Guide RNA categorization enables target site choice in Tn7-CRISPR-Cas transposons

Michael T. Petassi*, Shan-Chi Hsieh*, and Joseph E. Peters

Department of Microbiology, Cornell University, Ithaca, NY 14853, USA

Corresponding author: joe.peters@cornell.edu

* these authors contributed equally

Running header: CRISPR array categorization

Keywords: CRISPR regulation; Guide RNA categorization; RNA-directed transposition

CRISPR array categorization

Summary

CRISPR-Cas defense systems have been coopted multiple times in nature for guide RNA-directed transposition by Tn7-like elements. Prototypic Tn7 uses dedicated proteins for two targeting pathways, one targeting a neutral and conserved attachment site in the chromosome and a second directing transposition into mobile plasmids facilitating cell-to-cell transfer. We show that Tn7-CRISPR-Cas elements evolved a system of guide RNA categorization to accomplish the same two-pathway lifestyle. Selective regulation of specialized guide RNAs allows long-term memory for access to chromosomal sites upon entry into a new host, while conventional CRISPR features maintain the ability to continually acquire guide RNAs to new plasmid and phage targets. Transposon-encoded guide RNAs are also privatized to be recognized only by the transposon-adapted system working with selective regulation to guard against toxic self-targeting by endogenous CRISPR-Cas defense systems. This information reveals new avenues to engineer guide RNAs for enhanced CRISPR-Cas functionality for genome modification.

Introduction

CRISPR-Cas systems are widespread in bacteria and archaea providing an efficient defense system from predation from bacteriophages and the burden imposed by other mobile elements (Makarova et al., 2020). CRISPR-Cas systems have also been repurposed in nature for other functions that benefit from programmable guide RNA-based sequence recognition (Faure et al., 2019a). One compelling collection of systems involves the occurrence on multiple occasions where a specialized class of transposons called Tn7-like elements have coopted different CRISPR-Cas systems for guide RNA-targeted transposition (Faure et al., 2019b; Peters et al., 2017b). Beyond providing an intriguing example of the propensity of evolution to mix and match useful functions, these systems have also been repurposed in the laboratory as promising tools for genome modification (Klompe et al., 2019a; Strecker et al., 2019). Multiple questions remain with the basic functioning of Tn7-CRISPR-Cas elements, especially with how they interact with other canonical CRISPR-Cas systems. We have discovered mechanisms used by Tn7-CRISPR-Cas elements to categorize guide RNAs, tailoring them for specific functions. This curation of guide RNAs allows regulated pathway choice to recognize highly conserved chromosomal sites while maintaining the ability to continually acquire guide RNAs to new targets.

Tn7-like elements coopted CRISPR-Cas systems on at least four occasions for guide RNA-directed transposition (Peters, 2019). The largest group of Tn7-CRISPR-Cas elements adapted a specific subtype of CRISPR-Cas systems within the type I-F group for guide RNA-directed transposition. Canonical type I-F CRISPR-Cas systems, called type I-F1 systems, use four proteins (Cas8, Cas5, Cas7, and Cas6) as an effector complex called cascade to mature precrRNAs transcribed from a CRISPR array into functional guide RNA complexes. The repeats

CRISPR array categorization

in the CRISPR array encode the guide RNA handles that are bound by Cas proteins, while target specificity is encoded in the spacers. Cas1 and Cas2 are proteins found across CRISPR-Cas systems for acquiring new spacers (also called adaptation) which are inserted into one end of the array adjacent to a special leader region. Type I CRISPR-Cas systems degrade targets recognized by the system using a helicase-nuclease protein Cas3. The Tn7-CRISPR-Cas elements that derived from the canonical I-F1 systems are called I-F3 systems (Makarova et al., 2020). In the I-F3 Tn7-CRISPR-Cas system the Cas8 and Cas5 proteins are naturally fused. These elements lack the Cas1 Cas2/3 spacer acquisition system found in canonical I-F1 systems, something found true in all four families of independently evolved Tn7-CRISPR-Cas elements. Tn7-CRISPR-Cas elements collect new targeting information *in trans* using functions borrowed from canonical CRISPR-Cas systems. (Peters et al., 2017b). This dependency on canonical CRISPR-Cas systems for acquiring new guide RNAs would also render them vulnerable to the targeted degradation capacity of these same systems.

Prototypic Tn7 and Tn7-like elements are known for the control they have over target site selection. Prototypic Tn7 uses five element-encoded proteins for two pathways that direct transposition into different classes of transposition targets (Waddell and Craig, 1988), a specific neutral chromosomal attachment (*att*) site and mobile genetic elements capable of cell-to-cell transfer. A core machinery used for all transposition involves a heteromeric transposase, TnsA+TnsB, for the breaking and joining functions that underlie transposition that is controlled by a AAA+ regulator protein, TnsC (Bainton et al., 1991; Bainton et al., 1993; Stellwagen and Craig, 1998). TnsABC must function with one of two target site selection proteins, TnsD/TniQ or TnsE, to recognize the different types of transposition targets using distinct mechanisms. Tn7 transposition mediated by the sequence specific DNA binding protein TnsD/TniQ allows high

CRISPR array categorization

frequency transposition into its *att* site downstream of the *glmS* gene. Because TnsD/TniQ recognizes the coding region of the essential and highly conserved *glmS* gene, the Tn7 TnsD/TniQ *att*-site pathway virtually ensures a place for the transposon to integrate in a new bacterial host (Mitra et al., 2010). Tn7 or diverged families of Tn7-like elements have been identified in 10-20% of sequenced bacteria with homologs of the TnsA, TnsB, TnsC and TnsD/TniQ proteins (Peters et al., 2017b). Tn7-like elements are presumed to have evolved new TnsD/TniQ DNA binding specificities to recognize other coding-regions observed in the expanded collection of *att* sites in different families of Tn7-like elements (Peters, 2019). Prototypic Tn7 will also preferentially recognize mobile plasmids and bacteriophages capable of transfer between bacteria using a second transposition pathway carried out by TnsABC + TnsE (Finn et al., 2007; Wolkow et al., 1996). TnsE recognizes specific features found during DNA replication associated with plasmid mobilization to facilitate transfer of the element between bacteria (Parks et al., 2009; Peters and Craig, 2001; Shi et al., 2015). Unlike TnsD/TniQ, TnsE is not conserved across the other major families of Tn7-like elements and presumably other mechanisms are used to facilitate transfer of the element to new bacterial hosts.

We present a comprehensive bioinformatic analysis of I-F3 Tn7-CRISPR-Cas elements that reveals mechanisms that allowed the evolution of guide RNA-directed transposition involving categorization of guide RNAs. This updated analysis indicates that all I-F3 Tn7-CRISPR-Cas insertion events are explained by guide RNAs encoded in CRISPR arrays within the element. A form of curation allows the I-F3 elements to maintain different classes of guide RNAs to mirror the two-pathway lifestyle found with prototypic Tn7, but with a guide-RNA-only system. Guide RNA-directed transposition into the chromosome occurs via CRISPR arrays that are under the control of a specialized transcriptional regulation system that directs pathway

CRISPR array categorization

choice or using an atypical CRISPR repeat structure that allows the guide RNA to be private to the Tn7-CRISPR-Cas transposon. The system of guide RNA categorization found in I-F3 Tn7-CRISPR-Cas elements helps explain how they interact with related type I-F CRISPR-Cas systems, such as the ability to tolerate self-targeting guide RNAs that would otherwise allow canonical CRISPR-Cas systems to degrade the host chromosome.

Results

I-F3 Tn7-CRISPR-Cas element targeting is explained by spacers in atypical CRISPR array configurations

We conducted an updated bioinformatics analysis of the major family of Tn7-CRISPR-Cas elements which use the I-F3 variant of CRISPR-Cas systems (Experimental procedures) (Peters, 2019; Peters et al., 2017b). An analysis of over 53,000 genomes from gamma proteobacteria identified 801 Tn7-like elements that encode the type I-F3 CRISPR-Cas system found primarily across the orders *Vibrionales* and *Aeromonadales* in two branches (Figure 1). As noted previously one branch, which we refer to as I-F3a (Figure 1), primarily uses attachment sites adjacent to the *yciA* and *guaC* (IMPDH) genes. A second branch, which we refer to as I-F3b, is primarily found in an attachment site downstream of the *ffs* gene encoding the signal recognition particle (Figure 1). Analysis of the elements in the I-F3b branch revealed a subbranch where elements resided in a different attachment site downstream of the *rsmJ* gene (Figure 1). As part of this analysis we reexamined CRISPR arrays and made a striking finding that altered our understanding of how transposition is targeted across all of the I-F3 elements. Previously it was assumed that Tn7-CRISPR-Cas elements evolved to recognize new *att* sites by adapting the sequence-specific DNA binding ability of the TnsD/TniQ protein. However, we can now show

CRISPR array categorization

that the insertion position of all elements can be explained by guide RNA-directed transposition; for essentially all of the I-F3 elements we can identify a spacer within element-encoded CRISPR arrays that matches a region ~48 bp from the right end of the element (Figures 1, 2a, and 2b)(Supplementary Table S1). In each of these cases the spacer in the array matches the same protospacer in the *yciA*, *guaC*, *ffs*, or *rsmJ* genes (Figure 2b). The specific position of the protospacer that is recognized also appears important. In addition to being at one of the ends of the gene to direct transposition just outside the reading frame, the spacers matching the *yciA*, *guaC*, and *rsmJ* genes are all found in the same reading frame register that aligns the variable wobble position of the codons with every sixth position in the guide RNA, a position known to flip out and not required to match the protospacer (Fineran et al., 2014; Jackson et al., 2014; Mulepati et al., 2014; Zhao et al., 2014). While it has been previously proposed that guide RNA-directed transposition allowed Tn7-CRISPR-Cas elements to primarily target mobile plasmids to allow movement between bacteria (Peters, 2019; Peters et al., 2017b), we now find that transposition directed into the major chromosomal *att* sites used by the element is also guide RNA-directed.

Of additional interest, the spacers that recognize each one of the four major *att* sites are located in a specific position in the element-encoded CRISPR array(s). There are obvious trends with this position and the configuration of the CRISPR arrays that differed in the two major branches of I-F3 elements. In the branch of I-F3a elements, the spacer that matched the *yciA* and *guaC* *att* sites was located after a 70-90 bp gap in the array found immediately downstream of the *tmiQ*, *cas8/5*, *cas7*, *cas6* operon (Figures 2a and 2c)(see below). In these cases where the CRISPR array was not contiguous it was not clear if the array was transcribed as a single precrRNA or if all of the spacers were capable of being matured into functional guide RNA

CRISPR array categorization

complexes (addressed below). In the I-F3b branch of elements that recognizes *att* sites associated with the *ffs* and *rsmJ* genes the *att*-site specific spacer tended to be found in the single CRISPR array located downstream of the *tniQ-cas* operon, but always as the last spacer in the array (Figures 1 and 2a).

Around five percent of Tn7-CRISPR-Cas insertions identified in bacterial genomes are not located in one of the four major *att* sites (Supplemental Table S1). However, even with the insertions outside the major *att* sites we could still identify a spacer in the array that was specific to a protospacer ~48bp from the right end of the element (Figure 1). In all cases, the spacer that directed transposition via a guide RNA into the chromosome was still the last in the array suggesting this placement was significant for the pathway choice with transposition.

Only one transposition event was identified in a plasmid rather than in the chromosome in our analysis. The Tn7-CRISPR-Cas element in *Aeromonas salmonicida* S44 was on a large plasmid in the strain (pS44-1) that was predicted to be a mobile plasmid based on the presence of genes with known roles in conjugal DNA transfer (*tra* genes). Transposition into the site on the plasmid could still be explained by a guide RNA-encoded in the array, however, in this case the spacer was at the leader-proximal position in the array (Supplemental Figure S1a-c). Interestingly, a near-identical Tn7-CRISPR-Cas element found in the *ffs att* site in *Aeromonas hydrophila* AFG_SD03 had a spacer that recognized the same plasmid-encoded gene, but at a different position (Supplemental Figure S1b-c), suggesting a possible plasmid vector important for the dispersal of these elements from *Aeromonas*.

In addition to their distinctive position at the end of or distal to the CRISPR array, the *att* spacers were flanked by repeats with distinctive and novel sequences. New spacers are added to a CRISPR array at the leader-proximal end of the array in a process that duplicates the leader-

CRISPR array categorization

proximal repeat (Xiao et al., 2017). Therefore, although repeats can diverge over time, the first and second repeats start out identical in CRISPR arrays. In I-F3 Tn7-CRISPR-Cas elements the terminal spacer that was used for guide RNA-directed transposition into the chromosome was invariably flanked by repeats that were highly diverged from the leader-proximal repeat (Figure 2c and Supplemental Table S1). We call the highly diverged repeats, “atypical” repeats, and a guide RNA formed from these sequences atypical guide RNAs. Based on the high divergence from the leader-proximal repeat, it was not clear if the diverged repeats were still functional for guide RNA-directed transposition or if the diversification had rendered them nonfunctional.

Highly diverged atypical repeat-spacer units can form functional guide RNA complexes

To help understand the unique nature of the CRISPR array structure found in I-F3 Tn7-CRISPR-Cas elements we established guide RNA-directed transposition in a heterologous and genetically tractable system, *E. coli*. Previous studies trying to establish Tn7-CRISPR-Cas transposition in a heterologous host relied heavily on indirect PCR-based techniques to assess transposition, techniques that are vulnerable to artifacts (Rice et al., 2020; Strecker et al., 2020). To get a complete picture of Tn7-CRISPR-Cas transposition, we used an assay that monitored full transposition events. In this assay, candidate transposition targets to be tested were crossed by homologous recombination onto a conjugal F plasmid. Transposition events were monitored by mating the conjugal plasmid into a tester strain after inducing expression of the components of the guide RNA-directed transposition system (Supplemental Figure S2a).

We used informatics to identify Tn7-CRISPR-Cas elements that show clear evidence of robust crRNA-guided transposition, reasoning that such systems would likely show high activity in the laboratory setting. Elements identified in a plasmid in *Aeromonas salmonicida* S44 and in

CRISPR array categorization

the *ffs* attachment site in *Aeromonas hydrophila* AFG_SD03 were of particular interest because they were near-identical, but found in different species at distinct insertion points suggesting they were functional recently (Supplemental Figure S1a). For the transposition (Tns) and Cas proteins we utilized a coding sequence configuration we predicted to be highly active for transposition by looking for consensus across multiple elements found in *Aeromonas* (Experimental procedures). To monitor transposition using a modular system, we constructed the *tnsABC*, *tniQ-cas8/5,7,6*, and the CRISPR array in separate expression vectors. A mini Tn7-CRISPR-Cas element was constructed with *cis*-acting transposon end sequences predicted by putative TnsB-binding sites (Peters, 2015) flanking an antibiotic resistance determinant that was situated in the chromosome, the donor site for transposition in our assays.

Initially we analyzed candidate guide RNAs produced from the wild-type configuration of the CRISPR array found in *A. salmonicida* S44. In this configuration the leader-proximal spacer was a perfect match to a mobile plasmid-encoded gene from the native host and the second/terminal spacer had a degenerate match to the *ffs* protospacer with 10 mismatches (Figure 3a, Supplemental Figure S1c). Some of the mismatches between the spacer and the protospacer in the target were at every sixth position and therefore would not impact recognition of the *ffs* guide RNA target (Supplemental Figure S1c). Monitoring transposition following expression of the native array configuration confirmed that functional guide RNAs were produced both from the spacer with the canonical repeat structure at the leader-proximal position and the terminal spacer flanked by highly diverged atypical repeats (Figure 3b). Interestingly, guide RNA-mediated transposition occurred at a higher frequency with the *ffs*-specific spacer even though it contained mismatches and was flanked by atypical repeats (Figure 3b).

CRISPR array categorization

To test the individual contributions of the spacer, protospacer, and repeat sequences, we designed CRISPR array constructs with leader-proximal (typical) or terminal (atypical) flanking repeats as a single guide RNA expression construct and tested various native and synthetic spacer sequences individually. Not only were the guide RNAs with the atypical repeats functional, but also consistently allowed a higher frequency of transposition when compared to the typical repeat when tested with three different spacers (Figure 3c). Additionally the *ffs*-specific spacer showed a higher frequency of transposition than the spacer directed at the plasmid target even though the plasmid spacer had a perfect match to its target and the *ffs* spacer had 10 mismatches to its target (several that were not at the sixth positions that are predicted to be flipped out) (Figure 3c). Altering the native *ffs*-specific spacer so that it was a perfect match to the *ffs* protospacer consistently allowed a modestly higher frequency of transposition (Figure 3c-d).

Guide RNA complexes were also designed using spacers matching different positions in *lacZ* (Figures 3d-e). We found that transposition frequency varied as much as 10-fold with different spacers, even though the sequences recognized all had the same candidate PAM sequence, a result that was not explained by the DNA strand that was targeted in the highly expressed *lacZ* gene (Figures 3d-e). However, regardless of the spacer tested, a modestly higher transposition frequency was consistently found with guide RNAs with the atypical repeats when compared with typical repeats in the Tn7-CRISPR-Cas system from *A. salmonicida* S44 (Figures 3c-d). These experiments confirmed that a functional guide RNA complex could be produced from atypical repeats and hinted that the functionality of these complexes may show important differences from typical repeats. We found that multiple different positions could also be

CRISPR array categorization

targeted in the *E. coli* chromosome using guide RNA-directed transposition supporting a view that this was not a plasmid-specific process (Supplemental Figure S1d-e).

Atypical repeats form functionally distinct guide RNA complexes with the Tn7-CRISPR-Cas system from *A. salmonicida* S44

Our experiments suggested that the guide RNA produced from atypical repeats was functional and appeared to allow enhanced transposition activity with the Tn7-CRISPR-Cas system derived from *A. salmonicida* S44. To get a better understanding of the relevance of differences in the repeat sequences, we compared the sequences of the leader-proximal typical repeats with the atypical repeats flanking the terminal spacers to look for common trends across the two branches (Figure 4a). In both branches there were common trends in the final repeat encoding the 3' handle of the guide RNA with a tendency to lose the typical GTG (positions 1-3), a loss of conservation of the region cleaved from the final guide RNA (positions 21-28), and a general enrichment for adenines in the loop (Figure 4a). Functional differences from changes with the typical and atypical repeats were examined with the *A. salmonicida* S44 element by making changes to the repeat regions encoding the 5' and 3' handles of guide RNAs (Figure 4b). While not subject to an extensive analysis, changing the GUG region to an AUU in the 3' handle of a typical repeat resulted in a significant change in the frequency of guide RNA-directed transposition (Figure 4b). Changing the AUU to a GUG in an atypical array did not show the same significant effect (Figure 4b).

