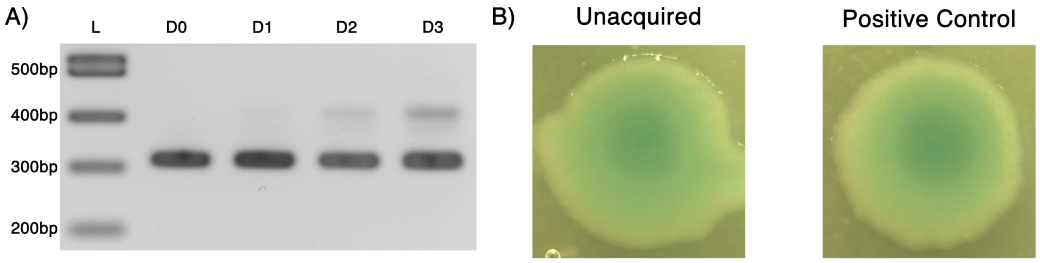

Figure 6.6: Overview of assays testing the VanR-based universal reporter.
(A) A SPIN assay was conducted. Cas1-Cas2 expression in BW25113 was induced from a pBAD plasmid at 0.2% L-arabinose. Cultures were grown overnight and diluted 1:1000 daily. Day 0 refers to an initial culture grown in the absence of an inducer. Colony PCRs were performed on day 0 to day 3 cultures (D0-D3). PCRs were visualised via agarose gel electrophoresis. Bands at 419 bp derive from unexpanded arrays. L refers to an NEB 100 bp ladder.
(B) The VanR-based universal reporter was transformed into the Δ*lacZ* strain BW25113 and grown on LB-agar supplemented with X-gal. A positive control was further supplemented with 100 μM vanillic acid. Following incubation at 37 °C for 5 days, representative colonies are shown.

6.2.3 Troubleshooting of TetR- and VanR-based Universal Reporters

TetR- and VanR-based reporters failed to generate phenotypes distinct from their respective positive controls when grown on LB-agar with X-gal. Their *lacZ* reporter genes were replaced with *YFP* to permit the quantification of expression. This was to determine the causes of the absence of a difference in phenotypes between reporters and their respective positive controls. Fluorescence produced by reporters and their positive controls was measured following overnight growth [Figure 6.7].

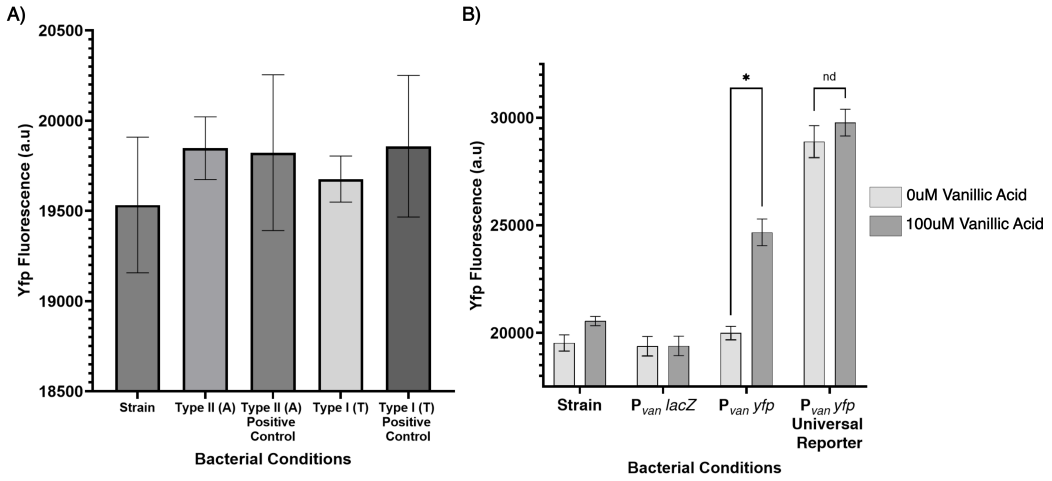

Figure 6.7: Fluorescence induction assays testing the TetR- and VanR-based reporters expressing *YFP*.
(A) The *YFP* gene was cloned into the Type II (A) and I (T) variants of the TetR-based reporter, alongside their positive controls. Plasmids were transformed into BW25113. Following overnight growth, O.D 600 was normalised to 1.0. The fluorescence was determined using an Amersham Typhoon with a Cy2 filter set.
(B) The *YFP* gene was cloned into the VanR-based reporter and its parental plasmid lacking an N-terminal CRISPR array. Plasmids were transformed into BW25113. Following overnight growth in the presence or absence of 100 μM vanillic acid inducer, O.D 600 was normalised to 1.0. The fluorescence was determined using an Amersham Typhoon with a Cy2 filter set.
Six replicates were performed per conditions. Means are plotted and error bars are SEM. Data values are found in the Supplementary Table S11.

Both the Type I (T) and II (A) variants of TetR-based reporter and their respective positive controls did not fluoresce greater than the host strain, BW25113, lacking

YFP [Figure 6.7a]. This suggests the P_{tetA} promoter is non-functional following the oblation of the O_1 site. I did not troubleshoot this reporter further due to this issue alongside the lack of detectable acquisition at the CRISPR locus.

The stepwise construction of the VanR-based reporter enabled analysis of how introducing an N-terminal CRISPR array affected the system [Figure 6.7b]. VanR was functional without an N-terminal CRISPR array, as induction at 100 μM vanillic acid caused a significant increase in fluorescence. The introduction of the reporter CRISPR array caused a loss of sensitivity to vanillic acid and full induction in the absence of an inducer. This shows that the presence of the N-terminal peptide on VanR prevents repression of P_{van}. Subsequent construction of a universal reporter required the identification of a repressor that could tolerate an N-terminal CRISPR array fusion.

The marionette collection is a set of tools designed to provide modular control over gene expression in *E. coli* [214]. Seven repressors within the collection were identified as candidates for the construction of a universal reporter. Activators and poorly performing repressors from the collection were excluded; LacI was also excluded as its derepression during lactose supplementation would inhibit its use in a papillation assay. The CRISPR array of MLS990, the positive control strain of a published YFP-based reporter for spacer acquisition, was placed upstream of selected repressors on their respective multi-copy P15A-*ori* inducible expression plasmids from the marionette collection [213]. The MLS990 CRISPR array was present as an in-frame N-terminal fusion to each respective marionette repressor. The fluorescence of a *YFP* reporter gene was then measured in the parental strain, in the absence of a repressor gene, in the presence of a repressor gene, and in the presence of the repressor gene expressed with an N-terminal CRISPR array [Figure 6.8].

The wildtype PcaU repressor did not repress *YFP* expression. When N-terminal CRISPR arrays were introduced, PhlF and BetI became completely non-functional.

Repression was disrupted but not oblated for the remaining proteins (CymR, VanR, TetR and TtgR). CymR most potently retained repression of *YFP* when an N-terminal CRISPR array was present. CymR was therefore selected for the further development of a universal reporter [Figure 6.9].

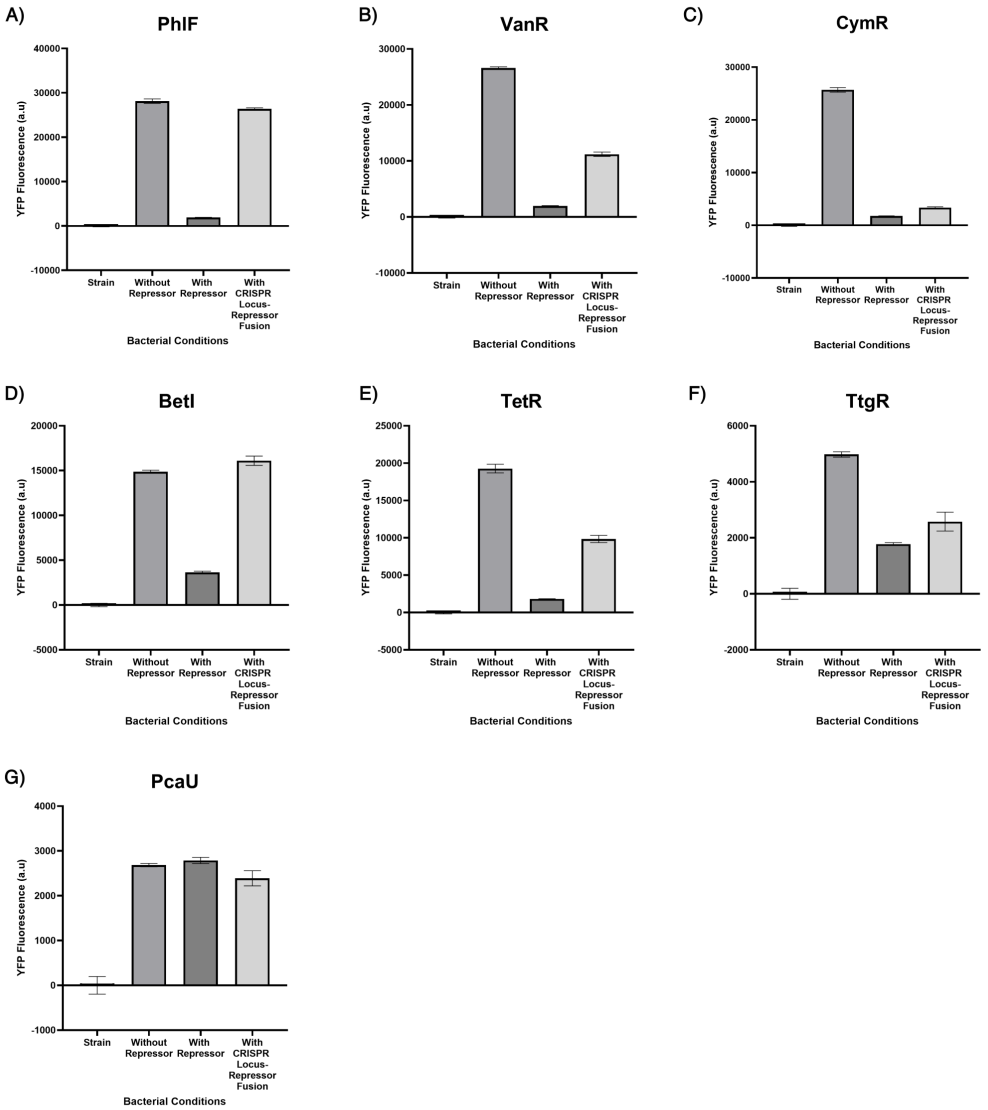

Figure 6.8: Overview of fluorescence induction with various iterations of repressors from the marionette collection [214].
Respective multi-copy P15A-*ori* plasmids were cloned and transformed into the strain BW25113. Following overnight growth at 37 °C, O.D 600 was normalised to 1.0. Fluorescence was determined using an Amersham Typhoon with a Cy2 filter set. Fluorescence was measured in the parental strain, in the absence of a repressor gene, in the presence of a repressor gene, and in the presence of a repressor gene expressed with an N-terminal CRISPR array. Fluorescence was measured for the following repressors:
(A) PhlF (B) VanR (C) CymR (D) BetI (E) TetR (F) TtgR (G) PcaU
Six replicates were performed per condition. Mean values are plotted, and error bars are SEM. Data values can be found in Supplementary Table S11

6.2.4 Construction and Testing a CymR-based Universal Reporter

Figure 6.9 outlines a CymR-based genetic circuit containing an N-terminal CRISPR array-CymR fusion protein, shown to retain repression of *YFP* [Figure 6.8]. CymR, acquired from *Pseudomonas putida*, acts to maintain cysteine homeostasis through repression of biosynthesis and uptake genes [263]. It forms a dimer, using a helix-turn-helix domain to bind at natural or synthetic operator sites (CuO) and induce repression of downstream genes [264]. Several CymR fusion proteins have been previously developed, with a peptide bound to the C-terminal dimerisation domain [265][266][267]. Therefore, Figure 6.8 represents the first example of CymR remaining functional with a peptide fused specifically to the N-terminal DNA binding domain [264]. N-terminal fusions have been shown to affect repressor behaviour [261].

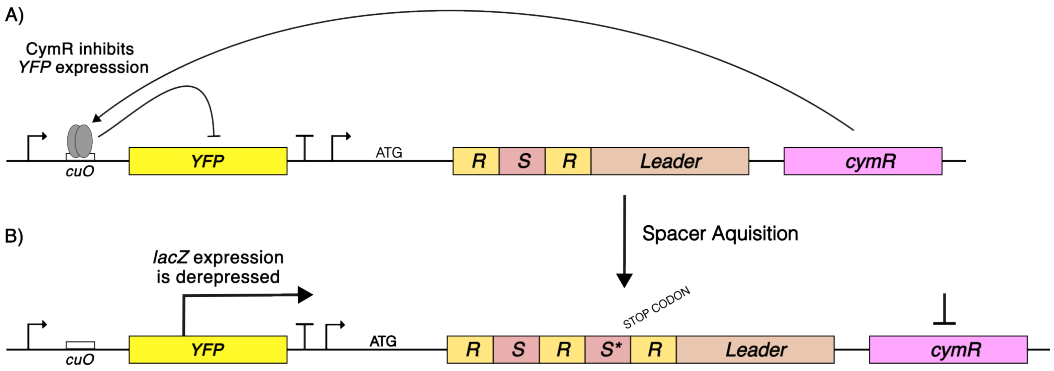

Figure 6.9: CymR-based universal reporter schematic.
(A) The expression of *cymR* is under the control of its evolved marionette promoter [214]. Expression of *YFP* is controlled by the P_{cu} promoter, with repression by CymR facilitated by the presence of its respective operator site (*cuO*) within P_{cu}. The *E. coli* Type II (A) CRISPR locus is expressed as an N-terminal fusion to CymR.
(B) Acquiring a spacer containing a stop codon induces premature termination of *cymR*. Therefore, the repression of *YFP* is lifted, generating a measurable fluorescent phenotype.
A black arrow depicts gene expression, whilst gene repression is depicted as a black line ending in a black bar.

The induction of CymR by cuminic acid was assayed to determine how the N-terminal CRISPR array affects behaviour [Figure 6.10]. Assays were carried out in a multi-copy P15A-*ori* and a single-copy F-*ori* plasmid, expressing CymR with or without an N-terminal CRISPR-array fusion peptide present. CymR repression remained titratable in both single- and multi-copy plasmids. However, the addition of an N-terminal fusion peptide caused a loss of titraitability in both plasmids, with *YFP* fully induced at 5μM cuminic acid. This demonstrated that the addition of an N-terminal fusion altered the behaviour of CymR at the P_{cu} promoter. Furthermore, moving the reporter circuit to a single-copy plasmid caused an increase in the leaky expression of *YFP* from the P_{cu} promoter. However, there was still a significant difference between leaky *YFP* expression and fully induced *YFP* expression. I determined these phenotypes were sufficiently distinct to continue the development of a CymR-based genetic circuit.

Positive controls were constructed for the *F*-ori CymR-based universal reporter. The first contained the CRISPR array from the published YFP reporter strain MLS989, which possesses a -1 frameshift relative to the MLS990 array [213]. The second contained a pair of stop codons within the spacer in the MLS990 CRISPR array but no frameshift. Fluorescence of the reporter was measured against that of controls [Figure 6.11].

The reporter had significantly greater fluorescence induction than a reporter lacking an N-terminal CRISPR array, reinforcing prior evidence of leaky expression when the reporter is present on a single-copy F-*ori* plasmid. However, both positive controls were significantly more fluorescent than the reporter and equally as fluorescent as a reporter lacking a repressor. This suggests introducing a frameshift **or** stop codons inactivates CymR, producing a phenotype distinguishable from the parental population.

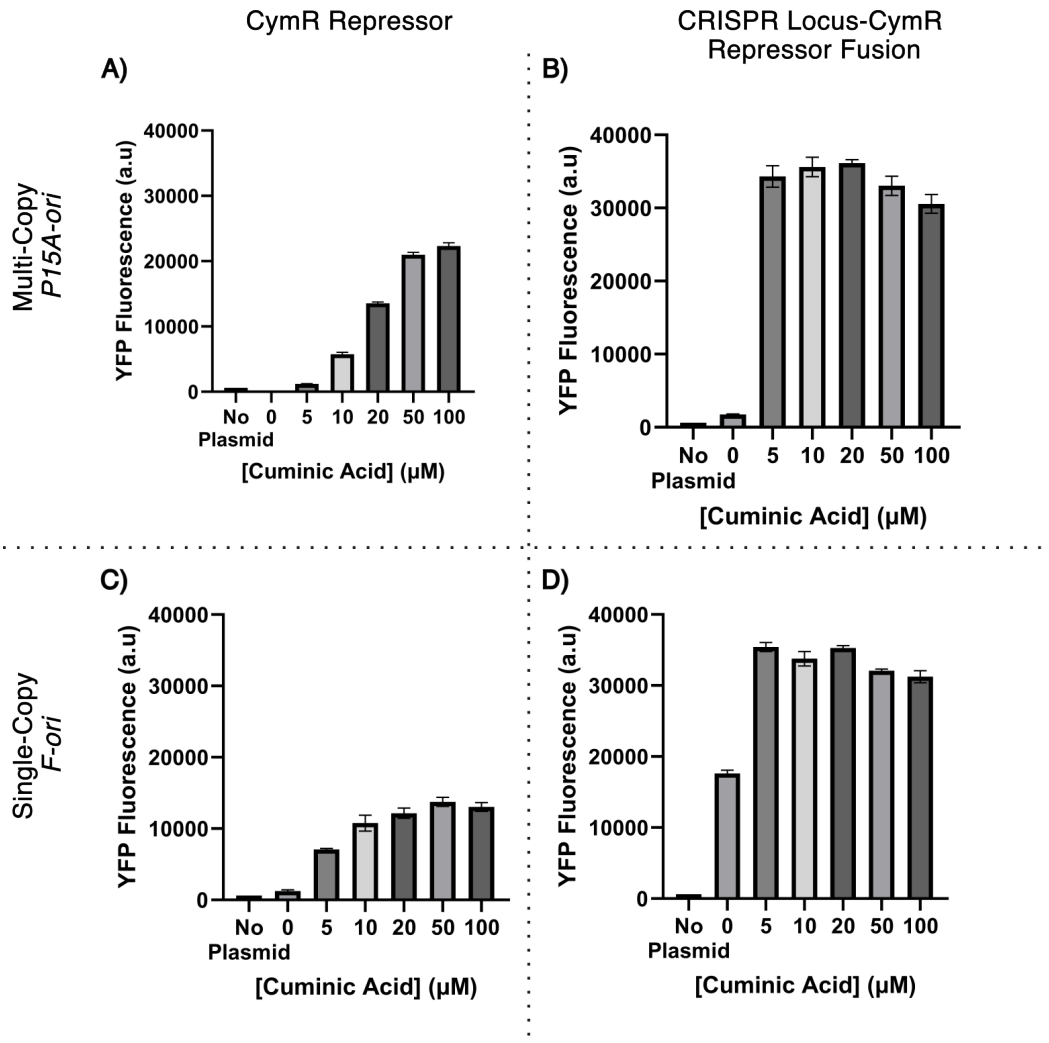

Figure 6.10: CymR cuminic acid titration assay in the presence and absence of N-terminal CRISPR arrays.
Respective plasmids were (where necessary) cloned and transformed into the strain BW25113. Following overnight growth, O.D 600 was normalised to 1.0. The fluorescence was determined using an Amersham Typhoon with a Cy2 filter set. Fluorescence was measured across a range of cuminic acid concentrations.
(A) YFP induction across a cuminic acid titration for a *P15A-ori* based plasmid expressing the CymR repressor.
(B) YFP induction across a cuminic acid titration for a *P15A-ori* based plasmid expressing the CRISPR locus-CymR repressor fusion.
(C) YFP induction across a cuminic acid titration for an F-*ori* based plasmid expressing the CymR repressor.
(D) YFP induction across a cuminic acid titration for an F-*ori* based plasmid expressing the CRISPR locus-CymR repressor fusion.
Six replicates were performed per condition for (A, B). Three replicates were performed per condition (C, D). Mean values are plotted, and error bars are SEM. Data is found in Supplementary Table S12.

