AcrIIA22 is a novel anti-CRISPR that impairs SpyCas9 activity by relieving DNA torsion of target plasmids

Kevin J. Forsberg[a,b,*], Danica T. Schmidtke[a], Rachel Werther[a], Deanna Hausman[a], Barry L. Stoddard[a], Brett K. Kaiser[a,c], Harmit S. Malik[a,b]

[a]Division of Basic Sciences & [b]Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, WA, 98109 USA; [c]Department of Biology, Seattle University, 901 12th Avenue, Seattle, WA 98122, USA

Running title: Altering DNA torsion can protect plasmids from Cas9

*Address correspondence to: Kevin J. Forsberg, 1100 Fairview Avenue N. A2-205, Seattle WA 98109; (206) 667-4512; email: kforsber@fredhutch.org

Abstract

To overcome CRISPR-Cas defense systems, many phages and mobile genetic elements encode CRISPR-Cas inhibitors called anti-CRISPRs (Acrs). Nearly all mechanistically characterized Acrs directly bind their cognate Cas protein to inactivate CRISPR immunity. Here, we describe AcrIIA22, an unconventional Acr found in hypervariable genomic regions of *Clostridial* bacteria and their prophages from the human gut microbiome. Uncovered in a functional metagenomic selection, AcrIIA22 does not bind strongly to SpyCas9 but nonetheless potently inhibits its activity against plasmids. To gain insight into its mechanism, we obtained an X-ray crystal structure of AcrIIA22, which revealed homology to PC4-like nucleic-acid binding proteins. This homology helped us deduce that *acrIIA22* encodes a DNA nickase that relieves torsional stress in supercoiled plasmids, rendering them less susceptible to SpyCas9, which is highly dependent on negative supercoils to form stable R-loops. Modifying DNA topology may provide an additional route to CRISPR-Cas resistance in phages and mobile genetic elements.

Introduction

CRISPR-Cas systems in bacteria and archaea confer sequence-specific immunity against invading phages and other mobile genetic elements (MGEs)[1,2]. In response, MGEs can circumvent CRISPR-Cas systems by evading CRISPR immunity. In its simplest form, evasion requires only a single mutation to a CRISPR target site, which allows a phage or MGE to escape immune recognition[3]. However, CRISPR-Cas systems routinely acquire new spacer sequences to target new sites within phage and MGE genomes[1]. This means that any single-site evasion strategy is likely to be short-lived. Thus, phages also employ forms of CRISPR-Cas evasion that are less easily subverted. For instance, some jumbophages assemble a proteinaceous, nucleus-like compartment around their genomes upon infection, allowing them to overcome diverse bacterial defenses, including CRISPR-Cas and restriction-modification (RM) systems[4,5]. Similarly, other phages decorate their DNA genomes with diverse chemical modifications, which can prevent Cas nucleases from binding their target sequence, such as the glucosylated cytosines used by phage T4 of *Escherichia coli*[6].

MGEs may also overcome CRISPR-Cas systems by inactivating, rather than evading, CRISPR immunity. MGEs encode diverse CRISPR-Cas inhibitors called anti-CRISPRs (Acrs), which allow them to overcome CRISPR-Cas systems and infect otherwise immune hosts[7]. Most known Acrs bind Cas proteins and inhibit Cas activity by either restricting access to target DNA, preventing necessary conformational changes, or inactivating critical CRISPR-Cas components[8,9]. The direct inactivation of Cas proteins by Acrs has proven an effective and widespread strategy for overcoming CRISPR immunity[10].

Recent genetic, bioinformatic, and metagenomic strategies have identified many Acrs that independently target the same CRISPR-Cas system[7-10]. Yet, most CRISPR-Cas systems are not inhibited by known Acrs[10]. Thus, many undiscovered strategies to inhibit or evade CRISPR-Cas systems probably exist in nature. Indeed, over half of the genes in an average phage genome have no known function[11]. To uncover new counter-immune strategies, we recently devised a high-throughput functional metagenomic selection to find genes that protect a target plasmid from *Streptococcus pyogenes* Cas9 (SpyCas9), the variant used most frequently for genome editing[12]. Our selection strategy was designed to reveal any gene capable of overcoming SpyCas9 activity in this system, regardless of mechanism. With this approach, we previously described a new

phage inhibitor of SpyCas9, called AcrIIA11, which acts via a novel mechanism and is prevalent across human gut microbiomes[12].

Here, we describe AcrIIA22, which was the second most common Acr candidate recovered from our original functional selection. AcrIIA22 encodes a 54 amino acid protein that impairs SpyCas9 activity. We observe that homologs of *acrIIA22* are found in hypervariable loci in phage and bacterial genomes. Unlike most other Acrs, AcrIIA22 does not bind strongly to SpyCas9 *in vitro*. Instead, guided by an X-ray crystal structure of AcrIIA22, we show that AcrIIA22 encodes a DNA nickase. By nicking a supercoiled plasmid substrate and relieving its torsional stress, AcrIIA22 renders the target less susceptible to SpyCas9 activity. AcrIIA22 thus represents a novel mechanism of SpyCas9 evasion, which capitalizes on SpyCas9's uniquely stringent requirement for negative supercoils to form a productive R-loop[13-16]. Such a resistance mechanism could be accessible to diverse MGEs, providing a route to CRISPR-Cas tolerance in many genetic contexts.

Results

Functional selection reveals a novel anti-CRISPR protein, AcrIIA22

We recently carried out a functional selection for SpyCas9 antagonism, recovering clones from metagenomic libraries that could potently inhibit SpyCas9[12]. In this two-plasmid setup, we used an inducible SpyCas9 on an expression plasmid to cleave the *kanamycin resistance* (*Kan^R*) gene of a second 'target' plasmid. We then grew cultures in SpyCas9-inducing conditions and measured the proportion of colony forming units (cfus) that remained kanamycin resistant (Figure 1A). This proportion is a measure of how many clones retained their target plasmid and thus how effectively that plasmid withstood SpyCas9 attack. In our previously published work, we describe AcrIIA11, a novel anti-CRISPR from a metagenomic clone named F01A_2 (Genbank ID MK637582.1), which was the most abundant from functional selection of a human fecal microbiome[12]. This functional selection also revealed a second protective clone, F01A_4 (Genbank ID MK637587.1). Together, these two contigs (F01A_2 and F01A_4) accounted for >96% of the normalized read coverage and were the most abundant clones recovered from this library.

The F01A_4 contig is 685 bp long, encodes three potential open reading frames (ORFs), and confers complete protection against SpyCas9, with plasmid retention equaling that of an uninduced SpyCas9 control (Figure 1B). To determine the genetic basis for SpyCas9 antagonism in this contig, we introduced an early stop codon into each of the three potential ORFs and analyzed how these mutations affected the contig's ability to protect a target plasmid from SpyCas9. We found that an early stop codon in *orf_1* reduced the proportion of *Kan^R* cfus by a factor of 10^5, matching the value observed for an empty vector control (Figure 1B). Furthermore, expression of *orf_1* alone was also sufficient for SpyCas9 antagonism (Figure 1C), protecting a target plasmid from SpyCas9 cleavage as well as the potent SpyCas9 inhibitor, AcrIIA4. In this assay, *orf_1* was slightly toxic when singly expressed in *E. coli*, reducing growth rate by 7% (Supplemental Figure 1). Combined, our results indicate that *orf_1* completely accounts for the SpyCas9 protection phenotype of contig F01A_4.

One trivial mechanism by which *orf_1* could apparently antagonize SpyCas9 in our functional assay would be to lower its expression. To address this possibility, we carried out two experiments. First, we swapped the *spycas9* gene for *gfp* in our expression vector and asked whether *orf_1* induction impacted fluorescence output. We saw no change in fluorescence upon

orf_1 induction, indicating that *orf_1* neither suppressed transcription from our expression vector nor altered its copy number (Supplemental Figure 2). Second, we used Western blots to test whether *orf_1* expression impacted SpyCas9 protein levels through the course of a plasmid protection assay. We used a crRNA that did not target our plasmid backbone to ensure that *orf_1* expression remained high and its potential impact on SpyCas9 expression levels would be most evident. We observed that *orf_1* expression had no meaningful effect on SpyCas9 expression at any timepoint. Thus, we conclude that *orf_1* does not impact SpyCas9's translation or degradation rate (Supplemental Figure 2). Therefore, *orf_1* must act via an alternative mechanism to inhibit SpyCas9 activity. Based on these findings, we conclude that *orf_1* encodes a *bona fide* anti-CRISPR protein and hereafter refer to it as *acrIIA22*.

Next, we investigated whether *acrIIA22* could also allow phages to escape from SpyCas9 immunity (Supplemental Figure 3). We measured SpyCas9's ability to protect *E. coli* from infection by phage Mu, in the presence or absence of *acrIIA22*. As a control, we carried out similar phage infections in the presence or absence of the well-established SpyCas9 inhibitor, *acrIIA4*. As anticipated, SpyCas9 significantly impaired Mu when targeted to the phage's genome but not if a non-targeting CRISPR RNA (crRNA) was used. Consistent with previous findings[12], phage Mu could infect targeting and non-targeting strains equally well when we expressed *acrIIA4*, indicating that SpyCas9 immunity was completely abolished by this *acr*. However, *acrIIA22* could only partially restore the infectivity of phage Mu across multiple experimental conditions (Supplemental Figure 3). We therefore conclude that *acrIIA22* only weakly protects Mu phage from SpyCas9 whereas it strongly protects plasmids against SpyCas9 cleavage.

AcrIIA22 homologs are present in hypervariable regions of bacterial and prophage genomes

AcrIIA22 is 54 amino acids in length and has no sequence homology to any protein of known function, including all previously described Acrs. We examined the distribution of *acrIIA22* homologs in NCBI's NR and WGS databases but found just seven hits, limiting our ability to make evolutionary inferences about its origins or prevalence. We therefore expanded our search to include IMG/VR, a curated database of cultured and uncultured DNA viruses[17], and assembly data from a meta-analysis of 9,428 diverse human microbiome samples[18]. With additional homologs from these databases in hand, we found that the majority of *acrIIA22* homologs exist in either of two genomic contexts: prophage genomes or small, bacterial genomic islands (Figures 2A, 2B). The original metagenomic DNA fragment from our selection, F01A_4, shared perfect nucleotide identity with one of these genomic islands (Figure 2B).

Because most *acr*s are found in phage genomes, we first examined the prophages that encoded AcrIIA22 homologs. These prophages were clearly related, based on many homologous genes and a similar genome organization (Figure 2A). Despite their similarity, we found these prophages inserted into several different bacterial loci, including one site between the bacterial genes *purF* and *radC* (locus #3, Figure 2A). This prophage insertion site is notable because it is nearly identical to the highly conserved sequences that flanked *acrIIA22*-encoding bacterial genomic islands (Figure 2B). Due to their common genomic loci, we hypothesized that the apparently bacterial *acrIIA22* homologs in these genomic islands diverged from a common phage ancestor, encoded by a prophage that previously integrated at this locus. We speculate that the original *acrIIA22*-encoding bacterial genomic island was left behind following the incomplete excision of an ancestral, *acrIIA22*-encoding prophage. Supporting this hypothesis, *acrIIA22*

homologs are always found at the end of prophage genomes, near their junction with a host bacterial genome (Figure 2A).