Previous work with a different I-F3 Tn7-CRISPR-Cas systems found with the Tn6677 element from *Vibrio cholerae* HE-45 indicated that guide RNA complexes could be directed to programmed target sites in *E. coli* using guide RNAs (Klompe et al., 2019a). The Tn6677

CRISPR array categorization

element is in the I-F3a branch of elements and provided a good point of comparison for understanding differences between the two branches of I-F3 Tn7-CRISPR-Cas elements (Figure 1). Tn6677 naturally resides in the *att* site downstream of *guaC* and consistent with the trends we identified above, this element carries the *att* site targeting spacer in a noncontiguous array with an atypical repeat structure (Figure 2). The Tns, Cas, and CRISPR array modules from Tn6677 were constructed under lactose and arabinose expression systems and tested in the transposition assay used for the *A. salmonicida* S44 element above (Supplemental Figure S2a). Like the result found with the *A. salmonicida* S44 transposon, we found that transposition into the native *guaC* attachment site used by Tn6677 required the *guaC*-specific guide RNA encoded in the array and that the atypical repeats in Tn6677 were also functional (Figure 4c). We tested if there was any functional relevance with changes with the typical verses atypical array structure with this element from the I-F3a branch of Tn7-CRISPR-Cas transposons. However, with Tn6677 a similar frequency of transposition was found with the typical and atypical arrays (Figure 4c) or with modest changes in the typical and atypical repeats (Figure 4d). Naturally occurring Tn7-like and Tn7-CRISPR-Cas elements control the left to right orientation with which they insert. Consistent with previous work, Tn6677 seemed to be somewhat relaxed for orientation control (Supplemental Figure S2b). The Tn7-CRISPR-Cas element from *A. salmonicida* S44 showed the expected bias for one orientation found with canonical Tn7 and found naturally when 24 independent insertions were analyzed (Supplemental Figure S2b)(Peters, 2015). For the *A. salmonicida* element, we were able to confirm the target site duplication expected when full transposition occurs and the slight wobble found in previous work with the other I-F3 system (Supplemental Figure S2c)(Klompe et al., 2019a).

CRISPR array categorization

Atypical repeats form private guide RNA complexes with the Tn7-CRISPR-Cas system from *A. salmonicida* S44

A question not previously addressed with Tn7-CRISPR/Cas systems involves possible cross-talk between CRISPR arrays with other type I-F CRISPR-Cas systems. If the CRISPR array from a Tn7-CRISPR-Cas element with the guide RNAs specific to the chromosome could be used by a standard I-F1 system, the chromosome *att* site would be a target for degradation. This could limit the spread of I-F3 Tn7-CRISPR-Cas elements if it entered a new host that encoded a standard I-F1 CRISPR-Cas system. We investigated if the typical and atypical CRISPR arrays from I-F3 CRISPR-Cas systems could be accessed by a I-F1 system using a transformation-based assay previously established by others (Chowdhury et al., 2017a). In our work testing the *P. aeruginosa* system we co-expressed the Cas proteins and a single spacer CRISPR array with arabinose and T7 expression systems (Vorontsova et al., 2015). Repeats from the type I-F1 system from *P. aeruginosa*, the type I-F3b system from *A. salmonicida*, or the type I-F3a *V. cholerae* Tn6677 were examined using a transformation efficiency assay examining plasmids with and without a protospacer. In control experiments we observed robust interference using the I-F1 CRISPR-Cas system from *P. aeruginosa* PA40 (Figure 5). Transformation was decreased over two orders of magnitude with the plasmid encoding a protospacer (spacer match) compared to a plasmid that lacked the protospacer. Similarly, the typical repeat of the I-F3b system from *A. salmonicida* S44 also allowed robust interference with the plasmid transformation assay. This was not unexpected because the repeats from the canonical I-F1 and I-F3 Tn7-CRISPR-Cas systems are similar (Supplemental Figure S3) and it is likely that the I-F3 Tn7-CRISPR-Cas systems rely on standard I-F1 systems for spacer acquisition (see Discussion). A different result was found when the spacer was situated with atypical repeats from the I-F3b system found with

CRISPR array categorization

the *A. salmonicida* Tn7-CRISPR-Cas system; guide RNA complexes formed with atypical repeats were not functional for interference in the plasmid transformation assay (Figure 5). The compromised use with the atypical repeats for interference was in contrast to the enhanced use we found for guide RNA-directed transposition with the I-F3b system from *A. salmonicida* (Figures 3 and 4). This result is of particular interest because it provides a mechanism that would allow chromosomal-targeting spacers to be tolerated in hosts with standard I-F1 CRISPR-Cas systems by allowing them to remain private to the I-F3b system. This privatization was absent in the Tn6677 I-F3a system from *V. cholerae*. With the I-F3a *Vc* system, robust interference was found with either the typical repeats or the highly diverged atypical array from this element. The results below suggest that I-F3a Tn7-CRISPR-Cas elements instead use a separate transcription network to tolerate self-targeting spacers (see below).

I-F3 elements utilize XRE-family transcriptional regulators to regulate CRISPR-Cas components

To better understand I-F3 Tn7-CRISPR-Cas element dissemination, we searched for genes conserved among diverse members of this group. One of the other genes found conserved across I-F3 Tn7-CRISPR-Cas elements were predicted XRE-family transcriptional regulators. The *xre* gene resides at a conserved position between the *tnsABC* and *tniQ-cas8/5,7,6* operons in nearly all I-F3 elements (Figure 2a). While each of the two branches of I-F3 elements have *xre* genes, the predicted regulatory gene in each branch segregated with phylogenetically distinct families of controller (C) proteins associated with restriction-modification systems. I-F3a elements have a 68 amino acid Xre protein related to *C.AhdI* and I-F3b elements have a ~100 amino acid Xre protein related to *C.Csp231I* (Supplemental Figure S4a). Candidate regulatory features could

CRISPR array categorization

also be identified with the *tmiQ-cas* and CRISPR arrays based on homology with the previously established systems (Supplemental Figure S4b, see below)(Streeter et al., 2004).

We analyzed putative promoter regions in I-F3a elements and discovered candidate sites for Xre-mediated regulation upstream of *xre* as well as directly upstream of the *att*-targeting spacer in Tn6677 and other members of this branch of elements (Figure 6a). To confirm physical interaction of element bound Xre regulators with these potential binding motifs, we purified Xre from two elements in the I-F3a branch, *V. cholerae* HE-45 Tn6677 (Vc) and *V. parahaemolyticus* RIMD221063 (Vp). Both Xre proteins showed sequence-specific interaction with the candidate promoter sequences upstream of *xre* and upstream of the *att*-targeting spacer as demonstrated by electromobility shift assays (Figure 6c). A functional role for this interaction was shown by a LacZ reporter assay testing expression from promoters pXre or pArray2 for both I-F3a elements tested in the absence of Xre or with varying expression levels using a titratable pBAD promoter from a plasmid. Xre was found to autoregulate its own pXre promoter, which allowed minimal transcription without Xre, was activated by low amounts of Xre, and repressed as the expression of Xre increased (Figure 6e). Meanwhile the pArray2 promoter identified for the *att*-targeting spacer was highly expressed when Xre was not present and increasingly repressed with increasing amounts of Xre induction (Figure 6e). As explained in greater detail in the Discussion, the regulation we observed would allow a burst of the atypical guide RNA that is specific to the *guaC* or *yciA att* sites with I-F3b elements upon entry into a new host via zygotic induction with a dual regulatory role as an inducer at low Xre concentration and a strong repressor at higher concentrations.

I-F3b elements were also surveyed for inverted repeat motifs to investigate the functional role of the conserved *C.Csp231I*-like XRE regulator. Like I-F3a elements, conserved motifs

CRISPR array categorization

were found in the promoter region of *xre* that were nearly identical to those used by *C.Csp231I* (Figure 6b, Supplemental Figure S4b)(McGeehan et al., 2011). Unlike the I-F3a elements, the conserved motif could not be identified upstream of the CRISPR arrays with the I-F3b elements, and instead we found a single copy of this motif upstream of the *tmiQ-cas8/5,7,6* operon (Figures 2a and 6b). To confirm a role for these motifs in regulation, we selected two I-F3b elements from *A. salmonicida* S44 (As) and *Vibrio* sp. 10N.286.45.B6 (VB6), and confirmed binding to these sites with EMSA and purified Xre regulator. Binding to the two predicted motifs in the upstream region of Xre could be visualized as two separately migrating species (Figure 6d). Mutating the *xre*-proximal regulatory motif weakened interaction as demonstrated by the higher concentration of protein required to achieve a full mobility shift (Figure 6d). We additionally visualized interaction with the motif upstream of *tmiQ-cas8/5,7,6* and confirmed the sequence-specific nature of binding by utilizing a mutated motif which weakened interaction. LacZ reporter assays were again used to confirm a functional role in regulation. Xre regulator was shown to act as a repressor of its own pXre promoter (Figure 6f). Interestingly, mutation of the proximal binding site which impaired binding *in vitro* resulted in Xre regulator instead acting as an activator, suggesting interaction with the distal site activates transcription while interaction with the proximal site represses it (Figure 6f). Similar to the result with the I-F3a elements, we found that the XRE regulator was able to repress *tmiQ-cas8/5,7,6* expression and this repression was impaired by mutation of the conserved binding motif (Figure 6f). As elaborated on in the discussion, Xre regulation in the I-F3b systems would allow high expression of the TniQ-Cascade guide RNA system when the element entered a new host, providing a mechanism for encouraging transposition into a conserved host chromosome *att* site.

CRISPR array categorization

Discussion

Our findings have revealed that the major family of Tn7-CRISPR-Cas systems have evolved to be dedicated for guide RNA-directed transposition. We find that all transposition can be accounted for by spacers in the element-encoded CRISPR array (Figure 1) and that transposition with members of both major branches of these elements require guide RNAs for *att*-targeted transposition (Figures 3 and 4). We also show that the spacers used to target chromosomal sites show certain characteristics; they are flanked by repeats that are highly diverged from those normally used for the process of spacer acquisition and reside at the last position in a CRISPR array (Figures 1 and 2). The repeat divergence appears to be specifically beneficial for a type I-F3b element from *A. salmonicida* S44, these atypical guide RNA complexes are unusable for interference by a canonical I-F1 system (Figure 5), while allowing for a higher level of transposition than found with the typical guide RNAs (Figure 3). We show that each of the major branches of I-F3 Tn7-CRISPR-Cas systems have separately coopted XRE family regulators (Supplementary Figure S4). XRE family regulators are common in mobile elements allowing a coordinated response upon entry into a new host. In the adaption of this system to I-F3 Tn7-CRISPR-Cas elements the regulatory features will result in a burst of expression specifically favoring guide RNAs directing transposition into chromosomal *att* sites (Figure 6). Regulation of the functional diverged guide RNAs in I-F3 elements allows a private transcription network that is likely important for I-F3 elements to tolerate self-targeting spacers.

CRISPR array categorization

Atypical arrays can form private guide RNA complexes in the I-F3b system from *A.*

***salmonicida* S44**

A significant consequence of I-F3 Tn7-CRISPR-Cas elements maintaining a guide RNA-only directed lifestyle is the hazard of canonical I-F1 CRISPR-Cas systems using the self-targeting guide RNAs that are encoded in the element. Self-targeting spacers that are accessed by other canonical CRISPR-Cas systems would be a liability whenever the element entered a host that already had this family of CRISPR-Cas systems. To this point, we find that a I-F1 system can use typical guide RNAs encoded in I-F3 CRISPR arrays in interference (Figure 5). Standard I-F1 systems are found in *Vibrionales* and *Aeromonadales* and with our analysis we found multiple specific examples where I-F3 Tn7-CRISPR-Cas elements reside in strains that have standard I-F1 systems (Supplemental Table S1). Self-targeting spacers have been identified in the work of others and have even been used as a tool to help identify anti-CRISPR systems (Borges et al., 2017). However, established members of anti-CRISPR systems are not widely conserved in elements with Tn7-CRISPR-Cas systems in our analysis (Wiegand et al., 2020).

In the case of the I-F3b system from *A. salmonicida* S44, the atypical repeat appears to be a specific adaptation that allows a higher frequency of guide RNA targeted transposition (Figures 3 and 4) and privatization from a canonical I-F1 interference system (Figure 5). It is unresolved from our work how spacers in the I-F3 CRISPR arrays can be private from the I-F1 system, but could result from poor processing by the canonical system. A better understanding of what allows atypical guide RNAs to function differently in the I-F3b element derived from *A. salmonicida* may offer avenues for optimizing guide RNA-directed transposition in future work by modifying the repeats. Interestingly, the type I-F3a Tn7-CRISPR-Cas system from Tn6677 did not show enhanced transposition with the atypical array found in this system; the frequency

CRISPR array categorization

of transposition was the same with the typical and atypical repeat, even though it was highly diverged (Figures 4c and 4d). The highly diverged repeats from Tn6677 were also readily used by a canonical I-F1 interference system (Figure 5), suggesting that other mechanisms are required for tolerating self-targeting arrays in this element (see below). Highly diverged repeats around a certain spacer in arrays would provide a benefit by helping to protect against any method of spacer loss that depends on homology.

We suspect that there are multiple distinct ways that I-F3 Cas proteins have adapted to recognize specialized atypical repeats based on our bioinformatics. Interestingly, we notice that in one subbranch within the I-F3b elements the final spacer is reduced by 10 to 12 base pairs in length (Figure 1 and Supplemental Figure S5). These smaller spacers produce functional guide RNAs as predicted by the commensurate natural repositioning of the insertion closer to the protospacer (Supplemental Figure S5). Previous work in closely related CRISPR-Cas systems suggests that guide RNAs of this length will not be functional for targeting transposition nor robust interference (Klompe et al., 2019a; Kuznedelov et al., 2016). However, the ability of the I-F3b systems to accommodate shorter guide RNAs could provide another mechanism of privatization from other I-F CRISPR-Cas systems. Intriguingly, naturally occurring minimal type I-F2 CRISPR-Cas systems tested in the laboratory are not functional for interference with similarly truncated guide RNAs, but can still form complexes capable of forming R-loops to matching protospacers (Gleditsch et al., 2016).

XRE family proteins were coopted in different ways in the two branches of I-F3 CRISPR-Cas elements

CRISPR array categorization

XRE family regulators are common on mobile elements allowing a coordinated regulatory program to facilitate adaptation upon entry into a new host. These systems take advantage of “zygotic” induction where the gene program is reset when they enter a new host that lack the Xre protein. This type of regulatory scheme was investigated in restriction-modification systems which utilize XRE family regulator C proteins to control temporal regulation of methylation and restriction (Rodic et al., 2017). When the restriction-modification system enters a naïve host on a mobile element, methylation of host DNA must precede endonuclease expression to prevent self-degradation. Methylation activity must remain at a modest level to methylate newly replicated host DNA without protecting subsequent invading DNA elements while the endonuclease must be expressed to defend against other mobile elements. Based on previous work with related *C.AhdI* (Bogdanova et al., 2008; McGeehan et al., 2006) and *C.Csp231I* (Rezula et al., 2016; Shevtsov et al., 2015) proteins and our *in vivo* and *in vitro* analysis of the promoters and Xre regulators found in I-F3 systems, we can make predictions to suggest why these are core conserved features in Tn7-CRISPR-Cas elements.

Both branches of I-F3 elements likely benefit from high expression of Xre-controlled promoters when they enter a new host that lacks the repressor protein. Using the I-F3a elements tested in our experiments as an example, the *att*-targeting spacer, *guaC* or *yicA*, becomes derepressed upon entering a new host. Its high expression level would program the Tn7-CRISPR-Cas element to transpose into the neutral chromosomal *att* sites. As the Xre repressor increases in concentration the promoter expressing the *att*-targeting spacer is predicted to be shut off based on the tight repression found in our experiments even under the lowest induction level (Figure 6e). Xre regulation, which includes activation at low concentrations (Figure 6e), is predicted to delay repression of *att*-targeting spacer to allow transposition to occur while also,

CRISPR array categorization

presumably, ensuring a stable Xre concentration after transposition. The ability to categorize guide RNAs with differential expression from two arrays would allow high expression on entry, but repression upon establishment of the element, thereby helping guard against potential self-targeting with this guide RNA in the presence of endogenous I-F1 systems. Presumably a basal expression level from the primary CRISPR array would allow the production of sufficient guide RNA complexes to recognize mobile elements to allow horizontal transfer of I-F3a elements into new hosts. Careful controlled expression may be especially important in I-F3a elements, because both typical and atypical could be used for interference by a I-F1 system (Figure 5).

In the I-F3b elements, the TniQ-Cascade operon itself is subject to Xre regulation. Upon entry into a new host, high expression of the I-F3b machinery is expected until sufficient repressor protein accumulates (Figure 6). Presumably this bolus of expression would facilitate transposition into the chromosomal *ffs* or *rsmJ att* site. In CRISPR-Cas systems, the total number of effective guide RNA complexes will be a function of the amount of cascade components as related to the number of different spacers and their expression levels (Martynov et al., 2017). Therefore, the I-F3b systems like the ones found with *A. salamonica* S44 may draw additional benefit by swamping the cell with effector complexes to overwhelm any I-F1 systems potentially found in a new host. While I-F3b elements do not rely on transcriptional regulation to privatize the self-targeting spacer from any endogenous CRISPR-Cas system found in a new host, they have nonetheless coopted a similar regulatory pathway to activate transposition function in a new host.

The two-pathway lifestyle is maintained in diverse Tn7-like elements using different mechanisms

CRISPR array categorization

The prototypic Tn7 element maintains two transposition pathways which are directed by two dedicated target site selection protein (Peters, 2014). We find that the type I-F3 Tn7-CRISPR-Cas elements use a form of spacer categorization involving repeat diversification and transcriptional regulation as mechanisms to establish the same two-pathway lifestyle.

The outlier element found in *Parashewanella curva* C51 provides an interesting alternative hybrid two-pathway strategy from the other elements using the I-F3 CRISPR-Cas system (Figure 1). Unlike the other I-F3 family elements indicated in Figure 1, this element does not have a spacer that could form a guide RNA complex to recognize where it inserted in the chromosome. Interestingly, this element is from a small group of representatives that encode two TniQ proteins. One TniQ is associated with the Cascade operon and the second is similar to a TniQ protein found in other Tn7-like elements lacking CRISPR-Cas components that reside in an attachment site downstream of the *parE* gene encoding the essential and conserved topoisomerase IV, subunit B (Supplemental Figure S6). These elements are predicted to use guide RNA-directed transposition to recognize mobile elements, but presumably continue to rely on the mechanism used by prototypic Tn7 and most other Tn7-like elements (Peters, 2014), in this case with the TniQ protein recognizing the *parE* gene coding region. In two previously identified instances where type I-B CRISPR-Cas systems were captured by other Tn7-like members the elements have a similar configuration with two *tniQ* genes and likely function in a similar way to allow a two pathway life-style (Peters et al., 2017b). It is unclear how Tn7-CRISPR-Cas elements with a type V-K CRISPR-Cas systems control the use of two pathways given these elements only encode a single *tniQ* gene (Faure et al., 2019b).

The guide RNA complexes used by the major *att* pathways recognizing the *guaC*, *yciA*, *ffs*, and *rsmJ* have certain characteristics that favor their maintenance and utility across bacterial

CRISPR array categorization

species. In addition to recognizing a site close to the end of a gene the specific sequence recognized also always occurs in a specific register with the reading frame where the sixth position aligns with the wobble position of codons (Figure 2b). Given that every sixth position in the guide RNA does not make base pair contacts with protospacer DNA, this placement of the guide RNA provides a partial safeguard to differences in protospacers that can be encountered from codon bias in the genes that are recognized for the major attachment sites. While presumably any guide RNA would benefit from the ability to accommodate the wobble position, this is more important for the lifestyle of Tn7-CRISPR-Cas elements that rely on identifying a safe integration site whenever they enter a new host and one that must serve as a form of long-term memory. Similar considerations may be important in other cases where CRISPR-Cas has been coopted for alternative activities involving self-targeting guide RNAs (Faure et al., 2019a). Presumably there is a barrier to initially obtaining and fixing these new spacers because we show that spacers with typical repeats can be accessed by a canonical I-F1 system (Figure 5). This may help explain why we only found one additional example outside of the *guaC*, *yicA*, *ffs* and *rsmJ*, sites where the same protospacer was recognized in more than one Tn7-CRISPR-Cas element; a spacer that recognizes a AraC-family protein of unknown function was found four times (Supplementary Figure S5b and Supplemental Table S1).

