RESEARCH ARTICLE SUMMARY

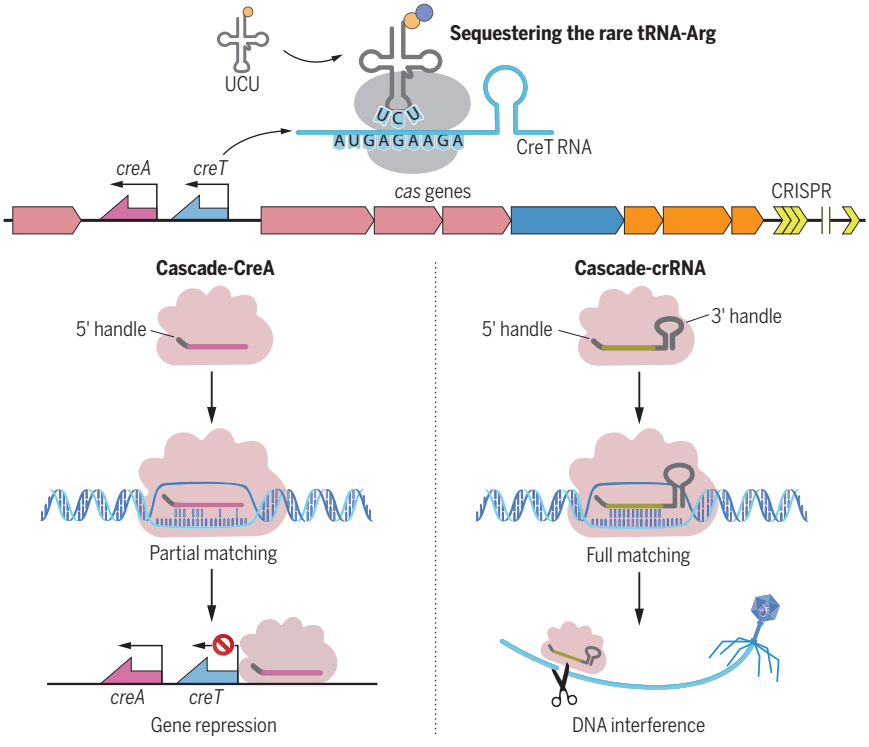
CRISPR

Toxin-antitoxin RNA pairs safeguard CRISPR-Cas systems

Ming Li†*, Luyao Gong†, Feiyue Cheng†, Haiying Yu, Dahe Zhao, Rui Wang, Tian Wang, Shengjie Zhang, Jian Zhou, Sergey A. Shmakov, Eugene V. Koonin, Hua Xiang*

INTRODUCTION: CRISPR-Cas systems efficiently protect bacteria and archaea from viruses and other types of foreign DNA, but, characteristically of defense systems, they also impart non-negligible fitness costs on the host, for example, the risk of autoimmunity and the repulsion to exogenous beneficial genes. Presumably, these costs result in frequent loss of CRISPR-Cas in bacteria, which is reflected in its patchy distribution, even among closely related bacterial strains. Nevertheless, in the current genome sequence databases, ~40% of bacterial and ~90% of archaeal genomes carry CRISPR-*cas* loci, suggesting the possibility that in addition to the direct benefits of adaptive immunity, mechanisms might exist that mitigate the costs of CRISPR systems and prevent their loss.

RATIONALE: We specifically looked into an archaeal type I-B CRISPR-Cas, where the genes encoding the subunits of the CRISPR effector Cascade cannot be deleted individually but can be readily deleted as a whole, including a 311–base pair intergenic region. These observations suggest that the Cascade gene cassette (*cas6-cas8-cas7-cas5*) includes a toxic component that makes it addictive to the host (elicits cellular toxicity once any of the *cascade* genes is deleted). We cloned and extensively analyzed the intergenic region between *cas6* and *cas8*, which allowed us to identify the Cascade-repressed toxin gene *creT*, along with an associating CRISPR repeat–like sequence that appears to be required for transcriptional repression of *creT*. We hypothesized that the repeat-like sequence is part of a CRISPR RNA–resembling antitoxin (CreA) RNA, which represses the toxin jointly with Cascade. We reasoned that CreTA would make the *cascade* genes addictive for the host.



Toxin-antitoxin RNA pair CreTA safeguards CRISPR-Cas. CRISPR effector (Cascade) is not only guided by CRISPR RNA to inactivate full-matching foreign nucleic acids but is also co-opted by CreA RNA to transcriptionally repress the toxin gene *creT* through partial complementarity between CreA and the *creT* promoter. When Cascade is inactivated, the derepressed CreT RNA sequesters the rare $tRNA^{UCU}$ that decodes a minor arginine codon and arrests cellular growth, thus making the CRISPR effector addictive to the host cell.

RESULTS: The intergenic sequence between *cas6* and *cas8* caused toxicity in cells lacking one or more *cascade* genes. By extensive mutational analysis, we identified the RNA toxin gene *creT* and its critical elements, namely, a combination of a strong Shine-Dalgarno motif, an efficient start codon, two minor arginine codons (AGA) located immediately downstream, and a stable stem-loop structure. Overexpression of $tRNA^{UCU}$ relieved the toxicity of CreT, supporting a mechanism whereby this RNA toxin arrests cellular growth by sequestering the rare arginine $tRNA^{UCU}$.

Mutational analysis of *creT* and its neighboring sequences revealed an adjacent CRISPR repeat–like sequence that is required to suppress the toxicity of CreT. This repeat-like sequence is immediately followed by a spacer-like sequence and a transcription terminator. By Northern blotting and RNA sequencing, we validated the expression of CreA RNA, a CRISPR RNA variant that lacks a 3′ handle. The spacer of CreA partially matches the promoter of *creT* (P_{creT}), and using a reporter gene, we confirmed that CreA, as a complex with Cascade, represses P_{creT}. Similar to CRISPR interference, repression of *creT* requires a protospacer adjacent motif (PAM) and the PAM-proximal base pairing. In cells lacking CreTA, the *cascade* genes become susceptible to disruption by transposable elements. Our bioinformatic analysis identified several CreTA analogs associated with diverse archaeal and bacterial CRISPR-*cas* loci and containing PAMs corresponding to those of the respective CRISPR systems. Notably, these CreTA analogs hold little conservation in nucleic acid sequence, suggesting that they have highly divergently evolved and, conceivably, exploited different toxicity mechanisms.

CONCLUSION: Our data unearth previously unnoticed toxin-antitoxin RNA pairs that prevent the loss of CRISPR-*cas* loci by making them addictive to the host cell. The naturally occurring reprogramming of CRISPR effectors for gene regulation highlights the multifunctionality of CRISPR-Cas in bacteria and archaea and illuminates the emerging topic of the evolution of antiviral defense and gene regulation. ■

The list of author affiliations is available in the full article online.
†These authors contributed equally to this work.
*Corresponding author. Email: lim_im@im.ac.cn (M.L.); xiangh@im.ac.cn (H.X.)
Cite this article as M. Li *et al.*, *Science* **372**, eabe5601 (2021). DOI: 10.1126/science.abe5601

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abe5601

Downloaded from https://www.science.org at Shanghai Normal University on October 08, 2021

RESEARCH ARTICLE

CRISPR

Toxin-antitoxin RNA pairs safeguard CRISPR-Cas systems

Ming Li[1,2,3]†*, Luyao Gong[1,3]†, Feiyue Cheng[1,3]†, Haiying Yu[1], Dahe Zhao[1], Rui Wang[1,3]‡, Tian Wang[4], Shengjie Zhang[1,3], Jian Zhou[1], Sergey A. Shmakov[5], Eugene V. Koonin[5], Hua Xiang[1,3,6]*

CRISPR-Cas systems provide RNA-guided adaptive immunity in prokaryotes. We report that the multisubunit CRISPR effector Cascade transcriptionally regulates a toxin-antitoxin RNA pair, CreTA. CreT (Cascade-repressed toxin) is a bacteriostatic RNA that sequesters the rare arginine $tRNA^{UCU}$ (transfer RNA with anticodon UCU). CreA is a CRISPR RNA–resembling antitoxin RNA, which requires Cas6 for maturation. The partial complementarity between CreA and the *creT* promoter directs Cascade to repress toxin transcription. Thus, CreA becomes antitoxic only in the presence of Cascade. In CreTA-deleted cells, *cascade* genes become susceptible to disruption by transposable elements. We uncover several CreTA analogs associated with diverse archaeal and bacterial CRISPR-*cas* loci. Thus, toxin-antitoxin RNA pairs can safeguard CRISPR immunity by making cells addicted to CRISPR-Cas, which highlights the multifunctionality of Cas proteins and the intricate mechanisms of CRISPR-Cas regulation.

Highly diversified CRISPR-Cas systems provide adaptive immunity in prokaryotes (*1*–*4*). Adaptation complexes incorporate segments of foreign DNA (spacers) into CRISPR arrays, and small CRISPR RNAs (crRNAs) guide a multisubunit effector complex (class 1) or a single-protein effector (class 2) to cleave the cognate foreign DNA or RNA at sequences complementary to the spacers (*5*). By disrupting their nucleolytic activity, DNA-cleaving CRISPR effectors have been engineered to develop versatile gene regulators (*6*, *7*). However, regulatory functions of CRISPR-Cas in bacteria and archaea are poorly understood. A role of the type II Cas9 effector of *Francisella novicida* in repressing a virulence-related regulon through limited complementarity between a non-canonical RNA guide and the target gene has been recently demonstrated (*8*). Furthermore, recent preliminary results reveal widespread autoregulation of transcription by the Cas9 effector, which uses a natural single guide RNA with partial complementarity to the *cas9* gene promoter (*9*).

Here, we report that some type I-B CRISPR effectors are natural gene regulators that transcriptionally repress a dormancy-inducing toxin-antitoxin (TA) system, hereafter CreTA, after Cascade-repressed toxin-antitoxin. CreTA is encoded within the respective CRISPR-*cas* loci and, unlike previously characterized TA modules that all encode a protein toxin (*10*–*12*), consists of two RNA molecules. Thus, CreTA functions as an addiction module that prevents the loss of the genes encoding the multisubunit I-B CRISPR effector complex, Cascade.

[1]State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China. [2]CAS Key Laboratory of Microbial Physiological and Metabolic Engineering, Chinese Academy of Sciences, Beijing, China. [3]College of Life Science, University of Chinese Academy of Sciences, Beijing, China. [4]College of Life Sciences, Sichuan Normal University, Chengdu, China. [5]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA. [6]Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao, China.
***Corresponding author. Email: lim_im@im.ac.cn (M.L.); xiangh@im.ac.cn (H.X.)**
†These authors contributed equally to this work. ‡Present address: Non-coding RNA and Drug Discovery Key Laboratory of Sichuan Province, Chengdu Medical College, Chengdu, Sichuan, China.

CreT is a bacteriostatic toxin suppressed by Cascade

We initially identified a CreTA module within the 311–base pair (bp) intergenic region between *cas6* and *cas8* in *Haloarcula hispanica* (Fig. 1A). In our previous study, we failed to delete any of the *cascade* genes (*cas5*, *cas6*, *cas7*, and *cas8*) individually from the wild-type (WT) strain but managed to delete them simultaneously (Δ*cas5-8*) (*13*). However, starting from a strain lacking the intergenic 311 bp (ΔTA), mutants ΔTAΔ*cas5*, ΔTAΔ*cas6*, ΔTAΔ*cas7*, and ΔTAΔ*cas8* were easily obtained (fig. S1). In these *cas* mutants and Δ*cas5-8*, a plasmid carrying the intergenic 311 bp (pTA) consistently showed toxicity, i.e., a marked reduction in transformation efficiency (by ~10^4-fold), compared with the empty vector (Fig. 1B). By contrast, this toxic effect was not observed in WT or ΔTA cells that encode a complete set of the Cascade subunits. We inferred that the Cascade complex represses a cryptic toxin, which we named CreT for "Cascade-repressed toxin," and the *creT* gene is embedded within the *cascade* gene cassette.