To better understand *acrIIA22*'s gene neighborhood, we again searched the raw assemblies of over 9,400 human microbiomes for more examples of these genomic islands[18], but did not include *acrIIA22* in our second search criteria. Instead, we focused on the recent evolutionary history of these bacterial genomic islands by only considering contigs with ≥98% nucleotide identity to *purF* and *radC*, the conserved genes that flanked the genomic islands. This search yielded 258 contigs. Aligning these sequences revealed that each contig encoded a short, hypervariable region of small ORFs which was flanked by conserved genomic sequences (Figure 2B). In total, we observed 128 unique examples of these hypervariable loci, which displayed considerable gene turnover, resulting in 54 distinct gene arrangements among the 128 unique loci. Despite not including them in our search strategy, *acrIIA22* homologs were universally conserved in all 128 unique genomic islands whereas no other gene was present in more than two-thirds of the 54 distinct gene arrangements (Figure 2C). Based on this finding, we infer that the arrival of *acrIIA22* preceded the diversification seen at this locus and that its homologs have been retained since, despite the considerable gene turnover that has occurred subsequently.

Though most ORFs in these islands were of unknown function, many had close homologs in the genomes of nine representative *acrIIA22*-encoding phage (dashed boxes in Figure 2A, phage icons in Figure 2C). This suggests that phages continue to supply the genetic diversity seen at these hypervariable genomic loci. These rapid gene gains and losses probably occur as they do in other genomic islands, via recombination between this locus and related MGEs that infect the same host bacterium without the MGE necessarily integrating into the locus[19]. Taken together, our data suggest that an incomplete prophage excision event left *acrIIA22* behind in a bacterial genomic locus, which then diversified via gene exchange with additional phage genomes (Figure 2D).

In prophage genomes, *acrIIA22* homologs were found in hypervariable regions, near the junction with the host bacterial genome (Figure 2A). Both features imply these loci are subject to higher than average rates of recombination. Despite this, we could find no gene consistently present within or outside of these genomic islands that could account for their hypervariable nature (*e.g.* an integrase, transposase, recombinase, or similar function that is typically associated with genomic islands[20]). Instead, *acrIIA22* was the only gene conserved at this locus. If it could somehow promote recombination, either alone or with other factors, this could account for the high rates of gene exchange observed adjacent to *acrIIA22* in phage and bacterial genomes (Figures 2A, 2B).

In total, we identified 30 unique *acrIIA22* homologs, 25 of which were predicted to originate from the unnamed *Clostridial* genus, CAG-217 (Figure 3A). Because *acr*s are only beneficial to phages if they inhibit CRISPR-Cas activity, they are typically found only in taxa with a high prevalence of susceptible Cas proteins[9]. If AcrIIA22 functions naturally as an Acr, we would predict that Cas9-encoding, type II-A CRISPR-Cas systems like SpyCas9 would be common in CAG-217 bacteria. To test this idea, we examined 779 draft assemblies of CAG-217 genomes and found that 179 of the 181 predicted CRISPR-Cas systems in CAG-217 genomes were Cas9-encoding, type II-A systems. This enrichment for Cas9 is particularly striking for a *Clostridial* genus, as *Clostridia* rarely encode Cas9. Instead, they typically encode other CRISPR-Cas defenses[21]. Thus, the distribution of CRISPR-Cas systems in CAG-217 genomes supports our hypothesis that *acrIIA22* functions natively as an *acr*.

We also found evidence that Cas9 is active in CAG-217 bacteria. Prophages from CAG-217 encode 78 type II-A Acrs (homologs of AcrIIA7, AcrIIA17, and AcrIIA21), suggesting they are actively engaged in an arms race with Cas9-based defenses in these bacteria. We even found one example where homologs of *acrIIA17* and *acrIIA22* were located within one kilobase of each other in a prophage genome (Supplemental Figure 4)[22]. Since phages often aggregate *acr*s in the same genomic locus[23], this observation independently supports our hypothesis that CAG-217 prophages encode *acrIIA22* homologs to inhibit type II-A CRISPR-Cas systems.

We next tested whether the ability to inhibit type II-A CRISPR-Cas systems was a shared property of *acrIIA22* homologs from CAG-217 bacteria. To do so, we selected *acrIIA22* homologs that spanned the phylogenetic diversity present among CAG-217 genomes (Figure 3A) and tested their ability to protect a target plasmid from SpyCas9 elimination. These analyses revealed that each *acrIIA22* homolog from CAG-217 could antagonize SpyCas9 activity at least partially (Figure 3B). This conservation of anti-SpyCas9 activity among divergent AcrIIA22 homologs (for example, sharing only 56.9% identity), suggests that they may broadly inhibit Cas9. Broad inhibition has been seen for some other type II-A Acrs[12] and can occur either by targeting a conserved feature of Cas9 or by inhibiting Cas9 via an indirect mechanism that it cannot easily evade.

AcrIIA22 functions via a non-canonical mechanism

Almost all characterized Acrs inhibit their cognate Cas proteins via direct binding without the involvement of additional co-factors; as a result, they exhibit strong inhibitory activity when tested *in vitro* (Supplemental Table 1). To determine if this was the case for AcrIIA22, we purified it from *E. coli* and asked whether it could bind and inhibit SpyCas9. To test for binding, we asked whether a tagged AcrIIA22 co-precipitated with SpyCas9 when mixed as purified proteins. Unlike AcrIIA4, which binds strongly to SpyCas9 and inhibits its activity *in vitro*, we could detect little to no binding between AcrIIA22 and SpyCas9, regardless of whether a single-guide RNA (sgRNA) was included or not (Supplemental Figure 5). We also observed that AcrIIA22 had no impact on SpyCas9's ability to cleave linear, double-stranded DNA (dsDNA), even when AcrIIA22 was included at substantial molar excess over SpyCas9 (Supplemental Figure 6). These results suggest that, at least in isolation, AcrIIA22 cannot bind and inhibit SpyCas9. Thus, AcrIIA22 lacks the predominant biochemical activities exhibited by previous Acrs that have been mechanistically characterized.

We therefore considered the possibility that AcrIIA22 encodes an unconventional anti-CRISPR that acts via a non-canonical mechanism. However, AcrIIA22 homologs had no sequence homology to other characterized proteins, which would have provided clues about AcrIIA22 activity and biochemical mechanisms. Anticipating that structural homology might provide some insight, we solved AcrIIA22's structure using X-ray crystallography. We first built a homology model from AcrIIA22's primary sequence with Robetta. We then used this model for molecular replacement to solve its structure at 2.80Å resolution (PDB:7JTA). The asymmetric unit in AcrIIA22's crystal comprises two monomers stacked end-to-end, with each monomer folding into a four-stranded β-sheet (Figure 4A, Table 1). A DALI structure-structure search revealed that the AcrIIA22 monomer is similar to members of the newly recognized PC4-like structural fold (Figure 4B, Supplemental Table 2). PC4-like proteins have independently evolved in all domains of life, typically adopt a β-β-β-β-α topology, and often homodimerize to bind diverse RNA and DNA species using variably positioned β-sheets[24].

Despite crystallizing as a homodimer, AcrIIA22 migrated at a size substantially larger than its molecular weight by size exclusion chromatography (SEC) (Figure 4C). This suggested that AcrIIA22 may oligomerize *in vivo*. Structural evidence supported the same conclusion, as AcrIIA22 was predicted to form a stable tetramer when analyzed with PISA, a tool for inferring macromolecular assembles from crystalline structure[25] (Figures 4D, 4E). This putative tetramer has a molecular mass consistent with that observed by SEC and comprises pairs of outward-facing, concave β-sheets, as is seen in other PC4-like proteins[24]. Interestingly, many PC4-like proteins bind nucleic acids using similar concave β-sheets and, in some instances, form higher-order oligomers as a necessary step for binding DNA or RNA[24]. Consistent with this possibility, adjacent β-sheets along each outward face of the putative AcrIIA22 tetramer form a groove that could potentially accommodate a nucleic acid substrate (Figure 4E). Thus, even though AcrIIA22 lacks the alpha-helix typically seen in PC4-like proteins, its other structural and functional attributes led us to suspect that AcrIIA22 also interacted with nucleic acids.

Our tetramer model predicts that a four amino acid interface at the C-terminus of AcrIIA22 is required for adjacent β-sheets to bind one another and form a grooved, oligomeric structure (Figures 4D, 4F). We predicted that a two-residue, C-terminal truncation of AcrIIA22 would disrupt this interface (Figure 4F). To test this prediction, we examined the oligomeric state of this 2-aa AcrIIA22 deletion mutant. Consistent with our hypothesis, we found that these AcrIIA22 complexes migrated at about half the size of their wild-type counterparts by SEC (Figure 4C), suggesting that this C-terminal interface is required to progress from a two to four-membered oligomer. Moreover, we found that the 2-aa deletion mutant was also impaired for SpyCas9 antagonism in our plasmid protection assay (Figure 4G). Thus, this C-terminal motif is necessary for protection from SpyCas9 and higher-order oligomerization, suggesting that oligomerization may be necessary for AcrIIA22's anti-SpyCas9 activity.

AcrIIA22 is a DNA nickase that relieves torsion of supercoiled plasmids

Our structural analyses indicated that AcrIIA22 is a PC4-like nucleic acid-interacting protein. Like AcrIIA22, many PC4-like proteins are encoded in phage genomes. Among these is AcrIIA22's closest structural relative in the PC4 family: a predicted single-stranded binding (SSB) protein from phage T5 (Figure 4B)[26]. This putative SSB protein has been predicted to directly stimulate recombination during the recombination-dependent replication of phage T5's genome[27]. This prediction, together with our inference from genomic analyses (Figure 2), led us to hypothesize that AcrIIA22 may have similar recombination-stimulating activity. Indeed, other PC4-like proteins have been observed experimentally to unwind duplex DNA, a function consistent with their proposed roles in transcription and recombination[24,28]. Therefore, we investigated whether AcrIIA22 might interact with duplexed DNA in a manner consistent with its putative recombinogenic properties.

We first asked whether we could detect any biochemical effect of *acrIIA22* on a double-stranded DNA (dsDNA) target plasmid *in vivo*. In this experiment, we considered three *acrIIA22* genotypes: the wild-type sequence, a null mutant with a single base pair change to create an early stop codon, and the 2-aa truncation mutant that we previously showed was defective for oligomerization (Figure 4C) and SpyCas9 antagonism (Figure 4G). We then grew overnight cultures of plasmids expressing each genotype, purified plasmid DNA, and analyzed its topology using gel electrophoresis (Figure 5A). As is typical for plasmid purifications from *E. coli*, the plasmid encoding the null mutant was predominantly recovered in a supercoiled form. In contrast, AcrIIA22 expression shifted much of the target plasmid to a slowly migrating form, consistent with

an open-circle conformation. These findings suggest that AcrIIA22 expression could relieve plasmid supercoiling, hinting at a potential DNA nickase activity. We also found that the 2-aa truncation mutant was impaired for this putative nickase activity, consistent with this mutant's compromised oligomerization and anti-Cas9 activities (Figure 4G).