It will be interesting to learn how the adaptations found in I-F3 Tn7-CRISPR-Cas systems extend to canonical CRISPR-Cas systems and in other cases where CRISPR-Cas systems have been coopted for new functions. The examples shown here join numerous others where it is now clear that CRISPR-Cas activity is regulated to offset potential negative impacts of these potent systems (Hoyland-Kroghsbo et al., 2017; Patterson et al., 2016; Westra et al., 2010). Transposition levels with the I-F3b system developed in our work is robust, occurring in

CRISPR array categorization

as much as one third of the cell population, suggesting that it should be a good candidate for future application-based work. The new information with atypical guide RNAs also suggests additional avenues for altering guide RNA activity for practical applications in Tn7-CRISPR-Cas systems or potentially CRISPR-Cas systems in general.

Acknowledgements

This work was supported by National Institutes of Health Grant R01 GM129118. Shan-Chi Hsieh was supported by a training fellowship from the Taiwan government. We thank Nancy Craig and Ailong Ke for comments on the manuscript. Cornell University has filed patent applications related to this work with JEP as inventor. We are grateful for reagents provided by John Cronan, Tobias Doerr, Elizabeth Raleigh, Ailong Ke, Addgene, and the *E. coli* Genetic Stock Center.

CRISPR array categorization

Experimental Procedures

Bacterial Strain and Plasmids

Strains and plasmids used in this study are listed in Tables 1 and 2. Primers and synthesized gene fragments are listed in Tables 3 and 4. Standard molecular cloning techniques were used to make the vectors described below using vendor instructions.

pMTP112 was constructed by ligating gBlock1 into the NotI site of pMS26 following digestion with NotI. The clone used has *A. salmonicida* left end proximal to the Tn7 right end. pMTP113 was constructed by assembling two PCR products amplified from pSL0527 (pDonor) (JEP1858+JEP1859 and JEP1860+JEP1861), one PCR product amplified from gBlock1 (JEP1862+JEP1863) and pMS26 digested with NotI using NEBuilder Hifi (NEB). pMTP114 was constructed by assembling two PCR products amplified from F plasmid (JEP1398+1340 and JEP1341+1399, GenBank: AP001918.1), one PCR product amplified from pMTP150 (JEP1343+JEP1344), one PCR product amplified from pBAD322S (JEP1345+JEP1346, GenBank: DQ131584.1) and pTSC29 digested with EcoRV using NEBuilder Hifi. pMTP115 was constructed by inserting a PCR product amplified from EMG2 (JEP1663+JEP1664, GenBank: U00096.3) into pMTP114 following digestion with BsaI using golden gate cloning (Engler et al., 2008). pMTP116 was constructed by inserting annealed oligos (JEP1485+JEP1486) into pMTP114 following digestion with BsaI using golden gate cloning. pMTP117 was constructed by inserting annealed and extended oligos (JEP1481+JEP1482) into pMTP114 following digestion with BsaI using golden gate cloning. pMTP118 was constructed by inserting annealed and extended oligos (JEP1878+JEP1879) into pMTP114 following digestion with BsaI using golden gate cloning. pMTP130 was constructed by assembling gBlock2, gBlock3 and a PCR product amplified from pTA106 (JEP1146+JEP1467) digested by DraII with

CRISPR array categorization

577 3,800 bp fragment gel purified using NEBuilder Hifi. pMTP140 was constructed by assembling
578 gBlock4, gBlock5, gBlock6 and pBAD322G digested with NcoI and HindIII using NEBuilder
579 Hifi. pMTP150 was constructed by assembling two PCR products amplified from pBAD33
580 (JEP1766+JEP1767 and JEP1768+JEP1769) with gBlock7 and gBlock8 using NEBuilder
581 Hifi. pMTP151 was constructed by inserting annealed oligos (JEP1477+JEP1478) into pMTP150
582 following digestion with BsaI using golden gate cloning. pMTP160 was constructed by
583 assembling two PCR products amplified from pBAD33 (JEP1766+JEP1767 and
584 JEP1768+JEP1769) with gBlock7 and one PCR product amplified from gBlock8
585 (JEP1475+JEP1773) using NEBuilder Hifi. pMTP161-165 were constructed by ligating
586 annealed oligos (JEP1477+JEP1478, pMTP161; JEP1776+JEP1777, pMTP162;
587 JEP1778+JEP1779, pMTP163; JEP1669+JEP1670, pMTP164; JEP1671+JEP1672, pMTP165).
588 pMTP170 was constructed by assembling two PCR products amplified from pBAD33
589 (JEP1766+JEP1767 and JEP1770+JEP1769) with one PCR product amplified from gBlock7
590 (JEP1774+JEP1474) and one PCR product amplified from gBlock8 (JEP1475+JEP1775) using
591 NEBuilder Hifi. pMTP171-183 were constructed by inserting annealed oligos
592 (JEP1784+JEP1785, pMTP171; JEP1780+JEP1781, pMTP172; JEP1782+JEP1783, pMTP173;
593 JEP1794+JEP1795, pMTP174; JEP1796+JEP1797, pMTP175; JEP1786+JEP1787, pMTP176;
594 JEP1788+JEP1789, pMTP177; JEP1798+JEP1799, pMTP178; JEP1800+JEP1801, pMTP179;
595 JEP1808+JEP1809, pMTP180; JEP1810+JEP1811, pMTP181; JEP1816+JEP1817, pMTP182;
596 JEP1818+JEP1819, pMTP183) into pMTP170 following digestion with BsaI using golden gate
597 cloning. pMTP190 was constructed by assembling two PCR product amplified from pBAD33
598 (JEP1766+JEP1767 and JEP1771+JEP1769) using NEBuilder Hifi. pMTP191 and pMTP192
599 were constructed by annealing four oligos (JEP1928, JEP1929, JEP1930, JEP1931 : pMTP191;

CRISPR array categorization

600 JEP1932, JEP1933, JEP1934, JEP1935 : pMTP192) and ligating with pMTP190 digested with
601 XmaI and BsaI.

602 pMTP230 was constructed by assembling one PCR product amplified from pBAD33
603 (JEP1864+JEP1865), one PCR product amplified from pMTP130 (JEP1866+JEP1867) and
604 pSL0284 digested with NcoI and PflFI with 3,707bp fragment gel purified using NEBuilder Hifi.

605 pMTP240 was constructed by assembling a PCR product amplified from pBAD322
606 (JEP1868+JEP1869) with pSL0284 digested with NdeI and BglI with 5,152bp fragment gel
607 purified using NEBuilder Hifi. pMTP250 was constructed by assembling a PCR product
608 amplified from pCDFDuet-1 (JEP1838+JEP1839), a PCR product amplified from pBAD322
609 (JEP1834+JEP1835) and a PCR product amplified from pBBR1MCS-3 (JEP1836+JEP1837)
610 using NEBuilder Hifi. pMTP260 and pMTP270 were constructed by annealing four oligos
611 (JEP1870, JEP1871, JEP1872, JEP1873 : pMTP260; JEP1908, JEP1909, JEP1910, JEP1911 :
612 pMTP270) and ligating with pMTP250 digested with XmaI and BsaI. pMTP261-264 were
613 constructed by inserting annealed oligos (JEP1914+JEP1915, pMTP161; JEP1912+JEP1913,
614 pMTP162; JEP1880+JEP1881, pMTP163; JEP1882+JEP1883, pMTP164) into pMTP260
615 following digestion with BsaI using golden gate cloning. pMTP271-274 were constructed by
616 inserting annealed oligos (JEP1914+JEP1919, pMTP271; JEP1912+JEP1917, pMTP272;
617 JEP1880+JEP1916, pMTP273; JEP1882+JEP1917, pMTP27) into pMTP270 following digestion
618 with BsaI using golden gate cloning. pMTP275 and pMTP276 were constructed by annealing
619 four oligos (JEP1920, JEP1921, JEP1922, JEP1923 : pMTP275; JEP1924, JEP1925, JEP1926,
620 JEP1927 : pMTP276) and ligating with pMTP250 digested with XmaI and BsaI.

621 All F derivatives were made by using recombineering(Datsenko and Wanner, 2000) to replace a
622 large region of plasmid F from strain EMG2 (GenBank: AP001918.1) with PCR fragments

CRISPR array categorization

623 amplified from pMTP114 derivatives (JEP1376+1386. pMTP115, $F\Delta(finO-fxsA)::lacZ\ specR$;
624 pMTP116, $F\Delta(finO-fxsA)::cysH^{As}\ specR$; pMTP117, $F\Delta(finO-fxsA)::ffs^{As}\ specR$; pMTP118,
625 $F\Delta(finO-fxsA)::guaC^{Vc}\ specR$).
626 pOPO01 was constructed by ligating a PCR product amplified from gBlock9
627 (JEP1657+JEP1757) digested with NdeI and HindIII into pBAD33 digested with the same
628 enzymes. The resulting construct was digested with NdeI and XbaI and ligated with
629 phosphorylated annealed oligos (JEP1842+JEP1843). pOPO02 was constructed by assembling a
630 PCR product amplified from gBlock10 (JEP1764+JEP1765) with pBAD33 digested with NdeI
631 and HindIII using NEBuilder Hifi. The resulting construct was digested with NdeI and XbaI and
632 ligated with phosphorylated annealed oligos (JEP1842+JEP1843). pOPO03 was constructed by
633 ligating a PCR product amplified from gDNA of *V. parahaemolyticus* RIMD221063 (kindly
634 provided by Tobias Doerr) (JEP1952+JEP1960) digested with NdeI and HindIII and
635 phosphorylated annealed oligos (JEP1842+JEP1843) into pBAD33 digested with NdeI and
636 HindIII. pOPO04 was constructed by ligating a PCR product amplified from gBlock11
637 (JEP1555+JEP1556) digested with SpeI and HindIII into pBAD33 digested with XbaI and
638 HindIII. pOPO05 was constructed by ligating a PCR product amplified from pBAD24
639 (JEP1759+JEP1760) digested with BsaI and XhoI with a PCR product amplified from EMG2
640 (JEP1761+JEP1762) digested with the same enzymes. pOPO06-11, pOPO14, and pOPO15 were
641 constructed by ligating fragments from gBlock10 or gBlock11 digested with XhoI and StuI
642 (gBlock10 : pOPO06-09; gBlock11 : pOPO10, pOPO11, pOPO14, pOPO15) into pOPO05
643 digested with XhoI and SmaI. pOPO12 and pOPO13 were constructed by ligating PCR products
644 amplified from gDNA of *V. parahaemolyticus* RIMD221063 (JEP1956+JEP1957, pOPO12;
645 JEP1954+JEP1955, pOPO13) digested with XhoI and StuI into pOPO05 digested with XhoI and

CRISPR array categorization

646 SmaI. pOPO16-19 were constructed by ligating PCR products amplified (from gBlock9,
647 JEP1675+JEP1758, pMTP016; from gBlock10, JEP1556+JEP1764, pMTP017; from gDNA of
648 *V. parahaemolyticus* RIMD221063, JEP1952+JEP1953, pMTP018; from gBlock11,
649 JEP1950+1951, pMTP019) digested with NdeI and XhoI into pET22b(+) digested with the same
650 enzymes. pOPO21 was constructed by ligating annealed oligos (JEP1906+JEP1907) into a PCR
651 product amplified from pCOLADuet-1 (JEP1902+JEP1903) digested with SapI. pOPO23 was
652 constructed by ligating a PCR product amplified from gBlock12 (JEP1892+JEP1893) digested
653 with BsaI, annealed oligos (JEP1778+JEP1779), annealed oligos (JEP1894+JEP1895), and a
654 PCR product amplified from pCDFDuet-1 (JEP1577+JEP1891) digested with BsaI. pOPO25
655 was constructed by ligating a PCR product amplified from gBlock12 (JEP1896+JEP1897)
656 digested with BsaI, annealed oligos (JEP1782+JEP1783), annealed oligos (JEP1898+JEP1899),
657 and a PCR product amplified from pCDFDuet-1 (JEP1577+JEP1891) digested with BsaI.
658 pOPO26 was constructed by ligating annealed oligos (JEP1940+JEP1941), annealed oligos
659 (JEP1942+JEP1943), and a PCR product amplified from pCDFDuet-1 (JEP1577+JEP1891)
660 digested with BsaI. pOPO27 was constructed by ligating a PCR product from gBlock13
661 (JEP1937+JEP1938) digested with BsaI, annealed oligos (JEP1948+JEP1949), and a PCR
662 product amplified from pCDFDuet-1 (JEP1577+JEP1891). pOPO29 was constructed by ligating
663 a PCR product amplified from gBlock13 (JEP81+JEP82) digested with BsaI, annealed oligos
664 (JEP1778+JEP1779), annealed oligos (JEP1946+JEP1947), and a PCR product amplified from
665 pCDFDuet-1 (JEP1577+JEP1891) digested with BsaI. pOPO30 was constructed by assembling a
666 PCR product amplified from pCas1_pCas2/3 (JEP1889+JEP1890) and pACYCDuet-1 digested
667 with NcoI and AvrII using NEBuilder Hifi.

668

CRISPR array categorization

Identifying type I-F CRISPR-guided Tn7-like transposons

Annotated protein fasta files, genomic sequences and feature tables of gamma proteobacteria were downloaded from National Center for Biotechnology Information (NCBI) FTP site on September 22, 2019. In total, there were 53079 genomes for analysis.

Profile HMMs associated with TnsA (PF08722,PF08721), TnsB (PF00665), TnsC (PF11426,PF05621), TniQ(PF06527), Cas5f(PF09614), Cas6f(PF09618), Cas7f(PF09615) and XRE family proteins(PF01381), which can be downloaded from The European Bioinformatics Institute (EMBL-EBI) Pfam database, were used for detecting homologs with hmmsearch (HMMER3, <http://hmmer.org/> (Johnson et al., 2010)).

Candidate proteins were grouped into *tnsABC* operons and *tniQ-cas* operon based on their orientation and proximity. Then each *tnsABC* operon was grouped with its downstream *tniQ-cas* operon into one transposon functional unit. The Xre/HTH (helix turn helix) proteins situated between the two operons and are homologous to restriction controller proteins (blastp, identity >40%) were defined as candidate regulators.

CRISPR array detection

Manually curated CRISPR repeats of CRISPR-guided transposons were used to create a DNA sequence profile, which was used as a query for nhmmscan searches (Wheeler and Eddy, 2013) to find CRISPR repeats in the downstream 20-kb region of *cas6*. Putative repeats were grouped into arrays by their distances to each other. The distance between repeats was required to be >55 bp and <65 bp, the bit-score threshold is -1. The distance between last repeat and previous repeat

CRISPR array categorization

was allowed to be between 43 bp and 55 bp, but in such cases its bit-score had to be ≥ 0.3 . The sum of bit-scores of repeats in an array cannot be lower than 6.0. The longest non-overlapping arrays are collected as putative CRISPR arrays. All repeats besides the final repeat from the first array downstream of *cas6* were used to create an updated repeat profile, and the CRISPR detection procedure was repeated with the new profile twice.

Protospacer detection

To detect protospacers that match the transposon-associated CRISPR spacers, each spacer was converted into position-specific scoring matrix (PSSM) and used to search upstream 1-kb DNA of *tnsA* for matches with biopython (threshold=11.0)(Cock et al., 2009). Because every 6th base of spacers is flipped out in type I CRISPR Cascade complex, all 6th positions of the matrix are set to have equal weight on all four bases.

Except for *ffs* (SRP-RNA) gene, the major attachment site genes that containing the candidate protospacers are classified with the annotations provided in NCBI. The attachment site SRP-RNA gene (*ffs*) is often poorly annotated, so it was reannotated using cmsearch (infernol, (Nawrocki and Eddy, 2013)) and SRP-RNA profile (RF00169) available on RFAM (<https://rfam.xfam.org/>).

Constructing similarity trees

The TnsA, TniQ and Xre proteins were clustered using the Cd-hit (Li and Godzik, 2006), with identity threshold set to 90%. Multiple alignments of the representatives were done with MUSCLE (Edgar, 2004). Similarity trees were made with FastTree (Price et al., 2009) using

CRISPR array categorization

WAG evolutionary model and the discrete gamma model with 20 rate categories as previously described (Peters et al., 2017a). The visualization of the trees, major attachment sites, CRISPR arrays and matched spacers was done with ETEToolkit (Huerta-Cepas et al., 2016).

Searching shared promoter motifs of *xre* and CRISPR-Cas genes

The transposons were classified into two groups based on associated *xre* lengths (68 a.a. for I-F3a or ~100 a.a. for I-F3b) and similarities to C.AhdI and C.Csp231I. For each group, the 100bp upstream of *xre*, second CRISPR array, and *tniQ-cas* operon were collected and deduplicated with dedupe.sh (BBTools, (Bushnell)) with threshold of 70% identity or 30 edit distance. The sequences were then sent to MEME (Bailey et al., 2009) for motif detection and comparison.

Comparing consensus CRISPR repeat sequences of chromosome targeting spacers to those of other spacers

To make consensus sequences of CRISPR repeats, the transposon representatives with non-redundant TniQ were selected with Cd-hit and separated into two groups based on their attachment sites being *ffs/rsmJ* or *guaC/yciA*. The upstream and downstream CRISPR repeats of the chromosome targeting spacers were collected and turned into sequence logo using WebLogo 3 (Crooks et al., 2004) then compared to the sequence log made of other repeats.

Transposition assays

All transposition assays were performed in MTP1191, or one of MTP997 or MTP1196 with an F plasmid derivative as indicated in Table 2.

CRISPR array categorization

For the elements derived from *A. salmonicida* S44, strains used to monitor transposition were made competent by standard chemical methods (Peters, 2007) and transformed with pMTP130, pMTP140, and a derivative of pMTP150, pMTP160, pMTP170, or pMTP190 as indicated in Table 2 onto LB + 100 µg/mL carbenicillin, 10 µg/mL gentamicin, 30 µg/mL chloramphenicol, .2% w/v glucose. After 16 hours incubation at 37°C, several hundred transformants were washed up in M9 maltose and diluted to a calculated OD = 0.2 in M9 maltose + 100 µg/mL carbenicillin, 10 µg/mL gentamicin, 30 µg/mL chloramphenicol, .2% w/v arabinose, .1 mM IPTG to induce transposition.

For experiments monitoring transposition frequency through loss of sugar metabolism on MacConkey's media, induction pools were incubated for 24 hours with shaking at 30°C before being serially diluted and plated on MacConkeys 1% w/v lactose, sorbitol, or galactose. Plates were incubated at 37°C for 16 hours before colonies were counted.

For experiments monitoring transposition frequency by the mate out assay (Supplemental Figure 2a), after 24 hours incubation with shaking at 30°C, a portion of induced cultures were washed once and resuspended in LB + .2% w/v glucose. After 2 hours incubation at 37°C induced pools were mixed with prepared mid-log CW51 recipient strain at a ratio of 1:5 donor:recipient and incubated with gentle agitation for 90 minutes at 37°C to allow mating. After incubation cultures were vortexed, placed on ice, then serially diluted and plated on LB + 20 µg/mL nalidixic acid, 100 µg/mL rifampicin, 100 µg/mL spectinomycin, 50 µg/mL X-gal, with or without 50 µg/mL kanamycin to sample the entire transconjugant population or select for transposition respectively. Plates were incubated at 37°C for 36 hours before colonies were counted.

CRISPR array categorization

760

761 *V. cholerae* Tn6677 transposition assays were performed as above with function plasmids

762 pMTP230, pMTP240, and a derivative of pMTP250, pMTP260, or pMTP270 as indicated in

763 Table 2 with the exception of 8 µg/mL tetracycline replacing gentamicin when present to

764 accommodate the different vector set.