We tested the toxic effect of a series of truncated variants of pTA and identified a 132-bp region that reduced transformation efficiency in ΔTAΔ*cas6*, ΔTAΔ*cas5*, ΔTAΔ*cas7*, and ΔTAΔ*cas8* cells (fig. S2 and pTA07 in Fig. 1C). The 132-bp sequence contained the archaeal promoter elements BRE (TF-IIB recognition) and TATA-box (*14*) (Fig. 1A). As expected, the toxic effect of pTA in ΔTAΔ*cas6* cells was no longer observed when the predicted TATA-box was mutated (pTTm in Fig. 1C). Based on the positions of the promoter elements, we predicted a 78–nucleotide (nt) *creT* transcript produced from the 132-bp region in pTA07 (Fig. 1A). Using a strong promoter (*15*), we constructed a plasmid overexpressing this 78-nt RNA (pOE) and observed very low transformation efficiency (~10 CFU/μg; CFU, colony-forming unit) in both ΔTA and ΔTAΔ*cas6* cells (Fig. 1C); by contrast, high transformation efficiency (10^4 to 10^5 CFU/μg) was observed when this strong promoter was mutated (pOEm). Thus, the 78-nt *creT* transcript was clearly toxic. The 78-nt transcript contained a pair of inverted repeats (10 nt each) (Fig. 1A), which allows this RNA to fold into a stem-loop structure (fig. S3). When we mutated one of the inverted repeats to disrupt this folding potential, the transformation efficiency of pTA in ΔTAΔ*cas6* cells was recovered to ~10^5 CFU/μg (pIRm), which was then markedly reduced (to <10 CFU/μg) when we complementarily mutated the other repeat to restore the stem-loop (pIRcm) (Fig. 1C). We next expressed an inactive CreT mutant in ΔTA and found that its abundance did not substantially change as a result of the stem-loop disruption (fig. S3). The stem-loop structure appears to be critical for the function rather than the stability of CreT RNA.

To characterize the toxic effect (bacteriostatic or bactericidal), we controlled *creT* expression using a tryptophan-inducible promoter (*16*) and introduced it into another haloarchaeon *Haloferax volcanii*. (The inducible promoter does not work in *H. hispanica*.) Compared with the *creT*– strain (containing the empty vector), the *creT*+ strain showed a growth defect in the inducing medium (Fig. 1D). By plating the inducing cultures onto noninducing plates (Fig. 1E), we measured their CFU curves. Notably, the CFU of the *creT*+ strain rose very slowly but did not decline (Fig. 1D), indicating that CreT suppressed cell multiplication. These results indicate that CreT is a bacteriostatic toxin that inhibits a cellular process(es) conserved (at least) among different species of haloarchaea.

CreT is a small RNA that sequesters $tRNA^{UCU}$

We noticed that the 5′ end of CreT RNA contains an 8-nt sequence that fully matches the 3′ end of 16*S* ribosomal RNA (rRNA) (Fig. 2A). When we modulated this complementarity to neighboring sequences on 16*S* rRNA (see sites 1 and 2 in Fig. 2A), plasmids carrying the mutated *creT* were not toxic (high transformation efficiency was observed in ΔTAΔ*cas6*

Downloaded from https://www.science.org at Shanghai Normal University on October 08, 2021

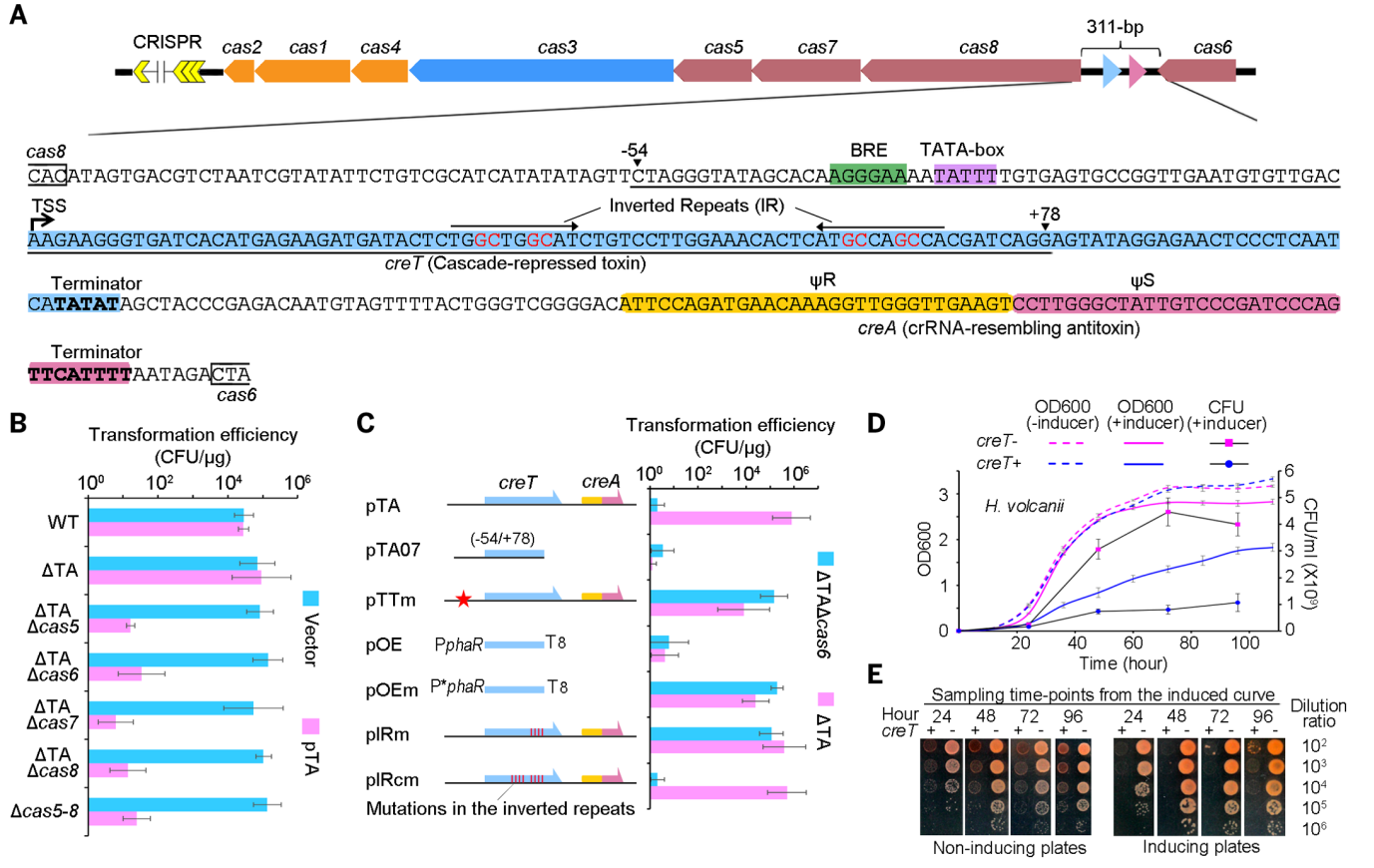


Fig. 1. CreT is a bacteriostatic toxin repressed by Cascade. (**A**) The *creT-creA* operon. BRE and TATA-box are promoter elements. TSS indicates the transcription start site. Red nucleotides within IRs were mutated to modulate complementarity. ψR and ψS of *creA* are analogous to the repeat and spacer of CRISPR, respectively. Positions are relative to the TSS of *creT*. (**B**) Transformation of *cas* mutants by pTA (carrying *creTA*). Vector is the empty pWL502. (**C**) Mutational analysis of pTA. pTA07 carried the underlined sequence in (A). pTTm contained a mutated TATA box (indicated with a red star). CreT was overexpressed using P_{phaR} (pOE) or its mutant P^*_{phaR} (pOEm); T8 is a terminator of eight thymine nucleotides. One of the IRs was mutated in pIRm, and the other was further complementarily mutated in pIRcm. (**D**) Effect of CreT on the growth of *H. volcanii* cells. (**E**) Calculation of CFU by dilution plating. Error bars represent mean ± standard deviation (n = 3).

cells) (Fig. 2B). This suggests that the 8-nt sequence interacts with the 3′ terminus of 16*S* rRNA and likely acts as an efficient Shine-Dalgarno (SD) motif, which was observed to enhance translation efficiency in some haloarchaea (*17*, *18*). We further noticed that CreT contains a canonical AUG start codon and subjected it to saturation mutagenesis. CreT remained toxic (i.e., markedly impaired the plasmid transformation efficiency in ΔTAΔ*cas6*) only when AUG was mutated to GUG (Fig. 2C), another efficient start codon in haloarchaea. Notably, when AUG was mutated to UUG, a less efficient start codon in haloarchaea (*19*), CreT was inactivated. This suggests that the toxicity of CreT depends on strong translation initiation signals.

The translation initiation signals in CreT are immediately followed by two rare AGA arginine codons (usage frequency among all codons is 0.22% in *H. hispanica*) and then by an opal stop codon (UGA) (Fig. 2A). We first mutated the stop codon (UGA to CGA) and found that the mutated CreT remained toxic (Fig. 2D). Then, we made synonymous mutations in the two AGA codons and found that CreT lost toxicity when the two AGA codons were replaced by the more common arginine codons CGA (usage frequency among all codons is 0.87%), CGU (0.63%), CGC (2.14%), or CGG (2.11%) but remained toxic when replaced by the other rare arginine codon AGG (0.21%) (Fig. 2D). Thus, it seemed most likely that the mechanism of CreT toxicity that required the strong translation initiation signals involved sequestering and depleting $tRNA^{UCU}$ that decodes the rare arginine codons. Indeed, we showed that overexpression of $tRNA^{UCU}$, which decodes both AGA and AGG codons, relieved the toxicity of the WT (AGA) CreT and the AGG mutant, whereas overexpression of $tRNA^{CCU}$, which decodes AGG but not AGA, relieved the toxicity of the AGG mutant but not of the WT CreT (Fig. 2E). We further measured the level of $tRNA^{UCU}$ in the *H. volcanii* cells with or without an inducible *creT* gene but did not observe detectable differences at 6 or 12 hours after induction, which indicates that the overall amount of charged and uncharged $tRNA^{UCU}$ was not affected by CreT (fig. S4). We conclude that CreT acts by sequestering the arginine-charged $tRNA^{UCU}$ and thus hampering decoding of the rare arginine codons.