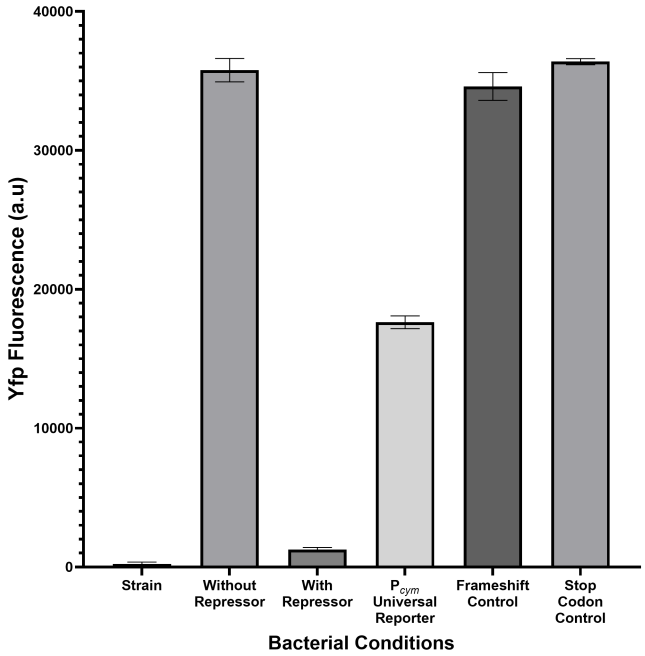

Figure 6.11: YFP fluorescence induction by the CymR-based reporter and controls.
Respective plasmids were (where necessary) cloned and transformed into the strain BW25113. All plasmids contain the single copy F-*ori*. Following overnight growth, O.D 600 was normalised to 1.0. The fluorescence was determined using an Amersham Typhoon with a Cy2 filter set.
Three replicates were performed per condition. Mean values are plotted, and error bars are SEM. Data is found in Supplementary Table S12.

Flow cytometry was performed on the reporter and stop codon control, with cultures of reporter and control titrated against one another [Figure 6.12]. The increase in fluorescence seen in the stop codon control enabled its detection as a separate population relative to the wildtype reporter. Therefore, stop codons within the reporter's CRISPR array can be detected at the level of the individual cell. However, the upper bound of fluorescence seen in the wildtype population overlapped with the lower bound of the control population. As CRISPR array expansion is principally of interest in the range of 0.1-10% of expanded arrays and rarely higher, gating conditions were selected to minimise the rates of false positives at these lower values. This increased the rate of false negatives reported, especially at higher rates of expanded loci ($\geq$50%) [Figure 6.12B]. This is reflected in the

deviation from the expected values line at higher YFP+ cell populations in Figure 6.12D. Populations with higher proportions of YFP+ cells underestimate values, whilst those with lower populations overestimate values. The point of inflection lies in the 0.1-0.5% YFP+ region [Table 6.1]. At $\leq$0.01% YFP+ cell populations, the assay becomes non-functional as rates of YFP+ cells become indistinguishable from the small population of false positives derived from the parental population. Therefore, flow cytometry with the CymR-based reporter is most accurate within the range of a 0.1-10% YFP+ population.

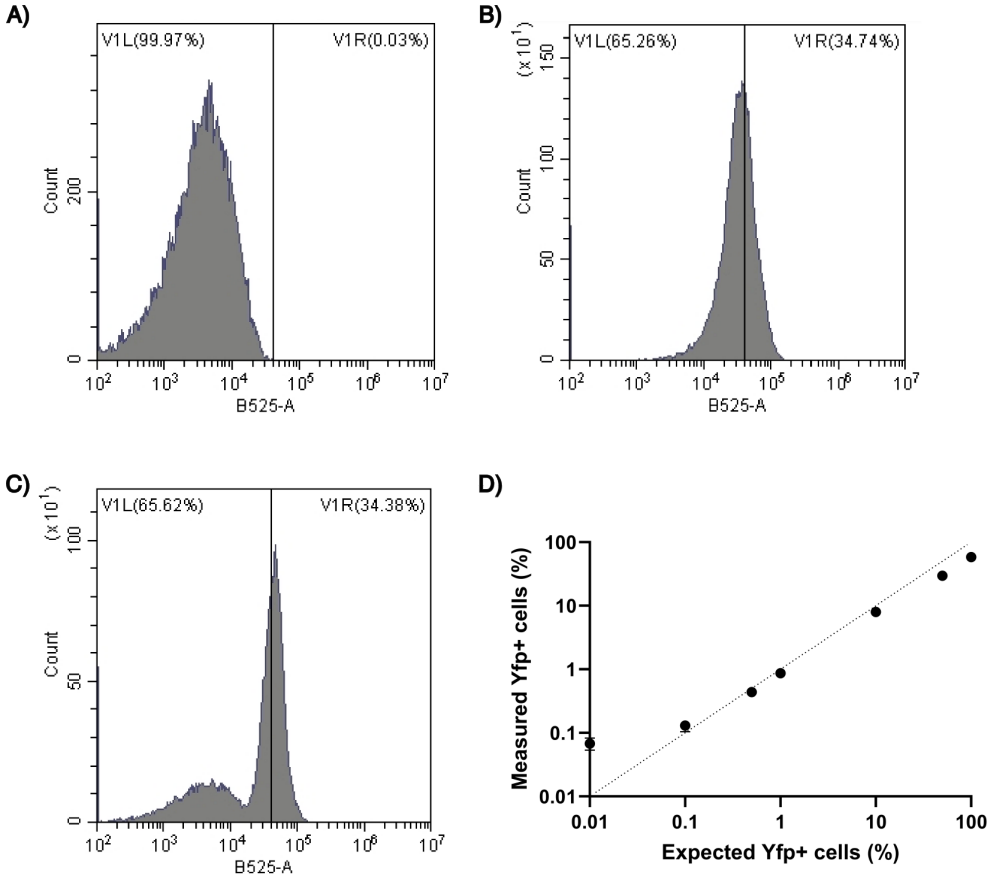

Figure 6.12: Sensitivity and accuracy of detections of the CymR-based universal reporter relative to a stop codon control.
(A) Flow cytometry fluorescence analysis of the parental CymR-based universal reporter. The black line indicates the gating conditions; the YFP- population is to the left, and the YFP+ population is to the right.
(B) Flow cytometry fluorescence analysis of the CymR-based universal reporter stop codon control.
(C) Flow cytometry fluorescence analysis of a 50:50 mixed population of CymR-based universal reporter and stop codon control.
(D) Detection rates of YFP+ cells by flow cytometry plotted against expected values for a stop codon control population titrated against a population containing the parental CymR-based reporter. The dotted line indicates expected values. Mean values can be found in table 6.1. Mean values were plotted. Five replicates were performed per condition. Data values are in Supplementary Table S8.

Expected YFP+ cells (%)	100	50	10	1	0.5	0.1	0.01	0
Mean detected YFP+ cells (%)	58.38	29.64	7.978	0.868	0.438	0.13	0.068	0.044
Accuracy (%)	58.38	59.28	79.78	86.8	87.6	130	680	-

Table 6.1: Values for detection rates of YFP+ cells via flow cytometry compared to expected values for the wildtype vs control reporter titration carried out in [Figure 6.12].

Flow cytometry was performed upon the CymR-based universal reporter transformed with a Cas1-Cas2 expression vector [Figure 6.13A]. Cultures were grown for 3 days in the presence of an inducer titration. No fluorescent cell population was detected for cultures grown at any arabinose concentration. SPIN assay of the day 3 cultures validated the absence of detectable acquisition [Figure 6.13B].

Cultures of BW25113 containing various iterations of the CymR-based reporter plasmid were plated to determine if spacer acquisition could be visualised at the level of a whole colony's fluorescence. The fully induced day 3 culture, which exhibited no spacer acquisition during flow cytometry, was plated on LB-agar containing appropriate antibiotics and incubated overnight. Control cultures of BW25113 containing a parental reporter plasmid, a stop codon-based positive control plasmid and a 50:50 reporter: control mixed population were plated likewise and incubated overnight. Colonies were imaged under visible light and a Cy2 fluorescence filter set [Figure 6.13C]. These results corroborated the absence of acquisition. Control colonies, either present on the control or mixed population plates, were visibly more fluorescent than those of wildtype. No colonies with increased fluorescence were present on the plate of the fully induced day 3 culture. This is equivalent to the wildtype plate. The CymR-based universal reporter, although conceptually functional, does not acquire novel spacers into its CRISPR array and cannot be used to quantify spacer acquisition.

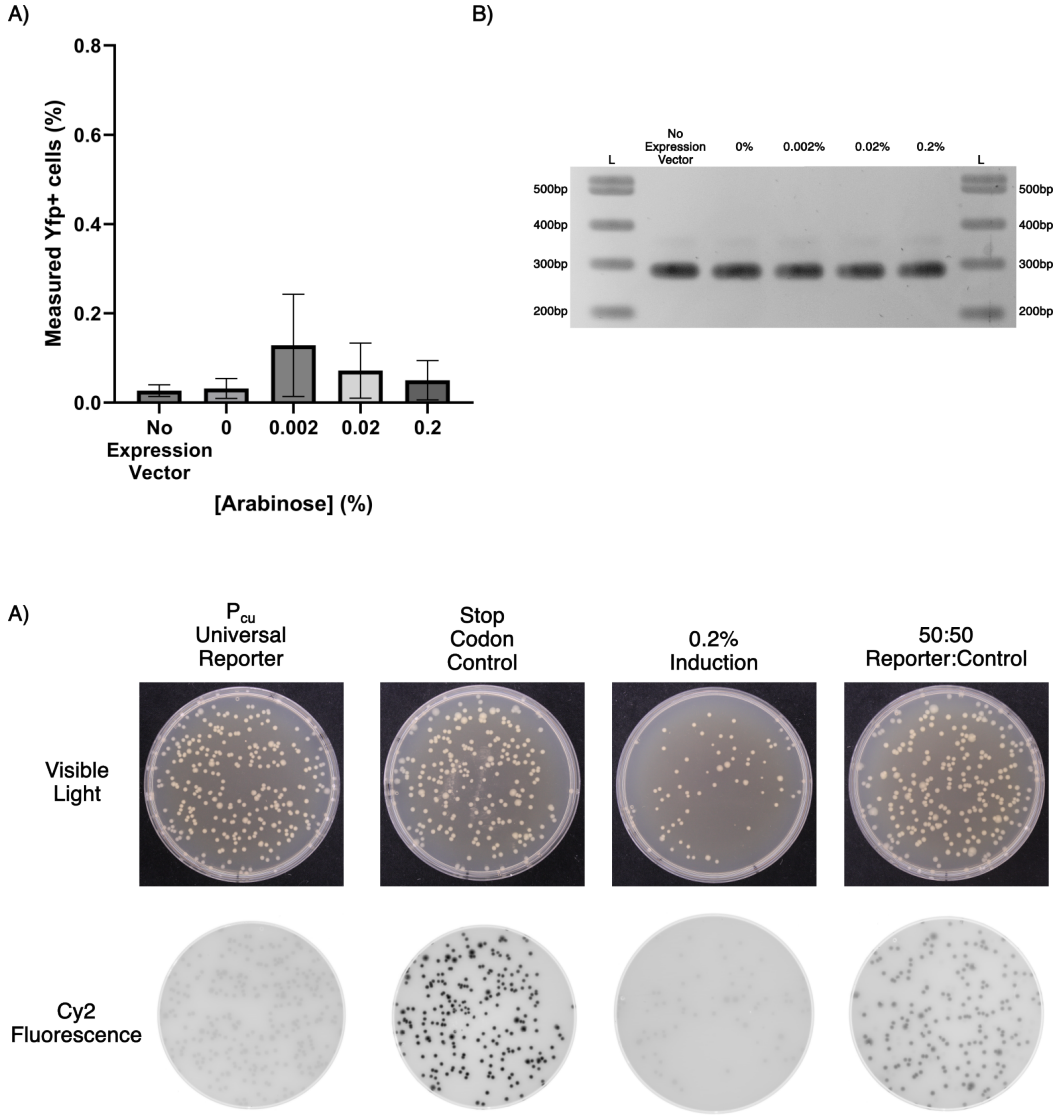

Figure 6.13: Overview of adaptation following Cas1-Cas2 induction in the CymR-based universal reporter.
(A) Flow cytometry fluorescence analysis of cultures of BW25113 transformed with the CymR-based universal reporter plasmid and a pBAD Cas1-Cas2 expression vector plasmid. Cultures were grown with arabinose concentrations titrated between 0-0.2% induction. Cultures were diluted 1:1000 daily over three days of growth. On day 3, cultures were diluted 1:100. The % YFP+ population was measured via flow cytometry. Mean values are plotted, and error bars are SEM. Six replicates were performed per condition. Data values are found in Supplementary Table S9.
Six replicates were performed per condition. Mean values are plotted, and error bars are SEM. Data can be found in Supplementary Table S9.
(B) Colony PCRs were performed on the CRISPR arrays of the CymR-based reporter plasmids in the day 3 cultures in (A). PCRs were visualised via agarose gel electrophoresis. Bands at 282 bp derive from unexpanded arrays. L refers to an NEB 100 bp ladder.
(C) Parental CymR-based universal reporter and stop codon control overnight cultures, alongside the 0.2% arabinose induction culture from (A) and the 50:50 reporter: control culture from [Figure 6.12C] in BW25113 were plated on LB-agar containing appropriate antibiotics and grown overnight. Plates were imaged under visible light and using an Amersham Typhoon with a Cy2 filter set. Representative plates are shown.

6.3 Chapter Summary

- This chapter described the construction and testing of various iterations of a genetic circuit for the universal detection of spacer acquisition.
- Repressors are not typically tolerant to the presence of the N-terminal CRISPR array fusions.
- CymR is tolerant to N-terminal fusions and can be applied to the construction of a universal reporter genetic circuit sensitive to the independent introduction of frameshifts and stop codons.
- Despite the CymR-based universal reporter proving functional in several preceding control assays, the reporter fails to acquire novel spacers. The cause of this flaw is unclear.

Chapter 7

Discussion and Future Work

7.1 Discussion

The work described in this thesis was undertaken to advance the understanding of CRISPR-Cas adaptation. Firstly, I investigated biases in spacer selection using a novel papillation spacer acquisition reporter strain. I then used this reporter system to identify hyperactive mutants in the Cas1-Cas2 integrase. Furthermore, I used this reporter system to screen metagenomic BAC libraries to look for anti-CRISPRs targeted against adaptation. Finally, I constructed and tested a range of genetic circuits to enable the generalised use of these assays in all CRISPR systems.

7.1.1 Sequencing of Spacers

I identified a bias for the acquisition of spacers from the plasmid into the reporter CRISPR array of the strain JB028. This finding contradicted previous work with the strain, which had identified an absence of bias in the source of acquired spacers [212]. The presence of a bias is concordant with the literature, in which a 100-1000 fold preference for acquisition from the plasmid has been reported [114]. I identified, at most, a 23-fold preference towards plasmid-derived DNA; this disparity may result from differences between genetic backgrounds and culturing conditions.

I assayed the source of spacers acquired by the papillation reporter strain under different culturing conditions. Cultures grown in liquid media exhibited a more pronounced bias towards acquisition from the plasmid when compared to cultures grown on LB-agar. Decreased induction of Cas1-Cas2 expression has been reported to lead to higher specificity of acquisition from exogenous DNA [114]. However, spacers were acquired by Cas1-Cas2 expressed from a pBAD vector that was induced at 0.2% in liquid culture assays compared to 0.002% on solid media assays. This contradicts the literature, suggesting higher induction may lead to an increased bias towards spacer acquisition from the plasmid. Cas1-Cas2 expression should be separately titred in liquid culture and solid media assays to

determine the relationship between induction and biases in the source of spacers, without the confounding variable of changing culture conditions.

Spacers that mapped to the plasmid exhibited a strong unidirectional bias commencing from the ColEI origin of replication, this was exhibited by spacers acquired under different culturing conditions. A bias towards acquisition proximal to the *ori* has been reported from the chromosome but not the plasmid [114]. This behaviour was more pronounced in spacers acquired during growth in liquid culture. I investigated whether a reduced fidelity in PAM selection in spacer acquired near the *ori* could account for this behaviour. Deep sequencing revealed no strong difference in PAM fidelity in spacers acquired from different regions of the plasmid.

The mapping of spacers to the chromosome identified the chromosomal origin of replication (*oriC*), the terminus, and CRISPR arrays as hotspots of spacer acquisition. The identification of hotspots at these sites is concordant with prior literature [114]. The Type I (T) CRISPR array was identified as a hotspot of spacer acquisition through the sequencing of spacers acquired into the distant Type II (A) CRISPR reporter array. This provides evidence that the stalling of replication forks at the CRISPR locus stimulates adaptation and that it is not enriched at CRISPR loci due to mass action promoting the acquisition of nearby sequences.

7.1.2 Hyperactive Cas1 Mutants

I screened a mutagenised Cas1-Cas2 library in a papillation reporter assay to identify Cas1 mutants that produced hyperactive phenotypes. Targeted mutational analysis of residues identified in an initial screen expanded the pool of hyperactive mutants identified, by exploring sequence space not present in the initial screen. In total, thirteen hyperactive mutants were identified [Table 7.1].

The mutant M17T was identified in a prior screen for Cas1-Cas2 integrase mutants with a reduced PAM specificity [268]. Decreased PAM fidelity increases the proportion of substrate that may be successfully integrated into the CRISPR array,

presenting a mechanism for hyperactivity. M17I, a similar mutation to M17V, appeared in the same prior screen for reduced PAM specificity. Therefore, it is likely that M17V produces a hyperactive phenotype through the same mechanism. It is surprising that M17I was not identified in my screen for hyperactivity, as several different point mutations in the M17 codon would produce M17I.

Mutation	**Wildtype residue**	**Mutant residue**	**Change in residue properties**
M17T	Methionine	Threonine	Non-polar to polar, neutrally charged
M17V	Methionine	Valine	Remains non-polar
D26G	Aspartic Acid	Glycine	Negatively charged to non-polar
I28K	Isoleucine	Lysine	Non-polar to positively charged
I28S	Isoleucine	Serine	Non-polar to polar, neutrally charged
I28N	Isoleucine	Asparagine	Non-polar to polar, neutrally charged
I28R	Isoleucine	Arginine	Non-polar to positivity charged
D29N	Aspartic Acid	Asparagine	Negatively charged to polar, neutrally charged
D29H	Aspartic Acid	Histidine	Negatively charged to positively charged
D29K	Aspartic Acid	Lysine	Negatively charged to positively charged
D29A	Aspartic Acid	Alanine	Negatively charged to non-polar
Q90R	Glutamine	Arginine	Polar, neutrally charged to positively charged
Y101H	Tyrosine	Histidine	Polar, neutrally charged to positively charged

Table 7.1: Overview of Cas1 mutations with a hyperactive spacer acquisition phenotype.

Several mutations in the residues between D26 and D29 of Cas1 resulted in hyperactive phenotypes. I28 and D29 are located in a beta turn at the interface between Cas1b/b', Cas2 and the protospacer, with R groups extended towards the protospacer [Figure 7.1]. D26 is located in an adjacent beta sheet close to the protospacer. Mutations identified in the D26 – D29 cluster are generally characterised by a shift from negatively or neutrally charged R groups towards neutrally or positively charged R groups [Table 7.1]. Cas1b V27, D29, and G30 have been shown to form intermolecular contacts with the protospacer [76]. The mutations identified may form or strengthen interactions between Cas1b/b' and the incoming protospacer, causing a hyperactive phenotype.

Spacers acquired by Cas1 I28K exhibit a reduced fidelity towards possessing a consensus PAM [Tables 4.3 and 4.4]. I hypothesise that enhanced non-specific Cas1-protospacer interactions may stabilise the binding of protospacers possessing non-consensus PAMs, as the integrase complex specifically binds the PAM during the maturation of protospacers [269]. This would suggest that the protospacer binding interface of Cas1 has evolved a reduced binding affinity to aid the selection of consensus PAM sequences in the 3' ssDNA tail of prespacer substrates. Mutations that enhance the stability of the protospacer-integrase complex may be less specific in the selection of consensus PAM sequences, leading to a weaker immune response by the interference machinery.

The low rate of wildtype spacer acquisition compared to that attainable through individual point mutations suggests an absence of selection for highly active integrases. The most active mutation I identified, I28K, conducted adaptation at five-fold the rate of wildtype [Figure 4.6]. A prior study identified single mutations in *E. coli* Cas1 with up to seven-fold increases in rates of adaptation [218]. These studies suggest that highly active adaptation machinery could quickly evolve through single point mutations.