Because *acrIIA22* expression altered plasmid topology *in vivo*, we next asked whether purified AcrIIA22 had an impact on a plasmid DNA substrate *in vitro*. By gel electrophoresis, we observed that AcrIIA22 shifted a supercoiled plasmid to a slowly migrating form in a time and concentration-dependent manner (Figures 5B, 5C, Supplemental Figure 7). For comparison, we also treated a plasmid with the nickase Nb.BssSI, yielding a band that migrated at the same position as the putatively open-circle product generated via AcrIIA22 activity (Figure 5B). Extended incubation times and high concentrations of AcrIIA22 resulted in conversion of plasmids to a linearized DNA product, consistent with a nickase-like nuclease activity acting on both strands of DNA (Figure 5B, Supplemental Figure 7). AcrIIA22's nickase activity was strongly stimulated in the presence of Mn^{2+}, Co^{2+}, and Mg^{2+}, weakly with Ni^{2+} and Zn^{2+}, but not at all with Ca^{2+} (Supplemental Figure 8). Consistent with our *in vivo* observations, we found that the 2-aa deletion mutant was impaired for nickase activity, relative to wildtype AcrIIA22 (Figure 5D).

Our *in vitro* and *in vivo* findings suggest that AcrIIA22 is responsible for the change in plasmid topology observed in bacterial cells. To confirm that the observed gel-shift was the result of AcrIIA22 nickase activity and not protein-bound DNA, we purified an AcrIIA22-treated plasmid with phenol-chloroform and re-examined it by gel electrophoresis. We observed that the nicked form of the plasmid persisted through purification, establishing AcrIIA22 as a *bona-fide* nickase (supp figure 8). Therefore, based on both *in vitro* and *in vivo* findings, we conclude that that *acrIIA22* encodes a nickase protein that relieves the torsional stress of supercoiled plasmids.

AcrIIA22's nickase activity impairs SpyCas9

Having established that AcrIIA22 is a DNA nickase, we next investigated whether this biochemical activity correlated with its ability to inhibit SpyCas9. If this were the case, it would explain how AcrIIA22 protected plasmids from SpyCas9 without directly binding the Cas protein. We therefore tested the consequences of expressing AcrIIA22 on a target plasmid in the presence of SpyCas9. As before, we began by comparing overnight plasmid purifications of a target plasmid expressing AcrIIA22, a null mutant with an early stop codon, or the 2-aa AcrIIA22 truncation mutant. However, this time, we also subjected the plasmid to SpyCas9 targeting during bacterial growth. We were unable to recover the negative control target plasmid after overnight growth, implying that this target plasmid was eliminated by SpyCas9 (Figure 6A). The 2-aa truncation mutant was also eliminated by SpyCas9, indicating these residues are important for function. In contrast, SpyCas9 did not eliminate a target plasmid that expressed full-length AcrIIA22 (Figure 6A), consistent with AcrIIA22's previously established capacity to protect against SpyCas9 (Figure 1C).

To be effective, a CRISPR-Cas system must eliminate its target at a faster rate than the target can replicate[29]. Our findings raised the possibility that AcrIIA22 modifies a target plasmid into a SpyCas9-resistant conformation to win this 'kinetic race' against SpyCas9, potentially shifting the equilibrium to favor plasmid persistence instead of elimination. To test this kinetic race model, we asked whether a plasmid that had been pre-treated with AcrIIA22 could resist digestion by SpyCas9 *in vitro*. Therefore, we purified the open-circle plasmid that resulted from AcrIIA22 pre-treatment and determined how efficiently it was cleaved by SpyCas9 compared to an unmodified,

supercoiled plasmid (Figure 6B). SpyCas9 showed a clear preference for cleaving the supercoiled substrate versus the AcrIIA22-treated open-circle plasmid (Figure 6C, Supplemental Figure 9). An open-circle plasmid pre-treated with the nickase Nb.Bss.SI was similarly recalcitrant to SpyCas9 digestion. Taken together, our findings suggest that relieving DNA torsion provides the mechanistic explanation for AcrIIA22's ability to inhibit SpyCas9 *in vivo*.

Our findings also help explain why AcrIIA22 is more adept at protecting plasmids than phages from SpyCas9 in our system. Because plasmids are maintained as circular, extrachromosomal elements, they are more likely to undergo torsional change when nicked than phages or transposons, which are often linear or spend significant time integrated into their host's genome. Additionally, linear DNA experiences minimal torsional stress and is therefore less susceptible than supercoiled plasmids to cleavage by SpyCas9[15]. This likely explains why AcrIIA22 failed to protect a linear dsDNA substrate from SpyCas9 *in vitro* (Supplemental Figure 6), as there is very little torsional stress for it to relieve in this substrate. Importantly, *in vitro* experiments indicate that Cas9 requires a higher degree of negative supercoiling to provide the free energy needed for R-loop formation than type I CRISPR-Cas systems[13]. *In vivo* observations also show that DNA supercoiling affects the recruitment of SpyCas9 to its target site in bacteria[14]. This suggests that Cas9 may be particularly susceptible to changes in DNA torsion. Thus, factors like the AcrIIA22 nickase, which modify DNA torsion, may provide a general means to protect against Cas9.

Discussion

In this study, we identify and characterize *acrIIA22*, a previously undescribed gene that can antagonize SpyCas9. We show that AcrIIA22 homologs have proliferated in genomes of CAG-217 bacteria, which have a high prevalence of Cas9 homologs. Using a combination of structural and biochemical studies, we show that AcrIIA22 acts by nicking supercoiled DNA to relieve torsional stress on a target plasmid, thereby impairing SpyCas9 activity *in vivo* and *in vitro*. Taken together, our data suggest that DNA topology represents a new battleground in the evolutionary arms race between CRISPR-Cas systems and MGEs. Because Cas9 is more susceptible to evasion via changes to DNA topology than other CRISPR-Cas systems[13], it may be more disadvantaged than other bacterial defense systems in this arms race. Additionally, DNA topology is dynamically regulated in phages, plasmids and other MGEs. This means that topology-modifying factors already exist in diverse MGEs that could have secondary effects on CRISPR-Cas activity and thus prove useful in the context of a molecular arms race[30,31]. For instance, though not studied in the context of bacterial defense systems, the fitness of phage T4 is improved via the expression of an accessory protein that modifies DNA supercoiling and the propensity of R-loops to form[32]. Other phages, such as those in the T5-like family, incorporate regular nicks into their genome, the function of which has eluded description for over 40 years[33]. Based on our findings, we hypothesize that phages and MGEs targeted by Cas9 exploit factors that modify DNA topology as a tactic to evade host immunity.

Functional selections like ours are biased towards identifying genes that work well in a heterologous context. For example, even though AcrIIA22 is encoded on the genome of a genetically intractable bacterium, we could identify it using a functional metagenomic selection for SpyCas9 antagonism in *E. coli*. Although we have characterized its activities in *E. coli* and *in vitro*, it remains formally possible that AcrIIA22 functions differently in its native context. For instance, we cannot rule out the possibility that AcrIIA22 might interact with a Cas9 protein from a CAG-217 bacterium. Alternatively, AcrIIA22's anti-Cas9 activity might be related to a native recombinogenic function (Figure 2). As precedent for this idea, CRISPR-Cas evasion was recently

355 demonstrated for homologs of the recombination proteins Redβ and λExo in *Vibrio cholerae*[34].
356 Nevertheless, the heterologous behavior of AcrIIA22 in *E. coli* is clearly sufficient for SpyCas9
357 antagonism *in vivo* and its nickase activity can protect plasmids from SpyCas9 *in vitro*.
358 Furthermore, an AcrIIA22 mutant that is partially defective for nickase activity *in vitro* (Figure 5D)
359 is ~1,000-fold less effective at protecting a plasmid from SpyCas9 *in vivo* (Figure 4G). This
360 indicates that modest changes in nickase activity can have major consequences for plasmid
361 survival, which is consistent with our kinetic race model (Figure 6B) and previous observations
362 that non-linear equilibrium dynamics determine whether an MGE withstands CRISPR-Cas
363 immunity[29].

364 Our results suggest that proteins that affect DNA torsion may also enable Cas9 antagonism.
365 For example, in addition to AcrIIA22, the Nb.BssSI nickase was also capable of protecting a
366 plasmid from SpyCas9 *in vitro*. Yet, despite the regular occurrence of nickases in nature,
367 functional selections for anti-Cas9 activity have not previously recovered these enzymes[12,35]. We
368 speculate that AcrIIA22 treads a fine balance between activity and toxicity; its nickase activity is
369 high enough to antagonize SpyCas9 in a kinetic race, but not so high that it would be toxic to the
370 host cell (Supplemental Figure 1). Its oligomerization may represent an important mechanism to
371 control nickase activity and suppress host toxicity. Studies of other phage- and bacterial-encoded
372 nickase proteins may provide additional insight into whether AcrIIA22 proteins have additional
373 properties that render them to be especially well-suited to antagonize SpyCas9.

374 Is AcrIIA22 a true anti-CRISPR? AcrIIA22 lacks features that are typical of conventional Acrs,
375 such as the ability to bind Cas proteins or to inhibit CRISPR-Cas activity as a purified protein.
376 However, other Acr proteins also lack these features. For example, the well-characterized
377 SpyCas9 antagonist AcrIIA1 does not inhibit purified SpyCas9, but instead stimulates Cas9
378 degradation[36]. Similarly, AcrIIA7 does not appear to bind SpyCas9 but can nevertheless inhibit it
379 *in vitro* via an unknown mechanism[35]. Indeed, anti-CRISPR proteins are defined by a common
380 strategy and outcome rather than by a common biochemical mechanism. Our finding that
381 *AcrIIA22* is encoded by prophages as a single gene that strongly protects plasmids and weakly
382 protects phages from SpyCas9 (Figure 3B, Supplemental Figure 4) makes it much more similar
383 to other Acrs[23] and distinct from non-canonical CRISPR-Cas evasion strategies like DNA
384 glucosylation[6] or homologous recombination[34].

385 AcrIIA22 does not appear to provide the same potency of Cas9 inhibition as some other
386 characterized Acrs, particularly in protecting phage Mu. However, potent inhibition is not a pre-
387 requisite for effective anti-CRISPR activity. Indeed, selection can favor weak anti-CRISPRs over
388 strong ones in mixed phage populations[37]. Even in cases where mechanisms for Cas9 evasion
389 are weak (Supplemental Figure 3), they may nonetheless confer substantial benefit. For example,
390 slowing down Cas9 cleavage could increase the time and probability for escape mutants to arise
391 (*e.g.* Cas9 target-site variants[1], deletion mutants[34]), allow for additional Acr expression[38,39], or
392 permit further genome replication to overwhelm CRISPR-Cas immunity[29]. This phenomenon –
393 weak tolerance giving rise to long-term resistance – is reproducibly observed in cases of strong
394 selective pressure. For instance, in the context of antibiotic resistance, the expression of QNR
395 pentapeptide proteins in many human pathogens can provide low-level drug tolerance, extend
396 survival, and allow time for additional mutations to develop that completely resist quinolone
397 antibiotics[40].

398 As the use of functional metagenomics to study phage-bacterial conflicts grows more
399 common, many novel genes and mechanisms for CRISPR-Cas inhibition are likely to be

described[12,35]. Like AcrIIA22, which has no homology to any previously described anti-CRISPR and lacks other genetic signatures used for *acr* discovery (*e.g.,* linkage with helix-turn-helix transcription factors)[41,42], these new genes may not exhibit canonical Acr behaviors. It is inevitable that these discoveries will lead a more nuanced understanding of the arms race between CRISPR-Cas systems and MGEs. While they might blur the precise boundaries on what defines an anti-CRISPR, these findings will also reveal undiscovered strategies for molecular antagonism and new battlegrounds in the age-old conflict between bacteria and their phages.