Among the 3859 *H. hispanica* genes, 2025 contain at least one AGA/AGG codon (1075 genes have a single AGA/AGG codon and 950 genes have two or more). The speed and/or accuracy of translation of these genes could be substantially affected by the availability of $tRNA^{UCU/CCU}$. In *Escherichia coli*, a subset of essential genes contains the rare AGA/AGG codons preferentially within their first 25 codons, and even a single AGG codon in this region substantially reduces the protein expression level (*20*, *21*). Under the minor codon modulator hypothesis, the availability of the least abundant $tRNA^{UCU/CCU}$ globally regulates the translation of these essential genes and hence modulates key cellular functions. We similarly analyzed

Downloaded from https://www.science.org at Shanghai Normal University on October 08, 2021

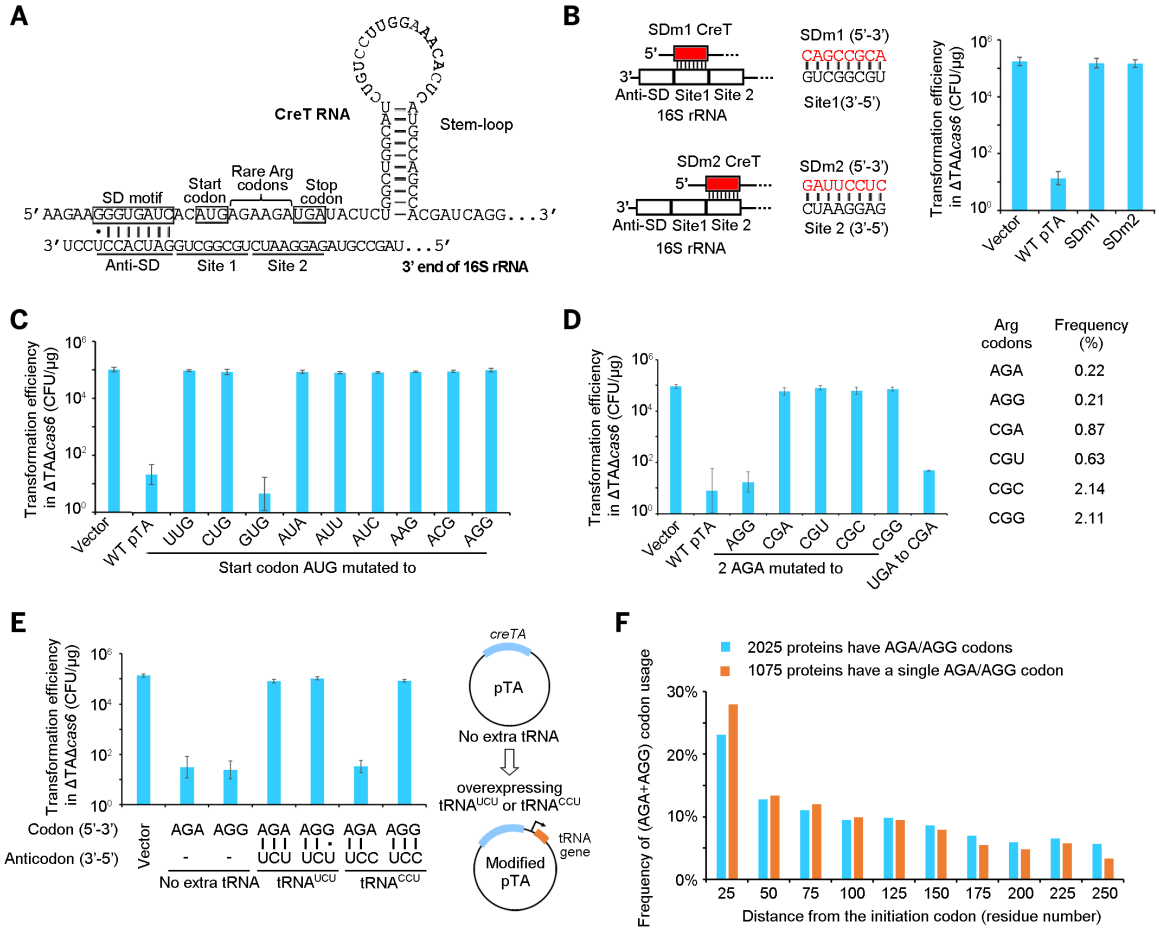


Fig. 2. CreT sequesters tRNA^{UCU}. (**A**) The structure of CreT RNA. (**B**) Mutational analysis of the SD motif. SD was mutated to match the 16S rRNA site 1 or 2 [indicated in (A)]. (**C**) Saturating mutagenesis of the start codon. (**D**) Synonymous mutation of the two rare Arg (arginine) codons and mutation of the stop codon (UGA to CGA). The usage frequency among all codons in *H. hispanica* is shown to the right of the plot. (**E**) Suppression of CreT by overexpression of tRNAs (tRNA^{UCU} or tRNA^{CCU}). The dot indicates a wobble base pair. Vector is the empty pWL502. (**F**) Usage frequency of AGA/AGG in *H. hispanica* genes. Starting from the initiation codon, the number of AGA/AGG within every 25-codon window was divided by the total number of AGA/AGG within the first 250 codons to calculate the frequency for each window. Error bars represent mean ± standard deviation (n = 3).

the 2025 *H. hispanica* genes containing AGA/AGG and found that these genes also preferentially used AGA/AGG within the first 25 codons, and this bias became more prominent when the 1075 genes containing a single AGA/AGG codon were specifically investigated (Fig. 2F). We further examined the set of genes with a single AGA/AGG codon within the first 25 codons and found that these genes are associated with various key cellular functions (table S1). Therefore, the minor codon modulator hypothesis is likely to apply also in *H. hispanica*, so that sequestration of the rare tRNA^{UCU} by CreT impairs the translation of some of these essential genes and thus inhibits cell growth and division.

CreA is an antitoxin RNA that resembles crRNA

We then asked how Cascade represses CreT. Canonically, Cascade is guided by crRNAs, but the only CRISPR array in *H. hispanica* can be deleted from the WT cells containing *creT* without eliciting toxicity (*22*), suggesting that crRNAs are not involved in CreT suppression. We noticed that pTA showed high transformation efficiency in ΔTA cells that retained all Cascade subunits, but mutants lacking the sequence immediately downstream of *creT* showed very low efficiency (Fig. 1C and fig. S2B). This finding implied that this sequence contained uncharacterized elements involved in the repression of the toxicity of CreT. Within the region downstream of *creT*, we detected a CRISPR repeat–like sequence (hereafter "ψR") (Fig. 1A). The 30-nt ψR shares 21 nucleotides with the CRISPR repeat from *H. hispanica* (Fig. 3A), and its transcript can be processed by Cas6 (fig. S5). ψR is directly followed by a ~33-bp "spacer" (hereafter "ψS") containing a T-rich sequence that might function as a transcription terminator (*14*, *23*) (Fig. 1A). Although type I crRNAs typically carry 5′ and 3′ handles derived from flanking repeat sequences, we have recently shown that replacing the downstream repeat with a transcription terminator produced functionally active noncanonical crRNA guides without the 3′ handle (*24*). Therefore, we predicted that the region downstream of *creT* encodes a crRNA-resembling antitoxin RNA (CreA), which consists of an 8-nt 5′ handle (remnant of ψR), the ~33-nt ψS, and no 3′ handle (Fig. 3A). Using a ψS-specific probe, we detected CreA RNA in WT but not in ΔTA cells unless these cells were transformed with pTA derivatives (Fig. 3B). In ΔTAΔ*cas6* cells, larger-sized precursors instead of mature CreA were detected, supporting the prediction that the CreA precursor RNA is processed by Cas6. When the promoter of *creT* (P_{creT}) was inactivated (pTTm), mature CreA and a ~90-nt precursor was detected in ΔTA and ΔTAΔ*cas6*, respectively (Fig. 3B), indicating that *creA* was expressed from its own promoter. When P_{creT} remained active (pIRm), we observed an additional, longer precursor likely corresponding to the *creT-creA* cotranscript (Fig. 3B).

We then used RNA sequencing (RNA-seq) to explore the production of CreA from pIRm (fig. S6). Because Cas6 cleavage generates a hydroxylated 5′ terminus that is inaccessible to adapter ligation during library construction, mature CreA molecules need be pretreated by polynucleotide kinase to add a 5′ monophosphate (see Materials and methods). As expected, RNA-seq revealed a high abundance of mature CreA in ΔTA but not in ΔTAΔ*cas6* cells (fig. S6), further validating the Cas6-dependent biogenesis of CreA. In addition, the predicted Cas6 cleavage site and an 8-nt 5′ handle of CreA were validated by RNA-seq.

Downloaded from https://www.science.org at Shanghai Normal University on October 08, 2021

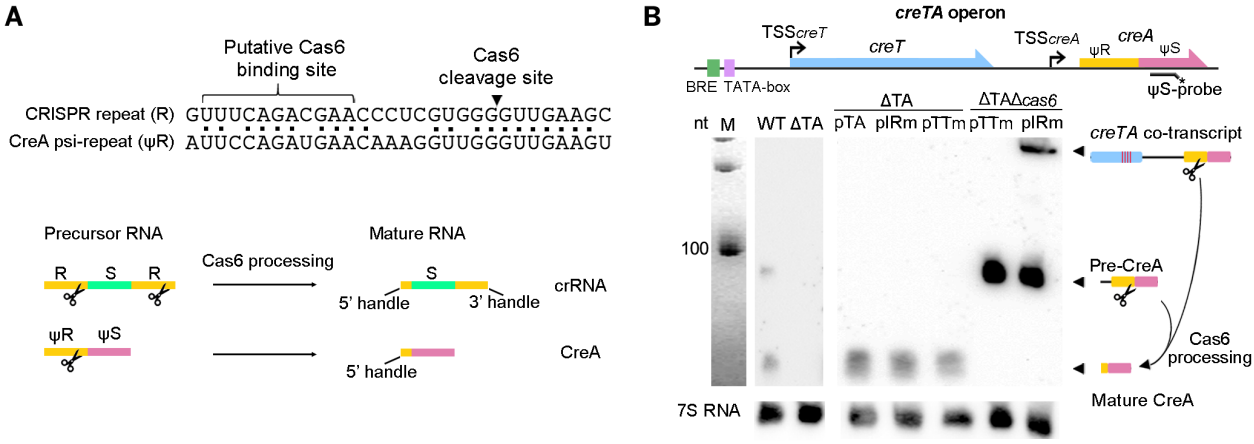


Fig. 3. CreA is analogous to crRNA and is processed by Cas6. (**A**) The nucleotide identity between the RNA transcripts from CRISPR repeat and CreA ΨR. A scheme depicting the Cas6-mediated processing of crRNA and CreA is given. S, CRISPR spacer; ΨS, the "spacer" of CreA. (**B**) Northern blot of CreA and its precursors using the ΨS-probe. 7*S* RNA served as the internal control. M is a 100-nt RNA marker.

CreA guides Cascade to transcriptionally repress *creT*

We noticed a partial match between the CreA RNA and the P_{creT} sequence (Fig. 4A). The target sequence (protospacer) of type I crRNAs is typically 5′ preceded by a protospacer adjacent motif (PAM), and the PAM-proximal base pairing between the spacer and the protospacer provide a critical "seed" during R-loop formation (*25*, *26*). Like this pattern, the first 11 nucleotides [except the sixth nucleotide, which is known to not be involved in base paring (*26*)] of ψS are complementary to a target sequence in P_{creT}, which is flanked by 5′-TTC-3′ [the PAM of *H. hispanica* CRISPR (*27*)] (Fig. 4A). Using scanning mutagenesis, we found that when the conserved 5′ handle and every seed nucleotide that base pairs with P_{creT} were individually mutated, the pTA derivatives transformed ΔTA cells consistently with a much lower efficiency (~10 CFU/μg) than the WT pTA (10^4-10^6 CFU/μg) (fig. S7), indicating that these elements of CreA are critical for its antitoxin activity; by contrast, the ψS nucleotides outside the seed were not important (fig. S7B). We then selected three seed mutants of CreA and verified that complementarily mutating P_{creT} restored their antitoxin activity and the high plasmid transformation efficiency in ΔTA (Fig. 4B). Therefore, the CreA-P_{creT} complementarity is critical for toxin repression.