765

766 In all experiments, non-target controls where the spacer did not match the target F plasmid were

767 used, with transposition frequency similar to non-target rate in Figure 3B for *A. salmonicida* S44

768 transposition, or Figure 4D for Tn6677 transposition.

770 *P. aeruginosa* CRISPR interference assays

771 All interference assays were performed in BL21-AI. BL21-AI was transformed with pOPO30,

772 pCsy complex, and a derivative of pCOLADuet-1 as indicated in Table 2 under repressing

773 conditions. Overnight cultures were diluted 1:50 in LB + 100 µg/mL carbenicillin, 50 µg/mL

774 kanamycin, and 30 µg/mL chloramphenicol and grown for one hour before addition of IPTG and

775 arabinose to a final concentration of .1 mM and .2% w/v respectively. Cultures were grown to

776 OD = 0.6 before electrocompetent cells were prepared by standard methods (Peters, 2007) and

777 transformed with various dilutions of pOPO21 or pCOLADuet-1. Cells were recovered in SOB

778 for one hour before being plated on LB + 100 µg/mL carbenicillin, 50 µg/mL kanamycin, 30

779 µg/mL chloramphenicol, and 100 µg/mL spectinomycin. Plates were incubated at 37°C for 16

780 hours before colonies were counted.

781

782 **Xre protein purification**

CRISPR array categorization

pOPO16, pOPO17, pOPO18 or pOPO19 were transformed into BL21 (DE3), which was cultured in Terrific Broth at 37°C and induced with 0.1mM IPTG during log-phase. Cells were cultured an additional 12-16 hours at 18°C before being collected with centrifugation and lysed by sonication in nickel buffer (20 mM HEPES–NaOH (pH 7.5), 500 mM NaCl, 30 mM imidazole, 5% (v/v) glycerol, 5 mM β-mercaptoethanol) supplemented with 0.15 mg/mL lysozyme. Lysate was cleared by centrifugation and loaded on Nickel-NTA column, washed with nickel buffer, and eluted over a 30 mM to 500 mM imidazole gradient. Selected purified fractions were pooled, dialyzed and buffer exchanged into storage buffer (20 mM HEPES–NaOH (pH 7.5), 100 mM KCl, 5% (v/v) glycerol, 1mM DTT). The purified proteins were snap-frozen with liquid nitrogen and stored at -80°C.

Electrophoretic mobility shift assay (EMSA)

The promoter fragments of putative Xre regulated genes and their mutated variants were PCR amplified and purified. 100 nM DNA was incubated with different amounts of purified Xre proteins in equilibrium buffer (50 mM Tris–HCl (pH 8.0), 1 mM DTT, 10 mM MgCl₂) at 25°C for 20min then mixed with glycerol (final concentration 6%). EMSAs were performed in 6% non-denaturing TBE PAGE (Polyacrylamide gel) with 0.5x TBE as running buffer, running at 80V for one hour at room temperature. The gels were EtBr stained and visualized with UV imager.

DNA substrates were produced as follows: ArapBAD was amplified from pBAD24 (JEP175+JEP1364), pXre(Vp) and pArray2(Vp) were amplified from *V. parahaemolyticus* RIMD221063 (JEP1956+JEP1957, pXre(Vp); JEP1954+JEP1955, pArray2(Vp)), pXre(Vc) and pArray2(Vc) were amplified from gBlock11 (JEP29+JEP30, pXre(Vc); JEP1553+JEP82,

CRISPR array categorization

pArray2(Vc)), pXre(As) was amplified from pOPO08 (JEP1321+JEP81), pTniQ(As) was amplified from pOPO09 (JEP1322+JEP81), pXre*(As) was amplified from pOPO10 (JEP1321+JEP81), pTniQ*(As) was amplified from pOPO11 (JEP1322+JEP81), pXre(VB6) was amplified from pOPO06 (JEP1553+JEP81), and pTniQ(VB6) was amplified from pOPO07 (JEP1554+JEP81).

In vivo promoter assay

To confirm and characterize the putative Xre-mediated regulation, the promoters of putative Xre-regulated genes were transcriptionally or translationally fused to a plasmid-encoded *lacZ* reporter gene. pOPO01, pOPO02, pOPO03, or pOPO04 and a derivative of pOPO05 as indicated in Table 2 were transformed into *E. coli* BW27783. Overnight cultures of transformed cells were diluted 100-times into LB broth supplied with different concentrations of sugar, and cultured for additional 20 hours at 30°C. The LacZ activities were measured with standard miller unit assay (Malke, 1993).

Table 1

Strains	Genotype	Reference/Construction
BW27783	$\Delta(araD-araB)567 \Delta(rhaD-rhaB)568$ $\Delta lacZ4787 (::rrnB-3) hsdR514 rph-1$ $\Delta(araFGH) \Phi(\Delta araEp P_{CP8}-araE)$	(Khlebnikov et al., 2002), CGSC#12119
MTP997	BW27783 $attTn7::miniTn7(miniAs(kanR))$	Donor module inserted into BW27783 via pMTP112 using standard methods

CRISPR array categorization

		(Sibley and Raleigh, 2012).
EMG2	F ⁺	(Malke, 1969), (GenBank: U00096.3), CGSC#4401
MTP1191	MTP997 Lac ⁺	MTP997 x P1 EMG2
MTP1196	BW27783 <i>attTn7::miniTn7(miniVc(kanR))</i>	Donor module inserted into BW27783 via pMTP113 using standard methods (Sibley and Raleigh, 2012).
BL21 (DE3)	<i>E. coli</i> str. B F ⁻ <i>ompT gal dcm lon hsdS_B(r_B⁻m_B⁻) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5])</i> <i>[malB⁺]_{K-12}(λ^S)</i>	Lab strain (GenBank: CP001509.3)

822

823 **Table 2**

Plasmids	Description
pMS26	Tn7 shuttle vector containing Tn7 end sequences flanking a cloning site and <i>tnsABCD</i> (Sibley and Raleigh, 2012).
pMTP112	pMS26 containing <i>A. salmonicida</i> S44 right (112 bp) and left (120 bp) end sequences flanking <i>kanR</i> cassette between Tn7 ends.
pMTP113	pMS26 containing Tn6677 right (127 bp) and left (145 bp) end sequences flanking <i>kanR</i> cassette between Tn7 ends.
pKD46	Temperature-sensitive plasmid containing RED recombination genes under arabinose inducible control (Datsenko and Wanner, 2000).

CRISPR array categorization

pTSC29	Temperature-sensitive plasmid with chloramphenicol resistance marker (Phillips, 1999).
pBAD322S	Expression vector with pBR322 replicon, arabinose expression system and spectinomycin resistance marker (Cronan, 2006) (GenBank: DQ131584).
pMTP114	pTSC29 containing two 1,000 bp fragments from F plasmid flanking a 747 bp BsaI-DHFR-BsaI cassette and 1,213 bp <i>specR</i> cassette.
pMTP115	pMTP114 with 3,310 bp <i>lacZ</i> including promoter inserted over BsaI-DHFR-BsaI cassette.
pMTP116	pMTP114 with 52 bp sequence upstream <i>cysH</i> from <i>A. salmonicida</i> S44 pS44-1 inserted over BsaI-DHFR-BsaI cassette.
pMTP117	pMTP114 with 68 bp bp 3' <i>ffs</i> sequence from <i>A. salmonicida</i> S44 inserted over BsaI-DHFR-BsaI cassette.
pMTP118	pMTP114 with 70 bp 3' <i>guaC</i> sequence from <i>V. cholerae</i> HE-45 inserted over BsaI-DHFR-BsaI cassette.
pTA106	Expression vector with pSC101 replicon, lactose promoter, ampicillin resistance marker (Peters and Craig, 2000).
pMTP130	pTA106 with codon-optimized <i>tnsABC</i> from <i>A. salmonicida</i> S44 (Figure 3B, 3C, 3D, 3E, 4B, Supplemental Figure 1D, 1E).
pBAD322G	Expression vector with pBR322 replicon, arabinose expression system and gentamycin resistance marker (Cronan, 2006) (GenBank: DQ119284.1).
pMTP140	pBAD322-Gen with codon-optimized <i>tniQ-cas8/5-cas7-cas6</i> operon

CRISPR array categorization

	from <i>A. salmonicida</i> S44 inserted (Figure 3B, 3C, 3D, 3E, 4B, Supplemental Figure 1D, 1E).
pBAD33	Expression vector with pACYC replicon, arabinose expression system and chloramphenicol resistance marker (Guzman et al., 1995).
pMTP150	pBAD33 with CRISPR array from <i>A. salmonicida</i> S44 with BsaI-DHFR-BsaI cassette replacing spacer one with additional BsaI downstream of MCS removed.
pMTP151	pMTP150 with first spacer from <i>A. salmonicida</i> S44 CRISPR array inserted, restoring wild-type CRISPR array (Figure 3B).
pMTP160	pBAD33 with CRISPR repeat one and CRISPR repeat two (typical configuration) from <i>A. salmonicida</i> S44 flanking BsaI-DHFR-BsaI cassette with additional BsaI downstream of MCS removed.
pMTP161	pMTP160 with first spacer from <i>A. salmonicida</i> S44 CRISPR array targeting plasmid <i>cysH</i> (Figure 3C).
pMTP162	pMTP160 with second spacer from <i>A. salmonicida</i> S44 CRISPR array targeting chromosomal <i>ffs</i> (<i>ffs</i> ^{WT}) (Figure 3C, 3D).
pMTP163	pMTP160 with second spacer from <i>A. salmonicida</i> S44 targeting chromosomal <i>ffs</i> , with mismatches to protospacer corrected (<i>ffs</i> ^{Exact}) (Figure 3C, 3D).
pMTP164	pMTP160 with <i>lacZ</i> -targeting spacer #3 (with TACC PAM) (Figure 3D).
pMTP165	pMTP160 with <i>lacZ</i> -targeting spacer #4 (with TACC PAM) (Figure 3D, 4B).
pMTP170	pBAD33 with CRISPR repeat two and CRISPR repeat three (atypical

CRISPR array categorization

	configuration) from <i>A. salmonicida</i> S44 flanking BsaI-DHFR-BsaI cassette with additional BsaI downstream of MCS removed.
pMTP171	pMTP170 with first spacer from <i>A. salmonicida</i> S44 targeting plasmid <i>cysH</i> (Figure 3C).
pMTP172	pMTP170 with second spacer from <i>A. salmonicida</i> S44 targeting chromosomal <i>ffs</i> (<i>ffs</i> ^{WT}) (Figure 3C, 3D).
pMTP173	pMTP170 with second spacer from <i>A. salmonicida</i> S44 targeting chromosomal <i>ffs</i> , with mismatches to protospacer corrected (<i>ffs</i> ^{Exact}) (Figure 3C, 3D).
pMTP174	pMTP170 with <i>lacZ</i> -targeting spacer #1 (with TACC PAM) (Figure 3E, Supplemental Figure 1D).
pMTP175	pMTP170 with <i>lacZ</i> -targeting spacer #2 (with TACC PAM) (Figure 3E, Supplemental Figure 1D).
pMTP176	pMTP170 with <i>lacZ</i> -targeting spacer #3 (with TACC PAM) (Figure 3D, 3E).
pMTP177	pMTP170 with <i>lacZ</i> -targeting spacer #4 (with TACC PAM) (Figure 3D, 3E, 4B, Supplemental Figure 1D).
pMTP178	pMTP170 with <i>lacZ</i> -targeting spacer #5 (with TACC PAM) (Figure 3E).
pMTP179	pMTP170 with <i>lacZ</i> -targeting spacer #6 (with TACC PAM) (Figure 3E).
pMTP180	pMTP170 with <i>galK</i> top strand-targeting spacer (with TACC PAM) (Supplemental Figure 1E).
pMTP181	pMTP170 with <i>galK</i> bottom strand-targeting spacer (with TACC PAM) (Supplemental Figure 1E).

CRISPR array categorization

pMTP182	pMTP170 with <i>srID</i> top strand-targeting spacer (with TTCC PAM) (Supplemental Figure 1E).
pMTP183	pMTP170 with <i>srID</i> bottom strand-targeting spacer (with TACC PAM) (Supplemental Figure 1E).
pMTP190	pBAD33 with additional BsaI site outside of MCS removed.
pMTP191	pMTP190 with <i>lacZ</i> -targeting spacer #4 flanked by <i>A. salmonicida</i> S44 typical repeats with changes at 3 nucleotide positions to corresponding atypical base (typical*) (Figure 4B).
pMTP192	pMTP190 with <i>lacZ</i> -targeting spacer #4 flanked by <i>A. salmonicida</i> S44 atypical repeats with changes at 3 nucleotide positions to corresponding typical base (atypical*) (Figure 4B).
pSL0527 (pDonor)	pUC19 with Tn6677 right end left ends flanking chloramphenicol resistance cassette. SL0527 (pDonor) was a gift from Samuel H. Sternberg (Addgene plasmid # 130634 ; http://n2t.net/addgene:130634 ; RRID:Addgene_130634) (Klompe et al., 2019b).
pSL0283 (pTnsABC)	pCOLADuet-1 with Tn6677 <i>tnsABC</i> . pSL0283 (pTnsABC) was a gift from Samuel H. Sternberg (Addgene plasmid # 130633 ; http://n2t.net/addgene:130633 ; RRID:Addgene_130633) (Klompe et al., 2019b).
pSL0284 (pQCascade_entry)	pCDFDuet-1 with Tn6677 <i>tniQ-cas8/5-cas7-cas6</i> and Tn6677 CRISPR repeat entry module. pSL0284 (pQCascade_entry) was a gift from Samuel H. Sternberg (Addgene plasmid # 130635 ; http://n2t.net/addgene:130635 ; RRID:Addgene_130635) (Klompe et al.,

CRISPR array categorization

	2019b).
pBAD322	Expression vector with pBR322 replicon, arabinose expression system and ampicillin resistance marker (Cronan, 2006) (GenBank: DQ119282).
pMTP230	pBAD33 with lac promoter and Tn6677 <i>tnsABC</i> inserted (Figure 4C, 4D).
pMTP240	pBAD322 with Tn6677 <i>tniQ-cas8/5-cas7-cas6</i> inserted (Figure 4C, 4D).
pBBR1MCS-3	Expression vector with pBBR replicon, lactose expression system and tetracycline resistance marker (Kovach et al., 1995).
pCDFDuet-1	Expression vector with pCDF replicon, tandem T7 promoters, and spectinomycin resistance marker (Novagen).
pMTP250	Expression vector with pCDF replicon, arabinose expression system and tetracycline resistance marker.
pMTP260	pMTP250 with two copies of Tn6677 repeat one (typical configuration) flanking BsaI sites.
pMTP261	pMTP260 with fourth spacer from Tn6677 CRISPR array targeting chromosomal <i>guaC</i> (<i>guaC</i> ^{WT}) (Figure 4C).
pMTP262	pMTP260 with fourth spacer from Tn6677 CRISPR array targeting chromosomal <i>guaC</i> , with mismatches to protospacer corrected (<i>guaC</i> ^{Exact}) (Figure 4C).
pMTP263	pMTP260 with <i>lacZ</i> -targeting spacer #3 (with TACC PAM) (Figure 4D).
pMTP264	pMTP260 with <i>lacZ</i> -targeting spacer #4 (with TACC PAM) (Figure 4D).
pMTP270	pMTP250 with Tn6677 repeat five and repeat six (atypical configuration) flanking BsaI sites.

CRISPR array categorization

pMTP271	pMTP270 with fourth spacer from Tn6677 CRISPR array targeting chromosomal <i>guaC</i> (<i>guaC</i> ^{WT}) (Figure 4C, 4D).
pMTP272	pMTP270 with fourth spacer from Tn6677 CRISPR array targeting chromosomal <i>guaC</i> , with mismatches to protospacer corrected (<i>guaC</i> ^{Exact}) (Figure 4C).
pMTP273	pMTP270 with <i>lacZ</i> -targeting spacer #3 (with TACC PAM) (Figure 4D).
pMTP274	pMTP270 with <i>lacZ</i> -targeting spacer #4 (with TACC PAM) (Figure 4D).
pMTP275	pMTP250 with <i>lacZ</i> -targeting spacer #4 flanked by Tn6677 typical repeats with changes at 3 nucleotide positions to corresponding atypical base (typical*) (Figure 4D).
pMTP276	pMTP250 with <i>lacZ</i> -targeting spacer #4 flanked by Tn6677 atypical repeats with changes at 3 nucleotide positions to corresponding typical base (atypical*) (Figure 4D).
FΔ(<i>finO</i> - <i>fxsA</i> :: <i>lacZ specR</i>	F plasmid (GenBank: AP001918.1) with a 44,962 bp section from <i>finO</i> to <i>fxsA</i> replaced with 3,310 bp <i>lacZ</i> including promoter (GenBank: U00096.3) and 1,213 bp <i>specR</i> cassette (GenBank: DQ131584.1) (Figure 3B, 3D, 3E, 4B, 4D).
FΔ(<i>finO</i> - <i>fxsA</i> :: <i>cysH</i> ^{As} <i>specR</i>	F plasmid (GenBank: AP001918.1) with a 44,962 bp section from <i>finO</i> to <i>fxsA</i> replaced with 52 bp upstream <i>cysH</i> sequence from <i>A. salmonicida</i> S44 pS44-1 including protospacer (GenBank: CP022176.1) and 1,213 bp <i>specR</i> cassette (GenBank: DQ131584.1) (Figure 3B, 3C).
FΔ(<i>finO</i> - <i>fxsA</i> :: <i>ffs</i> ^{As} <i>specR</i>	F plasmid (GenBank: AP001918.1) with a 44,962 bp section from <i>finO</i> to <i>fxsA</i> replaced with 68 bp 3' <i>ffs</i> sequence from <i>A. salmonicida</i> S44

CRISPR array categorization

	including protospacer (GenBank: CP022181.1) and 1,213 bp <i>specR</i> cassette (GenBank: DQ131584.1) (Figure 3B, 3C, 3D).
FΔ(<i>finO</i> - <i>fxsA</i> :: <i>guaC</i> ^{Vc} <i>specR</i>	F plasmid (GenBank: AP001918.1) with a 44,962 bp section from <i>finO</i> to <i>fxsA</i> replaced with 70 bp 3' <i>guaC</i> sequence from <i>V. cholerae</i> HE-45 including protospacer (GenBank: ALED01000027.1) and 1,213 bp <i>specR</i> cassette (GenBank: DQ131584.1) (Figure 4C).
pOPO01	pBAD33 with modified weaker RBS (iGEM BBa_B0033) and <i>xre</i> gene from <i>A. salmonicida</i> S44 inserted. (Figure 6F)
pOPO02	pBAD33 with modified weaker RBS (iGEM BBa_B0033) and <i>xre</i> gene from <i>Vibrio</i> sp. 10N.286.45.B6 inserted. (Figure 6F)
pOPO03	pBAD33 with modified weaker RBS (iGEM BBa_B0033) and <i>xre</i> gene from <i>V. parahaemolyticus</i> RIMD221063 inserted. (Figure 6E)
pOPO04	pBAD33 with modified weaker RBS (iGEM BBa_B0033) and <i>xre</i> gene from Tn6677 inserted. (Figure 6E)
pBAD24	Expression vector with pBR322 replicon lacking <i>rop</i> gene copy control, arabinose expression system and ampicillin resistance marker (Guzman et al., 1995) (GenBank: DQ119282).
pOPO05	pBAD24 with a promoterless, start codon-lacking <i>lacZ</i> inserted.
pOPO06	pOPO05 with <i>xre</i> promoter from <i>Vibrio</i> sp. 10N.286.45.B6 fused to <i>lacZ</i> . (Figure 6F)
pOPO07	pOPO05 with <i>tmiQ</i> promoter from <i>Vibrio</i> sp. 10N.286.45.B6 fused to <i>lacZ</i> . (Figure 6F)
pOPO08	pOPO05 with <i>xre</i> promoter from <i>A. salmonicida</i> S44 fused to <i>lacZ</i> .