Methods

Plasmid protection assay

All assays were done in *Escherichia coli* (strain: NEB Turbo). As described previously[12], SpyCas9 was expressed from a CloDF13-based plasmid marked with a spectinomycin resistance cassette. The SpyCas9 construct programed to eliminate a kanamycin-marked target plasmid was called pSpyCas9_crA (Supplemental Table 4). It eliminated a target vector that inducibly expressed a gene-of-interest via depression of the TetR transcription factor with doxycycline (named generically pZE21_tetR; Supplemental Table 4). IPTG was used in samples with the target vector to ensure high levels of TetR expression (which was driven by the lac promoter) and thus inducible control of our gene of interest. Cultures of each sample were grown overnight at 37C with shaking at 220 rpm in lysogeny broth (LB; 10 g/L casein peptone, 10 g/L NaCl, 5 g/L ultra-filtered yeast powder) containing spectinomycin 50 µg/ml, kanamycin 50 µg/ml, and 0.5mM IPTG. These growth conditions kept both SpyCas9 and the gene of interest in uninduced states. The next morning, overnight cultures were diluted 1:50 into LB broth containing spectinomycin 50 µg/ml, kanamycin 50 µg/ml, 0.5mM IPTG, and doxycycline 100 ng/ml to induce the gene of interest. Cultures were grown at 37C on a roller drum to mid-log phase (for approximately 1.5 hours to OD600 of 0.3-0.6). Once cells reached mid-log phase, they were diluted to OD600 value of 0.01 into two media types: (a) LB containing spectinomycin 50 µg/ml, 0.5mM IPTG, and doxycycline 100 ng/ml, and (b) LB containing spectinomycin 50 µg/ml, 0.5mM IPTG, doxycycline 100 ng/ml, and 0.2% (L) arabinose. These media induced either the gene of interest alone, or both the gene of interest and SpyCas9, respectively. Each sample was grown in triplicate in a 96 well plate in a BioTek Cytation 3 plate reader. After 6 hours of growth at 37^{o}C with shaking at 220 rpm, each sample was diluted ten-fold and plated on two types of media: (a) LB spectinomycin 50 µg/ml + 0.5mM IPTG or (b) LB spectinomycin 50 µg/ml, kanamycin 50 µg/ml, 0.5mM IPTG. Plates were incubated at 37C overnight. Then, colonies were counted to determine the fraction of colony forming units (cfus) that maintained kanamycin resistance (and thus the target vector). All figures depicting these data show the log-transformed proportion of Kan^{R}/total cfu, both with and without SpyCas9 induction. The growth curves in Supplemental Figure 1 match the experiment depicted Figure 1C for the uninduced SpyCas9 samples. For the uninduced *orf_1* sample, doxycycline was omitted from media throughout the experiment. Growth rates quoted in text were calculated using the slope of the OD600 growth curves during log phase, following a natural log transformation.

Impact of AcrIIA22 on GFP expression

We swapped *spyCas9* for *egfp* in our CloDF13-based plasmid and co-expressed AcrIIA22 to determine if AcrIIA22 impacted expression from this construct. We reasoned that if AcrIIA22 influenced CloDF13's copy number or the transcription of *spyCas9* it would also impact GFP levels in this construct (pCloDF13_GFP; Supplemental Table 4). To perform this experiment, we co-transformed pCloDF13_GFP and pZE21_tetR encoding *acrIIA22* into *E. coli* Turbo. Single colonies were picked into 4mL of LB containing spectinomycin 50 µg/ml ('spec50') and kanamycin 50 µg/ml ('kan50') and 0.5mM IPTG and grown overnight at 37^{o}C shaking at 220rpm. The next morning the overnight culture was diluted 1:50 into both LB spec50 Kan50 + 0.5mM IPTG with and without doxycycline (to induce *acrIIA22*) and grown at 37^{o}C for about 1.5 hours to mid-log phase (OD600 0.2-0.6). The OD600 was measured, and all samples were diluted to OD600 of 0.01 in two media types: (a) LB spec50 + kan50 + 0.5mM IPTG + 0.2% arabinose (inducing *gfp* only) or (b) LB spec50 + kan50 + 0.5mM IPTG + 0.2% arabinose + 100ng/ml doxycycline (inducing

gfp and *acrIIA22*). A volume of 200 µl of each sample was then transferred to a 96-well plate in triplicate and we measured GFP fluorescence every 15 minutes for 24 hours (GFP was excited using 485 nm light and emission detected via absorbance at 528 nm). In parallel, we included control samples that lacked the kanamycin-marked plasmid and varied whether doxycycline was added or not (at 100 ng/ml). In these control samples, we noticed that doxycycline slightly diminished GFP expression (sub-toxic levels of the antibiotic may still depress translation). Thus, we normalized GFP fluoresced measurements in our experiment with AcrIIA22 to account for this effect in all +doxycycline samples. These fluorescence measurements are depicted in Supplemental Figure 2B.

Western blots to AcrIIA22's impact on SpyCas9 expression

Overnight cultures of *E. coli* Turbo that expressed pSpyCa9_crNT and pZE21_tetR encoding a gene of interest (Supplemental Tables 4, 5) were grown in LB spec50 + kan50 + 0.5mM IPTG. The next morning, these cultures were diluted 1:100 in 4ml of either (a) LB spec50 + kan50 + 0.5mM IPTG or (b) LB spec50 + kan50 + 0.5mM IPTG + 100 ng/ml doxycycline (to induce the gene of interest). We included samples that expressed either *acrIIA22* or *gfp* as a gene of interest. In all SpyCas9 constructs, we used a crRNA that did not target our plasmid backbone (pSpyCa9_crNT) to ensure that *acrIIA22* expression remained high and its potential impact on SpyCas9 expression levels would be most evident. All samples were grown for two hours at 37ºC to reach mid-log phase (OD600 0.3 to 0.5) and transferred into media that contained 0.2% arabinose to induce SpyCas9. At transfer, volumes were normalized by OD600 value to ensure an equal number of cells were used (diluted to a final OD600 of 0.05 in the arabinose-containing medium). This second medium did or did not contain 100 ng/ml doxycycline to control expression of *acrIIA22* or *gfp,* as with the initial media. Throughout this experiment, we included a control strain that lacked pZE21_tetR and thus only expressed SpyCas9. Kanamycin and doxycycline were omitted from its growth media. For this control strain, we also toggled the addition of arabinose in the second growth medium to ensure positive and negative controls for SpyCas9 were included in our experiment. After three hours and six hours of SpyCas9 induction, OD600 readings were again taken and these values used to harvest an equal number of cells per sample (at three hours, OD600 values were between 0.76 and 0.93 and 0.75ml to 0.9ml volumes harvested; at six hours 0.4ml was uniformly harvested as all absorbance readings were approximately 1.6).

All samples were centrifuged at 4100g to pellet cells, resuspended in 100 µl of denaturing lysis buffer (12.5 mM Tris-HCl, pH 6.8; 4% SDS), and passed through a 25 gauge needle several times to disrupt the lysate. Samples were then boiled at 100ºC for 10 minutes, spun at 13,000 rpm at 4ºC for 15 minutes and the supernatants removed and frozen at -20ºC. The next day, 12 µl of lysate was mixed with 4 µl of 4x sample buffer (200 mM Tris-HCl, 8% SDS, 40% glycerol, 200 mM DTT, and 0.05% bromophenol blue) and boiled at 100ºC for 10 minutes. Then, 10 µl sample was loaded onto a BioRad Mini-Protean "any KD Stain Free TGX" gel (cat. #4569035) and run at 150V for 62 minutes. To verify that equivalent amounts of each sample were run, gels were visualized on a BioRad chemidoc for total protein content. Protein was then transferred to a 0.2 µM nitrocellulose membrane using the Bio-Rad Trans-Blot Turbo system (25 V, 1.3 A for 10 min). We then washed membranes in PBS/0.1% Triton-X before incubating them with a mixture of the following two primary antibodies, diluted in in Licor Odyssey Blocking Solution (cat. #927–40000): (i) monoclonal anti-SpyCas9, Diagenode cat. #C15200229-50, diluted 1:5,000; (ii) polyclonal anti-GAPDH, GeneTex cat. # GTX100118, diluted 1:5,000. The GAPDH antibody

served as a second check to ensure equal protein levels were run. Membranes were left shaking overnight at 4°C, protected from light. Then, membranes were washed four times in PBS/0.1% Triton-X (ten-minute washes) before they were incubated for 30 minutes at room temperature with a mixture of secondary antibodies conjugated to infrared dyes. Both antibodies were diluted 1:15,000 in LiCor Odyssey Blocking Solution. To detect SpyCas9, the following secondary antibody was used: IR800 donkey, anti-mouse IgG, Licor cat# 926–32212. To detect GAPDH, IR680 goat, anti-rabbit IgG, Licor cat# 926-68071 was used. Blots were imaged on a Licor Odyssey CLx after three additional washes.

Phage plaquing assay

Overnight cultures of *E. coli* Turbo that expressed pSpyCa9_crMu and pZE21_tetR encoding a gene of interest (Supplemental Tables 4, 5) were grown at 37°C in LB spec50 + kan50 + 0.5 mM IPTG. Genes of interest were either *acrIIA4*, *gfp*, or *acrIIA22*. The pSpyCas9 construct targeted phage Mu and was previously demonstrated to confer strong anti-phage immunity in this system[12]. A control strain expressing pZE21-tetR-*gfp* and SpyCas9_crNT (which encoded a CRISPR RNA that does not target phage Mu) was grown similarly. The next morning, all cultures were diluted 50-fold into LB spec50 + kan50 + 0.5 mM IPTG + 5 mM MgCl2 and grown at 37°C for three hours. Then, doxycycline was added to a final concentration of 100 ng/ml to induce the gene of interest. Two hours later, SpyCas9 was induced by adding a final concentration of 0.2% w/v arabinose. Two hours after that, cultures were used in soft-agar overlays on one of two media types, discordant for arabinose, to either maintain SpyCas9 expression or let it fade as arabinose was diluted in top agar and consumed by the host bacteria (per Supplemental Figure S2). Top and bottom agar media were made with LB spec50 + kan50 + 0.5 mM IPTG + 5 mM MgCl2. In cases where SpyCas9 expression was maintained, arabinose was also added at a final concentration of 0.02% to both agar types. Top agar was made using 0.5% Difco agar and bottom agar used a 1% agar concentration. For the plaquing assay, 100 µl of bacterial culture was mixed with 3 ml of top agar, allowed to solidify, and ten-fold serial dilutions of phage Mu spotted on top using 2.5 µl droplets. After the droplets dried, plates were overturned and incubated at 37°C overnight before plaques were imaged the subsequent day.

Identification of AcrIIA22 homologs and hypervariable genomic islands

We searched for AcrIIA22 homologs in three databases: NCBI nr, IMG/VR, and a set of assembled contigs from 9,428 diverse human microbiome samples[18]. Accession numbers for the NCBI homologs are indicated on the phylogenetic tree in Figure 3A. They were retrieved via five rounds of an iterative PSI-BLAST search against NCBI nr performed on October 2nd, 2017. In each round of searching, at least 90% of the query protein (the original AcrIIA22 hit) was covered, 88% of the subject protein was covered, and the minimum amino acid identity of an alignment was 23% (minimum 47% positive residues; e-value ≤ 0.001). Only one unique AcrIIA22 homolog was identified in IMG/VR (from several different phage genomes) via a blastp search against the July, 2018 IMG/VR proteins database (using default parameters). It is identical to the sequence of AcrIIA22b (Figure 3A).