To validate the P_{creT}-inhibiting effect of CreA, we introduced a P_{creT}-controlled green fluorescent protein gene (*gfp*) into WT and ΔTA cells (Fig. 4C). As expected, fluorescence was observed in ΔTA cells (lacking CreA) but not in WT cells (producing CreA) unless we mutated P_{creT} to disrupt the CreA-P_{creT} complementarity (Fig. 4C). We also tested the role of PAM in repressing P_{creT}. Because the PAM nucleotides 5′-TTC-3′ are located within the complement of the purine-rich BRE element of P_{creT} (see Fig. 4A), we preserved the purine-rich character of BRE by mutating PAM to 5′-CCC-3′ (PAM– in Fig. 4D). Nevertheless, only minimal fluorescence was detected in both WT and ΔTA cells, indicating that this mutation inactivated the BRE element. We next duplicated the BRE sequence (2BRE in Fig. 4D) to allow mutation of the PAM nucleotides without disrupting the promoter elements. As expected, the two-BRE P_{creT} drove fluorescence production in ΔTA but not in WT cells (producing CreA), indicating that the mutated P_{creT} was also repressed by CreA. When the PAM was further mutated (2BRE/PAM– in Fig. 4D), fluorescence was observed in both types of cells with equivalent intensity (which was weaker than that of the 2BRE mutant in ΔTA, presumably because of the direct effects of mutating the nucleotides next to BRE). Thus, the PAM motif, as well as the partial complementarity between CreA and P_{creT}, is required for P_{creT} repression. We then introduced the P_{creT}-controlled *gfp* into cells lacking *cas1*, *cas2*, *cas3*, *or cas4*, and notably, minimal fluorescence was observed in these cells unless we mutated P_{creT} to disrupt the CreA-P_{creT} complementarity (fig. S8). These results suggest that these four *cas* genes, which encode proteins that are not Cascade subunits, are not required for P_{creT} repression.

We also performed primer extension to analyze the activity of P_{creT} on the pIRm plasmid (fig. S9). This plasmid carries a mutated *creT* (see Fig. 1C), so we could test P_{creT} activity in cells lacking Cascade protein(s). In this setting, P_{creT}-driven transcription was not detected in ΔTA cells that encode a complete set of Cascade proteins but was derepressed in cells lacking Cas6 (the nuclease processing CreA) or Cas7 (the backbone subunit of Cascade) (fig. S9).

Taken together, our results demonstrate that Cascade transcriptionally represses *creT* based on the partial complementarity between CreA and P_{creT}. When we programmed the CRISPR array with a spacer identical to ψS (replacing the WT CRISPR array with a mini-CRISPR containing ψS), *creA* was no longer required to repress *creT* (fig. S10), indicating that canonical crRNAs were reprogrammed to regulate *creT* transcription and further supporting the guide RNA–dependent regulatory role of Cascade.

CreTA safeguards CRISPR immunity

Our previous studies have extensively demonstrated the adaptive CRISPR immunity in *H. hispanica* (*13*, *22*, *27*). Targeting of complementary DNA by the Cascade-crRNA complex results in the recruitment of the Cas3 helicase-nuclease to elicit interference (cleavage of the target DNA) and/or primed adaptation (acquisition of new CRISPR spacers from the target DNA). We asked how self-interference and self-adaptation are precluded when CreA directs Cascade binding to the P_{creT} DNA. By modifying the plasmid carrying the P_{creT}-controlled *gfp*, we coexpressed CreA mutants with differently extended (to 15, 20, 25, or 36 bp) complementarity to P_{creT} and tested three possible outcomes, i.e., plasmid interference, primed adaptation, or gene repression (fig. S11). All these self-targeting plasmids showed high transformation efficiency (~10^5 CFU/μg) in ΔTA cells. Because CreA lacks the 3′ handle, this result is compatible with our previous observation that deleting the entire 3′ handle from a virus-targeting crRNA resulted in no or little antivirus immunity (*24*). We then analyzed the CRISPR array from the transformants, which might have been expanded by new spacers if primed self-adaptation was occurring. CRISPR expansion was not observed

Downloaded from https://www.science.org at Shanghai Normal University on October 08, 2021

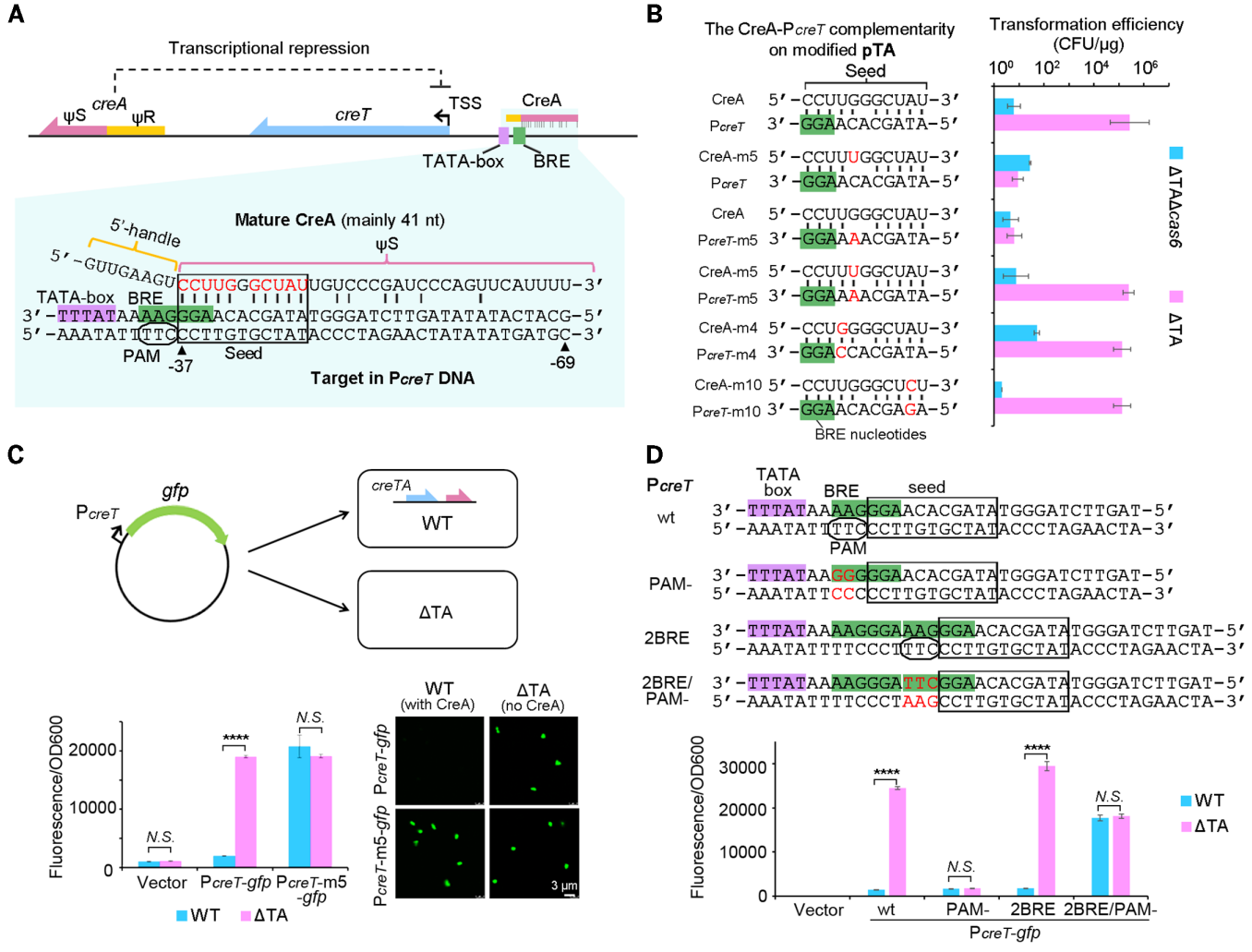


Fig. 4. CreA directs Cascade to repress P_{creT}. (**A**) Scheme illustrating the CreA-mediated P_{creT} repression and the partial complementarity between CreA and P_{creT}. Substituting any of the red nucleotides inactivated CreA (see fig. S7C). Positions are relative to the TSS of *creT*. (**B**) Nucleotide substitutions to alter the CreA-P_{creT} complementarity. (**C**) Fluorescence from a *gfp* gene controlled by P_{creT} or P_{creT}-m5 in WT or ΔTA cells. Vector is the empty pWL502. Representative microscopy images are provided. (**D**) Fluorescence from a *gfp* gene controlled by the WT P_{creT} (wt) or its PAM mutant (PAM–). The BRE element was duplicated (2BRE) to allow PAM mutation without disrupting the BRE element. Mutated nucleotides are highlighted in red. Error bars represent mean ± standard deviation (n = 3); two-tailed Student's t test [****P < 0.0001; *N.S.*, not significant (P > 0.05)].

until the CreA-P_{creT} complementarity was increased to 25 or 36 bp (fig. S11C), indicating that limited complementarity could not prime adaptation. We then measured the fluorescence from the transformant cells and observed a gradual decrease in fluorescence intensity as the CreA-P_{creT} complementarity increased (fig. S11D). These results indicate that the mismatches between CreA and its target preclude autoimmunity while allowing transcription regulation and might have been fine-tuned by selection to optimize the level of gene repression.

We then sought to investigate the impact of the CreTA toxicity modulation on CRISPR-Cas itself. Because *cascade* genes occupy a large genomic region that is susceptible to disruption by transposable element insertion and by deletion, CreTA that becomes toxic in the absence of Cascade might stabilize *cascade* genes by acting as an addiction module. Given that a transposition burst of the insertion (IS) element ISH27 (1390 bp) has been reported to occur in *H. hispanica* cells stored at 4°C (*28*), we investigated the WT and ΔTA strains that had been stored at 4°C for 2 years. The stored *H. hispanica* cultures were resuscitated and challenged with pTarget, which carries the target of spacer1 (the first spacer of the CRISPR array) (Fig. 5A). The survivors that failed to repel the target plasmid were screened on a selective medium, and their *cascade* genes were analyzed (Fig. 5B). We found that the *cascade* genes were interrupted by ISH27 in ~60% (17 of 28) of the ΔTA survivors but remained intact in nearly all the WT survivors (one lost the whole CRISPR-*cas* locus) (table S2). By contrast, *cas3*, which is required for plasmid interference but not for the CreT toxin repression, was occasionally disrupted in both the WT and ΔTA survivors. These findings show that CreTA is an addiction module that safeguards the genetic integrity of Cascade in *H. hispanica* (Fig. 5C).

Distribution of CreTA homologs and analogs across prokaryotes and among CRISPR subtypes

By sequence similarity search of the National Center for Biotechnology Information (NCBI) nucleotide genomic database, we identified only three *creTA* homologs, all located between the *cas6* and *cas8* genes in the I-B CRISPR-*cas* locus from haloarchaeal genomes (fig. S12).