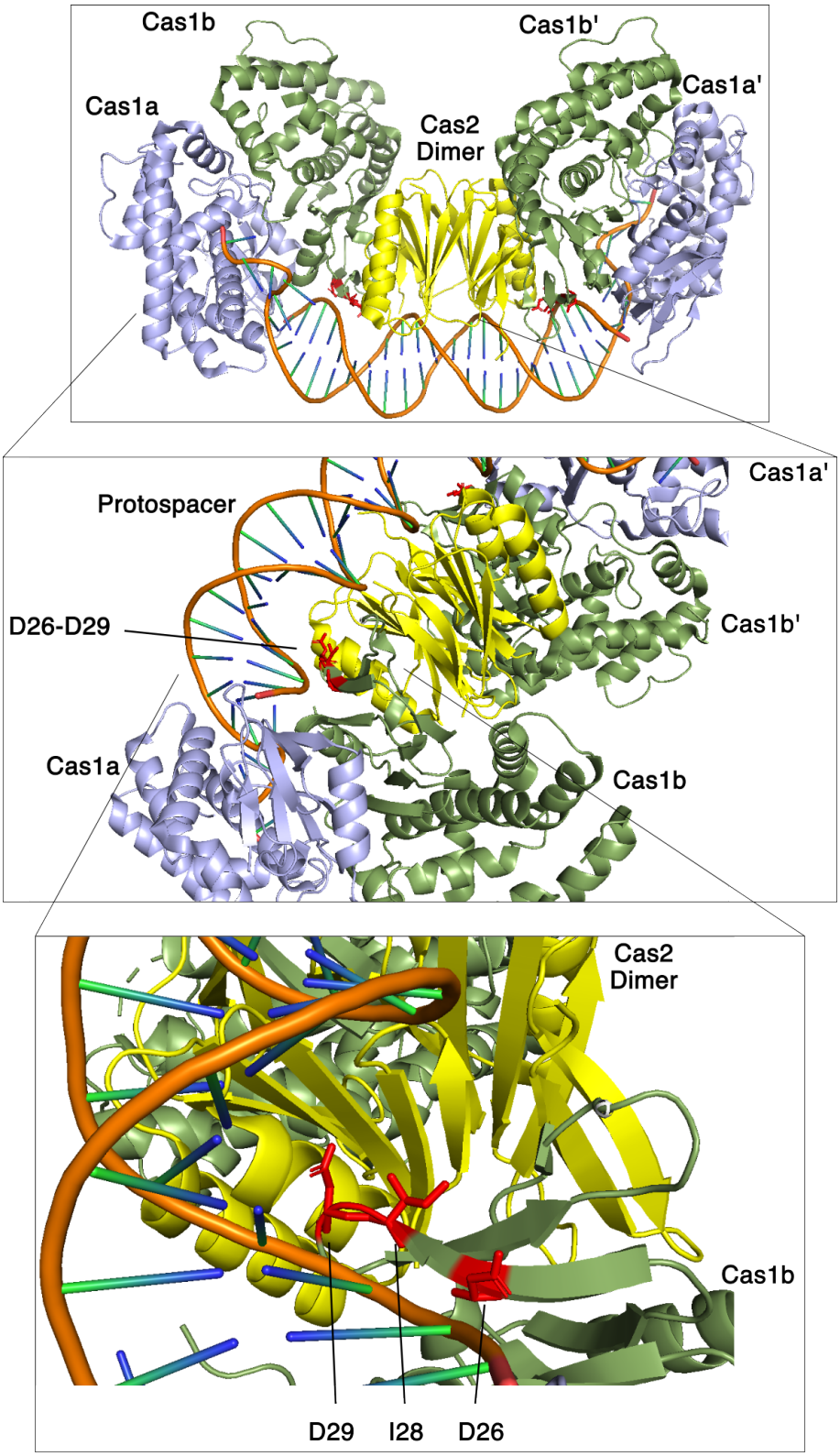

Figure 7.1: Cas1 D26-D29 region within the adaptation complex.
The crystal structure of the Cas1-Cas2 integrase bound to a protospacer is depicted (PDB: 5DS4). Residues mutated to confer hyperactive phenotypes within the D26-D29 region are highlighted (red) in Cas1b/b' (green). This region is found at the Cas1b-Protospacer-Cas2 interface.

Repression of Cas1-Cas2 expression by H-NS may have facilitated genetic drift away from a more active Cas1-Cas2 [127]. Analysis of the spacer content of *E. coli* Type I-E CRISPR arrays found in the stomachs of 42,000-year-old frozen mammoths suggest that only limited changes in spacer array content have occurred over this time scale [270]. Genetic drift may further arise as selective pressure for an active and efficient CRISPR system weakens when it is supported by the wide breadth of prokaryotic immune systems we are beginning to discover. Synergism between CRISPR-Cas and RM systems has been reported [271], and this synergism may be a general trend for bacterial immune systems [272].

An immediate fitness cost may be associated with an increased rate of spacer acquisition. The *S. pyogenes* Cas9 mutant I472F possesses an over 100-fold increase in the rate of adaptation, but incurs a fitness cost due to genotoxicity [273][274]. My results broadly support this hypothesis as most hyperactive phenotypes are associated with significant increases in the activation of the SOS response [Figures 4.7 and 4.10]. However, I28S and I28N are exceptions, with an increase in rates of spacer acquisition not associated with an increase in SOS activation. This suggests the presence of fitness costs associated with an increase in spacer acquisition rates may depend upon the mutation's mechanism.

The combination of mutations possessing hyperactive phenotypes was antagonistic towards rates of spacer acquisition [Figure 4.11]. There have been varying reports in the literature on the effect of integrating Cas1 mutations exhibiting hyperactive phenotypes. The combination of mutations that result in a 4- and 7-fold increase in the rate of spacer acquisition in *E. coli* are either synergistic or antagonistic depending on the validation method [218]. However, when multiple Cas1 mutations that moderately increase the rate of adaptation in the Type II system of *S. pyogenes* are combined, the effect is synergistic as rates of spacer acquisition increase 5-fold relative to the wild type [230]. Outcomes may depend on the nature of the mutations. When strong hyperactive mutations are combined, the effect may

be antagonistic to the spacer acquisition reaction. In comparison, the combination of moderate hyperactive mutants may favour synergism. This idea is supported by the effect that combining hyperactive mutations has on the induction of the SOS response [Figure 4.11]. Double mutants containing the I28K mutation exhibited a broad increase in the level of SOS induction relative to their corresponding single mutations.

7.1.3 Functional Metagenomic Screen for Anti-Adaptation Proteins

I conducted a functional metagenomic screen to identify inhibitors of naïve spacer acquisition. I used a papillation assay to screen 1.8 Gbp of environmental DNA. Of the 55,000 screened colonies, 2,203 were initially identified with reduced adaptation rates. Verification of candidates identified 49 colonies that maintained a hypopapillation phenotype. Further validation via a YFP reporter assay ruled out all of these as inhibitors of spacer acquisition.

A critical difference between the papillation and YFP reporter assays is the level of induction of Cas1-Cas2. In papillation assays, Cas1-Cas2 is expressed from a pBAD plasmid at 0.002% L-arabinose and produces a high rate of papillation [Figure 4.1]. In comparison, a stronger 0.2% L-arabinose concentration is used in YFP reporter assays to produce detectable rates of spacer acquisition [Figures 5.12 and 5.15]. Therefore, the amount of Cas1-Cas2 present to be inhibited by a candidate metagenomic clone in a YFP reporter assay is significantly higher relative to a papillation assay. I hypothesise this is why candidates could inhibit adaptation in the papillation assay but could not do so in YFP reporter assays.

Verification attempts using published methods of spacer acquisition may have challenged putative Acrs with a pool of Cas1-Cas2 beyond that which they have evolved to inhibit [213]. In the *E. coli* Type I-E system, native Cas1-Cas2 expression is inhibited by H-NS under laboratory conditions [127]. As H-NS binds to AT-rich re-

gions [275], it has been hypothesised that the introduction of phage containing such regions titrates H-NS away from the Cas1-Cas2 promoter to permit expression [127] [276]. Acr expression, increasing with copy number, would therefore be expected to scale with Cas1-Cas2 expression as an increase in phage copy number dilutes H-NS at the Cas1-Cas2 promoter. Therefore, it may be expected that a putative Acr expressed from its native promoter on a single copy F-*ori* plasmid cannot repress a pool of Cas1-Cas2 derived from a fully induced pBAD promoter [277].

Expression of homologs of genes ancillary to spacer acquisition may provide a mechanism for the inhibition of spacer acquisition. The three metagenomic clones with the lowest average spacer acquisition rate in their respective rounds of verification using the YFP reporter assays were sequenced. Two contained homologs of genes implicated in spacer acquisition or their interaction partners. Candidate 3 encoded an *ftsH* homolog; overexpression of *ftsH* has been shown to increase the rate of adaptation [229]. Candidate 8 encoded a *uvrC* homolog; knockout of UvrC's interaction partner, UvrB, has been shown to inhibit spacer acquisition [Unpublished, Jack Braithewaite]. These homologs may exhibit dominant negative phenotypes to inhibit adaptation, presenting a mechanism whereby MGEs may suppress or inhibit spacer acquisition without directly targeting the Cas1-Cas2 integrase.

The broad-spectrum anti-CRISPR AcrVA5, shown to inhibit adaptation in *Treponema denticola*'s Type II-A CRISPR-Cas system, demonstrated no activity in the *E. coli* Type I-E system. The effect of expression on spacer acquisition was measured via papillation and YFP reporter assay [Figure 5.16]. Successful inhibition would have validated my approach to identifying novel anti-adaptation Acrs. A titration of AcrVA5 exhibited no inhibition of the papillation assay, with expression toxic above inducer concentrations of $\geq$50 μM vanillic acid. Titration of AcrVA5 in a YFP reporter assay exhibited no consistent decrease in acquisition rates with increasing

inducer concentration. These results suggest that while AcrVA5 has been shown to disrupt adaptation machinery in multiple CRISPR systems, this behaviour is not exhibited in *E. coli* and therefore is not universal [205].

7.1.4 Universal Reporter of Spacer Acquisition

I constructed and tested several versions of a reporter designed for the universal detection of spacer acquisition. In contrast to frameshift-based reporter systems [212] [213] [219], these genetic circuits would detect adaptation through the acquisition of spacers containing stop codons. Each circuit tested failed to meet the requirements for a functional universal reporter. The TetR-based circuit did not conduct spacer acquisition and did not express a *YFP* reporter gene [Figures 6.4 and 6.7]. The VanR-based circuit did not repress either a *lacZ* or *YFP* reporter gene [Figures 6.6 and 6.7]. Finally, the CymR-based circuit did not conduct spacer acquisition [Figure 6.13].

Titration of the parental CymR-based universal reporter against a positive control demonstrated overlap in fluorescence between the normally distributed populations [Figure 6.12]. As a result, gating conditions necessitated underestimating the YFP+ population when present above 0.5% to mitigate false positives derived from the parental population. This overlap is not seen in the frameshift-dependent YFP reporter assay [213]. An increase in fluorescence in the parental population is concordant with the emergence of leaky YFP expression in the universal reporter following the replacement of the multi-copy P15A-*ori* with a single-copy F-*ori* [Figures 6.10 and 6.11]. Changing the origin of replication does not alter the ratio of repressor to repressor binding site in the cell.

I hypothesise that increased leaky expression of *YFP* when expressed from an F-*ori* plasmid derives from fluctuations in repressor expression from its promoter causing comparable fluctuations in cellular repressor concentrations. Introducing multiple copies of a repressor expression cassette in a cell would reduce overall

fluctuations and maintain more consistent repression. Therefore, a more reliable universal reporter would require the expression of a repressor from a highly stable promoter to ensure parental and acquired populations are distinct. This may be achieved through expression from a tandem natural minimal (TNM) promoter array, in which expression is driven from multiple promoters [278]. These promoters are each recognised by a different sigma factor, allowing expression to be maintained across various phases of bacterial growth.

The N-terminal fusion of a reporter CRISPR array to a repressor gene was tolerated to a low level generally [Figures 6.8 and 6.11]. All the tested marionette repressors exhibited increased fluorescence, ranging from mild leaky expression to complete inactivation. Furthermore, discrepancies in behaviour between the *vanR*-based universal reporter and the CRISPR array-VanR fusion derived from the marionette collection demonstrate that tolerance of CRISPR arrays varies depending upon the composition of CRISPR arrays [Figures 6.7 and 6.8]. The VanR-based universal reporter contained a leader and single repeat sequence and caused complete inactivation of VanR [Figure 6.7], meanwhile the marionette CRISPR array-VanR fusion contained a leader-repeat-spacer-repeat unit and maintained approx 40% repression of YFP [Figure 6.8]. This suggests repressor tolerance to N-terminal fusions is context-dependent. A universal repressor-based system cannot take this approach as it is unknown whether any given repressor will tolerate any given CRISPR array.

Inteins are segments of proteins which may excise themselves from a host protein and facilitate the ligation of the surrounding protein fragments in a process called protein splicing [279]. The posttranslational self-splicing of an intein between a CRISPR array and a repressor may facilitate the repressor's conditional translation depending upon the status of the CRISPR array, whilst ensuring repressor function is maintained in the absence of an N-terminal fusion affecting expression. This would mitigate the issues of repressor incompatibility and leaky expression

encountered in developing a universal reporter.

Repressors may better tolerate CRISPR arrays if they are located away from the N-terminus. Tested universal reporter systems have placed CRISPR arrays N-terminal to a repressor [Figures 6.3, 6.5 and 6.9]. However, the N-termini of the respective TetR, VanR, and CymR repressors encode DNA binding domains [280] [264] [281]. As such, the addition of CRISPR arrays into the N-termini may disrupt DNA binding, inhibiting repression unpredictably depending upon the amino acid composition of the translated array. CRISPR arrays may be better tolerated if introduced into a linker region of a repressor, between the N-terminal DNA binding domain and C-terminal regulatory domain.

This approach parallels lambda repressor DNA-binding domain fusions designed to study protein: protein interactions [282] [283]. A CRISPR array would be introduced into the lambda repressor's 40 residue flexible linker domain. Assuming it is tolerated, regulatory domains would exhibit cooperativity in the repression of a reporter gene [284]. Acquisition of a stop codon would induce premature termination before translation of the C-terminal dimerisation domain. Cooperativity would be lost, leading to the expression of a reporter gene [Figure 7.2].

A final issue that arose in the testing of the TetR- and CymR-based reporters, but not the VanR-based reporter, was the failure of the CRISPR arrays within these circuits to acquire spacers [Figures 6.4 and 6.13]. Sequencing revealed intact CRISPR arrays. This suggests that the +1 frameshift of these repressors may encode products toxic to the cell. This would prevent the use of these assays with the *E. coli* Type I-E CRISPR system and in other systems in which adaptation is concomitant with a frameshift. However, they may remain viable when applied to CRISPR systems lacking frameshifts, as is their intended use.

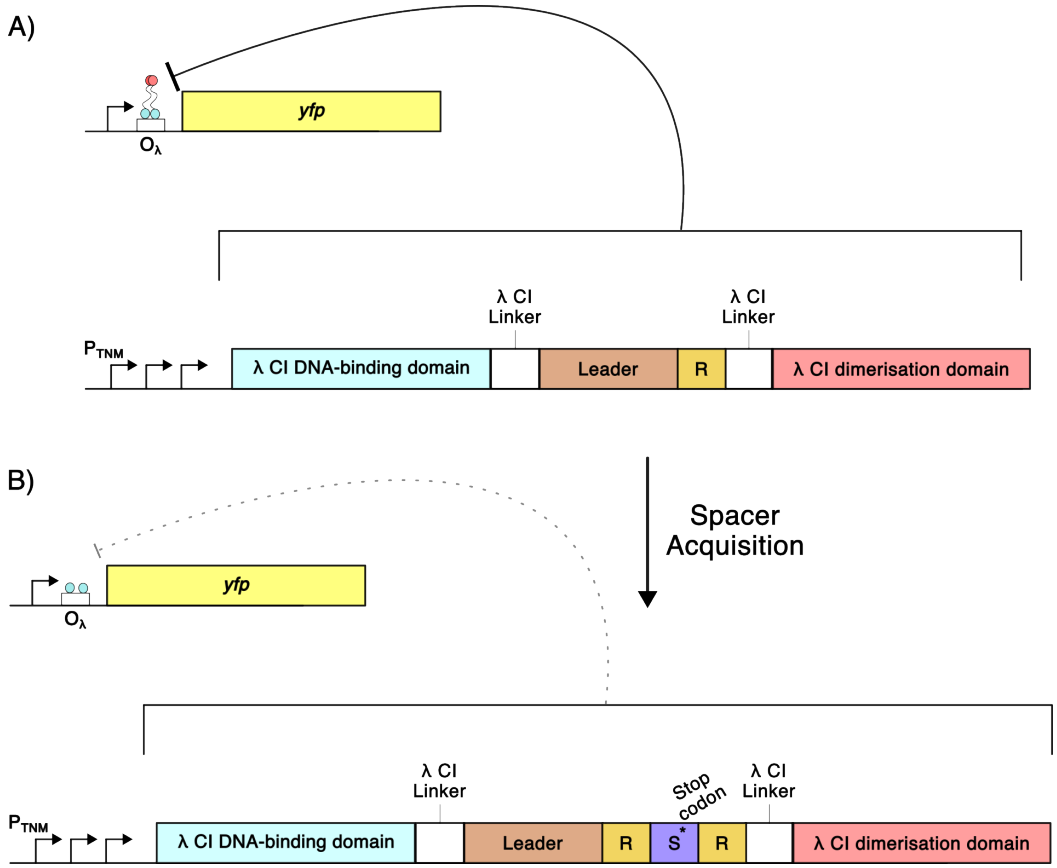

Figure 7.2: Genetic circuit of an updated universal reporter for the detection of spacer acquisition.
A) A minimal CRISPR array is present within the linker region of the lambda repressor. The array maintains the frame and does not possess stop codons. The CI dimerisation domain is transcribed and translated, conferring cooperativity to the repression of a reporter gene. The repressor is expressed from a tandem natural minimal promoter array, maintaining stable expression and mitigating leaky YFP expression.
B) Spacer acquisition introduces a stop codon into the CRISPR array. Translation is terminated before the CI dimerisation domain is expressed, causing a loss of cooperativity in *YFP repression*. An increase in *YFP* expression produces a detectable phenotype.

7.2 Future Work

7.2.1 Further Investigation of Hyperactive Cas1 Mutants

I hypothesise that the Cas1 D26-D29 mutations induce hyperactivity through enhanced binding with the incoming protospacer [Table 7.1]. I attempted to purify several such Cas1 mutants in complex with Cas2, but this resulted in poor yields

relative to wildtype, which prevented *in vitro* assays of integrase behaviour. This may be a result of increased toxicity associated with these mutations. However, mechanisms of hyperactivity may be independent of specific catalytic activity. Filamentation, associated with induction of the SOS response, does not occur with catalytically inactive Cas1 D221A [229]. This suggests that cloning Cas1 mutants into expression vectors containing Cas1 D221A may facilitate their purification without affecting behaviour in downstream biochemical assays. For example, electrophoretic mobility shift assays could be performed to investigate my hypothesis that D26-D29 mutations enhance Cas1-protospacer binding through a reduction in the concentration of negative charge in proximity to the protospacer's DNA backbone.

Multiple studies of *E. coli* Cas1-Cas2 have now identified highly active ($\geq$5-fold) adaptation mutants [Figure 4.6] [218]. In comparison, deep mutational scanning of *S. pyogenes* Cas1-Cas2-Csn2 identified point mutations with comparably minor hyperactivity phenotypes (<3-fold) [230]. The identification of highly active *E. coli* Cas1-Cas2 mutants through the assaying of only the sequence space available from taq mutagenesis or phage-based evolution suggests a rigorous investigation of the Cas1-Cas2 sequence space would identify a significantly greater number of highly active mutants [Figure 4.6] [218]. This may be achieved through the targeted mutagenesis of further residues of interest, as I have conducted for I28 and D29 [Figure 4.10], or the mutagenesis of the entire Cas1-Cas2 operon.

I would conduct a deep mutational scan of *E. coli* Cas1-Cas2 to identify further hyperactive mutants [Figure 7.3] [27]. The application of this method to *S. pyogenes* Cas1-Cas2 coupled spacer acquisition with expression of an antibiotic resistance gene to select for spacer acquisition [230]. While such a system exists for *E. coli* Cas1-Cas2 [219], selection for adaptation could also be performed through the plating of an induced overnight culture of the papillation reporter strain onto M9 minimal media supplemented with lactose. Next-generation sequencing of the

Cas1-Cas2 loci of colonies, combined with comparison to a control that would not undergo selection, would enable the determination of differential selection for each amino acid substitution at every residue. Prior mutagenesis studies could validate positively and negatively selected residues [Table 7.1] [218] [72].

The dataset of all hyperactive single residue mutations in *E. coli* Cas1-Cas2 could be compared to such a dataset for *S. pyogenes* Cas1-Cas2 [230]. The conservation of hyperactive mutations between the two distantly related complexes would provide evidence of strong selection for reduced Cas1-Cas2 activity. Furthermore, identifying additional mutations may facilitate the identification of novel combinations of mutations that exhibit synergy between their hyperactive adaptation phenotypes. This would enable the development of increasingly active Cas1-Cas2 complexes. These may be applied to Cas1-Cas2's use as a molecular recorder [268], especially as these systems are developed into tools for the non-invasive assessment of gut physiology [285].

Just as a Cas1-Cas2 deep mutational library can provide a high throughput method of identifying residue mutations affecting adaptation [Figure 7.3A], a high throughput approach can be applied to the same library to investigate how these mutations affect the induction of the SOS response [Figure 7.3B]. An overnight culture inducing Cas1-Cas2 in the presence of the GFP-SOS reporter plasmid could be subjected to fluorescent cell sorting. The binning of cells depending on SOS activation, and the subsequent next-generation sequencing of respective bins, would provide a broad picture of the interplay between hyperactivity and SOS response induction.