Most unique AcrIIA22 homologs were identified in the assembly data of over 9,400 human microbiomes performed by Pasolli and colleagues[18]. These data are grouped into multiple datasets: (i) the raw assembly data, and (ii) a set of unique species genome bins (SGBs), which was generated by first assigning species-level phylogenetic labels to each assembly and then selecting one representative genome assembly per species. We identified AcrIIA22 homologs

using several queries against both databases. First, we performed a tblastn search against the SGB database using the AcrIIA22 sequence as a query, retrieving 141 hits from 137 contigs. A manual inspection of the genome neighborhoods for these hits revealed that most homologs originated from a short, hypervariable genomic island but that some homologs were encoded by prophages. No phage-finding software was used to identify prophages; they were apparent from a manual inspection of the gene annotations that neighbored *acrIIA22* homologs (see the section entitled "Annotation and phylogenetic assignment of metagenomic assemblies" for details).

To find additional examples of AcrIIA22 homologs and of these genomic islands, we then queried the full raw assembly dataset. To do so without biasing for *acrIIA22*-encoding sequences, we used the *purF* gene that flanked *acrIIA22*-encoding genomic islands as our initial query sequence (specifically, we used the *purF* gene from contig number 1 in Supplemental Table 3; its sequence is also in Supplemental Table 5). To consider only the recent evolutionary history of this locus, we required all hits have ≥98% nucleotide identity and required all hits to be larger than 15 kilobases in length to ensure sufficient syntenic information. From these contigs, we further filtered for those that had ≥98% nucleotide identity to *radC*, the gene which flanked the other end of *acrIIA22*-encoding genomic islands (again, we used the variant from contig number 1 in Supplemental Table 3; its sequence is also in Supplemental Table 5). In total, this search yielded 258 contig sequences; nucleotide sequences and annotations for these contigs are provided in Supplementary Dataset 5. We then searched for *acrIIA22* homologs in these sequences using tblastn, again observing them in genomic islands and prophage genomes (these prophages were assembled as part of the 258 contigs). In total, this search revealed 320 *acrIIA22* homologs from 258 contigs. The 258 genomic islands from these sequences were retrieved manually by extracting all nucleotides between the *purF* and *radC* genes. These extracted sequences were then clustered at 100% nucleotide identity with the sequence analysis software geneious to identify 128 unique genomic islands.

Combined, our two searches yielded 461 AcrIIA22 sequences from these metagenomic databases that spanned 410 contig sequences. The 461 AcrIIA22 homologs broke down into 410 that clustered with the genomic island-like sequences (we specifically searched for genomic islands) and 51 that clustered with prophage-like homologs (we never directly searched for prophages). We then combined these 461 AcrIIA22 sequences with those from NCBI and IMG/VR and clustered the group on 100% amino acid identity to reveal 30 unique proteins. To achieve this, we used the software cd-hit[43] with the following parameters: -d 0 -g 1 -aS 1.0 -c 1.0. These 30 sequences were numbered to match their parent contig (as indicated in Supplemental Table 3) and used to create the phylogenetic tree depicted in Figure 3A. For AcrIIA22 homologs found outside NCBI, the nucleotide sequences and annotations their parent contigs can be found in Supplementary Datasets 1 and 2. This information can be retrieved for NCBI sequences via their accession numbers (which are shown in Figure 3A). The NCBI gene sequences also used in functional assays (Figure 3B) have been reprinted in Supplemental Table 5, for convenience.

Annotation and phylogenetic assignment of metagenomic assemblies

Contig sequences from IMG/VR, the Pasolli metagenomic assemblies, and some NCBI entries lacked annotations, making it difficult to make inferences about *acrIIA22's* genomic neighborhood. To facilitate these insights, we annotated all contigs as follows. We used the gene-finder MetaGeneMark[44] to predict open reading frames (ORFs) using default parameters. We then used their amino acid sequences in a profile HMM search with HMMER3[45] against TIGRFAM[46] and Pfam[47] profile HMM databases. The highest scoring profile was used to annotate

each ORF. We annotated these contigs to facilitate genomic neighborhood analyses for *acrIIA22* and not to provide highly accurate functional predictions of their genes. Thus, we erred on the side of promiscuously assigning gene function and our annotations should be treated with the appropriate caution. From these annotated contigs, we immediately observed several examples of *acrIIA22*-encoding prophages (we noticed 35-40 kilobase insertions within some contigs that contained mostly co-linear genes with key phage functions annotated). As a simple means to sample this phage diversity, we manually extracted nine examples of these prophage sequences (their raw sequences and annotated genomes can be found in Supplementary Datasets 3 and 4). Annotations were imported to in the sequence analysis suite Geneious Prime 2020 v1.1 for manual inspection of genome neighborhoods.

We used the genome taxonomy database (GTDB) convention for all sequences discussed in this manuscript[48]. In part, this was because all *acrIIA22* genomes are found in *Clostridial* genomes, which are notoriously polyphyletic in NCBI taxonomies (for instance, the NCBI genus appears in GTDB genera and 29 GTDB families)[49]. All SGBs that we retrieved from the Pasolli assemblies were assigned taxonomy as part of that work and were called Clostridium sp. CAG-217. Similarly, NCBI assemblies that encoded the most closely *acrIIA22* homologs to our original hit were assigned to the GTDB genus CAG-217[48,49]. The raw assembly data from the Pasolli database was not assigned a taxonomic label but was nearly identical in nucleotide composition to the CAG-217 contigs (Figure 2, Supplementary Datasets 1 and 2). Therefore, we also refer to these sequences as originating in CAG-217 genomes but take care to indicate which sequences have been assigned a rigorous taxonomy and which ones for which taxonomy has been inferred in this fashion (Supplemental Table 3).

Comparing genes in genomic islands to phage genomes

We first examined the annotated genes within each of the 128 unique genomic islands. Manual inspection revealed 54 unique gene arrangements (which differed in gene content and orientation). We then selected one representative from each arrangement and extracted all amino acid sequences from each encoded gene (n=506). Next, we collapsed these 506 proteins into orthologous groups by clustering at 65% amino acid using cd-hit with the following parameters: -d 0 -g 1 -aS 0.95 -c 0.65. These cluster counts were used to generate the histogram depicted in Figure 2C. To determine which protein families may also be phage encoded, the longest representative from each cluster with at least two sequences was queried against the database of nine CAG-217 phages described in the section entitled “Annotation and phylogenetic assignment of metagenomic assemblies”. We used tblastn with default parameters to perform this search, which revealed that some proteins in the CAG-217 genomic islands have homologs in prophage genomes that are out-of-frame with respect to the MetaGeneMark annotations depicted in Figure 2A.

Phylogenetic tree of AcrIIA22 homologs

The 30 unique AcrIIA22 homologs we retrieved were used to create the phylogeny depicted in Figure 3A. These sequences were aligned using the sequence alignment tool in the sequence analysis suite Geneious Prime 2020 v1.1. This alignment is provided as Supplementary Dataset 6. From this alignment, the phylogenetic tree in Figure 3A was generated using PhyML with the LG substitution model and 100 bootstraps. Coloration and tip annotations were then added in Adobe Illustrator.

Identification of CRISPR-Cas systems and Acrs in CAG-217 assemblies

To determine the type and distribution of CRISPR-Cas systems and Acrs in CAG-217 genomes, we downloaded all assembly data for the 779 SGBs assigned to CAG-217 in Pasolli *et. al*[18] (bin 4303). We then predicted CRISPR-Cas systems for all 779 assemblies in bulk using the command line version of the CRISPR-Cas prediction suite, cctyper[50]. Specifically, we used version 1.2.1 of cctyper with the following options: --prodigal meta --keep_tmp. To identify type II-A Acrs, we first downloaded representative sequences for each of the 21 experimentally confirmed type II-A Acrs from the unified resource for tracking anti-CRISPRs[51]. We then used tblastn to query these proteins against the 779 CAG-217 genome bins and considered any hit with e-value better than 0.001 (which included all hits with >30% identity across 50% of the query). To check if these Acrs were present in *acrIIA22*-encoding phages, we performed an identical tblastn search, but this time used the set of nine *acrIIA22*-encoding prophages as a database.

Recombinant protein overexpression and purification

The AcrIIA22 protein and its mutants were codon optimized for *E. coli* (Genscript or SynBio Technologies) and the gene construct was cloned into the pET15HE[12] plasmid to contain an N-terminal, thrombin-cleavable 6XHistidine tag. Constructs were transformed and overexpressed in BL21 (DE3) RIL *E. coli* cells. A 10 mL overnight culture (grown in LB + 100 µg/mL ampicillin) was diluted 100-fold into the same media and grown at 37°C with shaking to an OD600 of 0.8, followed by induction with 0.5 mM IPTG. The culture was shaken for an additional 3 hours at 37°C. Cells were harvested by centrifugation and the pellet stored at -20°C until purification.

Cell pellets were resuspended in 25 mM Tris, pH 7.5, 300 mM NaCl, 20 mM imidazole (Lysis Buffer) and lysed by sonication on ice. The lysate was centrifuged in an SS34 rotor at 18,000 rpm for 25 minutes, followed by filtering through a 5 µm syringe filter (Millipore #SLSV025LS). The clarified lysate was bound using the batch method to Ni-NTA agarose resin (Qiagen) at 4°C for 1 hour. The resin was transferred to a gravity column (Biorad), washed with >50 column volumes of Lysis Buffer and eluted with 25 mM Tris, pH 7.5, 300 mM NaCl, 200 mM imidazole. The protein was diluted with 2 column volumes of 25 mM Tris, pH 7.5 and purified on a HiTrapQ column (GE Healthcare) using a 20 mL gradient from 150 mM to 1 M NaCl in 25 mM Tris, pH 7.5. Peak fractions were pooled, concentrated and buffer exchanged into 200 mM NaCl, 25 mM Tris, pH 7.5 using an Amicon Ultra centrifugal filter with a 3,000 molecular weight cutoff (Millipore, UFC900324), then cleaved in an overnight 4°C incubation with biotinylated thrombin (EMD Millipore). Streptavidin agarose slurry (Novagen) was incubated with cleaved protein at 4°C for 30 minutes to remove thrombin. The sample was then passed through a 0.22 µm centrifugal filter and loaded onto a HiLoad 16/60 Superdex 200 prep grade size exclusion column (Millipore Sigma) equilibrated in 25 mM Tris, pH 7.5, 200 mM NaCl. The peak fractions were confirmed for purity by SDS-PAGE. Figure 4C depicts size exclusion chromatography data generated for thrombin-cleaved AcrIIA22 variants generated using a Superdex75 16/60 (GE HealthCare) column with 25 mM Tris, pH 7.5, 200mM NaCl. Recombinant AcrIIA4 was purified similarly to other Acr proteins as previously described[12], but with the following deviations. First, the 6XHistidine-tagged AcrIIA4 gene was cloned into pET15B rather than pET15HE, which differs by only by a few bases just upstream of the N-terminal thrombin tag. IPTG was used at 0.2 mM and cells were harvested after 18 hours of induction at 18°C. Thrombin cleavage also occurred at 18°C. This untagged version was used to help generate Supplemental Figure 5. Peak fractions for all proteins were pooled, concentrated, flash frozen as single-use aliquots in liquid nitrogen, and stored at −80°C.