CRISPR array categorization

	(Figure 6F)
pOPO09	pOPO05 with <i>tniQ</i> promoter from <i>A. salmonicida</i> S44 fused to <i>lacZ</i> . (Figure 6F)
pOPO10	pOPO05 with <i>xre</i> promoter from <i>A. salmonicida</i> S44 with proximal putative Xre binding site mutated fused to <i>lacZ</i> . (Figure 6F)
pOPO11	pOPO05 with <i>tniQ</i> promoter from <i>A. salmonicida</i> S44 with putative Xre binding site mutated fused to <i>lacZ</i> . (Figure 6F)
pOPO12	pOPO05 with <i>xre</i> promoter from <i>V. parahaemolyticus</i> RIMD221063 fused to <i>lacZ</i> . (Figure 6E)
pOPO13	pOPO05 with the second array promoter from <i>V. parahaemolyticus</i> RIMD221063 fused to <i>lacZ</i> . (Figure 6E)
pOPO14	pOPO05 with <i>xre</i> promoter from <i>V. cholerae</i> Tn6677 fused to <i>lacZ</i> . (Figure 6E)
pOPO15	pOPO05 with the second array promoter from <i>V. cholerae</i> Tn6677 fused to <i>lacZ</i> . (Figure 6E)
pET22b(+)	Expression vector with pBR322 replicon, T7 expression system, and ampicillin resistance marker (Novagen).
pOPO16	pET22b(+) with C-terminal His6-tagged codon-optimized <i>xre</i> from <i>A. salmonicida</i> S44. (Figure 6D)
pOPO17	pET22b(+) with C-terminal His6-tagged <i>xre</i> from <i>Vibrio</i> sp. 10N.286.45.B6(Figure 6D)
pOPO18	pET22b(+) with C-terminal His6-tagged <i>xre</i> from <i>V. parahaemolyticus</i> RIMD221063(Figure 6C)

CRISPR array categorization

pOPO19	pET22b(+) with C-terminal His6-tagged <i>xre</i> from <i>V. cholerae</i> Tn6677(Figure 6C)
pCsy_complex	Carries <i>csy1</i> , <i>csy2</i> , <i>csy3</i> , <i>csy4</i> genes from <i>Pseudomonas aeruginosa</i> PA14 under T7 promoter. pCsy_complex was a gift from Blake Wiedenheft (Addgene plasmid # 89232 ; http://n2t.net/addgene:89232 ; RRID:Addgene_89232) (Chowdhury et al., 2017b).
pCas1_Cas2/3	Carries <i>cas1</i> , <i>cas2/3</i> of <i>Pseudomonas aeruginosa</i> PA14. The <i>cas1</i> has Strep-Tag II. pCas1_Cas2/3 was a gift from Blake Wiedenheft (Addgene plasmid # 89240 ; http://n2t.net/addgene:89240 ; RRID:Addgene_89240) (Rollins et al., 2017) (Figure 5).
pCOLADuet-1	Expression vector with pCOLA replicon, tandem T7 promoters, and kanamycin resistance marker (Novagen) (Figure 5).
pACYCDuet-1	Expression vector with pACYC replicon, tandem T7 promoters, and chloramphenicol resistance marker (Novagen).
pOPO21	pCOLADuet-1 with <i>ffs</i> gene protospacer sequence with CC PAM (Figure 5).
pOPO23	pCDFDuet-1 with CRISPR repeat one and CRISPR repeat two (typical configuration) from <i>A. salmonicida</i> S44 flanking <i>ffs</i> -targeting spacer (Figure 5).
pOPO25	pCDFDuet-1 with CRISPR repeat two and CRISPR repeat three (atypical configuration) from <i>A. salmonicida</i> S44 flanking <i>ffs</i> -targeting spacer (Figure 5).
pOPO26	pCDFDuet-1 with CRISPR repeat five and CRISPR repeat six (atypical

CRISPR array categorization

	configuration) from Tn6677 flanking <i>ffs</i> -targeting spacer (Figure 5).
pOPO27	pCDFDuet-1 with two copies of CRISPR repeat one (typical configuration) from Tn6677 flanking <i>ffs</i> -targeting spacer (Figure 5).
pOPO29	pCDFDuet-1 with two copies of CRISPR repeat one from <i>P. aeruginosa</i> PA14 flanking <i>ffs</i> -targeting spacer (Figure 5).
pOPO30	pACYCDuet-1 with <i>cas1</i> and <i>cas2/3</i> from <i>P. aeruginosa</i> PA14 (Figure 5).

824

825

826 **Table 3**

Primer s	Sequence
JEP13 98	ATCCCCCGGGCTGCAGGAATTCGATCCGGGCAGGAGCGTGACC
JEP13 40	TCTGTTATTACAAATCGTGGGTCTCACGAATCATCCTCTTGCGGCC
JEP13 41	AGAGGAAGAGTAAGGATGGAACGGG
JEP13 99	GGTCGACGGTATCGATAAGCTTGATCGGCACGTCAGCGCTCTC
JEP13 43	GAGACCCACGATTTGTAATAACAG
JEP13	GAGACCTGCTCAGGGATCACC

CRISPR array categorization

44	
JEP13	TTTCGGTGATCCCTGAGCAGGTCTCCCTTGCGCTCGTTCGCCA
45	
JEP13	CCCGTTCCATCCTTACTCTTCCTCTCTTGAGTTAAGCCGCGCCGC
46	
JEP16	GGCGCGGTCTCATTCGCCCAATACGCAAACCGCCTCTC
63	
JEP16	GGCGCGGTCTCGCAAGAGACATGGCCTGCCCGGTATTA
64	
JEP13	GGGCCCTAGGGATAACAGGGTAATACGGCACGTCAGCGCTCTC
76	
JEP13	CTTGCCACCAGTGATGCGG
86	
JEP14	CGTGGGTCTCATTCGTCAGGACCGGAAGGTAGCAGCCAAGGCGGGGGACT
81	CGAGTGCCG
JEP14	CAGCGGTCTCCCAAGGTGGTGGCCCCATCAGCCACATCCCGGCACTCGAGT
82	CCCCCGCC
JEP14	TTCGGGTGCTTACCCGGAAGGGGAGGGCACCCCTGCTTATTGATAATCAGGA
85	GGCCC
JEP14	CAAGGGGCCTCCTGATTATCAATAAGCAGGGTGCCCTCCCCTTCCGGGTAA
86	GCACC
JEP11	TAGAGCGCACGAATGAGGGC
46	

CRISPR array categorization

JEP14 67	TGAAATTGTTATCCGCTCACAATTC
JEP17 66	CGGTGAGCGTGGGACTCGCGGTATCATTGCAGCACTGGG
JEP17 67	ATCCCCGGGTACCGAGC
JEP17 68	TGCGAGAGTAGGGAACTGCCA
JEP17 69	GATACCGCGAGTCCCACGCTCACCGGCTCCAGAT
JEP14 77	GAAACGGAAGGGGAGGGCACCTGCTTATTGATAAT G
JEP14 78	TTCACATTATCAATAAGCAGGGTGCCCTCCCCTTCCG
JEP14 75	CAAGTAGATACACTACATATATCTACAATA
JEP17 73	TGCCTGGCAGTTCCCTACTCTCGCAAATCTTGGCTGCCTATACGGCA
JEP17 76	GAAAAGGACTGGAAGAAATCATCCAAGTTGGGGACTG
JEP17 77	TTCACAGTCCCCAACTTGGATGATTTCTTCCAGTCCT
JEP17	GAAAAGGACCGGAAGGTAGCAGCCAAGGCGGGGGACG

CRISPR array categorization

78	
JEP17	TTCACGTCCCCCGCCTTGGCTGCTACCTTCCGGTCCT
79	
JEP16	GAAATACGGGTAACAGTTTCTTTATGGCAGGGTGAAG
69	
JEP16	TTCACTTCACCCTGCCATAAAGAAACTGTTACCCGTA
70	
JEP16	GAAAACAGCGGATGGTTCGGATAATGCGAACAGCGCG
71	
JEP16	TTCACGCGCTGTTTCGCATTATCCGAACCATCCGCTGT
72	
JEP17	AGGCAGATATTATTTGCGAGAGTAGGGAACTGCCA
70	
JEP17	GAATTCGAGCTCGGTACCCGGGGATGTGAACTGCCGTATAGGCAGCCAAG
74	ATTTGAGAC
JEP14	TAGATATATGTAGTGTATCTACTTGATC
74	
JEP17	CTACTCTCGCA AATAATATCTGCCTTTACGGCAGAAAAT
75	GAGACCTGCTCAGGGATCAC
JEP17	GATTCGGAAGGGGAGGGCACCTGCTTATTGATAATA
84	
JEP17	AAAATATTATCAATAAGCAGGGTGCCCTCCCCTTCCG
85	

CRISPR array categorization

JEP17 80	GATTAGGACTGGAAGAAATCATCCAAGTTGGGGACTA
JEP17 81	AAAATAGTCCCCAACTTGGATGATTTCTTCCAGTCCT
JEP17 82	GATTAGGACCGGAAGGTAGCAGCCAAGGCGGGGGACA
JEP17 83	AAAATGTCCCCCGCCTTGGCTGCTACCTTCCGGTCCT
JEP17 94	GATTCAACTTAATCGCCTTGCAGCACATCCCCCTTTA
JEP17 95	AAAATAAAGGGGGATGTGCTGCAAGGCGATTAAGTTG
JEP17 96	GATTCGTAGGTAGTCACGCAACTCGCCGCACATCTGA
JEP17 97	AAAATCAGATGTGCGGCGAGTTGCGTGACTACCTACG
JEP17 86	GATTTACGGGTAAACAGTTTCTTTATGGCAGGGTGAAA
JEP17 87	AAAATTTACCCCTGCCATAAAGAAACTGTTACCCGTA
JEP17 88	GATTACAGCGGATGGTTTCGGATAATGCGAACAGCGCA
JEP17	AAAATGCGCTGTTCGCATTATCCGAACCATCCGCTGT

CRISPR array categorization

89	
JEP17	GATTGGCGATGAGCGAACGCGTAACGCGAATGGTGCA
98	
JEP17	AAAATGCACCATTCGCGTTACGCGTTCGCTCATCGCC
99	
JEP18	GATTCATCGCGTGGGCGTATTCGCAAAGGATCAGCGA
00	
JEP18	AAAATCGCTGATCCTTTGCGAATACGCCCACGCGATG
01	
JEP18	GATTCTGCCACTCACACCATTCAGGCGCCTGGCCGCA
08	
JEP18	AAAATGCGGCCAGGCGCCTGAATGGTGTGAGTGGCAG
09	
JEP18	GATTAGAGTGTCAATTTGCGGCACGGTGATTTCGAAA
10	
JEP18	AAAATTTTCGAAATCACCGTGCCGCAAATTGACACTCT
11	
JEP18	GATTTGTGCCACGGTCTGGCTGCCGAGGGGTATCGCA
16	
JEP18	AAAATGCGATACCCCTCGGCAGCCAGACCGTGGCACA
17	
JEP18	GATTCCACGAGAGAGCGCCAGAACGCTTTGCTCGCTA
18	

CRISPR array categorization

JEP18 19	AAAATAGCGAGCAAAGCGTTCTGGCGCTCTCTCGTGG
JEP17 71	GAATTCGAGCTCGGTACCCGG
JEP19 28	CCGGGTGAACTGCCGTATAGGCAGCCAAGAAAACA
JEP19 29	GTTCGCATTATCCGAACCATCCGCTGTTTTCTTGGCTGCCTATACGGCAGTT CAC
JEP19 30	GCGGATGGTTCGGATAATGCGAACAGCGCATTAACTGCCGTATAGGCAGA CAAGATT
JEP19 31	ATGGAATCTTGTCTGCCTATACGGCAGTTAATGCGCT
JEP19 32	CCGGGTGAACTGCCGTATAGGCAGCCAAGATTACA
JEP19 33	GTTCGCATTATCCGAACCATCCGCTGTAATCTTGGCTGCCTATACGGCAGTT CAC
JEP19 34	GCGGATGGTTCGGATAATGCGAACAGCGCGTGTTCTGCCGTAAAGGCAGCT ATTATT
JEP19 35	ATGGAATAATAGCTGCCTTTACGGCAGAACACGCGCT
JEP18 58	TGAATTCGAGCTCAGGCGCGCCAGCACATTTGTTGATACAACCATAAAATG
JEP18	TTACTAGGATCCGATATCATTTAAAGAAATGACCTCTTCATGA

CRISPR array categorization

59	
JEP18 60	GGGACGGCGGCTTTGTTGAATAAATCTTACTGCAGTAGTTTTGCTG
JEP18 61	GAAC TTTCCTCAGCGTTTAAACGCAATGTTGTTGATGCAACCATAAAGTG
JEP18 62	TTTAAATGATATCGGATCCTAG
JEP18 63	ATTTATTCAACAAAGCCG
JEP18 78	CGTGGGTCTCATTGACGAAGCGTACAACGTTTATCCGAGTTCAAGAGCAA GAGAACAA
JEP18 79	CAGCGGTCTCCCAAGTATGCGCTTACTCACGCCCGTAAACGTTGTTCTCTTG CTCTTGAA
JEP18 64	CTGAACCGACGACAAGCTGACGACCCGTTTAAGGGCACCAATAAC
JEP18 65	CGGGCTCATGAGCAAATA
JEP18 66	GATAAAATATTTGCTCATGAGCCCGGCGCCCAATACGCAAAC
JEP18 67	CTGAGGGCGTAGGTAAACTTGTCGCCATGAGCTGTTACCTCCTTGAAATTG T
JEP18 68	GCCGGCAGGCGCTCCATTGCCAGTGTACCCGGGGATCCTCTA

CRISPR array categorization

JEP18 69	TAAGGTTTAGGTCTTTGCAAAAACATGGTGAATTCCTCCTGCTA
JEP18 38	GCGCTGCGGACACATACAAAGT
JEP18 39	ATCCCTGCTTCGTCCATTTGACAGGACTGAAATCTAGAGCGGTTTCAGTAGA A
JEP18 34	CCTGTCAAATGGACGAAGCAGGG
JEP18 35	CGTGATACGCCTATTTTTATAGGTTCTCATGAGCGGATACATATTTGAATGT A
JEP18 36	AACCTATAAAAATAGGCGTATCACGAGG
JEP18 37	GTAACTTTGTATGTGTCCGCAGCGCAATTGATTGGCTCCAATTCTTGGAG
JEP18 70	CCGGGTGAACTGCCGAGTAGGTTAG
JEP18 71	CAGCGGTCTCGGTTATCAGCTACCTACTCGGCAGTTCAC
JEP18 72	CTGATAACCGAGACCGCTGGTCTCCGTGAACTGCCGAGTAGGTTAGCTGATA AC
JEP18 73	ATGGGTTATCAGCTACCTACTCGGCAGTTCACGGAGAC
JEP19	CCGGTCATTACTACTGCAAAGTAG

CRISPR array categorization

08	
JEP19	CAGCGGTCTCGGTTATCAGCTACTTTGCAGTAGTAATGA
09	
JEP19	CTGATAACCGAGACCGCTGGTCTCCCTTTACTGCTGAATAAGTAGATAACT
10	AC
JEP19	ATGGGTAGTTATCTACTTATTCAGCAGTAAAGGGAGAC
11	
JEP19	TAACAAACGAACCACCTTCATTCGAGTACAAGAGCA
14	
JEP19	TCACTGCTCTTGTACTCGAATGAAGGTGGTTCGTTT
15	
JEP19	TAACAAGCGTACAACGTTTATCCGAGTTCAAGAGCA
12	
JEP19	TCACTGCTCTTGAACTCGGATAAACGTTGTACGCTT
13	
JEP18	TAACTACGGGTAACAGTTTCTTTATGGCAGGGTGAA
80	
JEP18	TCACTTCACCCTGCCATAAAGAACTGTTACCCGTA
81	
JEP18	TAACACAGCGGATGGTTCGGATAATGCGAACAGCGC
82	
JEP18	TCACGCGCTGTTTCGCATTATCCGAACCATCCGCTGT
83	

CRISPR array categorization

JEP19 19	AAAGTGCTCTTGTACTCGAATGAAGGTGGTTCGTTT
JEP19 18	AAAGTGCTCTTGAACCTCGGATAAACGTTGTACGCTT
JEP19 16	AAAGTTCACCCTGCCATAAAGAACTGTTACCCGTA
JEP19 17	AAAGGCGCTGTTCGCATTATCCGAACCATCCGCTGT
JEP19 20	CCGGGTGAACTGCCGAGTAGGTAGCTGATAACACA
JEP19 21	GTTCGCATTATCCGAACCATCCGCTGTGTTATCAGCTACCTACTCGGCAGTT CAC
JEP19 22	GCGGATGGTTCGGATAATGCGAACAGCGCCTTAACTGCCGAGTAGGTAGA TGATAAC
JEP19 23	ATGGGTTATCATCTACCTACTCGGCAGTTAAGGCGCT
JEP19 24	CCGGTCATTACTACTGCAAAGTAGCTGATAACACA
JEP19 25	GTTCGCATTATCCGAACCATCCGCTGTGTTATCAGCTACTTTGCAGTAGTAA TGA
JEP19 26	GCGGATGGTTCGGATAATGCGAACAGCGCGTGTACTGCTGAATAAGTAGCT AACTAC
JEP19	ATGGGTAGTTAGCTACTTATTCAGCAGTACACGCGCT

CRISPR array categorization

27	
JEP13	GGGCCCTAGGGATAACAGGGTAATACGGCACGTCAGCGCTCTC
76	
JEP13	CTTGCCACCAGTGATGCGG
86	
JEP16	GTTTCCATATGACTAATCCATTACCCATTCG
57	
JEP17	GTTTCAAGCTTACTCTGCCGAGTCACGCG
57	
JEP17	GTTTCCTCGAGCTCTGCCGAGTCACGCGA
58	
JEP18	CTAGTCTAGAGTCACACAGGACTACTA
42	
JEP18	TATAGTAGTCCTGTGTGACTCTAGA
43	
JEP17	CTTTAAGAAGGAGATATACATATGCAGAAGGATAACCC
64	
JEP17	CTCATCCGCCAAAACAGCCAAGCTTTTAACTTCTCAGC
65	
JEP17	GTTTCCTCGAGACTTCTCAGCTTTTCCAACAACA
63	
JEP19	GTTTCCATATGAGAGACTTAGCAAAAAAGTTTGG
52	

CRISPR array categorization

JEP19 53	GTTTCCTCGAGTGGTAGAAGCTCTCGAATATCGC
JEP19 60	GTTTCAAGCTTATGGTAGAAGCTCTCGAATATCGC
JEP19 50	GTTTCCATATGAAAGACTTAGCGAAACGGT
JEP19 51	GTTTCCTCGAGCGGAAGAAGCTCTCTAACATCAC
JEP15 55	GCCTGAACTGGAACAAAAGC
JEP15 56	TGCTGGTTTACCGGTTTATTG
JEP17 59	AGGTCTCAGCCTGATACAGATTAAATCAGAACGC
JEP17 60	GTTTCCTCGAGGATAAGCTGTCAAACATGAGCAGA
JEP17 61	GTTTCCTCGAGCACCCGGGACCATGATTACGGATTCACTGG
JEP17 62	AGGTCTCAAGGCTTATTTTTGACACCAGACCAACTGG
JEP19 54	GTTTCCTCGAGTTTAAAAAATCAATACTATTACCATCTCCA
JEP19	GGAAGGCCTCATAGCTGTTTCCTGTGTGGTTATCAGCTACTTTTTAGTAGTA