The induction of the SOS response has been shown to facilitate adaptation in E. coli [229]. However, constitutive induction of the SOS response is toxic to the host cell and results in SulA-mediated cell lysis [286]. Identification of Cas1 mutants that have an increased rate of spacer acquisition but are not associated with increased induction of the SOS response have potential applications as molecular

recorders. The absence of enhanced SOS induction and potential cell lysis associated with hyperactive Cas1 phenotypes may improve information storage in these systems.

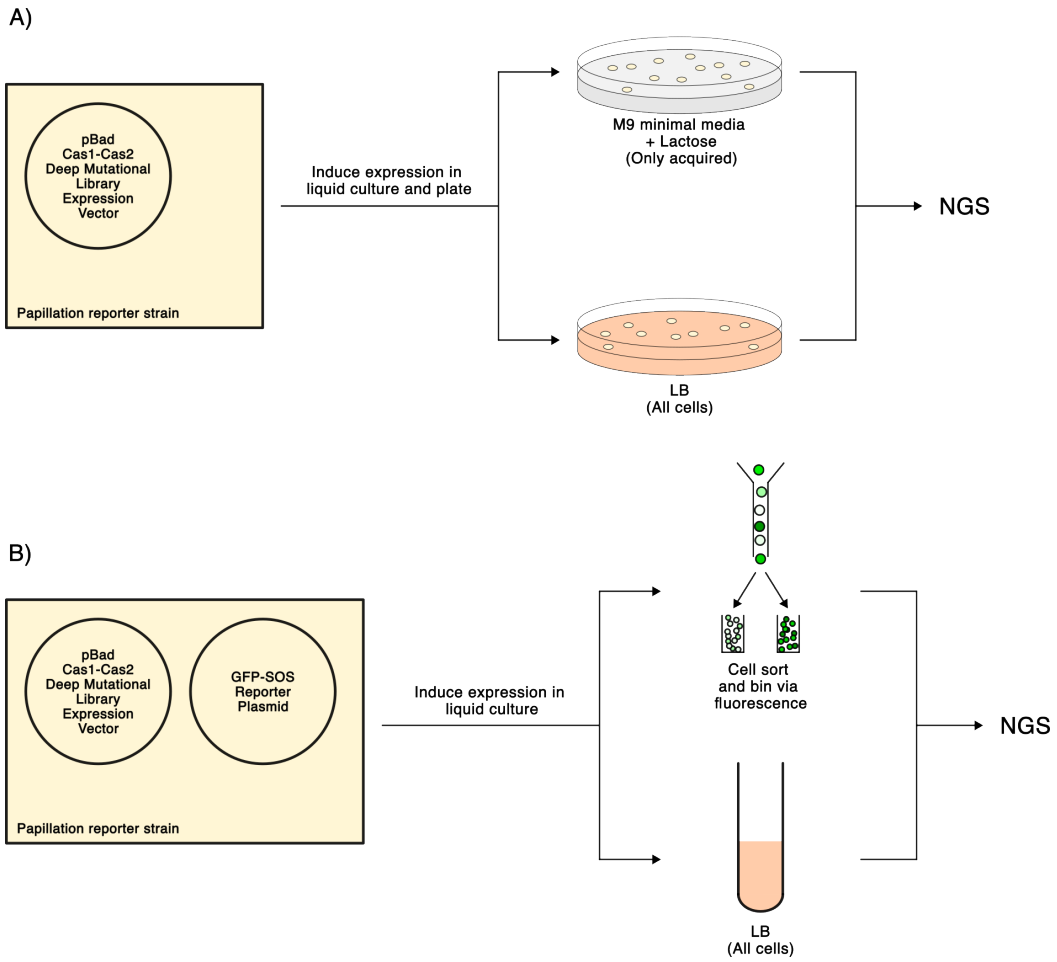

Figure 7.3: Deep mutational scanning of *E. coli* Cas1-Cas2 to investigate spacer acquisition and SOS response induction.
A) The JB028 papillation reporter strain containing a mutagenised Cas1-Cas2 library in a pBAD expression vector is grown overnight in the presence of L-arabinose inducer. The overnight culture is then plated onto LB agar or M9 minimal media agar containing lactose. Plates are incubated overnight at 37 C and separately pooled. Next-generation sequencing is conducted upon the Cas1-Cas2 loci of respective pools.
B) The JB028 papillation reporter strain containing a mutagenised Cas1-Cas2 library in a pBAD expression vector and GFP-SOS reporter plasmid is grown overnight in the presence of L-arabinose inducer. Cultures are cell sorted and binned depending on fluorescence. NGS is then conducted on the Cas1-Cas2 loci of both overnight and cell sorted cultures.

7.2.2 Screening Ancillary Adaptation Gene Homologs for Anti-Adaptation Phenotypes

The sequencing of metagenomic clones that caused a hypopapillating phenotype revealed that two of three clones possessed homologs of ancillary genes involved in spacer acquisition [Table S1]. These genes could be cloned into inducible vectors, and their expression titrated in papillation assays to demonstrate that it is the expression of these specific genes responsible for driving the hypoactive papillation phenotype. If the expression of ancillary gene homologs inhibits spacer acquisition, I would investigate whether similar ancillary homologs are found in phage genomes and inhibit adaptation. I would perform alignment searches of genes ancillary to spacer acquisition in the genomes of phages identified as infecting *E. coli* possessing a Type I-E CRISPR system. Candidates could then be cloned into inducible vectors to be tested via papillation assay.

7.2.3 Refined Functional Metagenomic Screen for Anti-Adaptation Acrs

Future functional metagenomic screens for identifying Acrs targeting adaptation should specifically utilise phage-derived metagenomic libraries to mitigate false positives. Sequences proximal to homologs of ancillary spacer acquisition genes suggest these candidates derive from bacterial genomes [Table S1]. While the identification of homologs to genes ancillary to spacer acquisitions suggests an indirect mechanism of inhibition, the functional metagenomic screen failed to identify any genes that directly inhibit the spacer acquisition process. From this perspective, bacterial-encoded ancillary homologs are false positives. Such homologs may be avoided by enriching library source DNA to contain phage-derived metagenomic DNA specifically. Virome library construction is complicated by the multiple log difference in the magnitude of recovered genetic biomass compared to standard metagenomic libraries. Therefore, virome libraries designed for functional

metagenomics require enrichment and amplification steps but would provide a method to screen relevant metagenomic DNA in the absence of false positives [287] [288].

A round of selection for the inhibition of interference may enrich a library to contain metagenomic clones encoding anti-defence islands [289]. Anti-defence islands are regions in MGE genomes in which inhibitors of host defences co-localise. Acrs have been shown to localise to anti-defence islands [186]. If an MGE were to encode co-localised inhibitors of adaptation and interference, selection for the inhibitor of interference may select for adjacent inhibitors of adaptation. Most phages exhibit species- or strain-specific host ranges [290]. Therefore, multiple Acrs within a genome frequently exhibit activity against the same CRISPR system [291]. Selection for the inhibition of interference of a particular system may simultaneously select for inhibitors of adaptation in that same system. These could then be identified in a subsequent round of screening.

Chapter 8

Appendix

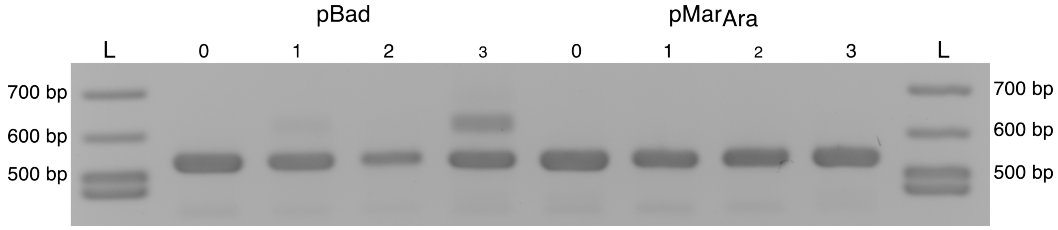

Figure S1: SPIN assay comparing the rates of spacer acquisition between pBAD and pMar$_{Ara}$ expression vectors for Cas1-Cas2.
JB028 was transformed with the pBAD and pMar$_{Ara}$ Cas1-Cas2 expression vectors pRC1656 and pRC2802, respectively. SPIN assay were conducted in cultures grown over three days in the presence of 0.2% L-arabinose inducer, diluted 1:1000 daily. Day 0 refers to an initial culture grown without an inducer. Colony PCRs were performed on cultures daily. Bands at 548 bp derive from unexpanded loci, bands at 609 bp derive from expanded loci. L refers to an NEB 100 bp Ladder.

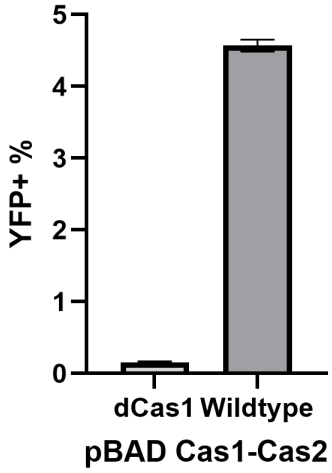

Figure S2: Comparison of spacer acquisition rates between wildtype and dCas1 using the YFP reporter assay.
MLS989 was transformed with pBAD wildtype Cas1-Cas2 and pBAD Cas1-Cas2 containing the inactivating D221A point mutation in Cas1. Cultures were grown over three days in the presence of 0.2% L-arabinose inducer, diluted 1:1000 each day. Flow cytometry was conducted on day three to determine the proportion of YFP+ cells in each culture. Five replicates were performed for each condition, with 20,000 cells analysed per replicate. Mean values are plotted, and error bars are SEM. Data values are in Supplementary Table S5.

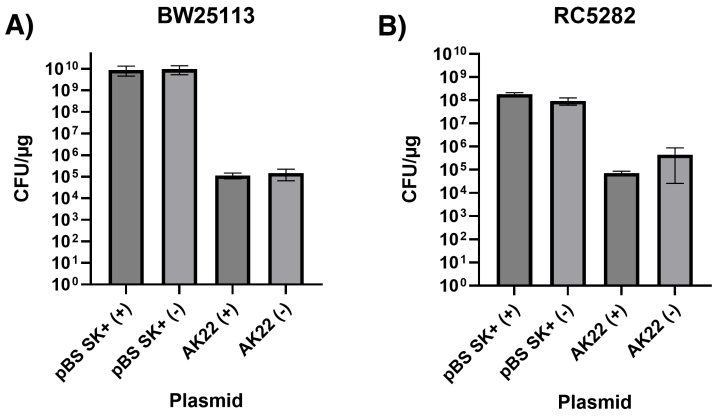

Figure S3: Transformation efficiencies of methylated and unmethylated AK22 and pBS SK+ plasmids in BW25113 and RC5282.
A) BW25113 was transformed with AK22 or pBS SK+ in triplicate. Plasmid was either methylated or unmethylated. Transformation efficiency was determined through titration of transformation cultures and plating on LB-agar containing appropriate antibiotics. Following overnight growth, colonies were counted to determine CFU/μg. Methylated plasmid was prepared from NEB5α whilst unmethylated plasmid was prepared from DH10β. (+) refers to methylated plasmids, and (-) refers to unmethylated plasmids.
B) Transformation efficiencies were determined for RC5282 as described above.
Conditions were measured in triplicate. Mean data values are plotted, and error bars are SEM. Data values are found in Supplementary Table S4.

Candidate	Gene	Organism	Identity	Percentage Match	Gene function summaries (Uniprot)
3	Tktl2	*Mus musculus* (Mouse)	54.60%	97.30%	Plays an essential role in total transketolase activity and cell proliferation in cancer cells
3	glpT	*Escherichia coli* (strain K12)	52.90%	91.60%	Responsible for glycerol-3-phosphate uptake
3	ftsH	*Escherichia coli* (strain K12)	56.80%	68.00%	ATP dependent metalloprotease which, as part of the FtsH/HflKC complex, degrades both soluble and inner membrane proteins.
3	IDH5	*Arabidopsis thaliana* (Mouse-ear cress)	56.70%	88.00%	Performs an essential role in the oxidative function of the citric acid cycle
3	moaB	*Bacillus cereus*	52.70%	95.90%	May be involved in the biosynthesis of molybdopterin
3	infA	*Carboxydothermus hydrogenoformans*	51.40%	97.20%	One of the essential components for the initiation of protein synthesis
3	dapE	*Serratia proteamaculans*	57.90%	14.10%	Catalyzes the hydrolysis of N-succinyl-L,L-diaminopimelic acid (SDAP), forming an essential component of bacterial cell walls

8	isfD2	*Chromohalobacter salexigens*	50.00%	98.40%	Catalyzes the formation of isethionate from 2-sulfoacetaldehyde in the deaminative pathway of taurine
8	ctaD	*Rhizobium leguminosarum*	50.00%	83.10%	Cytochrome c oxidase is the component of the respiratory chain that catalyzes the reduction of oxygen to water
8	uvrC	*Prosthecochloris aestuarii*	50.90%	99.20%	The UvrABC repair system catalyzes the recognition and processing of DNA lesions. UvrC both incises the 5' and 3' sides of the lesion
8	irrR	*Burkholderia pseudomallei*	51.10%	96.10%	Member of the two-component regulatory system IrrR/IrrS
8	petC	*Chlamydomonas reinhardtii*	50.00%	27.20%	Component of the cytochrome b6-f complex, which mediates electron transfer between photosystem II (PSII) and photosystem I (PSI)
8	vraR	*Staphylococcus aureus*	53.50%	20.60%	Member of the two-component regulatory system VraS/VraR involved in the control of the cell wall peptidoglycan biosynthesis.
44	yqkA	*Bacillus subtilis*	54.30%	13.40%	Protein inferred from homology

Table S1: List of genes in the sequencing of metagenomic candidates 3, 8 and 44.
Open reading frames were annotated via a BLASTp search of the Swissprot database. Summaries were derived from the UniProt database.

Figure	Distance from the ori	Spacer Count	Total Spacers (%)
3.4A	0-499	25	30.9
	500-999	19	23.5
	1000-1499	10	12.3
	1500-1999	5	6.2
	2000-2499	4	4.9
	2500-2999	5	6.2
	3000-3499	4	4.9
	3500-3999	2	2.5
	4000-4499	6	7.4
	4500-4999	0	0
	5000-5277	1	2.2
3.4B	0-499	17	22.7
	500-999	8	10.7
	1000-1499	4	5.3
	1500-1999	8	10.7
	2000-2499	8	10.7
	2500-2999	9	12
	3000-3499	8	10.7
	3500-3999	3	4
	4000-4499	4	5.3
	4500-4999	3	4
	5000-5277	3	7.2
3.6A	0-499	17764	25.57
	500-999	13659	19.66
	1000-1499	7223	10.40
	1500-1999	6960	10.02
	2000-2499	6874	9.89
	2500-2999	3676	5.29
	3000-3499	2336	3.36
	3500-3999	2452	3.53
	4000-4499	3859	5.55
	4500-4999	2616	3.77
	5000-5277	2062	2.97
3.6B	0-499	2705	16.92
	500-999	2140	13.39
	1000-1499	3867	24.19
	1500-1999	2887	18.06
	2000-2499	1335	8.35
	2500-2999	772	4.83
	3000-3499	440	2.75
	3500-3999	502	3.14
	4000-4499	588	3.68
	4500-4999	477	2.98
	5000-5277	271	1.70

Table S2: Total and percentage spacer counts at varying distances from the plasmid origin for Figures 3.4 and 3.6.
The plasmid was binned into 500 bp units, with the exception of the terminal 277 bp.

Sequence	Duplicate Count	RBS and Start Codon?	Gene
gaaattctgtgattgcgcataacttctcccaac	2180	+	*indR*
gttttaataaatgacattttgttctccacgtat	2115	+	*yfdR*
gaggatcgcattatgagcattatctccactaaa	1704	-	-
ggatgtcgcattttaattctccacgcttataag	1526	+	*tpiA*
gcaaaatcaaacatgatgttatcctcaaatcaa	1497	+	*xdhB*
gtgttcatctttacccccttcattgtccggtaat	1342	+	*ygcS*
gtaataatcttcattgcatctccagaaatcatg	1341	+	*fixA*
gaatgaaaaatttttgtcattccttatgctcct	1255	-	-
gttttcatgtttgctctcgtgtaggtaattaac	1180	+	*ykfI*
gcacgaactttcatttttactctccgtaacttc	1069	+	*rpmJ*

Table S3: The ten genomically-derived spacers with the highest duplicate counts from deep sequencing papillation assays CRISPR locus.

Figure	Strain	Plasmid	Rep 1	Rep 2	Rep 3	Mean
5.2	JB028	pBS SK+ (Methylated)	2.20E+08	1.76E+08	1.38E+08	1.78E+08
		pBS SK+ (Unmethylated)	1400000	1100000	1920000	1473333.33
		AK22 (Methylated)	22000	82000	78000	60666.67
		AK22 (Unmethylated)	0	0	0	0
	JB028 $\triangle$*hsdR*	pBS SK+ (Methylated)	7.60E+08	2.40E+08	1.32E+09	7.73E+08
		pBS SK+ (Unmethylated)	1.60E+09	1.36E+09	1.50E+09	1.49E+09
		AK22 (Methylated)	42000	100000	32000	58000
		AK22 (Unmethylated)	96000	84000	32000	70666.67
S3	BW25113	pBS SK+ (Methylated)	1.70E+10	2.24E+09	7.40E+09	8.88E+09
		pBS SK+ (Unmethylated)	6.40E+09	1.80E+10	4.20E+09	9.53E+09
		AK22 (Methylated)	86000	74000	180000	113333.33
		AK22 (Unmethylated)	140000	83333.3	280000	142777.77
	RC5282	pBS SK+ (Methylated)	1.40E+08	2.40E+08	1.64E+08	1.81E+08
		pBS SK+ (Unmethylated)	1.20E+08	2.80E+07	1.34E+08	9.40E+07
		AK22 (Methylated)	44000	84000	88000	72000
		AK22 (Unmethylated)	4000	48000	1300000	450666.67

Table S4: Transformation efficiency of different strains transformed with methylated and unmethylated pBS SK+ and AK22 plasmid. Efficiency is measured in CFU/μg in Figure 5.2 and Supplementary Figure S3

Figure	Condition	Rep 1	Rep 2	Rep 3	Rep 4	Rep 5	Mean	Comparison	P value	Summary
4.6	Wildtype	4.76	4.46	4.34	4.75	4.53	4.57	-	-	-
	M17V	9.74	8.20	9.53	13.74	-	10.30	Wildtype	0.001	***
	M17T	7.78	10.9	12.76	12.02	12.84	11.26	Wildtype	0.0001	***
	D26G	12.67	7.19	10.04	8.96	5.68	8.91	Wildtype	0.0069	**
	I28K	24.15	27.85	23.09	23.27	24.63	24.6	Wildtype	<0.0001	****
	D29N	13.46	11.67	13.04	12.06	13.08	12.66	Wildtype	<0.0001	****
	Q90R	20	20.13	19.03	19.63	19.67	19.49	Wildtype	<0.0001	****
	Y101H	8.42	8.29	7.83	7.74	7.89	8.03	Wildtype	<0.0001	****
4.10A	Wildtype	4.76	4.46	4.34	4.34	4.53	4.49	-	-	-
	I28K	24.15	27.85	23.09	23.27	24.63	24.6	Wildtype	<0.0001	****
	D29N	13.46	11.67	13.04	12.06	13.08	12.66	Wildtype	<0.0001	****
	I28S	13.26	17.82	14.6	19.95	13.45	15.82	Wildtype	<0.0001	****
	I28N	14.25	14.35	17.04	15.12	19.87	16.13	Wildtype	<0.0001	****
	I28R	15.58	18.67	19.08	16.56	18.11	17.6	Wildtype	<0.0001	****
	D29H	11.65	11.41	9.97	12.15	11.43	11.32	Wildtype	<0.0001	****
	D29K	13.32	13.99	12.93	12.77	-	13.25	Wildtype	<0.0001	****
	D29A	14.1	14.08	14.73	15.86	14.15	14.58	Wildtype	<0.0001	****
4.11A	Wildtype	4.76	4.46	4.34	4.75	4.53	4.57	-	-	-
	M17T	7.78	10.9	12.76	12.02	12.84	11.26	Wildtype	0.0001	***
	D26G	12.67	7.19	10.04	8.96	5.68	8.91	Wildtype	0.0069	**
	I28K	24.15	27.85	23.09	23.27	24.63	24.6	Wildtype	<0.0001	****
	Q90R	20	20.13	19.03	18.63	19.67	19.49	Wildtype	<0.0001	****
	M17T I28K	13.01	11.47	17.99	12.68	10.02	13.03	Wildtype	0.0002	***
	D26G I28K	2.09	2.28	5.62	8.65	3.08	4.34	Wildtype	0.8622	ns
	I28K Q90R	10.2	17.78	16.6	1.5	7.71	10.76	Wildtype	0.0723	ns
S2	Wildtype	4.76	4.46	4.34	4.75	4.53	4.57	-	-	-
	dCas1	0.14	0.19	0.14	0.14	0.18	0.16	Wildtype	<0.0001	****

Table S5: YFP reporter assay flow cytometry data. YFP+ percentages are given for replicates 1-5 for various conditions. T-test statistical comparisons are provided. Data for Figures 4.6, 4.10A, 4.11A and Supplementary Figure S2.