SpyCas9 was expressed in E. coli from plasmid pMJ806 (addgene #39312) to contain a TEV-cleavable N-terminal 6XHis-MBP tag and was purified as described previously[12]. Briefly, sequential steps of purification consisted of Ni-NTA affinity chromatography, TEV cleavage, Heparin HiTrap chromatography and SEC. The protein was stored in a buffer consisting of 200 mM NaCl, 25 mM Tris (pH 7.5), 5% glycerol, and 2 mM DTT.

To perform *in vitro* pulldown experiments, we purified AcrIIA22 and AcrIIA4 proteins with a C-terminal twin-strep tag. To achieve this, the Acrs were subcloned into pET15B which was previously engineered to contain a thrombin-cleavable C-terminal twin-strep tag. The protein was expressed as described above and purified according to the manufacturer's guidelines (IBA Inc.). Briefly, cell lysates were resuspended in Buffer W (150 mM NaCl, 100 mM Tris, pH 8.0, 1 mM EDTA) and lysed by sonication. Clarified lysates were then passed over Streptactin-Sepharose resin using a gravity filtration column. The flow through was passed over the resin an additional time. The column was washed with a minimum of 20 column volumes of buffer W, followed by elution in buffer E (150 mM NaCl, 100 mM Tris, pH 8.0 mM, 1 EDTA, 2.5 mM desthiobiotin). The eluted protein was purified over a HiTrap Q column, followed by SEC in 200 mM NaCl, 25 mM Tris, 7.5.

X-ray crystallography and structural analyses

An AcrIIA22 crystal was grown using 14mg/mL protein via the hanging drop method using 200mM ammonium nitrate, 40% (+/-)-2-methyl-2,4-pentanediol (MPD, Hampton Research), 10mM MgCl2 as a mother liquor. Diffraction data was collected at the Argonne National Laboratory Structural Biology Center synchrotron facility (Beamline 19BM). Data was processed with HKL2000 in space group P4332, then built and refined using COOT[52] and PHENIX[53]. The completed 2.80Å structure was submitted to the Protein Data Bank with PDB Code 7JTA. We submitted this finished coordinate file to the PDBe PISA server (Protein Data Bank Europe, Protein Interfaces, Surfaces and Assemblies; http://pdbe.org/pisa/) which uses free energy and interface contacts to calculate likely multimeric assemblies[25]. The server calculated tetrameric, dimeric and monomeric structures to be thermodynamically stable in solution. The tetrameric assembly matches the molecular weight expected from the size exclusion column elution peak and is the most likely quaternary structure as calculated by the PISA server. The tetramer gains -41.8 kcal/mol free energy by solvation when formed and requires an external driving force of 3.1 kcal/mol to disassemble it according to PISA ΔG calculations.

sgRNA generation

The single-guide RNA (sgRNA) for use in *in-vitro* experiments was generated as described previously[12]. It was transcribed from a double-stranded DNA (dsDNA) template by T7 RNA polymerase using Megashortscript Kit (Thermo Fisher #AM1354). We made the dsDNA template via one round of thermal cycling (98°C for 90 s, 55°C for 15 s, 72°C for 60 s) in 50 µl reactions. We used the Phusion PCR polymerase mix (NEB) containing 25 pmol each of the following two oligo sequences (the sequence that binds the protospacer on our pIDTsmart target vector is underlined):

(i) GAAATTAATACGACTCACTATAGG<u>TAATGAAATAAGATCACTAC</u>GTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGGCTAGTCCG

(ii) AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGC.

The dsDNA templates were then purified using an Oligo Clean and Concentrator Kit (ZymoResearch) before quantification via the Nanodrop. Reactions were then treated with DNAse, extracted via phenol-chloroform addition followed by chloroform, ethanol precipitated, resuspended in RNase free water, and frozen at −20°C. RNA was quantified by Nanodrop and analyzed for quality on 15% acrylamide/TBE/UREA gels.

Pulldown assay using strep-tagged AcrIIA22 and AcrIIA4

The same buffer was used for pulldowns and to dilute proteins, consisting of 200 mM NaCl, 25 mM Tris (pH 7.5). As a precursor to these assays, 130 pmol SpyCas9 and sgRNA were incubated together at room temperature for 15 minutes where indicated. SpyCas9, with or without pre-complexed sgRNA, was then incubated with 230 pmol AcrIIA4 or 320 pmol AcrIIA22 for 25 minutes at room temperature. Subsequently, 50 µl of a 10% slurry of Streptactin Resin (IBA biosciences #2-1201-002), pre-equilibrated in binding buffer, was added to the binding reactions and incubated at 4°C on a nutator for 45 minutes. Thereafter all incubations and washes were carried out at 4°C or on ice. Four total washes of this resin were performed, which included one tube transfer. Washes proceeded via centrifugation at 2000 rpm for one minute, aspiration of the supernatant with a 25-gauge needle, and resuspension of the beads in 100 µl binding buffer. Strep-tagged proteins were eluted via suspension in 40 µl of 1x BXT buffer (100 mM Tris-Cl, 150 mM NaCl, 1 mM EDTA, 50 mM Biotin, pH 8.0) and incubated for 15 min at room temperature. After centrifugation, 30 µl of supernatant was removed and mixed with 4X reducing sample buffer (Thermo Fisher). Proteins then separated by SDS PAGE on BOLT 4–12% gels in MES buffer (Invitrogen) and visualized by Coomassie staining.

SpyCas9 linear DNA cleavage assay

All SpyCas9 cleavage reactions using linear DNA were performed in the following cleavage buffer: 20mM Tris HCl (pH7.5), 5% glycerol, 100mM KCl, 5mM MgCl2, 1mM DTT. In preparation for these reactions, all proteins were diluted in 30 mM NaCl / 25 mM Tris, pH 7.4 / 2.7mM KCl, whereas all DNA and sgRNA reagents were diluted in nuclease-free water. Where indicated, SpyCas9 (0.36 µM) was incubated with sgRNA (0.36 µM) for 10 minutes at room temperature. Before use, sgRNA was melted at 95°C for five minutes and then slowly cooled at 0.1 °C/s to promote proper folding. SpyCas9 (either pre-complexed with sgRNA or not, as indicated in Supplemental Figure 6) was then incubated for 10 minutes at room temperature with AcrIIA4 (2.9 µM) or AcrIIA22 at the following concentrations: [23.2, 11.6, 5.8, and 2.9 µM]. As substrate, the plasmid pIDTsmart was linearized by restriction digest and used at a final concentration of 3.6 nM. The reaction was initiated by the addition of this DNA substrate in isolation or in combination with sgRNA (0.36 µM) as indicated in Supplemental Figure 6. Reactions were immediately moved to a 37^{o}C incubator and the reaction stopped after fifteen minutes via the addition of 0.2% SDS/100 mM EDTA and incubating at 75^{o}C for five minutes. Samples were then run on a 1.5% TAE agarose gel at 120V for 40 minutes. Densitometry was used to calculate the proportion of DNA cleaved by SpyCas9 via band intensities quantified using the BioRad ImageLab software v5.0.

In vivo assay to assess impact of AcrIIA22 on plasmid topology

In all experiments, cultures were first grown overnight at 37°C with shaking at 220 rpm in LB with 0.5mM IPTG, spectinomycin (at 50 µg/mL), and kanamycin (at 50 µg/mL). Then, these overnight cultures were diluted 1:50 into LB with 0.5mM IPTG, spectinomycin (at 50 µg/mL), and, where indicated, doxycycline (at 100 ng/mL, to induce *acr*s). Cultures were grown at 37°C with shaking at 220 rpm and, if indicated, 0.2% (L)-arabinose was added after two hours of growth to induce *spyCas9* expression. The next morning, cultures were centrifuged at 4100*g* and plasmids purified using a miniprep kit (Qiagen). The concentration of dsDNA in each miniprep was measured using the Qubit-4 fluorometer and the associated dsDNA high sensitivity assay kit (Invitrogen). For each sample with a SpyCas9-expressing plasmid, 150ng of DNA was digested with the restriction enzyme HincII (NEB) per manufacturer's recommendations, except that digests were incubated overnight before being stopped by heating at 65°C for 20 minutes. This restriction enzyme will cut once, only in the SpyCas9 plasmid, to linearize it. This allowed us to visualize the SpyCas9 plasmid as a single band, which served two purposes: (i) it allowed us to more easily identify bands from *acrIIA22*-encoding plasmids (which had not been digested), and (ii) it served as an internal control for plasmid DNA that is unaffected by SpyCas9 targeting or AcrIIA22 expression (Supplemental Figure 2). Following restriction digest, 30ng of sample was analyzed via gel electrophoresis using a 1% TAE-agarose gel run at 120V for between 45 and 60 minutes. In samples that lacked a SpyCas9-expressing plasmid, 30ng of purified plasmid was directly analyzed by gel electrophoresis, as described previously.

In vitro AcrIIA22 plasmid nicking assay

Except for the divalent cation experiment, all reactions were performed using NEB buffer 3.1 (100 mM NaCl, 50 mM Tris-HCl, pH 7.9, 10 mM MgCl2, 100 µg/mL BSA). To determine cation preference, the same reaction buffer was re-created, but MgCl2 was omitted. All proteins were diluted in 130 mM NaCl, 25 mM Tris, pH 7.4, 2.7 mM KCl. DNA was diluted in nuclease-free water. In the cation preference experiment, 60 µM AcrIIA22 and 6 nM of purified pIDTsmart plasmid DNA were used. All other reactions were set up with the AcrIIA22 final concentrations indicated in Figure 5 and Supplemental Figure 7. In the cation preference experiment, reactions were started by adding 10 mM of the indicated cation. All other reactions were initiated via the addition of 2 nM pIDTsmart plasmid DNA. In all cases, reactions were immediately transferred to a 37°C incubator. At 0.5, 1, 2, 4, 6, or 20-hour timepoints, a subset of the reaction was removed and run on a 1% TAE agarose gel at 120V for 40 minutes. For the cation preference experiment, only the 2-hour timepoint was considered and the reaction was stopped via the addition of NEB loading buffer and 100 mM EDTA. In this case, DNA was visualized on a 1% TBE gel run for 60 minutes at 110V. Densitometry was used to calculate the proportion of DNA in each topological form via band intensities quantified using the BioRad ImageLab software v5.0.

SpyCas9 cleavage kinetics assay

Except where indicated in Supplemental Figure 9B, all cleavage reactions were performed in the following cleavage buffer: 20mM Tris HCl (pH7.5), 5% glycerol, 100mM KCl, 5mM MgCl2, 1mM DTT. In preparation for these reactions, all proteins were diluted in 30 mM NaCl / 25 mM Tris, pH 7.4 / 2.7mM KCl, whereas all DNA and sgRNA reagents were diluted in nuclease-free

water. NEB Buffer 3.1 (100 mM NaCl, 50 mM Tris-HCl, pH 7.9, 10 mM MgCl2, 100 µg/mL BSA) was used as a reaction buffer in Supplemental Figure 9B.