CRISPR array categorization

55	ATGAGG
JEP19	GTTTCCTCGAGTTAGATATCTGCTAAGTCTCTCATTTCAAAAGT
56	
JEP19	GTGGTCTCCTCGATTAAGGCCTCACTTAGCAACTCTCGACAGT
57	
JEP17	GTAACAAAGCGGGACCAAAG
5	
JEP13	CAGTGAGCGAGGAAGCGGAACCAAAACAGCCAAGCTTGCATG
64	
JEP29	TAATACGACTCACTATAGGG
JEP30	GCTATTTAGGTGACACTATAG
JEP81	CGCCAGGGTTTTCCAGTCACGAC
JEP82	AGCGGATAACAATTTACACAGGA
JEP15	GGTTCGCCAATGCTGAAG
53	
JEP13	GCTCTGGCAGGATGTTTCGT
21	
JEP132	TCGCATCATTGTTGAGTTTTTC
2	
JEP157	TGGTCTCAACCTAGGCTGCTGCCACC
7	
JEP188	ACTTTAATAAGGAGATATACCATGGACGACATTTCTCCCAGC
9	

CRISPR array categorization

JEP189 0	TGCTCAGCGGTGGCAGCAGCTTATTATCAGTTGTATTTCTTGAACC
JEP189 1	TGGTCTCAATTTCTACAGGGGAATTGTTATCC
JEP189 2	TGGTCTCAA AATGGTCTATCCTTAGCAAGGT
JEP189 3	GGGTCTCATTTCTTGGCTGC
JEP189 4	TGAACTGCCGTATAGGCAGCCAAGATTCTATAGTGTCACCTAAATAGC
JEP189 5	AGGT GCTATTTAGGTGACACTATAGAATCTTGGCTGCCTATACGGCAG
JEP189 6	TGGTCTCAAAATGGTCTATCCTTAGCAAGGT
JEP189 7	TGGTCTCAAATCTTGGCTGCCTATACGGCA
JEP189 8	TTTTCTGCCGTAAAGGCAGATATTATTCTATAGTGTCACCTAAATAGC
JEP189 9	AGGTGCTATTTAGGTGACACTATAGAATAATATCTGCCTTTACGGCAG
JEP190 2	TGCTCTTCAGCCACCGCTGAGCAATAACT
JEP190	TGCTCTTCACAGGAGTCGCATAAGGGAGAG

CRISPR array categorization

3	
JEP190	CTGGGCCAGGACCGGAAGGTAGCAGCCAAGGCGGGGGAC
6	
JEP190	GGCGTCCCCCGCCTTGGCTGCTACCTTCCGGTCCT GGCC
7	
JEP194	AAATGGTCTCTCATTACTACTGCAAAGTAGCTGATAACAGGACCGGAAGG
0	
JEP194	GCTGCTACCTTCCGGTCCTGTTATCAGCTACTTTGCAGTAGTAATGAGAGA
1	CC
JEP194	TAGCAGCCAAGGCGGGGGACCTTTACTGCTGAATAAGTAGATAACTAC
2	
JEP194	AGGTGTAGTTATCTACTTATTCAGCAGTAAAGGTCCCCCGCCTTG
3	
JEP194	TGAACAGTCCCCAACTTGGATGATTTCTTCCAGTCCT
4	
JEP194	TGAACGTCCCCCGCCTTGGCTGCTACCTTCCGGTCCT
5	
JEP194	TTCAGTGCCGTGTAGGCAGCTAAGAACTATAGTGTCACCTAAATAGC
6	
JEP194	AGGTGCTATTTAGGTGACACTATAGTTTCTTAGCTGCCTACACGGCAG
7	
JEP194	TGAACTGCCGAGTAGGTAGCTGATAAC CTATAGTGTCACCTAAATAGC
8	

CRISPR array categorization

JEP194 9	AGGTGCTATTTAGGTGACACTATAGGTTATCAGCTACCTACTCGGCAG
JEP193 7	TGAGATGCGCAAGAAGGCC
JEP193 8	GTCCAGAACGGGCTTCAGG

827

828 **Table 4**

g	De	Sequence
B	scri	
lo	pti	
c	on	
k		
s		

CRISPR array categorization

1	mi	GGCCAGCGGCCGCACATTCACTGAACTTTGACGTACCGGCACATGTCGTCT
	niA	AAACCATAAGTTGACATATCTAAAGCATAAGTTGACATGATCTTGCATCAT
	s(k	AAGCTGACATAGCCTAAACCTGTAGTAAAGTAACAACAGGTTGGTGGTGT
	an	GTTTTTAAATGATATCGGATCCTAGTAAGCCACGTTGTGTCTCAAAATCTC
	R)	TGATGTTACATTGCACAAGATAAAAATATATCATCATGAACAATAAACT
		GTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACG
		GGAAACGTCTTGCTCGAGGCCGCGATTAAATTCCAACATGGATGCTGATTT
		ATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAA
		TCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATG
		GCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAAC
		TGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTA
		CTGATGATGCATGGTTACTCACCCTGCGATCCCCGGGAAAACAGCATTCC
		AGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGG
		CAGTGTTCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAA
		CAGCGATCGCGTATTTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGG
		TTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGA
		ACAAGTCTGGAAAGAAATGCATAAGCTTTTGCCATTCTCACCGGATTCAGT
		CGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAA
		ATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCA
		GGATCTTGCCATCCTATGGAAGTGCCTCGGTGAGTTTTCTCCTTCATTACAG
		AAACGGCTTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTG
		CAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGAATTGGTTAATTGGT
		TGTAACACTGGCAGAGCATTACGCTGACTTGACGGGACGGCGGCTTTGTTG
		AATAAATGATGCAGTTAAGTGCCTGCTGTGTATCTCGAACAGATTGCTGAA
		GTGGTAAAATAACAACACCGCAGGGGCAGGGCGTAGGGGTTGATGTC
		ACTATGTCAGCTTATGGGTAATTTTTATGTCAACTTATGGGTTATAAAACACT

CRISPR array categorization

2	<i>tns</i>	GAATTGTGAGCGGATAACAATTTCAAGGAGGTAACAGCTATGTATCGCCG
	<i>AB</i>	TCATCTGAAACACAGTCGCGTCAAGAACTTATTTAAGTTCGTGAGCGCTAA
	<i>C^{As}</i>	AATGAATACCGTCTTTACGGTAGAGTCTGCGCTTGAATTTGATACGTGTTT
	<i>gBl</i>	TCATCTGGAATACTCACCATCCGTGAAATTTTACGAAGCGCAGCCGGAGG
	<i>ock</i>	GTTTTTATTACGAGTTCGCTGGTCGTCAATGCCCATATACCCCTGACTTTTCG
	1	TTTGGTGGATCAGAACGACTCGGTTTCGTTTCTTGAAATCAAGCCCTCTGA CAAGGTGGCTGATCCAGATTTTCTGCATCGCTTTCCCCTGAAGCAGCAGCG CGCCATTGAACTTAGTTCACCATTAAAGCTTGTTACTGAAAAACAAATCCG CATTGATCCTATCCTGGGGAAGTTAAAGTTACTTCACCGCTATTCTGGCTTT CAATCCTTTACTCCTTTGCATATGCAATTATTGGGTCTGGTGCAAAAATTG GGTCGCGTATCACTGCTGCGTTTAAGTGACTCCATCGACGCACCACCTGAG GAGGTATTAGCAAGCGCGTTATCCTTAATCGCTCGTGGAATTATGCAGAGT GACTTAACGGTACAGAAGATTGGCATCAGTTCGTTTCGTGTGGGCTGGCGGT CATTGAGGCATTGACCATGGATAAGCATAACGGCGGTTTATTTGAAGACG AATTCGTAATTCCCCAACCATCCACATCTACCTACCCATTGACGCCATCC AAGCTGTCCTTCCGGCTACTGTTGATAGCTTTCTTATGTTTTAAAAGTTGA GGCGCTGCATCGTCGCGATTACATTCTTTGGGTGGAGAAGAATCTTGCCGG GGGTTGGACAGAGAAGAACCTTACTCCCTTGTTGGCAGATGCAGCGTTGGT TTTACCTCCCCCACGCCTAATTGGCGTACTCTGGCACGCTGGCGTAAAAT CTATATTCAACACGGTCGTAACTTGTCAGTTTAATCCCAAACATCAGGC AAAAGGTAACGCTCGCAGTCGCCTTCCACCCAGTGATGAGCTGTTTTTCGA GCAAGCAGTACATCGTTACTTAGTAGGAGAGCAGCCTTCAATTGCGTCGGC ATTCCAGTTGTACTCCGACTCAATCCGTATCGAGAATCTTGGTGTGGTGGA

CRISPR array categorization

		GAATCCTATTA AAAACTATCAGCTACATGGCTTTCTACAACCGCATTAAGAA ATTACCGGCGTATCAAGTAATGAAGAGCCGTAAGGGCTCCTATATTGCCG ACGTCGAGTTCAAAGCCATTGCCTCTCACAAGCCACCGAGTCGTATTATGG AGCGCGTGGAGATCGATCACACACCATTGGATTTACTTTTATTGGATGATG ACTTGCTTGTGCCACTTGGGCGTCCTTCCCTTACATTATTGATTGATGCCTA TTCGCATTGCGTGGTTCGGCTTTAATCTTAATTTCAATCAGCCTTCATACGAG TCCGTTTCGTAATGCGCTGTTATCGAGCATCTCCAAAAAAGATTATGTTAAA AACAAATACCCATCCATCGAACACGAATGGCCATGCTACGGTAAGCCAGA GACCCTTGTCGTGGATAACGGGGTTCGAGTTTTGGTCCGCTTCCCTTGCACA ATCGTGCCTTGA ACTGGGAATTAACATTCAGTACAATCCTGTTCGCAAACC TTGGTTAAAGCCGATGATCGAGCGTATGTTTCGGGATTATTAACCGTAAGCT TCTGGAACCCATTCCGGGCAAAACCTTCAGCAACATCCAGGAGAAGGGTG ATTATGATCCCCAAAAAGACGCTGTTATGCGCTTTAGCACATTTCTTGAAA TCTTTCATCACTGGGTCATTGACGTGTATCACTACGAACCTGACAGCCGTT ACCG
3	<i>tns</i> <i>AB</i> <i>C^{As}</i> <i>gBl</i> <i>ock</i> 2	ACTACGAACCTGACAGCCGTTACCGTTACATTCCGATTATCTCCTGGCAAC ACGGAAATAAAGACGCCCCACCTGCACCAATCATTGGTGATGATTTGACA AAGCTGGAAGTGATTTTAAGTCTTTCGTTACATTGTACCCATCGTCGTGGC GGTATCCAGCGCTACCATTTGCGCTATGATTCTGACGAGTTAGCATCATAT CGTATGAACTATCCAGATCAGACGCGTGGCAAACGTAAAGTTCTGGTCAA ACTTAATCCTCGTGATATCTCTTATGTGTACGTCTTCTTGGAGGATTTAGGG TCATACATTCGTGTCCCTTGTATCGACCCGATTGGTTACACCAAGGGTTTAT CTCTGCAAGAACACCAGATCAACGTCAAGTTGCATCGCGATTTTATTAATG

CRISPR array categorization

	<p> AGCAGATGGATGTGGTGAGTCTGAGCAAGGCACGTATTTACCTGAATGAT CGCATTAAAAACGAGCTTATCGAAGTGCGTCGCAACATCCGTCAGCGCAA CGTGAAAGGGGTAAATAAGATTGCCAAGTATCGTAATGTTGGATCTCACG CAGAAACATCCATTGTTTCATGAGCTGAACCATCCTGCGACGAATGAGGTC ATCTCTAAGATGGAGAGTGCCAGCCAGCCCGAACACTGTGACGATTGGGA TAATTTCACTTCTGGCTTGGAGCCCTACTGATGGACCTTTCGTGTCACGAC GCAGACAAATTGCGCTCGTTTATTGAATGTTATGTAGAAACCCCGCTTCTT CGCGCTATCCAGGAAGACTTCGATCGTCTTCGTTTCAACAAGCAATTTGCA GGTGAGCCGCAGTGTATGTTGCTTACTGGTGACACCGGCACCGGCAAGAG CTCCCTTATCCGTCATTACGCAGCCAAACATCCTGAACAGGTACGTCACGG TTTCATTACACAAGCCATTACTTGTTTCACGCATTCCGTCTCGTCCCACCCTT GAATCTACCATGGTGGAGTTGCTTAAAGATTTAGGTCAATTCGGCAGTTCT GACCGTATCCATAAAAGTTCCGCCGAAAGTTTAACCGAGGCTTTAATTAAA TGTCTTAAACGCTGTGAGACCGAGCTGATTATCATCGACGAGTTTCAAGAG CTTATTGAAAATAAGACCCGCGAGAAGCGCAACCAAATTGCGAATCGCTT GAAGTATATTTGCGAAACAGCTAAGATTCCGATTGTTTTAGTAGGTATGCC CTGGGCGACCAAGATTGCAGAAGAGCCACAATGGAGCTCTCGCTTGTTGA TTCGTGCTTCAATCCCCTATTTTAAGTTATCCGACGATCGTGAAAATTTTAT CCGCCTTATTATGGGATTGGCGAATCGCATGCCCTTTGAGACCCAAGCACG CCTTGAGACCAAACACACTATCTACGCTTTGTTTGCCGCGTGCTACGGGAG CCTTCGTGCACTGAAACAACCTGCTGGACGAATCAGTTAAGCAAGCGTTGG CCGCCCATGCCGAGACGCTGAAGCATGAGCACATTGCCGTGGCTTATGCAT TATTTTACCCGGATCAGGTCAATCCGTTTTTACAGCCGATTGACGAGATCA </p>
--	--

CRISPR array categorization

		<p>AGGCTTGTGAGGTTAAGCAATATAGTCGTTACGAAATCGATGCCGCCGGG</p> <p>AAAGAGGAAGTCCTGAACCCATTACAATTTACTGATAAAATCCCGATTTCT</p> <p>CAGCTGTTAAAAAACGATAATGACTCGAGGCCTCGTGATACGCCTATTTT</p> <p>TATAGG</p>
4	<i>tni</i> <i>Q^{As}</i>	<p>TTTTGGGCTAGCAGGAGGAATTCACCATGCACCTTCTTGTACGCCCCGGAAC</p> <p>CGTTTGCGGACGAGGCTTTGGAATCATACTTCTTACGTTTGTCACAGGAGA</p> <p>ATGGATTTGAGCGCTACCGCATTTTCTCCGGCTCCGTACAAGATTGGCTTC</p> <p>ATACAACAGATCACGCTGCCGCTGGCGCCTTCCCTCTTGAGCTGAGTCGTT</p> <p>TAAATATCTTTCACGCCTCTCGCAGTTCTGGTTTGCGCGTCCGTGCTCTTCA</p> <p>ATTGGTTGACCGTCTTACCGATGGCGCCCCCTTTCGCCTGCTTCAGCTGGC</p> <p>ATTATGTCATTCGGCAATTAGCTTCGGGAATCATTACAAGGCTGTTCATCG</p> <p>CAGTGGCGTGGATATCCCATATCTTTCATTCGTGTCCACCAAATTCCTTGC</p> <p>TGCCCCGATTGCCTTCGCGAGAGCGCTTATGTACGCCAATGTTGGCATT</p> <p>AAACCGTACGTGGGATGCCATCGCCACGGGGGCGCTTAATTTATTCATGT</p> <p>CCAGCATGTGGGGAGTCTTTGAACTATTTAGCCTCAGAATCGATTAATCAT</p> <p>TGCCAGTGCGGGTTCGACTTACGTACGGCCTCAACTGTTCCCGCACAGCCC</p> <p>GATGAAATTCAGTTGTCAGCCTTGGCCTACGGATGCTCGTTCGAGTCTTCC</p> <p>AATCCGCTTCTTGCGATCGGCTCTTTATCGGCCCGTTTCGGAGCTTTATACT</p> <p>GGTATCAGCAGCGTTACCTGTCGGATCACGAGGCTGTGCGTGACGATCGTG</p> <p>CTTTGACCAAGGCGATTGGTCACTTTACCGCTTGGCCCGATGCTTTCTGGC</p> <p>GTGAATTGCAACAGATGGTAGACGACGCCTTAGTACGCCAACTAAGCCG</p> <p>TTGAACCATACTGACTTTGTAGACGTCTTTGGCTCGGTTGTTGCTGATTGTC</p> <p>GTCAGATTCCCATGCGCAACACGGGCCAAAATTCATTTAAAAAACTTGA</p>

CRISPR array categorization

		TCGGATTTCTTACAGATTTGGTGGCACGCCACCCTCAGTGTCGCGTAGCAA ATGTCGGCGATCTTCTGCTGTCGGCAGTTGACGCGGCCACACTTCTTTCTA CTTCGGTCGAGCAGGTACGTCGTCTGCATCACGAGGGGTCTTACCCTTGT CTATCCGCCCTGCTAGTCGCAATACCGTCTCTCCGCATCGCGCAGTCTTCC ATTTACGCCACGTCGTCGAATTGCGTCAAGCACGCATGCAATCACATCATG ATCACAGTTCAACCTATCTACCTGCATGGTGACCA
5	Ca sca de ^A s gBl ock 1	AACCTATCTACCTGCATGGTGACCATCATGCACATCGAAGAGTTGTTGGAC ATCGAAGACCATGGGGAGCGTGATCGCCAATTGCGCCGTTACCTGGCGCC ATACTCTGCAGAGATCGGGGTCGATGGGGCCGAAAAGATGGCTTTAGTGG TCCTTTTGAATCTTACTCTGAAGCGCGACCGTGTTGAGAGCTTATGTGACG AGGGTCTGGCTCGTCAGTTACTGTCTGATGAAGGGCATATCACAACTGCC TGCACACGGTTCGTTGGTTGCACACACACAACTTAAAGTACCCGGATGCTC GTGTCTCAGGCGAGCGTTTAATTATTAACGCTCCGCCATTGATCCCGGGTG TAATCTCATCGGCAGGCTTACCGATGCGCATGGGGTGGGCCCATGACTCAA GCGATATCAACTTAGCCAAGCTGTTTCGGAACAAGTTTTCGCTATCGCGATG ACTCAACAAATTTAGCACTTCAGCTTGTGGCGCGTAGCAAACTTGGGAAC AAGCCTTAATTGGACTGGGACTGACTCAGCAGCAATTGGATATCTGGTGTC AACTTTTGGCATCTAATCTTGAAAATAACACATTTCCCACTGTTGTGTCCCC CTTTTCCAAACAGGTGCGCTTCTTGTATCAGGGTAACTATTGCGTGGTTAC GCCCCGTCGTCTCTCACGCCCTTCTTGCGCAATTGCAAAATGTGGTTCACGA AAAGAAGTTACAGTGTACGTACATTACCATGACCACCCCGCCTCAGTTGG CAGTTTAGTGGGGGCCTTGGGTGGTAAAGTCGCTGTATTAGATTACCCGCC CCCAGTTTCCCCGGATAAAGCTCGTAGCTTTAGCCAGGCCCGCAAACATCG