Figure	Condition	Rep 1	Rep 2	Rep 3	Mean	Comparison	Direction	P value	Summary
5.12B	Wildtype	4.79	3	2.75	3.51	-	-	-	-
	Hit 11	4.8	4.63	5.08	4.84	Wildtype	-	0.1137	ns
	Hit 54	5.55	5.23	5.6	5.46	Wildtype	Increase	0.0407	*
	Hit 91	6.63	5.98	5.89	6.17	Wildtype	Increase	0.0178	*
5.15A	Wildtype	4.79	3	2.75	3.51	-	-	-	-
	Hit 1	4.74	2.86	3.85	3.82	Wildtype	-	0.7366	ns
	Hit 2	2.82	4.52	2.8	3.38	Wildtype	-	0.8841	ns
	Hit 3	3.18	3.02	3.18	3.13	Wildtype	-	0.5809	ns
	Hit 4	5.8	4.66	3.47	4.64	Wildtype	-	0.2912	ns
	Hit 5	2.92	4.38	3.95	3.75	Wildtype	-	0.7753	ns
5.15B	Wildtype	1.49	1.24	1.45	1.39	-	-	-	-
	Hit 6	1.33	2.71	1.22	1.75	Wildtype	-	0.4997	ns
	Hit 7	1.06	1.03	1.32	1.14	Wildtype	-	0.0999	ns
	Hit 8	1.07	0.64	1.25	0.99	Wildtype	-	0.1078	ns
	Hit 9	0.83	1.43	1.85	1.37	Wildtype	-	0.9429	ns
	Hit 10	0.81	1.38	1.25	1.15	Wildtype	-	0.2621	ns
	Hit 11	1.25	0.83	1.26	1.11	Wildtype	-	0.158	ns
	Hit 12	1.22	2.31	1.4	1.64	Wildtype	-	0.5101	ns
	Hit 13		1.41	1.7	1.56	Wildtype	-	0.3526	ns
	Hit 14	2.03	1.19	1.42	1.55	Wildtype	-	0.5903	ns
	Hit 15	1.41	1.78	1.48	1.56	Wildtype	-	0.3004	ns
	Hit 16	1.46	1.54	1.86	1.62	Wildtype	-	0.1924	ns
	Hit 17	3.95	2	1.94	2.63	Wildtype	-	0.1364	ns
5.15C	Wildtype	2.15	2.62	2.36	2.38	-	-	-	-
	Hit 18	5.4	5.18	4.04	4.87	Wildtype	Increase	0.0049	**
	Hit 19	2.47	3.61	2.23	2.77	Wildtype	-	0.4284	ns
	Hit 20	2.99	3.32	1.86	2.72	Wildtype	-	0.4952	ns
	Hit 21	1.74	2.65	2.83	2.41	Wildtype	-	0.9382	ns
	Hit 22	8.56	5.05	3.18	5.6	Wildtype	-	0.1117	ns
	Hit 23	2.34	3.58	2.05	2.66	Wildtype	-	0.5972	ns
	Hit 24	3.19	3.26	5.32	3.92	Wildtype	-	0.0955	ns
	Hit 25	2.34	1.88	2.98	2.4	Wildtype	-	0.9496	ns
	Hit 26	2.67	2.92	3.44	3.01	Wildtype	-	0.0747	ns
	Hit 27	5.18	2.46	2.7	3.45	Wildtype	-	0.2909	ns
	Hit 28	3.12	2.16	2.27	2.52	Wildtype	-	0.6953	ns
	Hit 29	2.64	5.34	3.51	3.83	Wildtype	-	0.1462	ns
	Hit 30	3.73	7.74	3.71	5.06	Wildtype	-	0.1171	ns
5.15D	Wildtype	1.78	2.1	1.82	1.9	-	-	-	-
	Hit 31	3.6	1.79	2.2	2.53	Wildtype	-	0.3213	ns
	Hit 32	2.46	6.08	2.7	3.75	Wildtype	-	0.1905	ns
	Hit 33	2.58	1.65	2.34	2.19	Wildtype	-	0.3832	ns
	Hit 34	1.84	1.55	2.15	1.85	Wildtype	-	0.8033	ns
	Hit 35	2.16	1.92	2.29	2.12	Wildtype	-	0.2056	ns
	Hit 36	2.01	3.27	2.17	2.48	Wildtype	-	0.2266	ns
	Hit 37	2.18	2.1	1.81	2.03	Wildtype	-	0.4375	ns
	Hit 38	3.42	4.63	3.49	3.85	Wildtype	Increase	0.0086	**
	Hit 39	5.19	2.5	2.11	3.27	Wildtype	-	0.233	ns
	Hit 40	4.33	4.58	2.5	3.8	Wildtype	Increase	0.0455	*
	Hit 41	2.19	2.29	2.52	2.33	Wildtype	Increase	0.0366	*
	Hit 42	2.49	2.3	2.21	2.33	Wildtype	Increase	0.0291	*
	Hit 43	1.54	1.89	1.42	1.62	Wildtype	-	0.1773	ns
	Hit 44	2.04	4.53	1.92	2.83	Wildtype	-	0.3387	ns
	Hit 45	1.71	1.99	2.23	1.98	Wildtype	-	0.6934	ns
	Hit 46	2.7	2.11	3.18	2.66	Wildtype	-	0.0789	ns

Table S6: YFP reporter assay flow cytometry data for Figures 5.12B and 5.15. YFP reporter assay flow cytometry data. YFP+ percentages are given for replicates 1-3 for various candidates. T-test statistical comparisons are provided.

Figure	Condition	Rep 1	Rep 2	Rep 3	Rep 4	Rep 5	Rep 6	Mean	Comparison	P value	Summary
5.16B	No Plasmid	2.66	2.88	2.33	2.64	3.33	2.72	2.76	-	-	-
	0 uM Vanillic Acid	3.05	3.29	3.31	2.9	2.75	2.5	2.97	-	-	-
	5 uM Vanillic Acid	2.73	3	2.91	2.66	2.7	3.18	2.86	0 uM	0.5157	ns
	10 uM Vanillic Acid	2.48	2.73	3.25	2.65	2.98	2.78	2.81	0 uM	0.3822	ns
	20 uM Vanillic Acid	3.23	2.16	2.26	2.59	2.44	2.44	2.52	0 uM	0.0509	ns
	50 uM Vanillic Acid	2.62	2.31	1.74	2.47	2.27	2.24	2.28	0 uM	0.003	**
	100 uM Vanillic Acid	3	3.79	3.29	2.96	3.14	2.55	3.12	0 uM	0.4804	ns

Table S7: YFP reporter assay flow cytometry data. YFP+ percentages are given for replicates 1-6 for various conditions. T-test statistical comparisons are provided. Data for Figure 5.16B.

Figure	Condition (% YFP+ Cell Population)	Rep 1	Rep 2	Rep 3	Rep 4	Rep 5	Mean
6.12D	100	57.93	56.43	56.71	65.26	55.56	58.38
	50	34.28	29.12	30.01	24.95	29.84	29.64
	10	12.19	7.14	7.35	5.98	7.23	7.98
	1	1.04	0.93	0.85	0.59	0.93	0.87
	0.5	0.47	0.48	0.52	0.35	0.37	0.44
	0.1	0.11	0.08	0.11	0.23	0.12	0.13
	0.01	0.11	0.07	0.08	0.02	0.06	0.07
	0	0.06	0.03	0.03	0.06	0.04	0.04

Table S8: YFP reporter assay flow cytometry data. YFP+ percentages are given for replicates 1-5 for various conditions. Data for Figure 6.12D.

Figure	Condition	Rep 1	Rep 2	Rep 3	Rep 4	Rep 5	Rep 6	Mean	Comparison	P value	Summary
6.13A	No expression vector	0.01	0.09	0.01	0.03	0.01	0.01	0.03	-	-	-
	0%	0.14	0.01	0.01	0	0.03	0	0.03	-	-	-
	0.002%	0.02	0.03	0.01	0.01	0.7	0	0.13	0%	0.426	ns
	0.02%	0.01	0.01	0	0	0.38	0.03	0.07	0%	0.556	ns
	0.2%	0	0.01	0.02	0	0.27	0	0.05	0%	0.718	ns

Table S9: YFP reporter assay flow cytometry data. YFP+ percentages are given for replicates 1-6 for various conditions. % refers to L-arabinose inducer concentration. T-test statistical comparisons are provided. Data for Figure 6.13A.

Figure	Condition	Rep 1	Rep 2	Rep 3	Rep 4	Rep 5	Rep 6	Mean	Comparison	P value	Summary
4.7B	0 ug/ml Nalidixic Acid	179.61	-255.07	75.46	-	-	-	0	-	-	-
	0.2 ug/ml Nalidixic Acid	1242.6	1218	1287.83	-	-	-	1249.48	0 ug/ml Nalidixic Acid	0.0007	***
	0.4 ug/ml Nalidixic Acid	2287.82	2370.72	2593.56	-	-	-	2417.36	0 ug/ml Nalidixic Acid	0.0001	***
	0.8 ug/ml Nalidixic Acid	3102.45	3903.27	3334.62	-	-	-	3446.78	0 ug/ml Nalidixic Acid	0.0002	***
4.7C	Wildtype	8037.89	6858.92	5724.52	-	-	-	6873.78	-	-	-
	M17V	20750.52	20526.42	16571.58	-	-	-	19282.84	Wildtype	0.0012	**
	M17T	20523.39	19475.65	13995.85	-	-	-	17998.3	Wildtype	0.0064	**
	D26G	17937.89	17430	12524.48	-	-	-	15964.12	Wildtype	0.008	**
	I28K	13735.82	14263.37	9098.07	-	-	-	12365.76	Wildtype	0.0362	*
	D29N	20870.82	21651.08	15689.07	-	-	-	19403.66	Wildtype	0.0032	**
	Q90R	17248.09	17187.93	12729.93	-	-	-	15721.98	Wildtype	0.0057	**
	Y101H	14545.44	14180.36	10012.81	-	-	-	12912.87	Wildtype	0.0195	*
4.10B	Wildtype	4881.88	9867.33	10320.66	9551.22	8827.39	9018.75	8744.54	-	-	-
	I28K	17775.27	23119.29	21610.06	23229.99	19896.29	21284.21	21152.52	Wildtype	<0.0001	****
	I28S	9239.37	6100.93	10493.99	16113.09	10860.41	10996.19	10634	Wildtype	0.2509	ns
	I28N	13410.83	5967.48	12946.14	6712.73	10272.91	11821.33	10188.57	Wildtype	0.3666	ns
	I28R	24948.86	17140.27	21479.07	20611.5	18358.81	13449.36	19331.31	Wildtype	0.0002	***
	D29N	20036.22	16368.99	20461.59	16394.85	17662.74	14142.41	17511.13	Wildtype	<0.0001	****
	D29H	15599.26	21832.19	28381.37	18598.92	17778.6	18641.02	20138.56	Wildtype	0.0002	***
	D29K	17027.14	18337.61	20409.49	12288.11	16204.11	22408.77	17779.21	Wildtype	0.0003	***
	D29A	23563.91	7943.24	9948.46	12782.98	10599.75	12991.8	12971.69	Wildtype	0.1077	ns
4.11B	Wildtype	5965.37	5810.8	6291.73	-	-	-	6022.63	-	-	-
	M17T	27436.51	20766.65	24846.4	-	-	-	24349.85	Wildtype	0.0007	***
	D26G	25055.58	26662.76	23954.89	-	-	-	25224.41	Wildtype	<0.0001	****
	I28K	14067.97	20274.52	15465.09	-	-	-	16602.53	Wildtype	0.005	**
	Q90R	7053.31	7102.05	7322.18	-	-	-	7159.18	Wildtype	0.0023	**
	M17T I28K	45074.93	41207.36	40409.71	-	-	-	42230.67	Wildtype	<0.0001	****
	-	-	-	-	-	-	-	-	M17T	0.0018	**
	-	-	-	-	-	-	-	-	I28K	0.0004	***
	D26G I28K	21682.66	33297.22	52185.39	-	-	-	35721.76	Wildtype	0.0288	*
	-	-	-	-	-	-	-	-	D26G	0.3047	ns
	-	-	-	-	-	-	-	-	I28K	0.1031	ns
	I28K Q90R	36609.84	47607.92	55668.13	-	-	-	46628.63	Wildtype	0.0018	**
	-	-	-	-					I28K	0.0068	**
	-	-	-	-					Q90R	0.0018	**

Table S10: GFP SOS reporter assay data. Fluorescence in a.u is given for replicates 1-3/6 for various conditions. T-test statistical comparisons are provided. Data for Figures 4.7B, 4.7C, 4.10B and 4.11B.

Figure	Condition	Rep 1	Rep 2	Rep 3	Rep 4	Rep 5	Rep 6	Mean	Comparison	P value	Summary
6.7A	Strain	19112.13	19320.61	19174.98	19864.05	19744.56	19982.14	19533.08	-	-	-
	Type II (A)	19553.92	19799.7	19876.38	19846.8	20086.02	19922.55	19847.56	Strain	0.0926	ns
	Type II (A) +ve	19260.01	19815.34	19754.03	20059.94	20511.37	19538.31	19823.17	Strain	0.2429	ns
	Type I (T)	19605.9	19677.82	19579.12	19651.94	19927.16	19618.02	19676.66	Strain	0.3964	ns
	Type I (T) +ve	20023.25	19216.7	19569.65	20249.68	19938.22	20154.55	19858.68	Strain	0.1729	ns
6.7B	Strain 0uM Vanillic Acid	19112.13	19320.61	19174.98	19864.05	19744.56	19982.14	19533.08	-	-	-
	Strain 100uM Vanillic Acid	20409.47	20582.3	20438.92	20296.2	20700.57	20885.62	20552.18	Strain 0uM Vanillic Acid	0.0002	***
	Pvan lacZ 0uM Vanillic Acid	18614.35	19122.71	19517.82	19907.16	19568.05	19568.05	19383.02	Strain 0uM Vanillic Acid	0.5456	ns
	Pvan lacZ 100uM Vanillic Acid	18550.57	19438.26	19419.1	19906.26	19502.65	19566.34	19397.2	Pvan lacZ 0uM Vanillic Acid	0.9577	ns
	Pvan YFP 0uM Vanillic Acid	19580.45	19879.48	20111.8	20358.31	19728.59	20290.93	19991.59	Pvan lacZ 0uM Vanillic Acid	0.0218	*
	Pvan YFP 100uM Vanillic Acid	25169.38	25420.95	23714.7	24638.09	24834.06	24275.47	24675.44	Pvan YFP 0uM Vanillic Acid	<0.0001	****
	Pvan YFP Universal Reporter 0uM Vanillic Acid	28600.33	27713.94	29106.46	28678.06	29400.53	29869.35	28894.78	Pvan YFP 0uM Vanillic Acid	<0.0001	****
	Pvan YFP Universal Reporter 100uM Vanillic Acid	28686.15	29829.94	29525.43	30469.53	30093.3	30071.7	29779.34	Pvan YFP Universal Reporter 0uM Vanillic Acid	0.0495	*
6.8	Strain	667.6	321.58	105.3	-78.97	-314.51	-700.99	0	-	-	-
	PhlF - Without Repressor	27391.88	28294.84	27979.76	29447.53	26247.7	29366.76	28121.41	-	-	-
	PhlF - WIth Repressor	1818.15	1948.79	2077.45	1866.01	1802.01	1906.21	1903.1	PhlF - Without Repressor	<0.0001	****
	PhlF - With CRISPR Locus-Repressor Fusion	25970.2	26402.44	25466.51	26608.41	27266.69	26592.68	26384.49	PhlF - WIth Repressor	<0.0001	****
	CymR - Without Repressor	24177.26	25230.8	27395.28	25582.69	26060.95	25617.25	25677.37	-	-	-
	CymR - WIth Repressor	1824.27	1830.76	1678.03	1629.9	1836.79	1786.64	1764.4	CymR - Without Repressor	<0.0001	****
	CymR - With CRISPR Locus-Repressor Fusion	3014.71	3368.18	2826.36	3278.68	3763.34	3790.52	3340.3	CymR - WIth Repressor	<0.0001	****
	VanR - Without Repressor	26636.87	26340.28	26410.91	26967.1	27341.06	25840.7	26589.49	-	-	-
	VanR - WIth Repressor	1803.2	1882	2004.92	1872.27	2007.86	2190.56	1960.14	VanR - Without Repressor	<0.0001	****
	VanR - With CRISPR Locus-Repressor Fusion	10310.95	12177.54	11279.73	11115.3	9950.34	12311.01	11190.81	VanR - WIth Repressor	<0.0001	****
	TetR - Without Repressor	18149.52	18419.41	17803.09	19195.01	20867.42	21167.99	19267.07	-	-	-
	TetR - WIth Repressor	1782.66	1729.83	1828.09	1867.51	1882.75	1776.48	1811.22	TetR - Without Repressor	<0.0001	****
	TetR - With CRISPR Locus-Repressor Fusion	10662.42	10604.32	10934.59	9011.01	9998.62	7836.76	9841.29	TetR - WIth Repressor	<0.0001	****
	BetI - Without Repressor	14284.62	15435.61	14717.48	15239.4	14809.94	14734.34	14870.23	-	-	-
	BetI - WIth Repressor	4175.78	3182.01	3647.07	3827.38	3513.88	3400.89	3624.5	BetI - Without Repressor	<0.0001	****
	BetI - With CRISPR Locus-Repressor Fusion	16744.21	17318.31	16373.2	16679.98	13737.48	15742.58	16099.29	BetI - WIth Repressor	<0.0001	****
	TtgR - Without Repressor	4938.04	4643.66	4877.68	5106.85	4951.33	5340.01	4976.26	-	-	-
	TtgR - WIth Repressor	1653.47	1809.18	1716.39	1656.23	1989.39	1831.16	1775.97	TtgR - Without Repressor	<0.0001	****
	TtgR - With CRISPR Locus-Repressor Fusion	3720.11	3046.36	2860.98	2570.97	1645.4	1613.7	2576.25	TtgR - WIth Repressor	0.0408	*
	PcaU - Without Repressor	2606.08	2539.24	2762.89	2711.06	2717.11	2765.12	2683.58	-	-	-
	PcaU - WIth Repressor	2671.26	2704.1	2754.58	2652.6	2844.35	3101.16	2788.01	PcaU - Without Repressor	0.2109	ns
	PcaU - With CRISPR Locus-Repressor Fusion	2847.91	2860.42	2550.42	2034.32	1974.07	2068.6	2389.29	PcaU - WIth Repressor	0.054	ns

Table S11: YFP induction assay data. Fluorescence in a.u are given for replicates 1-6 for various conditions. T-test statistical comparisons are provided. Data for Figures 6.7 and 6.8.