Downloaded from https://www.science.org at Shanghai Normal University on October 08, 2021

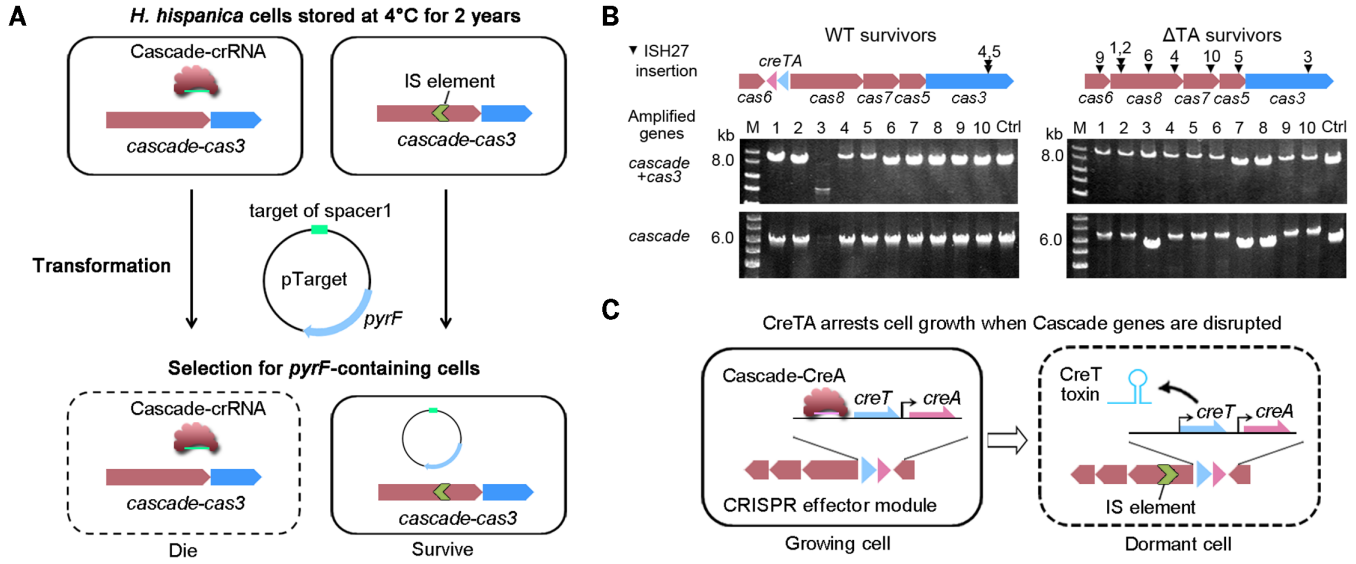


Fig. 5. Screening and analysis of cells with inactivated CRISPR-Cas. (**A**) Scheme illustrating the plasmid challenge assay. Storage at 4°C induced the transposition "burst" of the IS element (*28*). *pyrF* encodes the pyrimidine biosynthetic enzyme orotidine-5′-monophosphate decarboxylase. (**B**) Example clones surviving the challenge assay. Ctrl represents the culture before challenge; M is a double-stranded DNA ladder. Larger-sized PCR products indicate the insertion of IS element (ISH27). See table S2 for more information. (**C**) The model of CreTA-mediated addiction. *creT* is repressed by Cascade-CreA in WT cells and derepressed when Cascade is disrupted.

This paucity of detectable CreTA homologs is not surprising, given that small RNAs often show limited sequence conservation, and furthermore, ψR and ψS are likely to have divergently coevolved with Cas6 and P_{creT}, respectively. We examined the intergenic sequences flanking the *cas6* gene in other haloarchaeal genomes that encode type I-B CRISPR-Cas and identified five strains carrying *creA* analogs with limited sequence similarity to CreTA from *H. hispanica*, all of which were located upstream of *cas6* (Fig. 6A). Each *creA* analog contained a ψR sequence that was similar (60 to 80% identity) to the repeat from the co-occurring CRISPR. Using a probe against the putative ψS downstream of each ψR, we detected mature CreA RNA in four strains that were available in our laboratory (Fig. 6B). We predicted that the sequences upstream of each *creA* encompassed a *creT* gene and cloned these regions separately into a plasmid, which was then used to transform *H. hispanica* cells. These (putative) *creT*-containing plasmids consistently showed much lower transformation efficiency ($<10^2$ CFU/μg) than the empty control (~10^5 CFU/μg) (Fig. 6C), validating the toxicity of each *creT* analog. Like the *H. hispanica creT*, these toxic genes contained one or more pairs of inverted repeats and lacked open reading frames larger than 10 codons (fig. S13). The putative promoter of each *creT* contained a target sequence that was partially complementary to their cognate *creA*, and each target sequence was flanked by the 5′-TTC-3′ PAM (Fig. 6D and fig. S13), strongly suggesting that the cognate Cascade complexes bind to and repress these promoters. Thus, these *creTA* analogs are probably Cascade-regulated TA modules that safeguard the accompanying *cascade* genes, as in the case of *H. hispanica*. However, none of the *creT* RNAs in the CreTA-like modules contain a combination of a SD sequence, a start codon, and immediately following rare codons, suggesting distinct toxicity mechanisms. The I-B CRISPR loci in several haloarchaeal genomes closely related to those encoding CreTA or its analogs lack sufficiently long intergenic regions flanking *cas6* and do not contain readily detectable ψR sequence in the parts of the CRISPR loci, indicating that CreTA is a recurrent, but evolutionarily labile, accessory of CRISPR-Cas systems.

By searching for CRISPR repeat–like sequences flanking *cas6* genes in other archaea and bacteria, we predicted additional CreTA analogs, most of which are associated with a bacterial type I-B CRISPR (Fig. 7). All of these *cas6*-flanking intergenic regions contained both ψR and ψS with partial complementarity to a target sequence flanked by a PAM (Fig. 7). Type I-B CRISPRs are highly diversified, with at least 10 distinct subfamilies of *cas8b* encoding the large subunits of the effector complex (*29*). Six I-B CreTA analogs were associated with *cas8b1* and contained the cognate 5′-TCA-3′ or 5′-TTA-3′ PAM sequences (*30*), whereas two analogs were associated with *cas8b2* and contained the corresponding 5′-CCT-3′ PAM sequence (*31*). Thus, the CreTA-regulation apparently coevolved with CRISPR immunity according to their PAM specificity. Consistently, we also found a CreTA analog associated with an archaeal I-D CRISPR and containing the PAM (5′-GTG-3′) reported for this subtype (*32*, *33*) (Fig. 7). A putative CreTA was also predicted for the III-A CRISPR-Cas in a thermophilic bacterium. Like the haloarchaeal CreTA analogs, inverted repeats were frequently observed between the bacterial repeat-like sequences and their putative targets (fig. S14). It remains to be studied experimentally which of these and other repeat-like sequences found in CRISPR-*cas* loci, indeed, are components of CreTA-like modules. Nevertheless, these observations suggest that protection of *cas* genes by two-RNA TA modules, conceivably with different toxicity mechanisms, might be a widespread phenomenon.

Discussion

In this work, we demonstrated the dual functionality of the multisubunit subtype I-B CRISPR effectors (Cascades) in haloarchaea. The Cascade complex of *H. hispanica* is not only guided by canonical crRNAs to inactivate cognate foreign DNA but is also co-opted by a noncanonical guide (CreA) to down-regulate the transcription of a toxin RNA gene (*creT*) through the partial complementarity between the spacer-like sequence of CreA and the *creT* promoter. CRISPR-Cas systems efficiently protect bacteria and archaea from viruses and other types of foreign DNA, but characteristically of defense systems, they also impart non-negligible fitness costs on the host (*34*). Specifically, in the case of CRISPR, this fitness cost appears to come, primarily, from autoimmunity (*35*, *36*) or from targeting beneficial or addictive plasmids (*37*–*39*). Presumably, these costs result in frequent loss of CRISPR-Cas

Downloaded from https://www.science.org at Shanghai Normal University on October 08, 2021

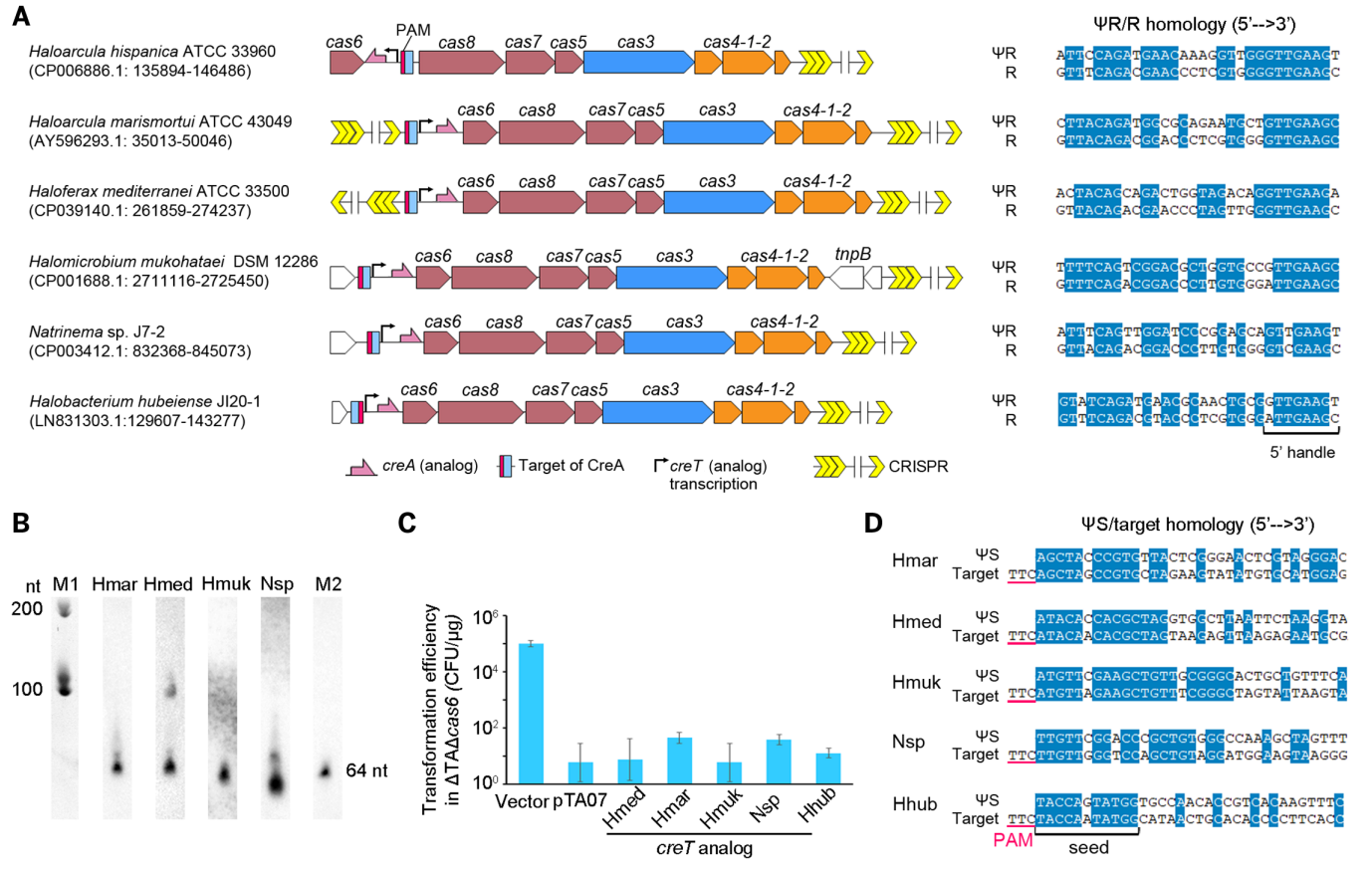


Fig. 6. CreTA-like modules associating with type I-B CRISPR-Cas. (**A**) Haloarchaeal CRISPR-*cas* loci associated with *creTA* analogs. The homology between CRISPR repeat (R) and *creA* ΨR is shown. (**B**) Northern blot of CreA analogs. M1 is a 100-nt RNA marker; M2 is a biotin-labeled 64-nt single-stranded DNA. (**C**) Transformation of *H. hispanica* with plasmids carrying a *creT* analog. Vector is the empty pWL502. Error bars represent mean ± standard deviation (n = 3). (**D**) The homology between *creA* ΨS and its target. Hmar, *Haloarcula marismortui*; Hmed, *Haloferax mediterranei*; Hmuk, *Halomicrobium mukohataei*; Nsp, *Natrinema* sp. J7-2; Hhub, *Halobacterium hubeiense*.

systems in bacteria, which is reflected in the patchy distribution of CRISPR-Cas even among closely related bacterial strains (*40*). Nevertheless, in the current genome sequence databases, ~40% of bacterial and ~90% of archaeal genomes carry CRISPR-*cas* loci, suggesting the possibility that in addition to the direct benefits brought about by adaptive immunity, mechanisms mitigating the costs of CRISPR systems and preventing their loss might exist. Here, we reveal one such mechanism whereby the *cas* genes encoding the CRISPR effector subunits become essential to the host, thanks to their ability to down-regulate the expression of a toxin, with the help of a natural guide RNA that serves as an antitoxin. The CreTA-like elements ensure the preservation of the effector *cas* genes but not the adaptation module of the CRISPR array. The presence of solo effector gene suites, without the adaptation module or CRISPR array, is not uncommon in bacterial and archaeal genomes. Thus, in the latest genomic census of CRISPR-Cas systems (*29*), among the 6254 identified CRISPR-*cas* loci, 500 were solo effectors, and more specifically, among the 669 I-B loci, 75 contained effector genes only. A comprehensive computational and experimental analysis of the solo effector genomic neighborhoods should show how often these genes are safeguarded by CreTA-like modules.