CRISPR array categorization

		<p>CTTGGCGAATGGTCAGAGCCTGTTTGATCGTTCCGTGTTTAACGACCATGT</p> <p>ATTCATCGACGCGTTAAAACATGTGATTAGTCGCCCCGGATTAACTCGCAA</p> <p>ACAACAACGTCAATTACGCCTTTCCGCATTGCGTTACCTTCGTCGCCAGTT</p> <p>GGCCATTTGGTTAGGCCCCATTATTGAGTGGCGTGATGAGATTGTTTCTAG</p> <p>CGGTCGTGGAGAACCAGGAAATCTGCCATCAGGAGGATTAGAGCTTGAGC</p> <p>TGATTACCCAACCAAAGAAGATGTTACCTGAGCTGATGCTTCAGGTTGCAG</p> <p>GGCGTTTTTCATTTAGAGCTTCAGAACCACAGTGCAGGACGCCGCTTCGCCT</p> <p>TCCATCCGGCGTTAATGGCTCCTATCAAATCTCAAATTCTTTGGCTGCTTCG</p> <p>CCAATTGGCCGACGATGAAGAAAAGGATGAACCTCACCCCCCACCAGTT</p> <p>GCTATTATTTACACCTTTCAGGGTTGACTGTGTATGACGCTTCAGCATTGGC</p> <p>GAACCCGTACCTGTGTGGAATCCCATCTTTGAGCGCTCTGGCTGGTTTCTG</p> <p>TCATGACTATGAGCGTCGCTTACAGTCCTTAATCGGGCAGAGCGTATATTT</p> <p>CCGTGGGTTGGCGTGGTACTTGGGGCGCTACTCCCTTGTTACCGGCAAGCA</p> <p>TTTACCCGAACCGTCGAAATCAGCTGACCCTAAGTCTGTTTCAGCGATTTCG</p> <p>CCGCCCAGGCTTGTTGGATGGGCGTTACTGCGACCTGGGAATGGACCTTAT</p> <p>TATCGAAGTTCACATTCCCACCGGCGGATCGTTGCCGTTACGACATGTCT</p> <p>GGACTTATTGCGTGTAAGCCCTGCCTGCACGTTTCGCTGGCGGCTGCCTTCA</p> <p>TCCCCCGTCCTTGACGAGGAATATAATTGGTGCACTGTCTACCAAGACAA</p> <p>GAGTACGCTGTTTACAGTCCTGTCACGTCTTCCGCGCTACGGCTGTTGGAT</p> <p>TTACCCCAAGTGATGCAGACTTGCGCAGCTTTGAAGAGCTGTCAGAAGC</p>
6	Ca sca de ^A	<p>GCAGCTTTGAAGAGCTGTCAGAAGCGTTAGCCTTGGACCGTCGCTTACGTC</p> <p>CGGTTGCTACAGGCTTTGTGTTTTTAGAAGAACCCGTGGAACGTGCAGGAT</p> <p>CGATCGAAGGCCAACACGTTTATGCCGAAAGTGCGATCGGAACCTGCCTTG</p>

CRISPR array categorization

s	<p>TGCATTAATCCAGTGGAGATGCGTCTGGCTGGTAAAAAGCGCTTCTTCGGC</p> <p>gBl GCAGGTTTTTGGCAGTTGAACGACGCTAAAGGTGCCATCCTGATGAATGGT</p> <p>ock TCAGCAAACACGGGGTAAATGCATGGAATTATGTACTCACTTGTCATATTC</p> <p>2 CCGCAGCCTGTCACCCGGAAGGCCGTCTTCTTCTACAAAACGGCCGAGTC</p> <p>AGACTTCGTCCCCTTACGTATCGAGGTTGCTAAGATCTCCGGACAAAAATG</p> <p>TGGCTATACAGAAGGTTTTGATGCAAATTTGAAGCCAAAAAACATTGAGC</p> <p>GCTACGAGTTGGCGTACAGTAACCCCCAACTATCGAGGCTTGTTACGTCC</p> <p>CGCCGAACGTAGATGAACTGTATTGCCGTTTCTCGCTTCGCGTCGAAGCTA</p> <p>ATTCCATGCGTCCCTATGTTTGTAGTAATCCAGATGTACTTCGTGTGATGAT</p> <p>TGGCCTGGCGCAGGCCTACCAGCGTCTGGGAGGCTACAACGAATTAGCTC</p> <p>GCCGTTATAGCGCGAATGTATTACGTGGAATTTGGTTATGGCGCAACCAAT</p> <p>ATACTCAGGGGACCAAAATTGAAATCAAAACATCATTGGGTTCTACCTATC</p> <p>ACATCCCAGACGCTCGCCGCCTTTCTTGGTCGGGCGACTGGCCTGAACTGG</p> <p>AACAAAAGCAACTTGAGCAGTTGACCTCAGAGATGGCGAAAGCTCTTTTCG</p> <p>CAGCCTGACATCTTTTGGTTTGCGGACGTTACCGCGTCACTGAAGACGGGT</p> <p>TTTTGCCAGGAAATCTTTCCTTACAAAAATTTACAGAGCGTCCTGATGAT</p> <p>CACTCCGTTGCCTCCCGCCAGCTTGCCACCGTGGAATGTTTCAGATGGACAA</p> <p>CTGGCAGCCTGCATCAACCCGCAGAAGATCGGCGCGGCTTTGCAAAAGAT</p> <p>CGATGACTGGTGGGCGAATGACGCGGATCTGCCTCTGCGTGTCCACGAAT</p> <p>ACGGGGCTAATCATGAAGCATTAACAGCGTTACGCCACCCTGCAACGGGT</p> <p>CAAGATTTCTACCACCTTCTTACTAAGGCCGAACAATTTGTGACGGTGCTT</p> <p>GAATCCTCTGAAGGAGGGGGCGTTGAGCTGCCGGGAGAGGTACACTATTT</p> <p>GATGGCAGTGTTGGTGAAGGGTGGATTGTTTCAGAAAGGGAAGGGGAGAT</p>
---	--

CRISPR array categorization

		<p>GACCGAAAATCGCTACTTTTTTCGCCATCCGTTATTTGAGTGACGATGTGGA</p> <p>CTGTGGTCTTTTAGCTGGACGTTGCATTTCGATTCTTCATGGGTTTCGCCAA</p> <p>GCGCACCTGGCATTCAAATCGGTGTCGCTTTCCCCGAATGGAGTGATCGC</p> <p>GACCTGGGTCGTAGCATCGCTTTTGTTCGACGAACAAGAGTCTGTTAGAA</p> <p>CGTTTCCGTGAGCGTTCCTACTTTCAGGTCATGCAGGCGGACAATTTCTTC</p> <p>GCGCTGTCTTTGGTACTGGAAGTCCCCGATACTTGTCAGAACGTTTCGCTTT</p> <p>ATCCGTAATCAAACTTAGCCAAGTTGTTTCGTCGGCGAACGCCGTCGTCGT</p> <p>TTGGCGCGTGCCAAGCGCCGCGCTAAGGCCCGCGGTGAAGCCTTCCAACC</p> <p>CCATATGCCGGATGAGACAAAAGTAGTTGGTGTCTTCCACTCCGTATTCAT</p> <p>GCAATCAGCAAGCTCTGGTCAGTCTTACATCCTGCATATTCAGAAACATCG</p> <p>TTATGAGCGTTCGGAGGATTTCGGGTTACTCGAGTTATGGCTTAGCAAGTAA</p> <p>CGACTTATATACAGGTTATGTCCCAGATCTTGGTGCAATCTTCTCTACGCTG</p> <p>TTTAAAGCTTGGCTGTTTTGGCGGATGAGAGAAG</p>
7	CR	GAATTGAGCTCGGTACCCGGGGATGTGAACTGCCGTATAGGCAGCCAAG
	ISP	AAATGAGACCCACGATTTGTAATAACAGAGTGTCTTGTATTTTAAAGAAA
	R	GTCTATTTAATAACAAGTGATTATATTAATTAACGGTAAGCATCAGCGGGTG
	arr	ACAAAACGAGCATGCTTACTAATAAAATGTTAACCTCTGAGGAAGAATTG
	ay	TGAAACTATCACTAATGGTAGCTATATCGAAGAATGGAGTTATCGGGAAT
	As	GGCCCTGATATTCCATGGAGTGCCAAAGGTGAACAGCTCCTGTTTAAAGCT
	gBl	ATTACCTATAACCAATGGCTGTTGGTTGGACGCAAGACTTTTGAATCAATG
	ock	GGAGCATTACCCAACCGAAAGTATGCGGTCGTAACACGTTCAAGTTTTACA
	1	<p>TCTGACAATGAGAACGTATTGATCTTCCATCAATTAAAGATGCTTTAACC</p> <p>AACCTAAAGAAAATAACGGATCATGTCATTGTTTCAGGTGGTGGGGAGAT</p>

CRISPR array categorization

		ATACAAAAGCCTGATCGATCAAGTAGATACACTACATATATCTA
8	CR	CAAGTAGATACACTACATATATCTACAATAGACATCGAGCCGGAAGGTGA
	ISP	TGTTTACTTTTCTGAAATCCCCAGCAATTTTAGGCCAGTTTTTACCCAAGAC
	R	TTCGCCTCTAACATAAATTATAGTTACCAAATCTGGCAAAGGGTTAACAA
	arr	GTGGCAGCAACGGATTTCGCAAACCTGTCACGCCTTTTGTGCCAAAAGCCGC
	ay	GCCAGGTTTTCGATCCGCTGTGCCAGGCGTTAGGCGTCATATGAAGATTTC
	As	GGTGATCCCTGAGCAGGTCTCGTGAAGTGGCGTATAGGCAGCCAAGATTA
	gBl	GGACTGGAAGAAATCATCCAAGTTGGGGACTATTTTCTGCCGTAAAGGCA
	ock	GATATTATTATGTAGCCCAGTCTTTGCGAGAGTAGGGAAGTGGCAGGCA
	2	
9	<i>xre</i>	GAGGCGCGGCCGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAG
	As	GCTTTACACTTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGAT
		AACAATTTCAAGGAGGTAACAGCTATGACTAATCCATTACCCATTCGCCTT
		AAAGCCGCTCGCAAGGCTACTGGACTGACTCAACAACAGTTAGGTATCCG
		CTTGGGAATGGAACAATCAACCGCATCTGCCCCGCATGAACCAGTACGAGA
		AAGGGAAACATGCGCCTGACTATCAAACCATGCAACGCATCGCCCAGGAG
		CTGGGCTATCCCGTTGCGTACTTCTATTGTGATGACGAGCTTCTGGCGGAA
		CTGATCTGCATGATGGCCAACTTTCCGAGGAGAAACAGCGCGAATTGCTT
		CAACAGTTGAGTGTAAGTGAATATGCAGAGTCGCGTGACTCGGCAGAGTG
		ATAAGCGGCCGCGGGCG
1	<i>xre</i>	CGCCAGGGTTTTCCAGTCACGACCTCGAGGCTCTGGCAGGATGTTTCGTA
0	VB6	ATTTCCGCTGAGCATTAAAACTAAGTATTGCTTAGTACACTTGTTTCGTAT
	,	AAGACTAAGAGAATCTTAGCTACCTGATCCTCAGCTAGCAGAGACCCCAT

CRISPR array categorization

	pX	GACAAACCCATTACCCATCAGAAGGCCTCATAAGAACACCGAAACTAAAT
	re ^V	ATTACTTAGTTTTTAGGTGTAGCACATATCCTTGGTATCACCATTTCGAGCA
	B ⁶ ,	GACACACTTTTAGCTCAGCAGCACACCGAAAACTCAACAATGATGCGAC
	pT	TCGAGGGTTCGCCAATGCTGAAGTGACTCTTACAGGCAATAAAAACTTAG
	ni	GATTGCTAAGTATACAAAAATCAGTAAACTTAGACAATCTAAGTTATCTA
	Q ^V	AAGGTTGAAAAGAAATGTGTAAGTGCAAAAAGACAACCCCATCCCAGCGA
	B ⁶ ,	GGCCTCACATGCCACCACTAAACTTAGAGTATCTAAGTTTTAGTATTGCC
	pX	GTTAATCCCGATAATTCAAATTCTATGGAGCCAAATTGTGAACTTTAATAA
	re ^A	ACTAACTATTGCGTGGTTACGCCCTCGAGTCCTGTGTGAAATTGTTATCCG
	s,	CTGCCTGAACTGGAACAAAAGCAAGCTTTTAACTTCTCAGCTTTTCCAACA
	pT	ACACCGCCTTCTCTTCATCACTCATCTTATCAATTGCGCACGCTAACTCAGC
	ni	GCTCAATTGCTCCTACAGAAAAAGTAGTTTAGCGGCACATCTAACTCTTC
	Q ^{As}	AGCCATGCGCTCAAGCGTACCAATATCAGGTACATGCCTGCCTTTCTCGTA
		ATGATTCATACGACCGCTAGCAGAACTTTGTTCCATACCAATTTTCACCCC
		TAAATCTTTTTGGGTGATCTTGGCTTTTTTACGCGCAGCTTTGAGCCTCGCC
		GGGATGGGGTTATCCTTCTGCATATGCAATAAACCGGTAAACCAGCA
1	<i>xre</i>	CGCCAGGGTTTTCCCAGTCACGACCTCGAGGCTCTGGCAGGATGTTTCGTA
1	V ^c ,	ATTTCCGCTGAGCATTAAAACTAAGTATTGCTTAGTACACTTGTTTCGTAT
	pX	AAGAAGAAGAGAATCTTAGCTACCTGATCCTCAGCTAGCAGAGACCCCAT
	re ^V	GACAAACCCATTACCCATCAGAAGGCCTCATAAGAACACCGAAACTAAAT
	c,	ATTACTTCTTTTTTAGGTGTAGCACATATCCTTGGTATCACCATTTCGAGCAG
	pA	ACACACTTTTAGCTCAGCAGCACACCGAAAACTCAACAATGATGCGACT
	rra	CGAGGGTTCGCCAATGCTGAAGCACACATCAACACCGTTACTTTCCAAACA

CRISPR array categorization

	y2	GCCACTCGGCTGTTTGTTCACCCCTAAGTGAGGATTATAATCCTCATTACT
	V _c ,	ACTGCAAAGTAGCTGATAACCACACAGGAAACAGCTATGAGGCCTTCCTG
	pX	TGTGAAATTGTTATCCGCTTAATACGACTCACTATAGGGCTCGAGTTAGAT
	re*	ATCCATTGGTCACATTCAAAGTGTTTTTTTTTGC GGATTATAGTCCGTGGCAA
	As,	TGTTTGCATCAAGCGTAGATAGTTTTTCATTTTGTGATTTCGATTAAATGAG
	pT	GCCTTAATCGATGAGACCCTATAGTGTCACCTAAATAGCTGCTGGTTTACC
	ni	GGTTTATTGGGTCTCACTAGTCTAGAGTCACACAGGACTACTATATGAAAG
	Q*	ACTTAGCGAAACGGTTTGGCACCAATCTACGCAATAAACGCCGAGATAAA
	As,	GGTGTGTCCCAAGACCAACTAGCGTTGCAAGCAAACATTGACAGAAGCTA
		TGTTGGGCGAATTGAACGAGGAGAAGTAAACATCACACTCGAAAAAGCCT
		ATCAATTAGCGCATGTACTTGGGTGTGATGTTAGAGAGCTTCTTCCGTAAG
		CTTGCTTTTGTTCAGTTCAGGC
1	CR	AATGGTCTATCCTTAGCAAGGTAAAACAAGCAGTTAGAGAAACACGCAGA
2	ISP	ATGATGGTCAAATTACAGCGATGCTATTTTTCCCTTTTCATTCAAAGTATTA
	R	CGTTTGCATGCGCACAGTGAAGTCCGTATAGGCAGCCAAGAAATGAGAC
	arr	CCACGATTTGTAATAACAGAGTGTCTTGTATTTTTAAAGAAAGTCTATTTA
	ay	ATACAAGTGATTATATTAATTAACGGTAAGCATCAGCGGGTGACAAAACG
	(wi	AGCATGCTTACTAATAAAATGTAAACCTCTGAGGAAGAATTGTGAAACTAT
	th	CACTAATGGTAGCTATATCGAAGAATGGAGTTATCGGGAATGGCCCTGAT
	lea	ATTCCATGGAGTGCCAAAGGTGAACAGCTCCTGTTTAAAGCTATTACCTAT
	der	AACCAATGGCTGTTGGTTGGACGCAAGACTTTTGAATCAATGGGAGCATTA
	)	CCCAACCGAAAGTATGCGGTCGTAACACGTTCAAGTTTTACATCTGACAAT
	As	GAGAACGTATTGATCTTCCATCAATTAAAGATGCTTTAACCAACCTAAAG

CRISPR array categorization

	gBl ock 1	AAAATAACGGATCATGTCATTGTTTCAGGTGGTGGGGAGATATACAAAAG CCTGATCGATCAAGTAGATACACTACATATATCTA
1 3	CR ISP R lea der PA1 4, CR ISP R lea der Vc	TGAGATGCGCAAGAAGGCCGGTCTCAAATCCCATTTTTCTGGCTTGTTA CAACCATATGAAATCATTAATGTTTTAATATACTGTAAGAAAAAGGGAAA AAAGCACTTCTTAACATCTAACTAGTTGAAAATAATGAGCTAAACTTTAAG TGGTAACGGTGAAGTACCGAGTAGGTAGCTGATAACAGGACCGGAAGGTA GCAGCCAAGGCGGGGACGTGAATGAGACCCCTGAAGCCCGTTCTGGACC GCCAGGGTTTTCCCAGTCACGACGGTCTCAAATTCGACCCCTTTTTTCGG ACGATTTCTTACGCCCTTATAAATCAGCAAGTTACGAGGCCTCGAAAAAAG AGGGTTTCTGGCGGGAAAACTCGGTATTTCTTTTTCCTTCAAATGGTTAT AGGTTTTTCGGAGCTAGTTCACTGCCGTGTAGGCAGCTAAGAAATGAGACCT CCTGTGTGAAATTGTTATCCGCT

829

830

CRISPR array categorization

Figure legends

Figure 1 - Tn7-like elements with I-F3 CRISPR-Cas systems found in gamma proteobacteria.

A TnsA protein similarity tree representing 801 elements indicated by host strain. Elements that were >90 percent identical are indicated with a single representative. A similarity score was calculated for repeats (Experimental procedures) and indicated as black as the highest score and lighter shades of grey as the score decreases. Spacers matching the protospacer found adjacent to the point of insertion are indicated by color. Width of rectangle indicating the spacers on the figure is based on predicted spacer length (Supplementary Table S1). See text for details.

Figure 2 – Selected representatives from four *att*-site families of Tn7-like elements with I-F3 CRISPR-Cas systems.

Representatives for three major families (*att* sites; *yciA*, *guaC*, and *ffs*) and one minor family (*rsmJ*) are indicated by host. (a) Transposition genes (*tnsA*, *tnsB*, *tnsC*, and *tniQ/tnsD*), Cas genes *cas6*, *cas7*, and *cas8/5*, and regulator *xre* are indicated. The CRISPR arrays are indicated as in Figure 1. The left (L) and Right (R) ends of the elements are indicated and putative Xre binding sites (red asterisks). (b) Matches between the guide RNA and protospacer are shown in detail with the gene (colored block arrow), right end of the element (grey box), with host indicated. Distance from the protospacer to the target site duplication (TSD, bold red) is shown. (c) CRISPR array is indicated with the leader region and spacer (S#) and repeats (R#) indicated showing the sequence of the repeats. Sequence differences from the first repeat (red), noting changes maintaining the stem (light blue) and the inverted repeats that makes the stem (boxed

CRISPR array categorization

orange) are indicated. The size of the gap in the array is indicated noting the putative Xre regulatory site.