Figure	Condition	Rep 1	Rep 2	Rep 3	Rep 4	Rep 5	Rep 6	Mean	Comparison	P value	Summary
6.10	No Plasmid	-146.46	-303.25	-390.6	-156.77	-63.14	1060.22	0	-	-	-
	P15A-*ori* CymR 0uM Cuminic Acid	-3.15	-81.27	-66.5	26.24	-171.36	39.44	-42.77	No Plasmid	0.8497	ns
	P15A-*ori* CymR 5uM Cuminic Acid	1114.36	1088.83	1100.61	1335.81	1325.16	1235.58	1200.06	P15A-*ori*CymR 0uM Cuminic Acid	<0.0001	****
	P15A-*ori* CymR 10uM Cuminic Acid	4795.38	5197.5	6384.26	5271.77	6781.35	6061.12	5748.56	P15A-*ori*CymR 0uM Cuminic Acid	<0.0001	****
	P15A-*ori* CymR 20uM Cuminic Acid	13552.2	12595.21	13620.47	14090.38	13347.83	14045.04	13541.86	P15A-*ori*CymR 0uM Cuminic Acid	<0.0001	****
	P15A-*ori* CymR 50uM Cuminic Acid	20374.61	19899.55	21062.95	21699.56	20555.42	22320.45	20985.42	P15A-*ori*CymR 0uM Cuminic Acid	<0.0001	****
	P15A-*ori* CymR 100uM Cuminic Acid	21005.03	22098.5	22838.65	20880.05	23507.31	23598.79	22321.39	P15A-*ori*CymR 0uM Cuminic Acid	<0.0001	****
	No Plasmid	-146.46	-303.25	-390.6	-156.77	-63.14	1060.22	0	-	-	-
	P15A-*ori*CymR Fusion 0uM Cuminic Acid	1691.67	1743.37	1605.49	1798.47	1608.15	2052.46	1749.94	No Plasmid	<0.0001	****
	P15A-*ori*CymR Fusion 5uM Cuminic Acid	34047.9	38271.78	38086.14	35043.51	31215.41	29236.71	34316.91	P15A-*ori*CymR Fusion 0uM Cuminic Acid	<0.0001	****
	P15A-*ori*CymR Fusion 10uM Cuminic Acid	41001.75	34544.4	37196.2	35825.95	33227.03	31892.4	35614.62	P15A-*ori*CymR Fusion 0uM Cuminic Acid	<0.0001	****
	P15A-*ori*CymR Fusion 20uM Cuminic Acid	37321.2	35911.3	36372.08	36149.69	37007.43	34052.79	36135.75	P15A-*ori*CymR Fusion 0uM Cuminic Acid	<0.0001	****
	P15A-*ori*CymR Fusion 50uM Cuminic Acid	36885.37	34449.5	34975.93	30394.42	33499.88	28014.56	33036.61	P15A-*ori*CymR Fusion 0uM Cuminic Acid	<0.0001	****
	P15A-*ori*CymR Fusion 100uM Cuminic Acid	35065.91	28620.24	33735.73	30454.48	27960.97	27557.94	30565.88	P15A-*ori*CymR Fusion 0uM Cuminic Acid	<0.0001	****
	No Plasmid	-706.39	227.76	478.63	-	-	-	0	-	-	-
	F-*ori*CymR 0uM Cuminic Acid	1552.19	1193.73	1008.14	-	-	-	1251.35	No Plasmid	0.0337	*
	F-*ori*CymR 5uM Cuminic Acid	6817.65	7041.95	7343.86	-	-	-	7067.82	F-*ori*CymR 0uM Cuminic Acid	<0.0001	****
	F-*ori*CymR 10uM Cuminic Acid	8769.13	12608.45	10948.89	-	-	-	10775.49	F-*ori*CymR 0uM Cuminic Acid	0.0011	**
	F-*ori*CymR 20uM Cuminic Acid	13301.12	12337.17	10848.89	-	-	-	12162.39	F-*ori*CymR 0uM Cuminic Acid	<0.0001	****
	F-*ori*CymR 50uM Cuminic Acid	13008.49	14990.2	13220.95	-	-	-	13739.88	F-*ori*CymR 0uM Cuminic Acid	<0.0001	****
	F-*ori*CymR 100uM Cuminic Acid	12453	14276.75	12376.63	-	-	-	13035.46	F-*ori*CymR 0uM Cuminic Acid	<0.0001	****
	No Plasmid	-706.39	227.76	478.63	-	-	-	0	-	-	-
	F-*ori*CymR Fusion 0uM Cuminic Acid	16742.67	18270.09	17866.59	-	-	-	17626.45	No Plasmid	<0.0001	****
	F-*ori*CymR Fusion 5uM Cuminic Acid	34152.89	35906.99	36163.08	-	-	-	35407.65	F-*ori*CymR Fusion 0uM Cuminic Acid	<0.0001	****
	F-*ori*CymR Fusion 10uM Cuminic Acid	32549.45	32995.67	35763.8	-	-	-	33769.64	F-*ori*CymR Fusion 0uM Cuminic Acid	<0.0001	****
	F-*ori*CymR Fusion 20uM Cuminic Acid	34863.9	35129.45	35900.67	-	-	-	35298.01	F-*ori*CymR Fusion 0uM Cuminic Acid	<0.0001	****
	F-*ori*CymR Fusion 50uM Cuminic Acid	31622.46	32429.4	32184.21	-	-	-	32078.69	F-*ori*CymR Fusion 0uM Cuminic Acid	<0.0001	****
	F-*ori*CymR Fusion 100uM Cuminic Acid	29536.22	31999.05	32172.77	-	-	-	31236.01	F-*ori*CymR Fusion 0uM Cuminic Acid	<0.0001	****
6.11	Strain	-706.39	227.76	478.63	-	-	-	0	-	-	-
	CymR - Without Repressor	34208.79	37102.12	36025.95	-	-	-	35778.95	Strain	<0.0001	****
	CymR - With Repressor	1552.19	1193.73	1008.14	-	-	-	1251.35	CymR - Without Repressor	<0.0001	****
	CymR Universal Reporter	16742.67	18270.09	17866.59	-	-	-	17626.45	CymR - WIth Repressor	<0.0001	****
	CymR Universal Reporter Frameshift Control	33270.5	36549.09	33985.83	-	-	-	34601.81	CymR Universal Reporter	0.0001	***
	CymR Universal Reporter Stop Codon Control	36097.34	36821.16	36243.64	-	-	-	36387.38	CymR Universal Reporter	<0.0001	****

Table S12: YFP induction assay data. Fluorescence in a.u is given for replicates 1-3/6 for various conditions. T-test statistical comparisons are provided. Data for Figures 6.10 and 6.11.

Figure	Source	Licence Number
Figure 1.3	Isaev *et al*., 2021	Creative Commons
Figure 1.4	van der Oost *et al*., 2014	5916020606836
Figure 1.5	Nuñez *et al*., 2015)	5916021148329
Figure 1.6	Nunez *et al*., 2015	5916021308281
Figure 1.7	Rath *et al*., 2015	Creative Commons
Figure 1.8	Zheng *et al*., 2020	Creative Commons
Figure 1.9	Zheng *et al*., 2020	Creative Commons
Figure 1.10	Xue and Sashital, 2019	Creative Commons
Figure 1.11	Makarova *et al*., 2020	5916030110900
Figure 1.12	Mayo-Muñoz *et al*., 2024	5916030234818
Figure 1.13	Bi *et al*., 2024	Creative Commons
Figure 3.5	Yosef *et al*., 2012	Creative Commons
Figure 5.3	Philips *et al*., 2019	Creative Commons

Table S13: Licence numbers for the reproduction of published figures in this work.

Acknowledgements

I would like to thank Ronald Chalmers for his supervision and support over the past years. I would also like to thank Dr. Harry Edwards, your steadfast patience in the face of my incessant bullshit has been instrumental. Thank you to Dr Jack Braithewaite for helping to show me the ropes both in the lab and at the JA. Finally, thank you to Colin for the many hours spent pouring plates together, I hope one day you'll find a PhD student worthy of your references.

Thank you to my friends from home for always being around to catch up and run it down, notably Jack for his help with Figure 6.1. To my friends from uni, thank you for only occasionally asking how many years it takes to finish a three year PhD.

Finally, I would like to thank my parents for a lifetime of ceaseless support.

I also gratefully acknowledge the financial support of the University of Nottingham and the Leverhulme Trust.

Bibliography

1. Dion, M. B., Oechslin, F. & Moineau, S. Phage diversity, genomics and phylogeny. *Nature Reviews Microbiology* **18,** 125–138 (Mar. 2020).
2. Koonin, E. V., Senkevich, T. G. & Dolja, V. V. The ancient Virus World and evolution of cells. *Biology Direct* **1,** 29 (2006).
3. Tesson, F. *et al.* Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nature Communications* **13,** 2561 (May 10, 2022).
4. Mayo-Muñoz, D., Pinilla-Redondo, R., Camara-Wilpert, S., Birkholz, N. & Fineran, P. C. Inhibitors of bacterial immune systems: discovery, mechanisms and applications. *Nature Reviews Genetics* **25,** 237–254 (Apr. 2024).
5. Partridge, S. R., Kwong, S. M., Firth, N. & Jensen, S. O. Mobile Genetic Elements Associated with Antimicrobial Resistance. *Clinical Microbiology Reviews* **31,** e00088–17 (Oct. 2018).
6. Karaolis, D. K. R. *et al.* A *Vibrio cholerae* pathogenicity island associated with epidemic and pandemic strains. *Proceedings of the National Academy of Sciences* **95,** 3134–3139 (Mar. 17, 1998).
7. Luria, S. E. & Human, M. L. A nonhereditary, host-induced variation of bacterial viruses. *Journal of Bacteriology* **64,** 557–569 (Oct. 1952).
8. Bertani, G. & Weigle, J. J. Host controlled variation in bacterial viruses. *Journal of Bacteriology* **65,** 113–121 (Feb. 1953).
9. Krüger, D. H. & Bickle, T. A. Abortive infection of Escherichia coli F+ cells by bacteriophage T7 requires ribosomal misreading. *Journal of Molecular Biology* **194,** 349–352 (Mar. 1987).
10. Thisted, T., Sørensen, N., Wagner, E. & Gerdes, K. Mechanism of post-segregational killing: Sok antisense RNA interacts with Hok mRNA via its 5-end single-stranded leader and competes with the 3-end of Hok mRNA for binding to the mok translational initiation region. *The EMBO Journal* **13,** 1960–1968 (Apr. 1994).
11. Yamaguchi, Y., Park, J.-H. & Inouye, M. Toxin-Antitoxin Systems in Bacteria and Archaea. *Annual Review of Genetics* **45,** 61–79 (Dec. 15, 2011).
12. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **315,** 1709–1712 (Mar. 23, 2007).
13. Makarova, K. S., Wolf, Y. I., Snir, S. & Koonin, E. V. Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems. *Journal of Bacteriology* **193,** 6039–6056 (Nov. 2011).
14. Gao, L. *et al.* Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* **369,** 1077–1084 (Aug. 28, 2020).
15. Doron, S. *et al.* Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359,** eaar4120 (Mar. 2, 2018).
16. Millman, A. *et al.* An expanded arsenal of immune systems that protect bacteria from phages. *Cell Host & Microbe* **30,** 1556–1569.e5 (Nov. 2022).
17. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nature Reviews Microbiology* **8,** 317–327 (May 2010).
18. Stokar-Avihail, A. *et al.* Discovery of phage determinants that confer sensitivity to bacterial immune systems. *Cell* **186,** 1863–1876.e16 (Apr. 2023).

19. Roberts, R. J., Vincze, T., Posfai, J. & Macelis, D. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Research* **38,** D234–D236 (Jan. 2010).
20. Kong, H. Functional analysis of putative restriction-modification system genes in the Helicobacter pylori J99 genome. *Nucleic Acids Research* **28,** 3216–3223 (Sept. 1, 2000).
21. Pingoud, A. Structure and function of type II restriction endonucleases. *Nucleic Acids Research* **29,** 3705–3727 (Sept. 15, 2001).
22. Tock, M. R. & Dryden, D. T. The biology of restriction and anti-restriction. *Current Opinion in Microbiology* **8,** 466–472 (Aug. 2005).
23. Kruger, D. H. Bacteriophage Survival: Multiple Mechanisms for Avoiding the Deoxyribonucleic Acid Restriction Systems of Their Hosts. *MICROBIOL. REV.* **47** (1983).
24. Bandyopadhyay, P. K., Studier, F. W., Hamilton, D. L. & Yuan, R. Inhibition of the Type I Restriction-modification Enzymes EcoB and EcoK by the Gene O-3 Protein of Bacteriophage T7. *J. Mol. Biol.* **182,** 567–578 (Oct. 1984).
25. Sumby, P. & Smith, M. C. M. Genetics of the phage growth limitation (Pgl) system of *Streptomyces coelicolor* A3(2). *Molecular Microbiology* **44,** 489–500 (Apr. 2002).
26. Fukuyo, M., Sasaki, A. & Kobayashi, I. Success of a suicidal defense strategy against infection in a structured habitat. *Scientific Reports* **2,** 238 (Jan. 30, 2012).
27. Fowler, D. M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nature Methods* **11,** 801–807 (Aug. 2014).
28. Schmitt, C. K. & Molineux, I. J. Expression of gene 1.2 and gene 10 of bacteriophage T7 is lethal to F plasmid-containing Escherichia coli. *Journal of Bacteriology* **173,** 1536–1543 (Feb. 1991).
29. Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G. & Siksnys, V. A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science* **357,** 605–609 (Aug. 11, 2017).
30. Snyder, L. Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents? *Molecular Microbiology* **15,** 415–420 (Feb. 1995).
31. Durmaz, E. & Klaenhammer, T. R. Abortive Phage Resistance Mechanism AbiZ Speeds the Lysis Clock To Cause Premature Lysis of Phage-Infected *Lactococcus lactis*. *Journal of Bacteriology* **189,** 1417–1425 (Feb. 15, 2007).
32. Lau, R. K. *et al.* Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity. *Molecular Cell* **77,** 723–733.e6 (Feb. 2020).
33. Levitz, R. *et al.* The optional E. coli prr locus encodes a latent form of phage T4-induced anticodon nuclease. *The EMBO Journal* **9,** 1383–1389 (May 1990).
34. Landsmann, J., Kroger, M. & Hobom, G. The rex region of bacteriophage lambda: two genes under three-way control. *Gene* **20,** 11–24 (Nov. 1982).
35. Parma, D. H. *et al.* The Rex system of bacteriophage k: tolerance and altruistic cell death. *Genes & Development* **6,** 497–510 (Jan. 1992).
36. Shinedling, S., Parma, D. & Gold, L. Wild-type bacteriophage T4 is restricted by the lambda rex genes. *Journal of Virology* **61,** 3790–3794 (Dec. 1987).
37. Unterholzner, S. J., Poppenberger, B. & Rozhon, W. Toxin–antitoxin systems: Biology, identification, and application. *Mobile Genetic Elements* **3,** e26219 (Sept. 20, 2013).
38. Schuster, C. F. & Bertram, R. Toxin-antitoxin systems are ubiquitous and versatile modulators of prokaryotic cell fate. *FEMS Microbiology Letters* **340,** 73–85 (Mar. 2013).
39. Gerdes, K., Rasmussen, P. B. & Molin, S. Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. *Proceedings of the National Academy of Sciences* **83,** 3116–3120 (May 1986).
40. Keren, I., Shah, D., Spoering, A., Kaldalu, N. & Lewis, K. Specialized Persister Cells and the Mechanism of Multidrug Tolerance in *Escherichia coli*. *Journal of Bacteriology* **186,** 8172–8180 (Dec. 15, 2004).
41. Fineran, P. C. *et al.* The phage abortive infection system, ToxIN, functions as a protein–RNA toxin–antitoxin pair. *Proceedings of the National Academy of Sciences* **106,** 894–899 (Jan. 20, 2009).

42. Blower, T. R. *et al.* A processed noncoding RNA regulates an altruistic bacterial antiviral system. *Nature Structural & Molecular Biology* **18,** 185–190 (Feb. 2011).
43. Blower, T. R., Evans, T. J., Przybilski, R., Fineran, P. C. & Salmond, G. P. C. Viral Evasion of a Bacterial Suicide System by RNA–Based Molecular Mimicry Enables Infectious Altruism. *PLoS Genetics* **8** (ed Casadesús, J.) e1003023 (Oct. 18, 2012).
44. Cohen, D. *et al.* Cyclic GMP–AMP signalling protects bacteria against viral infection. *Nature* **574,** 691–695 (Oct. 31, 2019).
45. Whiteley, A. T. *et al.* Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature* **567,** 194–199 (Mar. 2019).
46. Davies, B. W., Bogard, R. W., Young, T. S. & Mekalanos, J. J. Coordinated Regulation of Accessory Genetic Elements Produces Cyclic Di-Nucleotides for V. cholerae Virulence. *Cell* **149,** 358–370 (Apr. 2012).
47. Hobbs, S. J. *et al.* Phage anti-CBASS and anti-Pycsar nucleases subvert bacterial immunity. *Nature* **605,** 522–526 (May 19, 2022).
48. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology* **13,** 722–736 (Nov. 2015).
49. Tamulaitis, G., Venclovas, Č. & Siksnys, V. Type III CRISPR-Cas Immunity: Major Differences Brushed Aside. *Trends in Microbiology* **25,** 49–61 (Jan. 2017).
50. Niewoehner, O. *et al.* Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. *Nature* **548,** 543–548 (Aug. 2017).
51. Bohmert, K. AGO1 defines a novel locus of Arabidopsis controlling leaf development. *The EMBO Journal* **17,** 170–180 (Jan. 1, 1998).
52. Swarts, D. C. *et al.* DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* **507,** 258–261 (Mar. 2014).
53. Swarts, D. C. *et al.* The evolutionary journey of Argonaute proteins. *Nature Structural & Molecular Biology* **21,** 743–753 (Sept. 2014).
54. Elbashir, S. M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes & Development* **15,** 188–200 (Jan. 15, 2001).
55. Lan, F. *et al.* S. pombe LSD1 Homologs Regulate Heterochromatin Propagation and Euchromatic Gene Transcription. *Molecular Cell* **26,** 89–101 (Apr. 2007).
56. Hammond, S. M., Bernstein, E., Beach, D. & Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* **404,** 293–296 (Mar. 2000).
57. Nakanishi, K. When Argonaute takes out the ribonuclease sword. *Journal of Biological Chemistry* **300,** 105499 (Jan. 2024).
58. Lisitskaya, L. *et al.* Programmable RNA targeting by bacterial Argonaute nucleases with unconventional guide binding and cleavage specificity. *Nature Communications* **13,** 4624 (Aug. 8, 2022).
59. Zhen, X. *et al.* Structural basis of antiphage immunity generated by a prokaryotic Argonaute-associated SPARSA system. *Nature Communications* **15,** 450 (Jan. 11, 2024).
60. Pradhan, B. *et al.* Loop-extrusion-mediated plasmid DNA cleavage by the bacterial SMC Wadjet complex. *Molecular Cell* **85,** 107–116.e5 (Jan. 2025).
61. Deep, A. *et al.* The SMC-family Wadjet complex protects bacteria from plasmid transformation by recognition and cleavage of closed-circular DNA. *Molecular Cell* **82,** 4145–4159.e7 (Nov. 2022).
62. Liu, H. W. *et al.* DNA-measuring Wadjet SMC ATPases restrict smaller circular plasmids by DNA cleavage. *Molecular Cell* **82,** 4727–4740.e6 (Dec. 2022).
63. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of Bacteriology* **169,** 5429–5433 (Dec. 1987).
64. Nakata, A., Amemura, M. & Makino, K. Unusual nucleotide arrangement with repeated sequences in the Escherichia coli K-12 chromosome. *Journal of Bacteriology* **171,** 3553–3556 (June 1989).

65. Mojica, F. J. M., Díez-Villaseñor, C., Soria, E. & Juez, G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular Microbiology* **36,** 244–246 (Apr. 2000).
66. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151,** 2551–2561 (Aug. 1, 2005).
67. Haft, D. H., Selengut, J., Mongodin, E. F. & Nelson, K. E. A Guild of 45 CRISPR-Associated (Cas) Protein Families and Multiple CRISPR/Cas Subtypes Exist in Prokaryotic Genomes. *PLoS Computational Biology* **1** (ed Eisen, J. A.) e60 (Nov. 11, 2005).
68. Makarova, K. S. A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Research* **30,** 482–496 (Jan. 15, 2002).
69. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct* **1,** 7 (Dec. 2006).
70. Brouns, S. J. J. *et al.* Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science* **321,** 960–964 (Aug. 15, 2008).
71. Marraffini, L. A. & Sontheimer, E. J. CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. *Science* **322,** 1843–1845 (Dec. 19, 2008).
72. Nuñez, J. K. *et al.* Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nature Structural & Molecular Biology* **21,** 528–534 (June 2014).
73. Yosef, I., Goren, M. G. & Qimron, U. Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Research* **40,** 5569–5576 (July 1, 2012).
74. Fineran, P. C. *et al.* Degenerate target sites mediate rapid primed CRISPR adaptation. *Proceedings of the National Academy of Sciences* **111** (Apr. 22, 2014).
75. Xue, C. & Sashital, D. G. Mechanisms of Type I-E and I-F CRISPR-Cas Systems in *Enterobacteriaceae*. *EcoSal Plus* **8** (eds Slauch, J. M. & Phillips, G.) 10.1128/ecosalplus.ESP–0008–2018 (Dec. 31, 2019).
76. Wang, J. *et al.* Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell* **163,** 840–853 (Nov. 2015).
77. Nuñez, J. K., Lee, A. S. Y., Engelman, A. & Doudna, J. A. Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature* **519,** 193–198 (Mar. 12, 2015).
78. Xiao, Y., Ng, S., Nam, K. H. & Ke, A. How type II CRISPR–Cas establish immunity through Cas1–Cas2-mediated spacer integration. *Nature* **550,** 137–141 (Oct. 2017).
79. Wright, A. V. *et al.* Structures of the CRISPR genome integration complex. *Science* **357,** 1113–1118 (Sept. 15, 2017).
80. Ivančić-Baće, I., Cass, S. D., Wearne, S. J. & Bolt, E. L. Different genome stability proteins underpin primed and naïve adaptation in *E. coli* CRISPR-Cas immunity. *Nucleic Acids Research* **43,** 10821–10830 (Dec. 15, 2015).
81. Achigar, R. *et al.* Ectopic Spacer Acquisition in Streptococcus thermophilus CRISPR3 Array. *Microorganisms* **9,** 512 (Mar. 1, 2021).
82. McGinn, J. & Marraffini, L. A. CRISPR-Cas Systems Optimize Their Immune Response by Specifying the Site of Spacer Integration. *Molecular Cell* **64,** 616–623 (Nov. 2016).
83. Rollie, C., Schneider, S., Brinkmann, A. S., Bolt, E. L. & White, M. F. Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *eLife* **4,** e08716 (Aug. 18, 2015).
84. Fagerlund, R. D. *et al.* Spacer capture and integration by a type I-F Cas1–Cas2-3 CRISPR adaptation complex. *Proceedings of the National Academy of Sciences* **114** (June 27, 2017).
85. Nuñez, J. K., Bai, L., Harrington, L. B., Hinder, T. L. & Doudna, J. A. CRISPR Immunological Memory Requires a Host Factor for Specificity. *Molecular Cell* **62,** 824–833 (June 2016).
86. Dillon, S. C. & Dorman, C. J. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature Reviews Microbiology* **8,** 185–195 (Mar. 2010).
87. Budhathoki, J. B. *et al.* Real-time observation of CRISPR spacer acquisition by Cas1–Cas2 integrase. *Nature Structural & Molecular Biology* **27,** 489–499 (May 2020).