The discovery of the role of CreTA modules as safeguards of *cas* genes complements the recent discovery of the widespread autorepression of transcription by Cas9, the type II CRISPR effector. Such autorepression is a distinct strategy of CRISPR cost mitigation in which a noncanonical guide RNA is used, similar to the case of CreTA (*9*). The use of guide RNAs is a general principle for delivery of protein effectors to distinct sites on DNA and RNA molecules. Inactivation of foreign nucleic acids is only one of the RNA-guided functions, even if this was the primary driving force in the evolution of CRISPR. It appears likely that the autoregulation of Cas9 and the regulation of CreTA-like modules are not the only cases of dual functionality of CRISPR effectors and, perhaps, other defense systems that might possess various regulatory capacities, in addition to their direct involvement in parasite inactivation. Because self-targeting CRISPR spacers that partially match the host DNA are widespread (*35*), it is possible that some canonical crRNAs have been selected to direct CRISPR effectors to regulate host genes.

Some Cas nucleases can be activated upon target recognition to degrade host RNA, which causes cell death and the abortion of phage infection or arrests cellular growth until the foreign targets are eliminated (*41*–*43*). CRISPR-Cas and other defense systems also frequently associate with diverse TA modules in bacterial and archaeal genomes, and it has been proposed that immune systems interact with the TA such that the latter induce dormancy or cell death when the former fail, for example, in the presence of virus-encoded anti-CRISPR proteins (*44*). Here, we describe a different type of such interaction where the immune system down-regulates the toxin expression, rendering the immunity genes addictive to the host. In such cases, the immune systems and the TA behave as a pair of symbiotic selfish genetic elements, in accord with the general

Downloaded from https://www.science.org at Shanghai Normal University on October 08, 2021

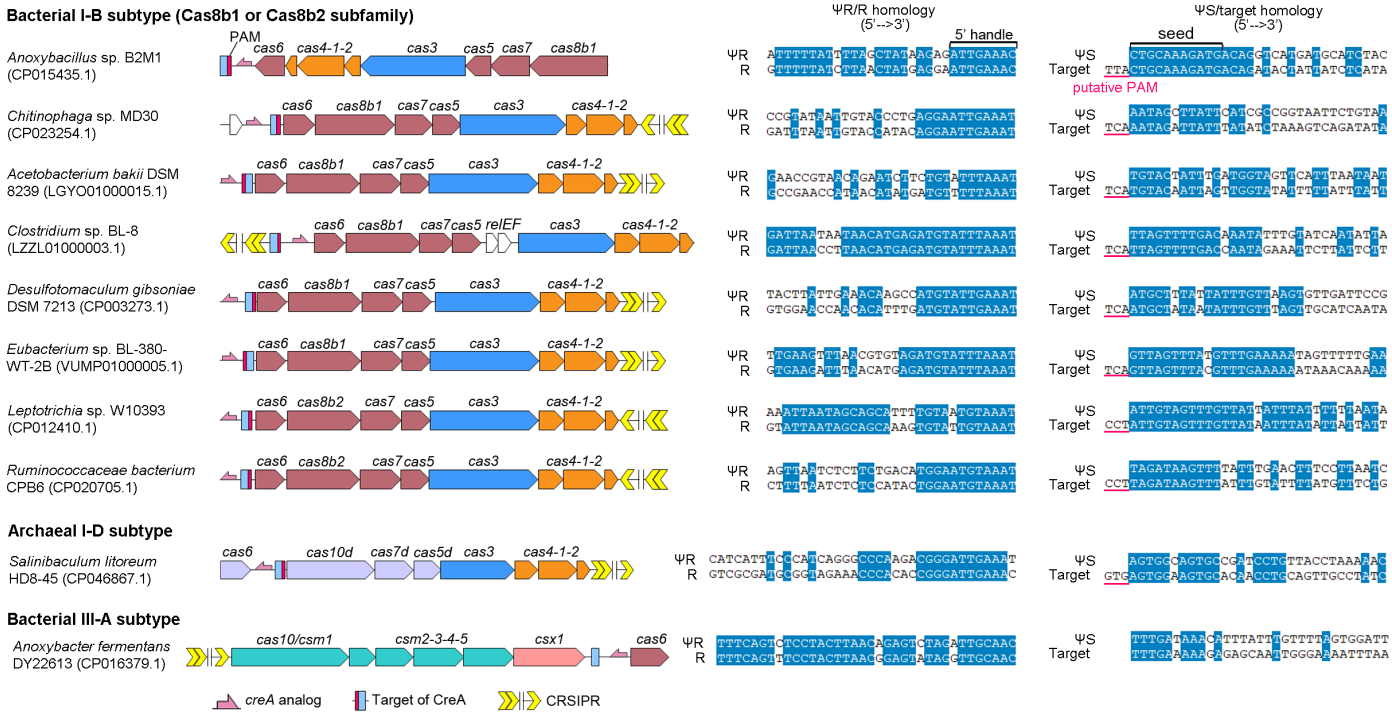


Fig. 7. CreA-like genes associated with other archaeal (I-D) or bacterial (I-B and III-A) CRISPR-*cas* loci. Alignments of the homologous regions between CRISPR repeat (R) and *creA* ΨR and between *creA* ΨS and its putative target are shown. The CRISPR-*cas* locus of *Clostridium* sp. BL-8 also encodes a type II toxin-antitoxin system (*relEF*).

paradigm of the evolutionary entanglement between defense mechanisms and mobile elements (*45*). However, it remains to be investigated whether CreTA-like TA modules could also be bifunctional, serving both as the last line of defense and the safeguard for the *cas* genes.

The CreTA module is a previously unknown type of TA, where both the toxin and the antitoxin are represented by small RNA molecules, although the antitoxin activity strictly depends on the Cascade protein complex. Given the small size and poor sequence conservation of both RNA components, it seems likely that such modules are more common than is shown in this work. The CRISPR-Cas systems appear to be specifically prone to spawn such TA modules because repeat propagation outside the CRISPR array is a common phenomenon that is thought to give rise, in particular, to the tracrRNA of type II and some type V CRISPR (*46*). Such ectopic repeats provide the scaffold for noncanonical guide RNAs that can evolve into antitoxins controlling either an RNA or a protein toxin. Discovery and structural and functional dissection of such CRISPR-regulated CreTA-like modules can be expected to reveal unknown facets of TA and CRISPR biology and, particularly, their behavior as selfish genetic elements.

Materials and methods

Strains and growth conditions

Haloarchaeal strains were cultivated at 37°C in AS-168 medium (per liter, 200 g NaCl, 20 g $MgSO_4{\cdot}7H_2O$, 2 g KCl, 3 g trisodium citrate, 1 g sodium glutamate, 50 mg $FeSO_4{\cdot}7H_2O$, 0.36 mg $MnCl_2{\cdot}4H_2O$, 5 g Bacto Casamino Acids, and 5 g yeast extract, pH 7.2), unless specified. *H. hispanica* ATCC 33960 Δ*pyrF* strain DF60 (*47*) or its derivatives were cultivated in AS-168 medium supplemented with uracil (at a final concentration of 50 mg/liter). The strains transformed by the pWL502 derivatives were cultivated in yeast extract–subtracted AS-168.

The *H. volcanii* H1424 strain was cultivated in the Hv-YPC medium (*48*) with some modifications [per liter, 144 g NaCl, 30 g $MgCl_2{\cdot}6H_2O$, 33 g $MgSO_4{\cdot}7H_2O$, 4.2 g KCl, 0.333 g $CaCl_2$, 5 g yeast extract, 1 g peptone (soya), 1 g Bacto Casamino acids, and 12 ml of 1 M Tris HCl (pH 7.5)] supplemented with uracil and thymidine (at final concentrations of 50 and 40 mg/liter, respectively). The strains transformed by the pTA1228 (*16*) derivatives were cultivated in Hv-YPC medium without adding uracil and thymidine. Tryptophan was added to a final concentration of 500 mg/liter to induce CreT expression.

Natrinema sp. J7-2 was cultivated in the 18% MGM medium (per liter, 144 g NaCl, 21 g $MgSO_4{\cdot}7H_2O$, 0.5 g $CaCl_2$, 18 g $MgCl_2{\cdot}6H_2O$, 4.2 g KCl, 3 g yeast extract, 5 g tryptone, pH 7.5).

E. coli JM109 was used for cloning and cultivated at 37°C in Luria-Bertani medium. Ampicillin was added to a final concentration of 100 mg/liter when needed.

Plasmid engineering and transformation

The *cas6-cas8* intergenic sequence (NC_015943.1: 145387-145697) was amplified from the *H. hispanica* genomic DNA using the high-fidelity KOD-Plus DNA polymerase (TOYOBO, Osaka, Japan), digested by BamHI and KpnI (New England Biolabs, MA, USA), and inserted into the predigested pWL502 (*47*) with T4 DNA ligase (New England Biolabs, MA, USA). Using a series of internal primer pairs (table S3), various truncated versions were generated for this fragment. When needed, the sequence of P_{phaR} (*15*) was directly designed on the primer. The point mutation was introduced using the overlap extension polymerase chain reaction (PCR) strategy. For example, primer pairs TA-F/TTm-R and TTm-F/TA-R were separately used to amplify the two moieties, and the products were mixed and used as the template for a second round of PCR reaction with the primer pair TA-F/TA-R. This DNA product contained a TATA box–mutated P_{creT}, because the mutation had been designed within the complementary part of the primers TTm-F and TTm-R. The cloning was performed using *E. coli* JM109, and the plasmid was extracted using the AxyPrep Plasmid Miniprep Kit (Corning, NY, USA). The insert was validated by DNA sequencing.

Haloarchaeal cells were transformed according to the online Halohandbook (https://haloarchaea.com/wp-content/uploads/2018/10/Halohandbook_2009_v7.3mds.pdf), and

Downloaded from https://www.science.org at Shanghai Normal University on October 08, 2021

the transformants were selected on the yeast extract–subtracted AS-168 plates or the Hv-YPC plates without adding uracil and thymidine. The transformation efficiency (CFU/μg) was log-transformed, and then the average and standard deviation were calculated for plotting. Two-tailed Student's *t* test was also performed based on the log-transformed data. Each transformation assay was double-checked by a different lab member.

Gene knockout and knockin

To knock out *creTA* or a *cas* gene, their upstream ~500 bp and downstream ~500 bp were amplified using the corresponding primer pair UF/UR or DF/DR, and the generated upstream and downstream fragments were connected by overlap extension PCR using the primer pair UF/DR. The final PCR products were digested and ligated into the suicide vector pHAR (*47*) and then introduced into the parental (WT or ΔTA) cells. The mutants were obtained through the previously described two-step screening (*47*) and further validated by colony PCR using the UF/DR primer pair and subsequent DNA sequencing. In a similar way, the mini-CRISPR containing the ψS spacer was constructed by overlap extension PCR and used to replace the WT CRISPR on the ΔTA chromosome.