In preparation for these reactions, purified pIDTsmart plasmid was pre-treated with either AcrIIA22, the nickase Nb.Bss.SI (NEB), or no enzyme. For the AcrIIA22 pre-treatment, 3.1 µg of plasmid was incubated with 230 µM AcrIIA22 and the plasmid nicked as described previously. Plasmid nicking with Nb.Bss.SI proceeded via manufacturer's recommendations (NEB). Both reactions were incubated at 37 ºC for 2 hours. To isolate the nicked plasmid, samples were then run on a 1.5% agarose gel for 2 hours and the open-circle form of the plasmid was excised and purified using the Zymo Research Gel DNA Recovery Kit. Untreated plasmid was also purified via gel extraction. Plasmid yield was quantified using a Nanodrop.

To determine SpyCas9's substrate preference, we incubated each pre-treated plasmid substrate with SpyCas9 and looked for the appearance of a linearized plasmid as indication of SpyCas9 digestion. In all cases, SpyCas9 was used at a final concentration of 31.2 nM. To begin the reaction, DNA substrate and sgRNA were added simultaneously to the reaction mix and the samples moved immediately from ice to 37 ºC and incubated for either 1 or 5 minutes. We noticed that the digestion reaction proceeded too quickly with NEB Buffer 3.1 to detect SpyCas9's substrate preference (i.e., the substrates were all rapidly linearized Supplemental Figure 9B). The cleavage buffer used in most reactions (detailed atop this section) was chosen because it slowed digestion kinetics so that we could detect SpyCas9's substrate preference. Before addition to the reaction, sgRNA was melted at 95°C for five minutes and then slowly cooled at 0.1 °C/s to promote proper folding. At each timepoint, 5 µl of the reaction was removed and the reaction was stopped using 0.2% SDS/100 mM EDTA, then incubating at 75ºC for 5 minutes. Samples were run on a 1.5% TAE gel at 120V for 40 minutes.

Acknowledgements

We thank Kaylee Dillard, Ilya Finkelstein, and Tera Levin for comments on the manuscript. Use of the Advanced Photon Source, an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the U.S. DOE under Contract No. DE-AC02-06CH11357. This work was supported by a Helen Hay Whitney Foundation postdoctoral fellowship awarded to KJF, a Seattle University summer faculty fellowship to BKK, NIH grant R01GM105691 and discretionary funding from the Fred Hutchinson Cancer Research Center to BLS, and grants from the G. Harold and Leila Y. Mathers Foundation and the Howard Hughes Medical Institute to HSM. The funders played no role in study design, data collection and interpretation, or the decision to publish this study. HSM is an Investigator of the Howard Hughes Medical Institute.

Competing Interests

All authors declare no significant competing financial, professional, or personal interests that might have influenced the performance or presentation of the work described in this manuscript.

References

1 Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712, doi:10.1126/science.1138140 (2007).

2 Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843-1845, doi:10.1126/science.1165771 (2008).

3 Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J Bacteriol* **190**, 1390-1400, doi:10.1128/JB.01412-07 (2008).

4 Mendoza, S. D. *et al.* A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases. *Nature* **577**, 244-248, doi:10.1038/s41586-019-1786-y (2020).

5 Malone, L. M. *et al.* A jumbo phage that forms a nucleus-like structure evades CRISPR-Cas DNA targeting but is vulnerable to type III RNA-based immunity. *Nat Microbiol* **5**, 48-55, doi:10.1038/s41564-019-0612-5 (2020).

6 Bryson, A. L. *et al.* Covalent Modification of Bacteriophage T4 DNA Inhibits CRISPR-Cas9. *mBio* **6**, e00648, doi:10.1128/mBio.00648-15 (2015).

7 Stanley, S. Y. & Maxwell, K. L. Phage-Encoded Anti-CRISPR Defenses. *Annu Rev Genet* **52**, 445-464, doi:10.1146/annurev-genet-120417-031321 (2018).

8 Trasanidou, D. *et al.* Keeping crispr in check: diverse mechanisms of phage-encoded anti-crisprs. *FEMS Microbiol Lett*, doi:10.1093/femsle/fnz098 (2019).

9 Davidson, A. R. *et al.* Anti-CRISPRs: Protein Inhibitors of CRISPR-Cas Systems. *Annu Rev Biochem* **89**, 309-332, doi:10.1146/annurev-biochem-011420-111224 (2020).

10 Wiegand, T., Karambelkar, S., Bondy-Denomy, J. & Wiedenheft, B. Structures and Strategies of Anti-CRISPR-Mediated Immune Suppression. *Annu Rev Microbiol* **74**, 21-37, doi:10.1146/annurev-micro-020518-120107 (2020).

11 Hatfull, G. F. Dark Matter of the Biosphere: the Amazing World of Bacteriophage Diversity. *J Virol* **89**, 8107-8110, doi:10.1128/JVI.01340-15 (2015).

12 Forsberg, K. J. *et al.* Functional metagenomics-guided discovery of potent Cas9 inhibitors in the human microbiome. *eLife* **8**, e46540, doi:10.7554/eLife.46540 (2019).

13 Szczelkun, M. D. *et al.* Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc Natl Acad Sci U S A* **111**, 9798-9803, doi:10.1073/pnas.1402597111 (2014).

14 Farasat, I. & Salis, H. M. A Biophysical Model of CRISPR/Cas9 Activity for Rational Design of Genome Editing and Gene Regulation. *PLoS Comput Biol* **12**, e1004724, doi:10.1371/journal.pcbi.1004724 (2016).

15 Ivanov, I. E. *et al.* Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling. *Proc Natl Acad Sci U S A* **117**, 5853-5860, doi:10.1073/pnas.1913445117 (2020).

16 Tsui, T. K. M., Hand, T. H., Duboy, E. C. & Li, H. The Impact of DNA Topology and Guide Length on Target Selection by a Cytosine-Specific Cas9. *ACS Synth Biol* **6**, 1103-1113, doi:10.1021/acssynbio.7b00050 (2017).

17 Paez-Espino, D. *et al.* IMG/VR v.2.0: an integrated data management and analysis system for cultivated and environmental viral genomes. *Nucleic acids research* **47**, D678-D686, doi:10.1093/nar/gky1127 (2019).

18 Pasolli, E. *et al.* Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* **176**, 649-662 e620, doi:10.1016/j.cell.2019.01.001 (2019).

19 Dobrindt, U., Hochhut, B., Hentschel, U. & Hacker, J. Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* **2**, 414-424, doi:10.1038/nrmicro884 (2004).

20 Juhas, M. *et al.* Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev* **33**, 376-393, doi:10.1111/j.1574-6976.2008.00136.x (2009).

21 Makarova, K. S. *et al.* Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol* **18**, 67-83, doi:10.1038/s41579-019-0299-x (2020).

22 Mahendra, C. *et al.* Broad-spectrum anti-CRISPR proteins facilitate horizontal gene transfer. *Nat Microbiol* **5**, 620-629, doi:10.1038/s41564-020-0692-2 (2020).

23 Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429-432, doi:10.1038/nature11723 (2013).

24 Janowski, R. & Niessing, D. The large family of PC4-like domains - similar folds and functions throughout all kingdoms of life. *RNA Biol* **17**, 1228-1238, doi:10.1080/15476286.2020.1761639 (2020).

25 Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797, doi:10.1016/j.jmb.2007.05.022 (2007).

26 Steigemann, B., Schulz, A. & Werten, S. Bacteriophage T5 encodes a homolog of the eukaryotic transcription coactivator PC4 implicated in recombination-dependent DNA replication. *J Mol Biol* **425**, 4125-4133, doi:10.1016/j.jmb.2013.09.001 (2013).

27 Werten, S. Identification of the ssDNA-binding protein of bacteriophage T5: Implications for T5 replication. *Bacteriophage* **3**, e27304, doi:10.4161/bact.27304 (2013).

28 Werten, S. *et al.* High-affinity DNA binding by the C-terminal domain of the transcriptional coactivator PC4 requires simultaneous interaction with two opposing unpaired strands and results in helix destabilization. *J Mol Biol* **276**, 367-377, doi:10.1006/jmbi.1997.1534 (1998).

29 Vink, J. N. A. *et al.* Direct Visualization of Native CRISPR Target Search in Live Bacteria Reveals Cascade DNA Surveillance Mechanism. *Mol Cell* **77**, 39-50 e10, doi:10.1016/j.molcel.2019.10.021 (2020).

30 Westra, E. R. *et al.* CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol Cell* **46**, 595-605, doi:10.1016/j.molcel.2012.03.018 (2012).

31 Dorman, C. J. & Ni Bhriain, N. CRISPR-Cas, DNA Supercoiling, and Nucleoid-Associated Proteins. *Trends Microbiol* **28**, 19-27, doi:10.1016/j.tim.2019.08.004 (2020).

32 Mattenberger, Y., Silva, F. & Belin, D. 55.2, a phage T4 ORFan gene, encodes an inhibitor of Escherichia coli topoisomerase I and increases phage fitness. *PLoS One* **10**, e0124309, doi:10.1371/journal.pone.0124309 (2015).

33 Johnston, J. V., Nichols, B. P. & Donelson, J. E. Distribution of "minor" nicks in bacteriophage T5 DNA. *J Virol* **22**, 510-519 (1977).

34 Roy, D., Huguet, K. T., Grenier, F. & Burrus, V. IncC conjugative plasmids and SXT/R391 elements repair double-strand breaks caused by CRISPR–Cas during conjugation. *Nucleic Acids Research*, doi:10.1093/nar/gkaa518 (2020).

35 Uribe, R. V. *et al.* Discovery and Characterization of Cas9 Inhibitors Disseminated across Seven Bacterial Phyla. *Cell Host Microbe* **25**, 233-241 e235, doi:10.1016/j.chom.2019.01.003 (2019).

36 Osuna, B. A. *et al.* Listeria Phages Induce Cas9 Degradation to Protect Lysogenic Genomes. *Cell Host Microbe* **28**, 31-40 e39, doi:10.1016/j.chom.2020.04.001 (2020).

37 Chevallereau, A. *et al.* Exploitation of the Cooperative Behaviors of Anti-CRISPR Phages. *Cell Host Microbe* **27**, 189-198 e186, doi:10.1016/j.chom.2019.12.004 (2020).

38 Borges, A. L. *et al.* Bacteriophage Cooperation Suppresses CRISPR-Cas3 and Cas9 Immunity. *Cell* **174**, 917-925 e910, doi:10.1016/j.cell.2018.06.013 (2018).

39 Landsberger, M. *et al.* Anti-CRISPR Phages Cooperate to Overcome CRISPR-Cas Immunity. *Cell* **174**, 908-916 e912, doi:10.1016/j.cell.2018.05.058 (2018).

40 Hooper, D. C. & Jacoby, G. A. Mechanisms of drug resistance: quinolone resistance. *Ann N Y Acad Sci* **1354**, 12-31, doi:10.1111/nyas.12830 (2015).

41 Pawluk, A., Davidson, A. R. & Maxwell, K. L. Anti-CRISPR: discovery, mechanism and function. *Nat Rev Microbiol*, doi:10.1038/nrmicro.2017.120 (2017).

42 Borges, A. L., Davidson, A. R. & Bondy-Denomy, J. The Discovery, Mechanisms, and Evolutionary Impact of Anti-CRISPRs. *Annu Rev Virol* **4**, 37-59, doi:10.1146/annurev-virology-101416-041616 (2017).