88. Carte, J., Wang, R., Li, H., Terns, R. M. & Terns, M. P. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes & Development* **22,** 3489–3496 (Dec. 15, 2008).
89. Jackson, R. N. *et al.* Crystal structure of the CRISPR RNA–guided surveillance complex from *Escherichia coli*. *Science* **345,** 1473–1479 (Sept. 19, 2014).
90. Xue, C., Zhu, Y., Zhang, X., Shin, Y.-K. & Sashital, D. G. Real-Time Observation of Target Search by the CRISPR Surveillance Complex Cascade. *Cell Reports* **21,** 3717–3727 (Dec. 2017).
91. Guo, T. W. *et al.* Cryo-EM Structures Reveal Mechanism and Inhibition of DNA Targeting by a CRISPR-Cas Surveillance Complex. *Cell* **171,** 414–426.e12 (Oct. 2017).
92. Dillard, K. E. *et al.* Assembly and Translocation of a CRISPR-Cas Primed Acquisition Complex. *Cell* **175,** 934–946.e15 (Nov. 2018).
93. Xiao, Y. *et al.* Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. *Cell* **170,** 48–60.e11 (June 2017).
94. Van Erp, P. B. G. *et al.* Conformational Dynamics of DNA Binding and Cas3 Recruitment by the CRISPR RNA-Guided Cascade Complex. *ACS Chemical Biology* **13,** 481–490 (Feb. 16, 2018).
95. Mulepati, S. & Bailey, S. In Vitro Reconstitution of an Escherichia coli RNA-guided Immune System Reveals Unidirectional, ATP-dependent Degradation of DNA Target. *Journal of Biological Chemistry* **288,** 22184–22192 (Aug. 2013).
96. Sinkunas, T. *et al.* Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system: Cas3 nuclease/helicase. *The EMBO Journal* **30,** 1335–1342 (Apr. 6, 2011).
97. Redding, S. *et al.* Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. *Cell* **163,** 854–865 (Nov. 2015).
98. Swarts, D. C., Mosterd, C., Van Passel, M. W. J. & Brouns, S. J. J. CRISPR Interference Directs Strand Specific Spacer Acquisition. *PLoS ONE* **7** (ed Mokrousov, I.) e35888 (Apr. 27, 2012).
99. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513,** 569–573 (Sept. 2014).
100. Weissman, J., Stoltzfus, A., Westra, E. R. & Johnson, P. L. Avoidance of Self during CRISPR Immunization. *Trends in Microbiology* **28,** 543–553 (July 2020).
101. Gleditzsch, D. *et al.* PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA Biology* **16,** 504–517 (Apr. 3, 2019).
102. Datsenko, K. A. *et al.* Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature Communications* **3,** 945 (July 10, 2012).
103. Sternberg, S. H., Haurwitz, R. E. & Doudna, J. A. Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA* **18,** 661–672 (Apr. 2012).
104. Nussenzweig, P. M., McGinn, J. & Marraffini, L. A. Cas9 Cleavage of Viral Genomes Primes the Acquisition of New Immunological Memories. *Cell Host & Microbe* **26,** 515–526.e6 (Oct. 2019).
105. Heler, R., Wright, A. V., Vucelja, M., Doudna, J. A. & Marraffini, L. A. Spacer Acquisition Rates Determine the Immunological Diversity of the Type II CRISPR-Cas Immune Response. *Cell Host & Microbe* **25,** 242–249.e3 (Feb. 2019).
106. Rusinov, I., Ershova, A., Karyagina, A., Spirin, S. & Alexeevski, A. Lifespan of restriction-modification systems critically affects avoidance of their recognition sites in host genomes. *BMC Genomics* **16,** 1084 (Dec. 2015).
107. Kupczok, A. & Bollback, J. P. Motif depletion in bacteriophages infecting hosts with CRISPR systems. *BMC Genomics* **15,** 663 (2014).
108. Dillingham, M. S. & Kowalczykowski, S. C. RecBCD Enzyme and the Repair of Double-Stranded DNA Breaks. *Microbiology and Molecular Biology Reviews* **72,** 642–671 (Dec. 2008).
109. Radovčić, M. *et al.* CRISPR–Cas adaptation in Escherichia coli requires RecBCD helicase but not nuclease activity, is independent of homologous recombination, and is antagonized by 5 ssDNA exonucleases. *Nucleic Acids Research* **46,** 10173–10183 (Sept. 5, 2018).

110. Handa, N., Bianco, P. R., Baskin, R. J. & Kowalczykowski, S. C. Direct Visualization of RecBCD Movement Reveals Cotranslocation of the RecD Motor after Recognition. *Molecular Cell* **17,** 745–750 (Mar. 2005).
111. Bianco, P. R. & Kowalczykowski, S. C. The recombination hotspot Chi is recognized by the translocating RecBCD enzyme as the single strand of DNA containing the sequence 5-GCTGGTGG-3. *Proceedings of the National Academy of Sciences* **94,** 6706–6711 (June 24, 1997).
112. Anderson, D. G. & Kowalczykowski, S. C. The Translocating RecBCD Enzyme Stimulates Recombination by Directing RecA Protein onto ssDNA in a x-Regulated Manner. *Cell* **90,** 77–86 (July 11, 1997).
113. Dupuis, M.-È., Villion, M., Magadán, A. H. & Moineau, S. CRISPR-Cas and restriction–modification systems are compatible and increase phage resistance. *Nature Communications* **4,** 2087 (July 2, 2013).
114. Levy, A. *et al.* CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **520,** 505–510 (Apr. 23, 2015).
115. Modell, J. W., Jiang, W. & Marraffini, L. A. CRISPR–Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. *Nature* **544,** 101–104 (Apr. 2017).
116. Shiimori, M. *et al.* Role of free DNA ends and protospacer adjacent motifs for CRISPR DNA uptake in Pyrococcus furiosus. *Nucleic Acids Research* **45,** 11281–11294 (Nov. 2, 2017).
117. Wei, Y., Terns, R. M. & Terns, M. P. Cas9 function and host genome sampling in Type II-A CRISPR–Cas adaptation. *Genes & Development* **29,** 356–361 (Feb. 15, 2015).
118. Staals, R. H. J. *et al.* Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR–Cas system. *Nature Communications* **7,** 12853 (Oct. 3, 2016).
119. Stringer, A. M., Cooper, L. A., Kadaba, S., Shrestha, S. & Wade, J. T. *Characterization of Primed Adaptation in the Escherichia coli type I-E CRISPR-Cas System* Feb. 12, 2020.
120. Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A. & Severinov, K. High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in *E. coli*. *RNA Biology* **10,** 716–725 (May 2013).
121. Richter, C. *et al.* Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Research* **42,** 8516–8526 (July 29, 2014).
122. Künne, T. *et al.* Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation. *Molecular Cell* **63,** 852–864 (Sept. 2016).
123. Wimmer, F. & Beisel, C. L. CRISPR-Cas Systems and the Paradox of Self-Targeting Spacers. *Frontiers in Microbiology* **10,** 3078 (Jan. 22, 2020).
124. Westra, E. R. *et al.* Parasite Exposure Drives Selective Evolution of Constitutive versus Inducible Defense. *Current Biology* **25,** 1043–1049 (Apr. 2015).
125. Hommais, F. *et al.* Large-scale monitoring of pleiotropic regulation of gene expression by the prokaryotic nucleoid-associated protein, H-NS. *Molecular Microbiology* **40,** 20–36 (Apr. 2001).
126. Dame, R. T. *et al.* DNA Bridging: a Property Shared among H-NS-Like Proteins. *Journal of Bacteriology* **187,** 1845–1848 (Mar. 2005).
127. Pul, Ü. *et al.* Identification and characterization of *E. coli* CRISPR- *cas* promoters and their silencing by H-NS. *Molecular Microbiology* **75,** 1495–1512 (Mar. 2010).
128. Forrest, D., Warman, E. A., Erkelens, A. M., Dame, R. T. & Grainger, D. C. Xenogeneic silencing strategies in bacteria are dictated by RNA polymerase promiscuity. *Nature Communications* **13,** 1149 (Mar. 3, 2022).
129. Westra, E. R. *et al.* H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. *Molecular Microbiology* **77,** 1380–1393 (Sept. 2010).
130. Yang, C.-D., Chen, Y.-H., Huang, H.-Y., Huang, H.-D. & Tseng, C.-P. CRP represses the CRISPR/Cas system in Escherichia coli: evidence that endogenous CRISPR spacers impede phage P1 replication. *Molecular Microbiology* **92,** 1072–1091 (June 2014).
131. Soberón-Chávez, G., Alcaraz, L. D., Morales, E., Ponce-Soto, G. Y. & Servín-González, L. The Transcriptional Regulators of the CRP Family Regulate Different Essential Bacterial

Functions and Can Be Inherited Vertically and Horizontally. *Frontiers in Microbiology* **8,** 959 (May 31, 2017).
132. Perez-Rodriguez, R. *et al.* Envelope stress is a trigger of CRISPR RNA-mediated DNA silencing in Escherichia coli: Envelope stress triggers CRISPR silencing. *Molecular Microbiology* **79,** 584–599 (Feb. 2011).
133. Nagasawa, S., Ishige, K. & Mizuno, T. Novel Members of the Two-Component Signal Transduction Genes in Escherichia coli1. *The Journal of Biochemistry* **114,** 350–357 (Sept. 1993).
134. Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology* **37,** 67–78 (June 2017).
135. Altae-Tran, H. *et al.* Uncovering the functional diversity of rare CRISPR-Cas systems with deep terascale clustering. *Science* **382,** eadi1910 (Nov. 24, 2023).
136. Makarova, K. S. *et al.* Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology* **18,** 67–83 (Feb. 2020).
137. Couvin, D. *et al.* CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Research* **46,** W246–W251 (W1 July 2, 2018).
138. Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S. & Sternberg, S. H. Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature* **571,** 219–225 (July 2019).
139. Strecker, J. *et al.* RNA-guided DNA insertion with CRISPR-associated transposases. *Science* **365,** 48–53 (July 5, 2019).
140. Koonin, E. V. & Makarova, K. S. Origins and evolution of CRISPR-Cas systems. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374,** 20180087 (May 13, 2019).
141. Goldberg, G. W., Jiang, W., Bikard, D. & Marraffini, L. A. Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature* **514,** 633–637 (Oct. 30, 2014).
142. Samai, P. *et al.* Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell* **161,** 1164–1174 (May 2015).
143. Rouillon, C. *et al.* Structure of the CRISPR Interference Complex CSM Reveals Key Similarities with Cascade. *Molecular Cell* **52,** 124–134 (Oct. 2013).
144. Marraffini, L. A. & Sontheimer, E. J. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463,** 568–571 (Jan. 2010).
145. Elmore, J. R. *et al.* Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR–Cas system. *Genes & Development* **30,** 447–459 (Feb. 15, 2016).
146. Pinilla-Redondo, R. *et al.* Type IV CRISPR–Cas systems are highly diverse and involved in competition between plasmids. *Nucleic Acids Research* **48,** 2000–2012 (Feb. 28, 2020).
147. McRobbie, A.-M. *et al. Staphylococcus aureus* DinG, a helicase that has evolved into a nuclease. *Biochemical Journal* **442,** 77–84 (Feb. 15, 2012).
148. Faure, G. *et al.* CRISPR–Cas in mobile genetic elements: counter-defence and beyond. *Nature Reviews Microbiology* **17,** 513–525 (Aug. 2019).
149. Yang, J. *et al.* Structural basis for the activity of the type VII CRISPR–Cas system. *Nature* **633,** 465–472 (Sept. 12, 2024).
150. Shmakov, S. *et al.* Diversity and evolution of class 2 CRISPR–Cas systems. *Nature Reviews Microbiology* **15,** 169–182 (Mar. 2017).
151. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nature Protocols* **8,** 2281–2308 (Nov. 2013).
152. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471,** 602–607 (Mar. 2011).
153. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337,** 816–821 (Aug. 17, 2012).
154. Jeon, Y. *et al.* Direct observation of DNA target searching and cleavage by CRISPR-Cas12a. *Nature Communications* **9,** 2777 (July 17, 2018).
155. Chen, J. S. *et al.* CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360,** 436–439 (Apr. 27, 2018).

156. Shmakov, S. *et al.* Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Molecular Cell* **60,** 385–397 (Nov. 2015).
157. Meeske, A. J. & Marraffini, L. A. RNA Guide Complementarity Prevents Self-Targeting in Type VI CRISPR Systems. *Molecular Cell* **71,** 791–801.e3 (Sept. 2018).
158. Behler, J. & Hess, W. R. Approaches to study CRISPR RNA biogenesis and the key players involved. *Methods* **172,** 12–26 (Feb. 2020).
159. Özcan, A. *et al.* Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. *Nature Microbiology* **4,** 89–96 (Nov. 5, 2018).
160. Workman, R. E. *et al.* A natural single-guide RNA repurposes Cas9 to autoregulate CRISPR-Cas expression. *Cell* **184,** 675–688.e19 (Feb. 2021).
161. Zegans, M. E. *et al.* Interaction between Bacteriophage DMS3 and Host CRISPR Region Inhibits Group Behaviors of *Pseudomonas aeruginosa*. *Journal of Bacteriology* **191,** 210–219 (Jan. 2009).
162. Heussler, G. E. *et al.* Clustered Regularly Interspaced Short Palindromic Repeat-Dependent, Biofilm-Specific Death of Pseudomonas aeruginosa Mediated by Increased Expression of Phage-Related Genes. *mBio* **6** (ed Sperandio, V.) e00129–15 (July 2015).
163. Tang, B. *et al.* Deletion of cas3 gene in Streptococcus mutans affects biofilm formation and increases fluoride sensitivity. *Archives of Oral Biology* **99,** 190–197 (Mar. 2019).
164. Solbiati, J., Duran-Pinedo, A., Godoy Rocha, F., Gibson, F. C. & Frias-Lopez, J. Virulence of the Pathogen Porphyromonas gingivalis Is Controlled by the CRISPR-Cas Protein Cas3. *mSystems* **5** (ed Bik, H.) e00852–20 (Oct. 27, 2020).
165. Li, M. *et al.* Toxin-antitoxin RNA pairs safeguard CRISPR-Cas systems. *Science* **372,** eabe5601 (Apr. 30, 2021).
166. Roberts, G. A. *et al.* Exploring the DNA mimicry of the Ocr protein of phage T7. *Nucleic Acids Research* **40,** 8129–8143 (Sept. 2012).
167. Walkinshaw, M. *et al.* Structure of Ocr from Bacteriophage T7, a Protein that Mimics B-Form DNA. *Molecular Cell* **9,** 187–194 (Jan. 2002).
168. Isaev, A. *et al.* Phage T7 DNA mimic protein Ocr is a potent inhibitor of BREX defence. *Nucleic Acids Research* **48,** 5397–5406 (June 4, 2020).
169. Went, S. C. *et al.* Structure and rational engineering of the PglX methyltransferase and specificity factor for BREX phage defence. *Nature Communications* **15,** 7236 (Aug. 22, 2024).
170. Chen, K. *et al.* ArdA proteins from different mobile genetic elements can bind to the EcoKI Type I DNA methyltransferase of E. coli K12. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1844,** 505–511 (Mar. 2014).
171. Belogurov, A. A., Delver, E. P. & Rodzevich, O. V. Plasmid pKM101 encodes two nonhomologous antirestriction proteins (ArdA and ArdB) whose expression is controlled by homologous regulatory sequences. *Journal of Bacteriology* **175,** 4843–4850 (Aug. 1993).
172. Serfiotis-Mitsa, D. *et al.* The structure of the KlcA and ArdB proteins reveals a novel fold and antirestriction activity against Type I DNA restriction systems in vivo but not in vitro. *Nucleic Acids Research* **38,** 1723–1737 (Mar. 2010).
173. Court, R., Cook, N., Saikrishnan, K. & Wigley, D. The Crystal Structure of -Gam Protein Suggests a Model for RecBCD Inhibition. *Journal of Molecular Biology* **371,** 25–33 (Aug. 2007).
174. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences* **97,** 6640–6645 (June 6, 2000).
175. Millman, A. *et al.* Bacterial Retrons Function In Anti-Phage Defense. *Cell* **183,** 1551–1561.e12 (Dec. 2020).
176. Bobonis, J. *et al. Bacterial retrons encode tripartite toxin/antitoxin systems* June 22, 2020.
177. Azam, A. H. *et al.* Viruses encode tRNA and anti-retron to evade bacterial immunity.
178. Yasuda, T. Inhibition of Escherichia coli RecA coprotease activities by DinI. *The EMBO Journal* **17,** 3207–3216 (June 1, 1998).
179. Petrova, V., Chitteni-Pattu, S., Drees, J. C., Inman, R. B. & Cox, M. M. An SOS Inhibitor that Binds to Free RecA Protein: The PsiB Protein. *Molecular Cell* **36,** 121–130 (Oct. 2009).

180. Bagdasarian, M. *et al.* An inhibitor of SOS induction, specified by a plasmid locus in Escherichia coli. *Proceedings of the National Academy of Sciences* **83,** 5723–5726 (Aug. 1986).
181. Maslowska, K. H., Makiela-Dzbenska, K. & Fijalkowska, I. J. The SOS system: A complex and tightly regulated response to DNA damage. *Environmental and Molecular Mutagenesis* **60,** 368–384 (May 2019).
182. Binnenkade, L., Teichmann, L. & Thormann, K. M. Iron Triggers So Prophage Induction and Release of Extracellular DNA in Shewanella oneidensis MR-1 Biofilms. *Applied and Environmental Microbiology* **80** (ed Spormann, A. M.) 5304–5316 (Sept. 2014).
183. Leavitt, A. *et al.* Viruses inhibit TIR gcADPR signalling to overcome bacterial defence. *Nature* **611,** 326–331 (Nov. 10, 2022).
184. Yirmiya, E. *et al.* Phages overcome bacterial immunity via diverse anti-defence proteins. *Nature* **625,** 352–359 (Jan. 11, 2024).
185. Samuel, B., Mittelman, K., Croitoru, S. Y., Ben Haim, M. & Burstein, D. Diverse anti-defence systems are encoded in the leading region of plasmids. *Nature* **635,** 186–192 (Nov. 7, 2024).
186. Pinilla-Redondo, R. *et al.* Discovery of multiple anti-CRISPRs highlights anti-defense gene clustering in mobile genetic elements. *Nature Communications* **11,** 5652 (Nov. 6, 2020).
187. Uribe, R. V. *et al.* Discovery and Characterization of Cas9 Inhibitors Disseminated across Seven Bacterial Phyla. *Cell Host & Microbe* **25,** 233–241.e5 (Feb. 2019).
188. Forsberg, K. J. *et al.* Functional metagenomics-guided discovery of potent Cas9 inhibitors in the human microbiome. *eLife* **8,** e46540 (Sept. 10, 2019).
189. Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493,** 429–432 (Jan. 17, 2013).
190. Srikant, S., Guegler, C. K. & Laub, M. T. The evolution of a counter-defense mechanism in a virus constrains its host range. *eLife* **11,** e79549 (Aug. 4, 2022).
191. Bondy-Denomy, J. *et al.* A Unified Resource for Tracking Anti-CRISPR Names. *The CRISPR Journal* **1,** 304–305 (Oct. 2018).
192. Malone, L. M. *et al.* A jumbo phage that forms a nucleus-like structure evades CRISPR–Cas DNA targeting but is vulnerable to type III RNA-based immunity. *Nature Microbiology* **5,** 48–55 (Dec. 9, 2019).
193. Nakamura, M. *et al.* Anti-CRISPR-mediated control of gene editing and synthetic circuits in eukaryotic cells. *Nature Communications* **10,** 194 (Jan. 14, 2019).
194. Dedrick, R. M. *et al.* Phage Therapy of *Mycobacterium* Infections: Compassionate Use of Phages in 20 Patients With Drug-Resistant Mycobacterial Disease. *Clinical Infectious Diseases* **76,** 103–112 (Jan. 6, 2023).
195. Naghavi, M. *et al.* Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. *The Lancet* **404,** 1199–1226 (Sept. 2024).
196. Harrington, L. B. *et al.* A Broad-Spectrum Inhibitor of CRISPR-Cas9. *Cell* **170,** 1224–1233.e15 (Sept. 2017).
197. He, F. *et al.* Anti-CRISPR proteins encoded by archaeal lytic viruses inhibit subtype I-D immunity. *Nature Microbiology* **3,** 461–469 (Mar. 5, 2018).
198. Bondy-Denomy, J. *et al.* Multiple mechanisms for CRISPR–Cas inhibition by anti-CRISPR proteins. *Nature* **526,** 136–139 (Oct. 1, 2015).
199. Osuna, B. A. *et al.* Listeria Phages Induce Cas9 Degradation to Protect Lysogenic Genomes. *Cell Host & Microbe* **28,** 31–40.e9 (July 2020).
200. Knott, G. J. *et al.* Broad-spectrum enzymatic inhibition of CRISPR-Cas12a. *Nature Structural & Molecular Biology* **26,** 315–321 (Apr. 2019).
201. Dong, L. *et al.* An anti-CRISPR protein disables type V Cas12a by acetylation. *Nature Structural & Molecular Biology* **26,** 308–314 (Apr. 2019).
202. Stanley, S. Y. *et al.* Anti-CRISPR-Associated Proteins Are Crucial Repressors of Anti-CRISPR Transcription. *Cell* **178,** 1452–1464.e13 (Sept. 2019).
203. Camara-Wilpert, S. *et al.* Bacteriophages suppress CRISPR–Cas immunity using RNA-based anti-CRISPRs. *Nature* **623,** 601–607 (Nov. 16, 2023).