Northern blot analysis

Haloarchaeal cells were collected from 3 ml of early-stationary culture by centrifugation, and the total RNA was extracted using TRIzol (Thermo Fisher Scientific, MA, USA) according to the standard protocol. RNA concentration was determined by a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific, MA, USA). A total of 8 μg of RNA was mixed with an equal volume of the RNA loading dye (Takara, Shiga, Japan), denatured by heating at 65°C for 10 min, and coelectrophoresed with the Century-Plus RNA ladder (Thermo Fisher Scientific, MA, USA) or a biotin-labeled 64-nt single-stranded DNA on an 8% polyacrylamide gel (containing 7.6 M urea). The ladder lane was excised and imaged after ethidium bromide staining. The RNA samples were electrotransferred onto a Biodyne B nylon membrane (Pall, NY, USA). The biotin-labeled probe was used for hybridization, and the signal was detected using the Chemiluminescent Nucleic Acid Detection Module Kit (Thermo Fisher Scientific, MA, USA), according to the manufacturer's protocol. The membrane was imaged using the Tanon 5200 Multi chemiluminescent imaging system (Tanon Science & Technology, Shanghai, China). Each blotting assay was performed with two biological replicates, and a representative result was provided.

RNA-seq analysis

The ΔTA or ΔTAΔ*cas6* colonies transformed by pIRm were picked to inoculate 10 ml of yeast extract–subtracted AS-168. After subinoculation and a 2-day cultivation, the late exponential culture was collected and total RNA was extracted. A total of 50 μg of RNA was treated with polynucleotide kinase (New England Biolabs, MA, USA) according to the manufacturer's protocol. The kinase was inactivated by incubating at 65°C for 20 min and then removed through the phenol-chloroform extraction. After precipitation with the same volume of isopropanol and 0.1 volume of 3M sodium acetate, the RNA sample was redissolved. RNA molecules ranging from 30 to 300 nt were selected to construct a small RNA library with the NEXTFLEX Small RNA-Seq Kit (Bioo Scientific, TX, USA) and then subjected to Illumina HiSeq sequencing (paired-end, 150-bp reads). The raw data was processed to remove adapters. The resulting reads were mapped to the *creA* sequence using custom Perl scripts (*49*).

Primer extension analysis

The 5′-FAM (6-carboxyfluorescein)–labeled primer+155 (table S3) was ordered from Thermo Fisher Scientific. A total of 2.5 μg of the labeled primer was mixed with 5 μg of the total RNA, and reverse transcription was performed using 200 enzyme units (U) of M-MLV reverse transcriptase (Promega, WI, USA). The extension products were screened using the ABI3730xl DNA Analyzer (Thermo Fisher Scientific, MA, USA), and the results were analyzed using GeneMapper 4.1.

Heterologous expression of CreT

The tryptophan-inducible promoter p.*tnaA* (*16*) was linked with the 78-bp *creT* gene using the overlap extension PCR strategy. The hybrid DNA was digested by BamHI and KpnI, inserted into the predigested expression vector pTA1228 (*16*), and then introduced into the *H. volcanii* H1424 cells. Single colonies of each transformant (by empty or modified pTA1228) were selected and separately inoculated into 100 ml of Hv-YPC medium in triplicate for growth-curve measurements. The inducing medium contained 500 mg/liter tryptophan for toxin induction. Optical density at 600 nm (OD600) was monitored using the Shimadzu UV-2550 spectrophotometer. At different time points (0, 24, 48, 72, and 96 hours) through the growth curve, the cell culture was sampled, serially diluted, and plated onto the Hv-YPC media for CFU calculation. The triplicates were measured for each treatment to get the mean and standard deviation.

Fluorescence measurement

The gene of a soluble-modified red-shifted GFP protein (*50*) was linked to the P_{creT} promoter (or its derepressed mutant) using the overlap extension PCR strategy. The hybrid DNA was introduced into the *H. hispanica* cells using the pWL502 vector. For each transformation assay, three individual colonies were selected and cultured to the late exponential phase, and their OD600 and fluorescence were simultaneously determined using the Synergy H4 Hybrid multimode microplate reader (BioTeck, VT, USA). The fluorescence/OD600 ratio was calculated for each of the three individual samples, and average and standard deviation were calculated. Two-tailed Student's *t* test was performed. The transformant cells were visualized using the Leica TCS SP8 confocal microscope in combination with Leica application suite software (Las X).

Plasmid challenge assay

Two complementary oligonucleotides that included the sequence of spacer1, a 5′-TTC-3′ trinucleotide as the PAM, and two sticky ends were ordered from Thermo Fisher Scientific. The oligonucleotides were mixed, denatured, annealed, and then inserted into pWL502 (predigested by BamHI and KpnI) to generate the target plasmid pTarget. *H. hispanica* strains stored at 4°C for two years were resuscitated in fresh medium and then subinoculated before transformation by pTarget. The survivors were randomly selected for colony PCR analyses. The forward primer ~500 bp upstream of *cas6* (cascade-F) and the backward primer ~500 bp downstream of *cas5* (cascade-R) were used to amplify the four *cascade* genes (table S3). The forward primer cascade-F and the backward primer ~500 bp downstream of *cas3* (cas3-R) were used to amplify the genomic fragment containing the *cascade* genes and *cas3*. The PCR products were further analyzed by DNA sequencing to get the information of the precise position of IS insertion events (table S2).

Codon usage analysis

The *H. hispanica* protein-coding genes were downloaded from the NCBI ftp site (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/223/905/GCA_000223905.1_ASM22390v1/). The usage of AGA/AGG codons within these genes was examined. For each of the 2025 AGA/AGG-containing genes (or the 1075 genes containing only one AGA or AGG codon), we calculated the frequency of (AGA+AGG) codon usage for every 25 codons, starting from its initiation codon to the 250th codon.

Bioinformatic analysis

The folding potential of RNA was analyzed using the RNAfold webserver (*51*). Sequence alignments were constructed using the T-Coffee webserver, and the results were visualized using the GeneDoc software (version 2.6.002). The base pairings between CreT and 16*S* rRNA were analyzed using the IntaRNA web server (*52*). To identify the target site of *H. hispanica* CreA, a regular expression ("CCTTG.GCTAT") was used

Downloaded from https://www.science.org at Shanghai Normal University on October 08, 2021

to search the genome sequence. The targets of CreA analogs were similarly predicted (see next section).

Search for creTA homologs and analogs

To find *creTA* homologs, the BLASTN program was run with default parameters against the NCBI nucleotide collection database or against the RefSeq Genome Database (taxonomy ID: 183963).

To identify potential *creTA* homologs or analogs in haloarchaea, the intergenic sequences flanking haloarchaeal *cas6* genes (downloaded from the NCBI database) were searched for a conserved 5′ handle sequence "NTTGAAGN." This motif, together with the upstream 22 bp, was assumed to represent the ΨR. The 30 to 40 bp downstream of this motif were assumed to represent the ΨS, and the *cas6*-flanking sequences were searched using a regular expression for a putative target matching the first 1 to 5 and 7 to 11 nucleotides of the ΨS. Only when the target was identified, the assumed ΨR and ΨS sequences were defined as a *creA* analog, and a *creT* analog was predicted to be located between *creA* and its target.

For the preliminary identification of *creTA* analogs in other archaea and bacteria, a database containing ~5000 archaeal and ~40,000 nonredundant bacterial genomes was downloaded from the NCBI. The Cas6 proteins were identified using the corresponding hidden Markov model profiles (*40*) and the HMMER suite (*53*), and CRISPR arrays were identified using the minCED tool (https://github.com/ctSkennerton/minced) with default parameters. For each *cas6* gene, the upstream and downstream sequences (500 bp from each side) were extracted and then searched for repeat-like elements (putative ΨR) with the BLASTN program (-task blastn-short), using the CRISPR repeat sequence from the same genome as the query (*49*). The sequences containing a single repeat-like sequence (to exclude from consideration of CRISPR arrays) were manually examined for putative *creTA* as described above.