Figure 3 – I-F3b element derived from *A. salmonicida* S44 allows RNA-guided transposition with typical and atypical repeats

Various transposition targets were tested using the *A. salmonicida* S44 native array or as individual repeat-spacer-repeat units. (a) Simplified representation of transposition/CRISPR-associated genes, CRISPR array (marked as in Figure 1) and the resulting typical and atypical guide RNAs with the 5' and 3' handles indicated. Position of Cas6 processing is indicated (scissors). (b) Frequency of transposition found with the native *A. salmonicida* S44 array with targets constructed into an F plasmid; *A. salmonicida* S44 plasmid pS44-1 (pS44-1), chromosomal *ffs att* site (*ffs*) or a negative control, *lacZ* gene. (*lacZ*) (c-e) Transposition frequency found with a single repeat-spacer-repeat unit in various combinations of spacers with typical or atypical repeats from *A. salmonicida* with the indicated targets constructed into an F plasmid. Transposition assay was as described in the Experimental procedures and the average of three experiments is indicated +/- one standard deviation (n=3).

Figure 4 – Analysis of atypical repeat sequences

(a) Consensus sequence of the typical and atypical as a function of position in the array. Symbols are as in previous Figures with the stem-loop indicated (top, n=85 for I-F3a, n=74 for I-F3b), (middle, n=51 for I-F3a, n=41 for I-F3b) or (bottom, n=51 for I-F3a, n=41 for I-F3b). (b) Frequency of transposition found with changes in the *A. salmonicida* atypical guide RNAs. (c-d) Frequency of transposition with various typical and atypical repeat combinations from Tn6677

CRISPR array categorization

with indicated spacers and their associated targets. Typical guide RNA were tested, comparing with naturally-occurring changes from atypical repeats (red) or engineered mutations (underlined) with the indicated targets constructed into an F plasmid. Transposition assay was as described in the Methods and the average of three experiments is indicated +/- one standard deviation (n=3).

Figure 5 – *P. aeruginosa* type I-F1 Cascade can utilize heterologous I-F3 CRISPR arrays in a plasmid interference assay, but I-F3b atypical guide RNAs are privatized

Expression of *P. aeruginosa* protein components with various arrays reduces transformation efficiency for protospacer containing plasmid (*ffs*), but not control (empty). Single unit arrays from *P. aeruginosa* PA14, *A. salmonicida* S44, and *V. cholerae* HE-45 Tn6677 were tested in various typical and atypical repeat configurations as indicated. Repeat configuration sequences are presented in Supplemental Figure S3b. All data indicate mean +/- standard deviation of three biological replicates.

Figure 6 – Xre proteins regulate components of RNA-guided transposition

(a-b) Consensus sequence for putative Xre binding motifs in I-F3a and I-Fb elements. (c-d) Xre-binding resolved by EMSA. DNA fragments with the transcription control regions were incubated with increasing amounts of Xre protein from the respective element before electrophoresis (100 nM DNA; protein:DNA ratios = 0,2,5,10,20:1). (e-f) Promoter function resolved by LacZ expression monitored by Miller units at various arabinose controlled Xre expression levels. Transcription control regions are indicated from *V. cholerae* HE-45 Tn6677

CRISPR array categorization

- 898 (Vc), *V. parahaemolyticus* RIMD221063 (Vp), *A. salmonicida* S44 (As), and *Vibrio* sp.
899 10N.286.45.B6 (VB6).

CRISPR array categorization

References

- Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research* 37, W202-W208.
- Bainton, R., Gamas, P., and Craig, N.L. (1991). Tn7 transposition in vitro proceeds through an excised transposon intermediate generated by staggered breaks in DNA. *Cell* 65, 805-816.
- Bainton, R.J., Kubo, K.M., Feng, J.-N., and Craig, N.L. (1993). Tn7 transposition: target DNA recognition is mediated by multiple Tn7-encoded proteins in a purified in vitro system. *Cell* 72, 931-943.
- Bogdanova, E., Djordjevic, M., Papapanagiotou, I., Heyduk, T., Kneale, G., and Severinov, K. (2008). Transcription regulation of the type II restriction-modification system AhdI. *Nucleic Acids Res* 36, 1429-1442.
- Borges, A.L., Davidson, A.R., and Bondy-Denomy, J. (2017). The Discovery, Mechanisms, and Evolutionary Impact of Anti-CRISPRs. *Annu Rev Virol* 4, 37-59.
- Bushnell, B. BBMap.
- Chowdhury, S., Carter, J., Rollins, M.F., Golden, S.M., Jackson, R.N., Hoffmann, C., Nosaka, L., Bondy-Denomy, J., Maxwell, K.L., Davidson, A.R., *et al.* (2017a). Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex. *Cell* 169, 47-57 e11.
- Chowdhury, S., Carter, J., Rollins, M.F., Golden, S.M., Jackson, R.N., Hoffmann, C., Nosaka, L., Bondy-Denomy, J., Maxwell, K.L., Davidson, A.R., *et al.* (2017b). Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex. *Cell* 169, 47-57.e11.

CRISPR array categorization

923 Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I.,
924 Hamelryck, T., Kauff, F., Wilczynski, B., *et al.* (2009). Biopython: freely available Python tools
925 for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422-1423.
926 Cronan, J.E. (2006). A family of arabinose-inducible *Escherichia coli* expression vectors having
927 pBR322 copy control. *Plasmid* 55, 152-157.
928 Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo
929 generator. *Genome research* 14, 1188-1190.
930 Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in
931 *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of*
932 *the United States of America* 97, 6640-6645.
933 Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high
934 throughput. *Nucleic Acids Research* 32, 1792-1797.
935 Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning
936 method with high throughput capability. *PloS one* 3, e3647-e3647.
937 Faure, G., Makarova, K.S., and Koonin, E.V. (2019a). CRISPR-Cas: Complex Functional
938 Networks and Multiple Roles beyond Adaptive Immunity. *J Mol Biol* 431, 3-20.
939 Faure, G., Shmakov, S.A., Yan, W.X., Cheng, D.R., Scott, D.A., Peters, J.E., Makarova, K.S.,
940 and Koonin, E.V. (2019b). CRISPR-Cas in mobile genetic elements: counter-defence and
941 beyond. *Nat Rev Microbiol* 17, 513-525.
942 Fineran, P.C., Gerritzen, M.J., Suarez-Diez, M., Kunne, T., Boekhorst, J., van Hijum, S.A.,
943 Staals, R.H., and Brouns, S.J. (2014). Degenerate target sites mediate rapid primed CRISPR
944 adaptation. *Proc Natl Acad Sci U S A* 111, E1629-1638.

CRISPR array categorization

945 Finn, J.A., Parks, A.R., and Peters, J.E. (2007). Transposon Tn7 directs transposition into the
946 genome of filamentous bacteriophage M13 using the element-encoded TnsE protein. *J Bacteriol*
947 *189*, 9122-9125.

948 Gleditsch, D., Muller-Esparza, H., Pausch, P., Sharma, K., Dwarakanath, S., Urlaub, H., Bange,
949 G., and Randau, L. (2016). Modulating the Cascade architecture of a minimal Type I-F CRISPR-
950 Cas system. *Nucleic Acids Res* *44*, 5872-5882.

951 Guzman, L.M., Belin, D., Carson, M.J., and Beckwith, J. (1995). Tight regulation, modulation,
952 and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* *177*,
953 4121-4130.

954 Hoyland-Kroghsbo, N.M., Paczkowski, J., Mukherjee, S., Broniewski, J., Westra, E., Bondy-
955 Denomy, J., and Bassler, B.L. (2017). Quorum sensing controls the *Pseudomonas aeruginosa*
956 CRISPR-Cas adaptive immune system. *Proc Natl Acad Sci U S A* *114*, 131-135.

957 Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: Reconstruction, Analysis, and
958 Visualization of Phylogenomic Data. *Molecular biology and evolution* *33*, 1635-1638.

959 Jackson, R.N., Golden, S.M., van Erp, P.B., Carter, J., Westra, E.R., Brouns, S.J., van der Oost,
960 J., Terwilliger, T.C., Read, R.J., and Wiedenheft, B. (2014). Structural biology. Crystal structure
961 of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science* *345*, 1473-
962 1479.

963 Johnson, L.S., Eddy, S.R., and Portugaly, E. (2010). Hidden Markov model speed heuristic and
964 iterative HMM search procedure. *BMC Bioinformatics* *11*, 431.

965 Khlebnikov, A., Datsenko, K., Skaug, T., Wanner, B., and Keasling, J. (2002). Homogeneous
966 expression of the PBAD promoter in *Escherichia coli* by constitutive expression of the low-
967 affinity high-capacity AraE transporter. *Microbiology (Reading, England)* *147*, 3241-3247.

CRISPR array categorization

968 Klompe, S.E., Vo, P.L.H., Halpin-Healy, T.S., and Sternberg, S.H. (2019a). Transposon-encoded
969 CRISPR-Cas systems direct RNA-guided DNA integration. *Nature* 517, 219-225.

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

972 Kovach, M.E., Elzer, P.H., Steven Hill, D., Robertson, G.T., Farris, M.A., Roop, R.M., and
973 Peterson, K.M. (1995). Four new derivatives of the broad-host-range cloning vector
974 pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* 166, 175-176.

975 Kuznedelov, K., Mekler, V., Lemak, S., Tokmina-Lukaszewska, M., Datsenko, K.A., Jain, I.,
976 Savitskaya, E., Mallon, J., Shmakov, S., Bothner, B., *et al.* (2016). Altered stoichiometry
977 Escherichia coli Cascade complexes with shortened CRISPR RNA spacers are capable of
978 interference and primed adaptation. *Nucleic Acids Res* 44, 10849-10861.

979 Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of
980 protein or nucleotide sequences. *Bioinformatics* 22, 1658-1659.

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

984 Malke, H. (1969). R. C. Clowes and W. Hayes (Eds.), *Experiments in Microbial Genetics*. 244
985 S., 14 Abb. Oxford - Edinburgh 1968: Blackwell Scientific Publications. 42 s. Zeitschrift für
986 allgemeine Mikrobiologie 9, 407-408.

987 Malke, H. (1993). Jeffrey H. Miller, *A Short Course in Bacterial Genetics – A Laboratory*
988 *Manual and Handbook for Escherichia coli and Related Bacteria*. Cold Spring Harbor 1992. Cold
989 Spring Harbor Laboratory Press. ISBN: 0-87969-349-5. *Journal of Basic Microbiology* 33, 278-
990 278.

CRISPR array categorization

991 Martynov, A., Severinov, K., and Ispolatov, I. (2017). Optimal number of spacers in CRISPR
992 arrays. *PLoS Comput Biol* 13, e1005891.

993 McGeehan, J.E., Papapanagiotou, I., Streeter, S.D., and Kneale, G.G. (2006). Cooperative
994 binding of the C.AhdI controller protein to the C/R promoter and its role in endonuclease gene
995 expression. *J Mol Biol* 358, 523-531.

996 McGeehan, J.E., Streeter, S.D., Thresh, S.J., Taylor, J.E., Shevtsov, M.B., and Kneale, G.G.
997 (2011). Structural analysis of a novel class of R-M controller proteins: C.Csp231I from
998 *Citrobacter* sp. RFL231. *J Mol Biol* 409, 177-188.

999 Mitra, R., McKenzie, G.J., Yi, L., Lee, C.A., and Craig, N.L. (2010). Characterization of the
1000 TnsD-attTn7 complex that promotes site-specific insertion of Tn7. *Mobile DNA* 1, 18.

1001 Mulepati, S., Heroux, A., and Bailey, S. (2014). Structural biology. Crystal structure of a
1002 CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science* 345, 1479-1484.

1003 Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches.
1004 *Bioinformatics* 29, 2933-2935.

1005 Parks, A.R., Li, Z., Shi, Q., Owens, R.M., Jin, M.M., and Peters, J.E. (2009). Transposition into
1006 replicating DNA occurs through interaction with the processivity factor. *Cell* 138, 685-695.

1007 Patterson, A.G., Jackson, S.A., Taylor, C., Evans, G.B., Salmond, G.P.C., Przybilski, R., Staals,
1008 R.H.J., and Fineran, P.C. (2016). Quorum Sensing Controls Adaptive Immunity through the
1009 Regulation of Multiple CRISPR-Cas Systems. *Mol Cell* 64, 1102-1108.

1010 Peters, J.E. (2007). Gene Transfer in Gram-Negative Bacteria. In *Methods for General and*
1011 *Molecular Microbiology*, Third Edition (American Society of Microbiology).

1012 Peters, J.E. (2014). Tn7. *Microbiology Spectrum* 2, 1-20.

CRISPR array categorization

1013 Peters, J.E. (2015). Tn7. In *Mobile DNA III*, L. Craig Nancy, P. Rice, A. Lambowitz, M. Gellert,
1014 and S.B. Sandmeyer, eds. (Washington DC: ASM Press), p. In Press.

1015 Peters, J.E. (2019). Targeted transposition with Tn7 elements: safe sites, mobile plasmids,
1016 CRISPR/Cas and beyond. *Mol Microbiol* 112, 1635-1644.

1017 Peters, J.E., and Craig, N.L. (2000). Tn7 transposes proximal to DNA double-strand breaks and
1018 into regions where chromosomal DNA replication terminates. *Mol Cell* 6, 573-582.

1019 Peters, J.E., and Craig, N.L. (2001). Tn7 recognizes target structures associated with DNA
1020 replication using the DNA binding protein TnsE. *Genes & Dev* 15, 737-747.

1021 Peters, J.E., Makarova, K.S., Shmakov, S., and Koonin, E.V. (2017a). Recruitment of CRISPR-
1022 Cas systems by Tn7-like transposons. *Proceedings of the National Academy of Sciences* 114,
1023 E7358.

1024 Peters, J.E., Makarova, K.S., Shmakov, S., and Koonin, E.V. (2017b). Recruitment of CRISPR-
1025 Cas systems by Tn7-like transposons. *Proc Natl Acad Sci U S A* 114, E7358-E7366.

1026 Phillips, G.J. (1999). New Cloning Vectors with Temperature-Sensitive Replication. *Plasmid* 41,
1027 78-81.

1028 Price, M.N., Dehal, P.S., and Arkin, A.P. (2009). FastTree: Computing Large Minimum
1029 Evolution Trees with Profiles instead of a Distance Matrix. *Molecular biology and evolution* 26,
1030 1641-1650.

1031 Rezulak, M., Borsuk, I., and Mruk, I. (2016). Natural C-independent expression of restriction
1032 endonuclease in a C protein-associated restriction-modification system. *Nucleic Acids Res* 44,
1033 2646-2660.

1034 Rice, P.A., Craig, N.L., and Dyda, F. (2020). Comment on "RNA-guided DNA insertion with
1035 CRISPR-associated transposases". *Science* 368.

CRISPR array categorization

1036 Rodic, A., Blagojevic, B., Zdobnov, E., Djordjevic, M., and Djordjevic, M. (2017).
1037 Understanding key features of bacterial restriction-modification systems through quantitative
1038 modeling. *BMC Syst Biol* *11*, 377.

1039 Rollins, M.F., Chowdhury, S., Carter, J., Golden, S.M., Wilkinson, R.A., Bondy-Denomy, J.,
1040 Lander, G.C., and Wiedenheft, B. (2017). Cas1 and the Csy complex are opposing regulators of
1041 Cas2/3 nuclease activity. *Proc Natl Acad Sci U S A* *114*, E5113-e5121.

1042 Shevtsov, M.B., Streeter, S.D., Thresh, S.J., Swiderska, A., McGeehan, J.E., and Kneale, G.G.
1043 (2015). Structural analysis of DNA binding by C.Csp231I, a member of a novel class of R-M
1044 controller proteins regulating gene expression. *Acta Crystallogr D Biol Crystallogr* *71*, 398-407.

1045 Shi, Q., Straus, M.R., Caron, J.J., Wang, H., Chung, Y.S., Guarne, A., and Peters, J.E. (2015).
1046 Conformational toggling controls target site choice for the heteromeric transposase element Tn7.
1047 *Nucleic Acids Res.*

1048 Sibley, M.H., and Raleigh, E.A. (2012). A versatile element for gene addition in bacterial
1049 chromosomes. *Nucleic Acids Res* *40*, e19.

1050 Stellwagen, A.E., and Craig, N.L. (1998). Mobile DNA elements: controlling transposition with
1051 ATP-dependent molecular switches. *Trends Biochem Sci* *23*, 486-490.

1052 Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J.L., Makarova, K.S., Koonin, E.V., and
1053 Zhang, F. (2019). RNA-guided DNA insertion with CRISPR-associated transposases. *Science*
1054 *365*, 48-53.

1055 Strecker, J., Ladha, A., Makarova, K.S., Koonin, E.V., and Zhang, F. (2020). Response to
1056 Comment on "RNA-guided DNA insertion with CRISPR-associated transposases". *Science* *368*.

CRISPR array categorization

1057 Streeter, S.D., Papapanagiotou, I., McGeehan, J.E., and Kneale, G.G. (2004). DNA footprinting
1058 and biophysical characterization of the controller protein C.AhdI suggests the basis of a genetic
1059 switch. *Nucleic Acids Res* 32, 6445-6453.

1060 Vorontsova, D., Datsenko, K.A., Medvedeva, S., Bondy-Denomy, J., Savitskaya, E.E., Pougach,
1061 K., Logacheva, M., Wiedenheft, B., Davidson, A.R., Severinov, K., *et al.* (2015). Foreign DNA
1062 acquisition by the I-F CRISPR-Cas system requires all components of the interference
1063 machinery. *Nucleic Acids Res* 43, 10848-10860.

1064 Waddell, C.S., and Craig, N.L. (1988). Tn7 transposition: two transposition pathways directed by
1065 five Tn7-encoded genes. *Genes Dev* 2, 137-149.

1066 Westra, E.R., Pul, U., Heidrich, N., Jore, M.M., Lundgren, M., Stratmann, T., Wurm, R., Raine,
1067 A., Mescher, M., Van Heereveld, L., *et al.* (2010). H-NS-mediated repression of CRISPR-based
1068 immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. *Mol*
1069 *Microbiol* 77, 1380-1393.

1070 Wheeler, T.J., and Eddy, S.R. (2013). nhmmer: DNA homology search with profile HMMs.
1071 *Bioinformatics* 29, 2487-2489.

1072 Wiegand, T., Karambelkar, S., Bondy-Denomy, J., and Wiedenheft, B. (2020). Structures and
1073 Strategies of Anti-CRISPR-Mediated Immune Suppression. *Annu Rev Microbiol*.

1074 Wolkow, C.A., DeBoy, R.T., and Craig, N.L. (1996). Conjugating plasmids are preferred targets
1075 for Tn7. *Genes & Dev* 10, 2145-2157.

1076 Xiao, Y., Ng, S., Nam, K.H., and Ke, A. (2017). How type II CRISPR-Cas establish immunity
1077 through Cas1-Cas2-mediated spacer integration. *Nature* 550, 137-141.

CRISPR array categorization

1078 Zhao, H., Sheng, G., Wang, J., Wang, M., Bunkoczi, G., Gong, W., Wei, Z., and Wang, Y.
 1079 (2014). Crystal structure of the RNA-guided immune surveillance Cascade complex in
 1080 Escherichia coli. *Nature* 515, 147-150.
 1081

B

C

Target plasmid

control
protospacer

Guide : empty
array