204. Liu, C. *et al. Phage CRISPR-like regulatory RNAs silence bacterial adaptive and innate immunity* Oct. 30, 2024.
205. Bi, M., Su, W., Li, J. & Mo, X. Insights into the inhibition of protospacer integration via direct interaction between Cas2 and AcrVA5. *Nature Communications* **15,** 3256 (Apr. 16, 2024).
206. Watters, K. E., Fellmann, C., Bai, H. B., Ren, S. M. & Doudna, J. A. Systematic discovery of natural CRISPR-Cas12a inhibitors. *Science* **362,** 236–239 (Oct. 12, 2018).
207. Kang, X. *et al.* Reversible regulation of Cas12a activities by AcrVA5-mediated acetylation and CobB-mediated deacetylation. *Cell Discovery* **8,** 45 (May 17, 2022).
208. Sahakyan, H., Makarova, K. S. & Koonin, E. V. Search for Origins of Anti-CRISPR Proteins by Structure Comparison. *The CRISPR Journal* **6,** 222–231 (June 1, 2023).
209. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology* **2,** 2006.0008 (Jan. 2006).
210. Liles, M. R. *et al.* Recovery, Purification, and Cloning of High-Molecular-Weight DNA from Soil Microorganisms. *Applied and Environmental Microbiology* **74,** 3302–3305 (May 15, 2008).
211. Allen, H. K., Moe, L. A., Rodbumrer, J., Gaarder, A. & Handelsman, J. Functional metagenomics reveals diverse b-lactamases in a remote Alaskan soil. *The ISME Journal* **3,** 243–251 (Oct. 2008).
212. Braithwaite, J. *Development of a High-Throughput Screen for the study of CRISPR Spacer Acquisition* PhD thesis (University of Nottingham, 2021). 227 pp.
213. Amlinger, L., Hoekzema, M., Wagner, E. G. H., Koskiniemi, S. & Lundgren, M. Fluorescent CRISPR Adaptation Reporter for rapid quantification of spacer acquisition. *Scientific Reports* **7,** 10392 (Sept. 4, 2017).
214. Meyer, A. J., Segall-Shapiro, T. H., Glassey, E., Zhang, J. & Voigt, C. A. Escherichia coli "Marionette" strains with 12 highly optimized small-molecule sensors. *Nature Chemical Biology* **15,** 196–204 (Feb. 2019).
215. Cui, L. & Bikard, D. Consequences of Cas9 cleavage in the chromosome of *Escherichia coli*. *Nucleic Acids Research* **44,** 4243–4251 (May 19, 2016).
216. Cherepanov, P. P. & Wackernagel, W. Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158,** 9–14 (Jan. 1995).
217. Shiriaeva, A., Fedorov, I., Vyhovskyi, D. & Severinov, K. Detection of CRISPR adaptation. *Biochemical Society Transactions* **48,** 257–269 (Feb. 28, 2020).
218. Yosef, I. *et al.* Highly active CRISPR-adaptation proteins revealed by a robust enrichment technology. *Nucleic Acids Research* **51,** 7552–7562 (Aug. 11, 2023).
219. Díez-Villaseñor, C., Guzmán, N. M., Almendros, C., García-Martínez, J. & Mojica, F. J. CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of Escherichia coli. *RNA Biology* **10,** 792–802 (May 2013).
220. Görke, B. & Stülke, J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nature Reviews Microbiology* **6,** 613–624 (Aug. 2008).
221. Sezonov, G., Joseleau-Petit, D. & D'Ari, R. *Escherichia coli* Physiology in Luria-Bertani Broth. *Journal of Bacteriology* **189,** 8746–8749 (Dec. 2007).
222. Huisman', O. & Kleckner, N. A New Generalizable Test for Detection of Mutations Affecting TnlO Transposition.
223. Hershfield, V., Boyer, H. W., Yanofsky, C., Lovett, M. A. & Helinski, D. R. Plasmid ColE1 as a Molecular Vehicle for Cloning and Amplification of DNA. *Proceedings of the National Academy of Sciences* **71,** 3455–3459 (Sept. 1974).
224. Fortuin, S., Nel, A. J., Blackburn, J. M. & Soares, N. C. Comparison between the proteome of Escherichia coli single colony and during liquid culture. *Journal of Proteomics* **228,** 103929 (Sept. 2020).
225. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Research* **14,** 1188–1190 (Jan. 2004).
226. Liu, D. & Chalmers, R. Hyperactive mariner transposons are created by mutations that disrupt allosterism and increase the rate of transposon end synapsis. *Nucleic Acids Research* **42,** 2637–2645 (Feb. 1, 2014).

227. Zayed, H., Izsvák, Z., Walisko, O. & Ivics, Z. Development of Hyperactive Sleeping Beauty Transposon Vectors by Mutational Analysis. *Molecular Therapy* **9,** 292–304 (Feb. 2004).
228. Voigt, F. *et al.* Sleeping Beauty transposase structure allows rational design of hyperactive variants for genetic engineering. *Nature Communications* **7,** 11126 (Mar. 30, 2016).
229. Edwards, H. *Characterising Host Factors Required for CRISPR Cas Adaptation in Escherichia coli* PhD thesis (University of Nottingham, 2022). 203 pp.
230. Hofmann, R., Herman, C., Mo, C. Y., Mathai, J. & Marraffini, L. A. *Deep mutational scanning identifies variants of Cas1 and Cas2 that increase spacer acquisition in type II-A CRISPR-Cas systems* Oct. 10, 2024.
231. Simmons, L. A., Foti, J. J., Cohen, S. E. & Walker, G. C. The SOS Regulatory Network. *EcoSal Plus* **3** (ed Slauch, J. M.) (Jan. 12, 2008).
232. Brent, R. & Ptashne, M. Mechanism of action of the lexA gene product. *Proceedings of the National Academy of Sciences* **78,** 4204–4208 (July 1981).
233. Little, J. W., Mount, D. W. & Yanisch-Perron, C. R. Purified lexA protein is a repressor of the recA and lexA genes. *Proc. Natt Acad. Sci.* **78,** 4199–4203 (July 1981).
234. Fernández De Henestrosa, A. R. *et al.* Identification of additional genes belonging to the LexA regulon in *Escherichia coli*. *Molecular Microbiology* **35,** 1560–1572 (Mar. 2000).
235. Radding, C. M. Helical interactions in homologous pairing and strand exchange driven by RecA protein. *Journal of Biological Chemistry* **266,** 5355–5358 (Mar. 1991).
236. Butala, M. *et al.* Interconversion between bound and free conformations of LexA orchestrates the bacterial SOS response. *Nucleic Acids Research* **39,** 6546–6557 (Aug. 2011).
237. Quillardet, P., Huisman, O., D'Ari, R. & Hofnung, M. SOS chromotest, a direct assay of induction of an SOS function in Escherichia coli K-12 to measure genotoxicity. *Proceedings of the National Academy of Sciences* **79,** 5971–5975 (Oct. 1982).
238. Huisman, O., D'Ari, R. & Gottesman, S. Cell-division control in Escherichia coli: specific induction of the SOS function SfiA protein is sufficient to block septation. *Proceedings of the National Academy of Sciences* **81,** 4490–4494 (July 1984).
239. Schoemaker, J. M., Gayda, R. C. & Markovitz, A. Regulation of cell division in Escherichia coli: SOS induction and cellular location of the sulA protein, a key to lon-associated filamentation and death. *Journal of Bacteriology* **158,** 551–561 (May 1984).
240. Trusca, D., Scott, S., Thompson, C. & Bramhill, D. Bacterial SOS Checkpoint Protein SulA Inhibits Polymerization of Purified FtsZ Cell Division Protein. *Journal of Bacteriology* **180,** 3946–3953 (Aug. 1998).
241. Sugino, A., Peebles, C. L., Kreuzer, K. N. & Cozzarelli, N. R. Mechanism of action of nalidixic acid: Purification of *Escherichia coli nalA* gene product and its relationship to DNA gyrase and a novel nicking-closing enzyme. *Proceedings of the National Academy of Sciences* **74,** 4767–4771 (Nov. 1977).
242. Leenay, R. T. *et al.* Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. *Molecular Cell* **62,** 137–147 (Apr. 2016).
243. Wong, T. S., Roccatano, D., Zacharias, M. & Schwaneberg, U. A Statistical Analysis of Random Mutagenesis Methods Used for Directed Protein Evolution. *Journal of Molecular Biology* **355,** 858–871 (Jan. 2006).
244. Nuñez, J. K., Harrington, L. B., Kranzusch, P. J., Engelman, A. N. & Doudna, J. A. Foreign DNA capture during CRISPR–Cas adaptive immunity. *Nature* **527,** 535–538 (Nov. 2015).
245. Goryshin, I. Y. & Reznikoff, W. S. Tn5 in Vitro Transposition. *Journal of Biological Chemistry* **273,** 7367–7374 (Mar. 1998).
246. Huiting, E. *et al.* Bacteriophages inhibit and evade cGAS-like immune function in bacteria. *Cell* **186,** 864–876.e21 (Feb. 2023).
247. Hooton, S. P. T. & Connerton, I. F. Campylobacter jejuni acquire new host-derived CRISPR spacers when in association with bacteriophages harboring a CRISPR-like Cas4 protein. *Frontiers in Microbiology* **5** (Jan. 5, 2015).
248. Philippe, C. *et al.* A truncated anti-CRISPR protein prevents spacer acquisition but not interference. *Nature Communications* **13,** 2802 (May 19, 2022).
249. Varble, A. *et al.* Prophage integration into CRISPR loci enables evasion of antiviral immunity in Streptococcus pyogenes. *Nature Microbiology* **6,** 1516–1525 (Nov. 24, 2021).

250. Marino, N. D., Pinilla-Redondo, R., Csörgő, B. & Bondy-Denomy, J. Anti-CRISPR protein applications: natural brakes for CRISPR-Cas technologies. *Nature Methods* **17,** 471–479 (May 2020).
251. D'Amato, R. *et al.* Anti-CRISPR Anopheles mosquitoes inhibit gene drive spread under challenging behavioural conditions in large cages. *Nature Communications* **15,** 952 (Feb. 1, 2024).
252. Williamson, L. L. *et al.* Intracellular Screen To Identify Metagenomic Clones That Induce or Inhibit a Quorum-Sensing Biosensor. *Applied and Environmental Microbiology* **71,** 6335–6344 (Oct. 2005).
253. Murray, N. E. Type I Restriction Systems: Sophisticated Molecular Machines (a Legacy of Bertani and Weigle). *Microbiology and Molecular Biology Reviews* **64,** 412–434 (June 2000).
254. Dryden, D., Cooper, L. & Murray, N. Purification and characterization of the methyltransferase from the type 1 restriction and modification system of Escherichia coli K12. *Journal of Biological Chemistry* **268,** 13228–13236 (June 1993).
255. Kan, N. C., Lautenberger, J. A., Edgell, M. H. & Hutchison, C. A. The nucleotide sequence recognized by the Escherichia coli K12 restriction and modification enzymes. *Journal of Molecular Biology* **130,** 191–209 (May 1979).
256. Phillips, Z. N., Husna, A.-U., Jennings, M. P., Seib, K. L. & Atack, J. M. Phasevarions of bacterial pathogens – phase-variable epigenetic regulators evolving from restriction–modification systems. *Microbiology* **165,** 917–928 (Sept. 1, 2019).
257. Kisker, C., Kuper, J. & Van Houten, B. Prokaryotic Nucleotide Excision Repair. *Cold Spring Harbor Perspectives in Biology* **5,** a012591–a012591 (Mar. 1, 2013).
258. Biswas, A., Staals, R. H., Morales, S. E., Fineran, P. C. & Brown, C. M. CRISPRDetect: A flexible algorithm to define CRISPR arrays. *BMC Genomics* **17,** 356 (Dec. 2016).
259. Grkovic, S., Brown, M. H. & Skurray, R. A. Regulation of Bacterial Drug Export Systems. *Microbiology and Molecular Biology Reviews* **66,** 671–701 (Dec. 2002).
260. Meier, I., Wray, L. V. & Hillen, W. Differential regulation of the Tn 10-encoded tetracycline resistance genes tetA and tetR by the tandem tet operators 01 and °2. *EMBO* **7,** 567–572 (1988).
261. Van Poppel, N. F., Welagen, J., Duisters, R. F., Vermeulen, A. N. & Schaap, D. Tight control of transcription in Toxoplasma gondii using an alternative tet repressor. *International Journal for Parasitology* **36,** 443–452 (Apr. 2006).
262. Reese, M. G. Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. *Computers & Chemistry* **26,** 51–56 (Dec. 2001).
263. Eaton, R. W. p-Cymene catabolic pathway in Pseudomonas putida F1: cloning and characterization of DNA encoding conversion of p-cymene to p-cumate. *Journal of Bacteriology* **179,** 3171–3180 (May 1997).
264. Shepard, W. *et al.* Insights into the Rrf2 repressor family – the structure of CymR, the global cysteine regulator of *Bacillus subtilis*. *The FEBS Journal* **278,** 2689–2701 (Aug. 2011).
265. Alexander, J. M. *et al.* Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *eLife* **8,** e41769 (May 24, 2019).
266. Mullick, A. *et al.* The cumate gene-switch: a system for regulated expression in mammalian cells. *BMC Biotechnology* **6,** 43 (2006).
267. Tanous, C. *et al.* The CymR Regulator in Complex with the Enzyme CysK Controls Cysteine Metabolism in Bacillus subtilis. *Journal of Biological Chemistry* **283,** 35551–35560 (Dec. 2008).
268. Shipman, S. L., Nivala, J., Macklis, J. D. & Church, G. M. Molecular recordings by directed CRISPR spacer acquisition. *Science* **353,** aaf1175 (July 29, 2016).
269. Yoganand, K. N., Muralidharan, M., Nimkar, S. & Anand, B. Fidelity of prespacer capture and processing is governed by the PAM-mediated interactions of Cas1-2 adaptation complex in CRISPR-Cas type I-E system. *Journal of Biological Chemistry* **294,** 20039–20053 (Dec. 2019).
270. Savitskaya, E. *et al.* Dynamics of *Escherichia coli* type I-E CRISPR spacers over 42 000 years. *Molecular Ecology* **26,** 2019–2026 (Apr. 2017).

271. Yang, Y. *et al.* CRISPR-Cas3 and type I restriction-modification team up against blaKPC-IncF plasmid transfer in Klebsiella pneumoniae. *BMC Microbiology* **24,** 240 (July 3, 2024).
272. Wu, Y. *et al.* Bacterial defense systems exhibit synergistic anti-phage activity. *Cell Host & Microbe* **32,** 557–572.e6 (Apr. 2024).
273. Heler, R. *et al.* Mutations in Cas9 Enhance the Rate of Acquisition of Viral Spacer Sequences during the CRISPR-Cas Immune Response. *Molecular Cell* **65,** 168–175 (Jan. 2017).
274. Kim, H. & Marraffini, L. A. Cas9 interaction with the tracrRNA nexus modulates the repression of type II-A CRISPR-cas genes. *Nucleic Acids Research* **52,** 10595–10606 (Sept. 23, 2024).
275. Navarre, W. W. *et al.* Selective Silencing of Foreign DNA with Low GC Content by the H-NS Protein in *Salmonella*. *Science* **313,** 236–238 (July 14, 2006).
276. Navarre, W. W., McClelland, M., Libby, S. J. & Fang, F. C. Silencing of xenogeneic DNA by H-NS—facilitation of lateral gene transfer in bacteria by a defense system that recognizes foreign DNA. *Genes & Development* **21,** 1456–1471 (June 15, 2007).
277. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of Bacteriology* **177,** 4121–4130 (July 1995).
278. Han, L. *et al.* Realization of Robust and Precise Regulation of Gene Expression by Multiple Sigma Recognizable Artificial Promoters. *Frontiers in Bioengineering and Biotechnology* **8,** 92 (Feb. 19, 2020).
279. Topilina, N. I. & Mills, K. V. Recent advances in in vivo applications of intein-mediated protein splicing. *Mobile DNA* **5,** 5 (2014).
280. Orth, P., Schnappinger, D., Hillen, W., Saenger, W. & Hinrichs, W. Structural basis of gene regulation by the tetracycline inducible Tet repressor–operator system. *Nature Structural Biology* **7,** 215–219.
281. Kwak, Y. M. *et al.* Crystal structure of the VanR transcription factor and the role of its unique -helix in effector recognition. *The FEBS Journal* **285,** 3786–3800 (Oct. 2018).
282. Hu, J. C. Repressor fusions as a tool to study protein–protein interactions. *Structure* **3,** 431–433 (May 1995).
283. Mariño-Ramírez, L., Campbell, L. & Hu, J. C. Screening Peptide/Protein Libraries Fused to the Repressor DNA-Binding Domain in E. coli Cells. *E. coli Gene Expression Protocols* **205.** In collab. with Vaillancourt, P. E., 235–250 (Oct. 22, 2002).
284. Stayrook, S., Jaru-Ampornpan, P., Ni, J., Hochschild, A. & Lewis, M. Crystal structure of the repressor and a model for pairwise cooperative operator binding. *Nature* **452,** 1022–1025 (Apr. 2008).
285. Schmidt, F. *et al.* Noninvasive assessment of gut function using transcriptional recording sentinel cells. *Science* **376,** eabm6038 (May 13, 2022).
286. Murata, M. *et al.* Cell Lysis Directed by SulA in Response to DNA Damage in Escherichia coli. *International Journal of Molecular Sciences* **22,** 4535 (Apr. 26, 2021).
287. Gu Liu, C., Thompson, B. E., Chang, J. D., Min, L. & Maresso, A. W. Construction and characterization of DNA libraries from cultured phages and environmental viromes. *Applied and Environmental Microbiology* **90** (ed Vives, M.) e01171–24 (Oct. 23, 2024).
288. Schmitz, J. E., Daniel, A., Collin, M., Schuch, R. & Fischetti, V. A. Rapid DNA Library Construction for Functional Genomic and Metagenomic Screening. *Applied and Environmental Microbiology* **74,** 1649–1652 (Mar. 2008).
289. Tesson, F. *et al. Exploring the diversity of anti-defense systems across prokaryotes, phages, and mobile genetic elements* Aug. 21, 2024.
290. Chung, K. M., Liau, X. L. & Tang, S. S. Bacteriophages and Their Host Range in Multidrug-Resistant Bacterial Disease Treatment. *Pharmaceuticals* **16,** 1467 (Oct. 16, 2023).
291. Pons, B. J., Van Houte, S., Westra, E. R. & Chevallereau, A. Ecology and evolution of phages encoding anti-CRISPR proteins. *Journal of Molecular Biology* **435,** 167974 (Apr. 2023).