REFERENCES AND NOTES

1. R. Barrangou *et al*., CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007). doi: 10.1126/science.1138140; pmid: 17379808
2. S. J. J. Brouns *et al*., Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008). doi: 10.1126/science.1159689; pmid: 18703739
3. B. Wiedenheft, S. H. Sternberg, J. A. Doudna, RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331–338 (2012). doi: 10.1038/nature10886; pmid: 22337052
4. F. Hille *et al*., The biology of CRISPR-Cas: Backward and forward. *Cell* **172**, 1239–1259 (2018). doi: 10.1016/j.cell.2017.11.032; pmid: 29522745
5. E. V. Koonin, K. S. Makarova, F. Zhang, Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* **37**, 67–78 (2017). doi: 10.1016/j.mib.2017.05.008; pmid: 28605718
6. L. S. Qi *et al*., Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013). doi: 10.1016/j.cell.2013.02.022; pmid: 23452860
7. M. L. Luo, A. S. Mullis, R. T. Leenay, C. L. Beisel, Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. *Nucleic Acids Res.* **43**, 674–681 (2015). doi: 10.1093/nar/gku971; pmid: 25326321
8. H. K. Ratner *et al*., Catalytically active Cas9 mediates transcriptional interference to facilitate bacterial virulence. *Mol. Cell* **75**, 498–510.e5 (2019). doi: 10.1016/j.molcel.2019.05.029; pmid: 31256988
9. R. E. Workman *et al*., A natural single-guide RNA repurposes Cas9 to autoregulate CRISPR-Cas expression. *Cell* **184**, 675–688.e19 (2021). doi: 10.1016/j.cell.2020.12.017; pmid: 33421369
10. R. Page, W. Peti, Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nat. Chem. Biol.* **12**, 208–214 (2016). doi: 10.1038/nchembio.2044; pmid: 26991085
11. L. R. Walling, J. S. Butler, Toxins targeting transfer RNAs: Translation inhibition by bacterial toxin-antitoxin systems. *Wiley Interdiscip. Rev. RNA* **10**, e1506 (2019). doi: 10.1002/wrna.1506; pmid: 30296016
12. N. Fraikin, F. Goormaghtigh, L. Van Melderen, Type II toxin-antitoxin systems: Evolution and revolutions. *J. Bacteriol.* **202**, e00763-e19 (2020). doi: 10.1128/JB.00763-19; pmid: 31932311
13. M. Li, R. Wang, D. Zhao, H. Xiang, Adaptation of the *Haloarcula hispanica* CRISPR-Cas system to a purified virus strictly requires a priming process. *Nucleic Acids Res.* **42**, 2483–2492 (2014). doi: 10.1093/nar/gkt1154; pmid: 24265226
14. M. Brenneis, O. Hering, C. Lange, J. Soppa, Experimental characterization of *Cis*-acting elements important for translation and transcription in halophilic archaea. *PLOS Genet.* **3**, e229 (2007). doi: 10.1371/journal.pgen.0030229; pmid: 18159946
15. S. Cai *et al*., A novel DNA-binding protein, PhaR, plays a central role in the regulation of polyhydroxyalkanoate accumulation and granule formation in the haloarchaeon *Haloferax mediterranei*. *Appl. Environ. Microbiol.* **81**, 373–385 (2015). doi: 10.1128/AEM.02878-14; pmid: 25344243
16. T. Allers, S. Barak, S. Liddell, K. Wardell, M. Mevarech, Improved strains and plasmid vectors for conditional overexpression of His-tagged proteins in *Haloferax volcanii*. *Appl. Environ. Microbiol.* **76**, 1759–1769 (2010). doi: 10.1128/AEM.02670-09; pmid: 20097827
17. S. Sartorius-Neef, F. Pfeifer, In vivo studies on putative Shine-Dalgarno sequences of the halophilic archaeon *Halobacterium salinarum*. *Mol. Microbiol.* **51**, 579–588 (2004). doi: 10.1046/j.1365-2958.2003.03858.x; pmid: 14756795
18. M. Huber *et al*., Translational coupling via termination-reinitiation in archaea and bacteria. *Nat. Commun.* **10**, 4006 (2019). doi: 10.1038/s41467-019-11999-9; pmid: 31488843
19. O. Hering, M. Brenneis, J. Beer, B. Suess, J. Soppa, A novel mechanism for translation initiation operates in haloarchaea. *Mol. Microbiol.* **71**, 1451–1463 (2009). doi: 10.1111/j.1365-2958.2009.06615.x; pmid: 19210623
20. G. F. T. Chen, M. Inouye, Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes. *Nucleic Acids Res.* **18**, 1465–1473 (1990). doi: 10.1093/nar/18.6.1465; pmid: 2109307
21. G. T. Chen, M. Inouye, Role of the AGA/AGG codons, the rarest codons in global gene expression in *Escherichia coli*. *Genes Dev.* **8**, 2641–2652 (1994). doi: 10.1101/gad.8.21.2641; pmid: 7958922
22. R. Wang, M. Li, L. Gong, S. Hu, H. Xiang, DNA motifs determining the accuracy of repeat duplication during CRISPR adaptation in *Haloarcula hispanica*. *Nucleic Acids Res.* **44**, 4266–4277 (2016). doi: 10.1093/nar/gkw260; pmid: 27085805
23. D. Dar, D. Prasse, R. A. Schmitz, R. Sorek, Widespread formation of alternative 3′ UTR isoforms via transcription termination in archaea. *Nat. Microbiol.* **1**, 16143 (2016). doi: 10.1038/nmicrobiol.2016.143; pmid: 27670118
24. L. Gong *et al*., Primed adaptation tolerates extensive structural and size variations of the CRISPR RNA guide in *Haloarcula hispanica*. *Nucleic Acids Res.* **47**, 5880–5891 (2019). doi: 10.1093/nar/gkz244; pmid: 30957847
25. E. Semenova *et al*., Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10098–10103 (2011). doi: 10.1073/pnas.1104144108; pmid: 21646539
26. H. Zhao *et al*., Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. *Nature* **515**, 147–150 (2014). doi: 10.1038/nature13733; pmid: 25118175
27. M. Li, R. Wang, H. Xiang, *Haloarcula hispanica* CRISPR authenticates PAM of a target sequence to prime discriminative adaptation. *Nucleic Acids Res.* **42**, 7226–7235 (2014). doi: 10.1093/nar/gku389; pmid: 24803673
28. F. Pfeifer, U. Blaseio, Transposition burst of the ISH27 insertion element family in *Halobacterium halobium*. *Nucleic Acids Res.* **18**, 6921–6925 (1990). doi: 10.1093/nar/18.23.6921; pmid: 2175883
29. K. S. Makarova *et al*., An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015). doi: 10.1038/nrmicro3569; pmid: 26411297
30. J. E. Walker *et al*., Development of both type I-B and type II CRISPR/Cas genome editing systems in the cellulolytic bacterium *Clostridium thermocellum*. *Metab. Eng. Commun.* **10**, e00116 (2019). doi: 10.1016/j.mec.2019.e00116; pmid: 31890588
31. P. Boudry *et al*., Function of the CRISPR-Cas system of the human pathogen *Clostridium difficile*. *mBio* **6**, e01112-155 (2015). doi: 10.1128/mBio.01112-15; pmid: 26330515
32. K. Osakabe *et al*., Genome editing in plants using CRISPR type I-D nuclease. *Commun. Biol.* **3**, 648 (2020). doi: 10.1038/s42003-020-01366-6; pmid: 33159140
33. J. Lin *et al*., DNA targeting by subtype I-D CRISPR-Cas shows type I and type III features. *Nucleic Acids Res.* **48**, 10470–10478 (2020). doi: 10.1093/nar/gkaa749; pmid: 32960267
34. J. Iranzo, J. A. Cuesta, S. Manrubia, M. I. Katsnelson, E. V. Koonin, Disentangling the effects of selection and loss bias on gene dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E5616–E5624 (2017). doi: 10.1073/pnas.1704925114; pmid: 28652353
35. A. Stern, L. Keren, O. Wurtzel, G. Amitai, R. Sorek, Self-targeting by CRISPR: Gene regulation or autoimmunity? *Trends Genet.* **26**, 335–340 (2010). doi: 10.1016/j.tig.2010.05.008; pmid: 20598393
36. R. B. Vercoe *et al*., Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLOS Genet.* **9**, e1003454 (2013). doi: 10.1371/journal.pgen.1003454; pmid: 23637624
37. D. Bikard, A. Hatoum-Aslan, D. Mucida, L. A. Marraffini, CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host Microbe* **12**, 177–186 (2012). doi: 10.1016/j.chom.2012.06.003; pmid: 22901538
38. W. Jiang *et al*., Dealing with the evolutionary downside of CRISPR immunity: Bacteria and beneficial plasmids. *PLOS Genet.* **9**, e1003844 (2013). doi: 10.1371/journal.pgen.1003844; pmid: 24086164
39. L. van Sluijs *et al*., Addiction systems antagonize bacterial adaptive immunity. *FEMS Microbiol. Lett.* **366**, fnz047 (2019). pmid: 30834930
40. K. S. Makarova *et al*., Evolutionary classification of CRISPR-Cas systems: A burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020). doi: 10.1038/s41579-019-0299-x; pmid: 31857715
41. A. J. Meeske, S. Nakandakari-Higa, L. A. Marraffini, Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. *Nature* **570**, 241–245 (2019). doi: 10.1038/s41586-019-1257-5; pmid: 31142834
42. J. T. Rostøl, L. A. Marraffini, Non-specific degradation of transcripts promotes plasmid clearance during type III-A CRISPR-Cas immunity. *Nat. Microbiol.* **4**, 656–662 (2019). doi: 10.1038/s41564-018-0353-x; pmid: 30692669
43. S. Song, T. K. Wood, A primary physiological role of toxin/antitoxin systems is phage inhibition. *Front. Microbiol.* **11**, 1895 (2020). doi: 10.3389/fmicb.2020.01895; pmid: 32903830
44. E. V. Koonin, F. Zhang, Coupling immunity and programmed cell suicide in prokaryotes: Life-or-death choices. *BioEssays* **39**, e201600186 (2017). doi: 10.1002/bies.201600186; pmid: 27896818
45. E. V. Koonin, K. S. Makarova, Y. I. Wolf, M. Krupovic, Evolutionary entanglement of mobile genetic elements and host defence systems: Guns for hire. *Nat. Rev. Genet.* **21**, 119–131 (2020). doi: 10.1038/s41576-019-0172-9; pmid: 31611667
46. G. Faure *et al*., Comparative genomics and evolution of trans-activating RNAs in class 2 CRISPR-Cas systems. *RNA Biol.* **16**, 435–448 (2019). doi: 10.1080/15476286.2018.1493331; pmid: 30103650
47. H. Liu, J. Han, X. Liu, J. Zhou, H. Xiang, Development of *pyrF*-based gene knockout systems for genome-wide manipulation of the archaea *Haloferax mediterranei* and *Haloarcula hispanica*. *J. Genet. Genomics* **38**, 261–269 (2011). doi: 10.1016/j.jgg.2011.05.003; pmid: 21703550
48. T. Allers, H. P. Ngo, M. Mevarech, R. G. Lloyd, Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the *leuB* and *trpA* genes. *Appl. Environ. Microbiol.* **70**, 943–953 (2004). doi: 10.1128/AEM.70.2.943-953.2004; pmid: 14766575
49. H. Yu, M. Li, Scripts accompanying "Toxin-antitoxin RNA pairs safeguard CRISPR-Cas systems," Version 1.0, Zenodo (2021); https://doi.org/10.5281/zenodo.4586956.

Downloaded from https://www.science.org at Shanghai Normal University on October 08, 2021

50. C. J. Reuter, J. A. Maupin-Furlow, Analysis of proteasome-dependent proteolysis in *Haloferax volcanii* cells, using short-lived green fluorescent proteins. *Appl. Environ. Microbiol.* **70**, 7530–7538 (2004). doi: 10.1128/AEM.70.12.7530-7538.2004; pmid: 15574956
51. R. Lorenz *et al.*, ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011). doi: 10.1186/1748-7188-6-26; pmid: 22115189
52. M. Mann, P. R. Wright, R. Backofen, IntaRNA 2.0: Enhanced and customizable prediction of RNA-RNA interactions. *Nucleic Acids Res.* **45**, W435–W439 (2017). doi: 10.1093/nar/gkx279; pmid: 28472523
53. R. D. Finn, J. Clements, S. R. Eddy, HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37 (2011). doi: 10.1093/nar/gkr367; pmid: 21593126

ACKNOWLEDGMENTS

We thank X. Chen for providing the *Natrinema* sp. J7-2 strain, X. Zhang for help with microscopy, and K. S. Makarova for help with the bioinformatics analysis. **Funding:** H.X. is supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA24020101), the National Key R&D Program of China (2020YFA0906800), and the National Natural Science Foundation of China (91751201). M.L. is supported by the National Natural Science Foundation of China (31771381, 32022003, and 31970544), the National Transgenic Science and Technology Program (2019ZX08010-001 and 2019ZX08010-003), the Youth Innovation Promotion Association of CAS (2020090), and the Young Elite Scientists Sponsorship Program by CAST (2017QNRC001). S.A.S. and E.V.K. are supported by the Intramural Research Program of the National Institutes of Health (National Library of Medicine). **Author contributions:** M.L. and H.X. designed experiments and supervised the project. M.L., L.G., and F.C. constructed pTA derivatives and performed the transformation assays with input from T.W. and J.Z. M.L. and L.G. performed the RNA-seq, Northern blotting, primer extension, and plasmid challenge assays. F.C. performed the fluorescence analyses and measured the *H. volcanii* growth curves. M.L. and H.Y. analyzed the RNA-seq data. R.W. constructed the *cas* knockout mutants. M.L. performed the bioinformatic analyses with input from H.Y., S.S., D.Z., and S.Z. M.L., E.V.K., and H.X. analyzed the data and wrote the manuscript, which was edited and approved by all authors. **Competing interests:** M.L., F.C., and H.X. have filed a related patent. **Data and materials availability:** All data are available in the main text or the supplementary materials. Reagents are available upon request from H.X. Perl scripts are available on Zenodo (*49*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6541/eabe5601/suppl/DC1
Figs. S1 to S14
Tables S1 to S3
MDAR Reproducibility Checklist

View/request a protocol for this paper from Bio-protocol.

16 September 2020; resubmitted 29 December 2020
Accepted 10 March 2021
10.1126/science.abe5601

Downloaded from https://www.science.org at Shanghai Normal University on October 08, 2021

Science

Toxin-antitoxin RNA pairs safeguard CRISPR-Cas systems

Ming LiLuyao GongFeiyue ChengHaiying YuDahe ZhaoRui WangTian WangShengjie ZhangJian ZhouSergey A. ShmakovEugene V. KooninHua Xiang

Science, 372 (6541), eabe5601.

Small RNAs guard CRISPR-Cas

The microbial adaptive immunity system CRISPR-Cas benefits microbes by warding off genetic invaders, but it also inflicts a fitness cost because of occasional autoimmune reactions, rendering CRISPR loci evolutionarily unstable. Li *et al.* identified previously unnoticed toxin-antitoxin RNA pairs embedded within diverse CRISPR-Cas loci. The antitoxin RNA mimics a CRISPR RNA and repurposes the CRISPR immunity effector to transcriptionally repress a toxin RNA that would otherwise arrest cell growth by sequestering a rare transfer RNA. These small RNAs thus form a symbiosis with CRISPR, rendering CRISPR addictive to the host despite its fitness cost. These findings reveal how CRISPR-Cas can operate as a selfish genetic element.

Science, this issue p. eabe5601

View the article online
https://www.science.org/doi/10.1126/science.abe5601
Permissions
https://www.science.org/help/reprints-and-permissions

Downloaded from https://www.science.org at Shanghai Normal University on October 08, 2021

Use of think article is subject to the Terms of service

Science (ISSN 1095-9203) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.
Copyright © 2021 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works