43 Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-1659, doi:10.1093/bioinformatics/btl158 (2006).

44 Zhu, W., Lomsadze, A. & Borodovsky, M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res* **38**, e132, doi:10.1093/nar/gkq275 (2010).

45 Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**, W29-37, doi:10.1093/nar/gkr367 (2011).

46 Haft, D. H. *et al.* TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res* **29**, 41-43 (2001).

47 Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res* **28**, 263-266 (2000).

48 Parks, D. H. *et al.* A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol*, doi:10.1038/s41587-020-0501-8 (2020).

49 Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* **36**, 996-1004, doi:10.1038/nbt.4229 (2018).

50 Russel, J., Pinilla-Redondo, R., Mayo-Muñoz, D., Shah, S. A. & Sørensen, S. J. CRISPRCasTyper: An automated tool for the identification, annotation and classification of CRISPR-Cas loci. *bioRxiv*, 2020.2005.2015.097824, doi:10.1101/2020.05.15.097824 (2020).

51 Bondy-Denomy, J. *et al.* A Unified Resource for Tracking Anti-CRISPR Names. *The CRISPR Journal* **1**, 304-305, doi:10.1089/crispr.2018.0043 (2018).

52 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).

53 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).

Table 1. Structural features of AcrIIA22.

Data collection	
Space Group	P4332
Cell Dimensions	
a, b, c (Å)	128.56, 128.56, 128.56
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å)	50.00 - 2.80
R_{merge}	0.106 (0.906)
I/σ_I	17.4 (2.6)
Completeness (%)	98.7 (100.0)
Redundancy	10.4 (10.7)
CC 1/2	0.837
Refinement	
No. Reflections	9334
R_{work} (R_{free}) (%)	22.2 (24.6)
No. Complex in ASU	2
No. atoms	
Protein	810
Heteroatoms	50
Water	3
B-factor	82.82
R.m.s deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.610
Ramachandran	
Preferred (%)	98.15
Allowed (%)	1.85
Outliers (%)	0

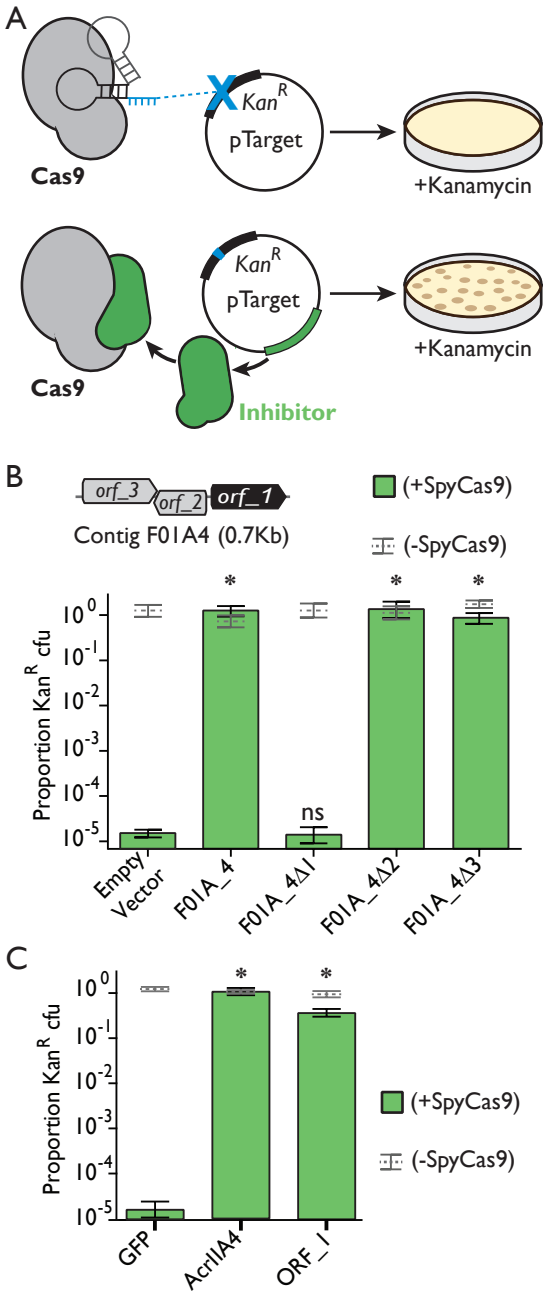


Figure 1. *Orf_1* from the metagenomic contig F01A_4 encodes a SpyCas9 inhibitor. (**A**) The plasmid protection assay used to reveal SpyCas9 inhibition. Plasmids without SpyCas9 inhibitors are cleaved by Cas9 and do not give rise to Kan^R colonies. Those with inhibitors withstand SpyCas9 attack and yield colonies. (**B**) An early stop codon in *orf_1 (Δ1),* but not *orf_2* or *orf_3 (Δ2* and *Δ3),* eliminates the ability of contig F01A_4 to protect a plasmid from SpyCas9. Asterisks depict statistically significant differences in plasmid retention between the indicated genotype and an empty vector control in SpyCas9-inducing conditions (Student's t-test, $p<0.002$, $n=3$); ns indicates no significance. All p-values were corrected for multiple hypotheses using Bonferroni's method. (**C**) Expression of *orf_1* is sufficient for SpyCas9 antagonism, protecting a plasmid as well as *acrIIA4*. Asterisks are as in panel B but relate to the GFP negative control rather than an empty vector.

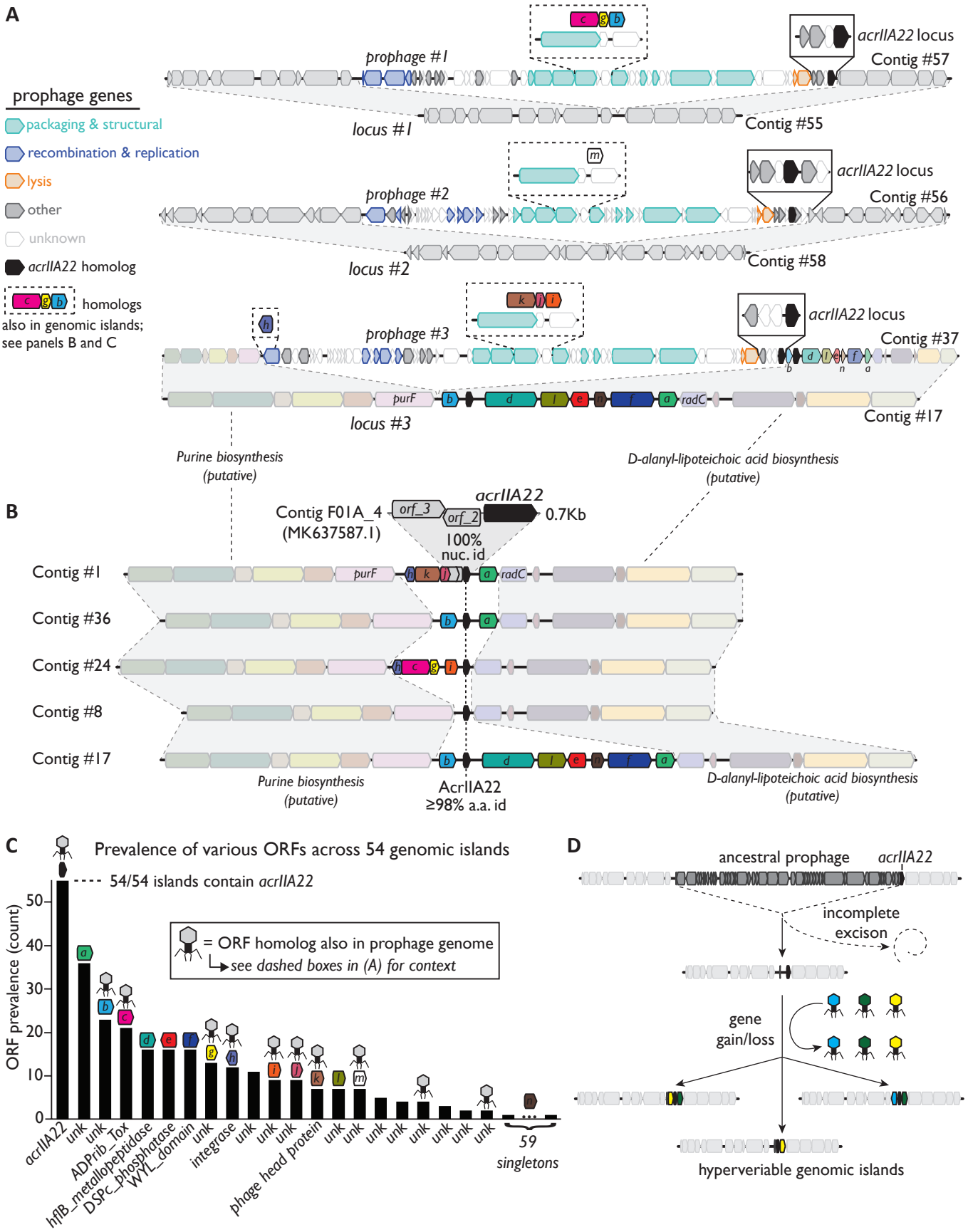


Figure 2. Caption on next page.

Figure 2. *AcrIIA22* homologs are found in hypervariable regions of prophage and bacterial genomes. (**A**) Homologs of *acrIIA22* are depicted in three related prophage genomes, integrated at three different genomic loci, revealed by a comparison of prophage-bearing contigs (#57, #56, #37) relative to unintegrated contigs (#55, #58, #17 respectively) that are otherwise nearly identical. Prophage genes are colored by functional category, according to the legend at the left of panel A. Genes immediately adjacent to *acrIIA22* (solid boxes) vary across phages, despite strong relatedness across much of the prophage genomes. Bacterial genes are colored gray, except for in contig #17, which is also depicted in panel B, below. (**B**) Homologs of *acrIIA22* are depicted in diverse genomic islands, including Contig #1, whose sequence has perfect nucleotide identity to the original metagenomic contig we recovered (F01A_4). All *acrIIA22* homologs in these loci are closely related but differ in their adjacent genes, which often have homologs in the prophages depicted in panel A (dashed boxes). Genomic regions flanking these hypervariable islands are nearly identical to one another and to prophage integration locus #3, as shown by homology to contig #17 from panel A. (**C**) The prevalence of various protein families (clustered at 65% amino acid identity) in a set of 54 unique genomic islands is shown. Each of these islands is flanked by the conserved genes *purF* and *radC* but contains a different arrangement of encoded genes. Domain-level annotations are indicated below each protein family (unk; unknown function). Gene symbols above each protein family are colored and lettered to indicate their counterparts or homologs in panels A and B. The phage capsid icon indicates sequences with homologs in prophage genomes. (**D**) An evolutionary model for the origin of the *acrIIA22*-encoding hypervariable genomic islands depicted in panel B is shown. We propose that *acrIIA22* moved via a phage insertion into a bacterial genomic locus, remained following an incomplete prophage excision event, and its neighboring genes subsequently diversified via horizontal exchange with additional phage genomes. Contigs are numbered to indicate their descriptions in Supplemental Table 3, which contains their metadata, taxonomy, and sequence retrieval information. All sequences and annotations may also be found in Supplementary Datasets 1 and 2.

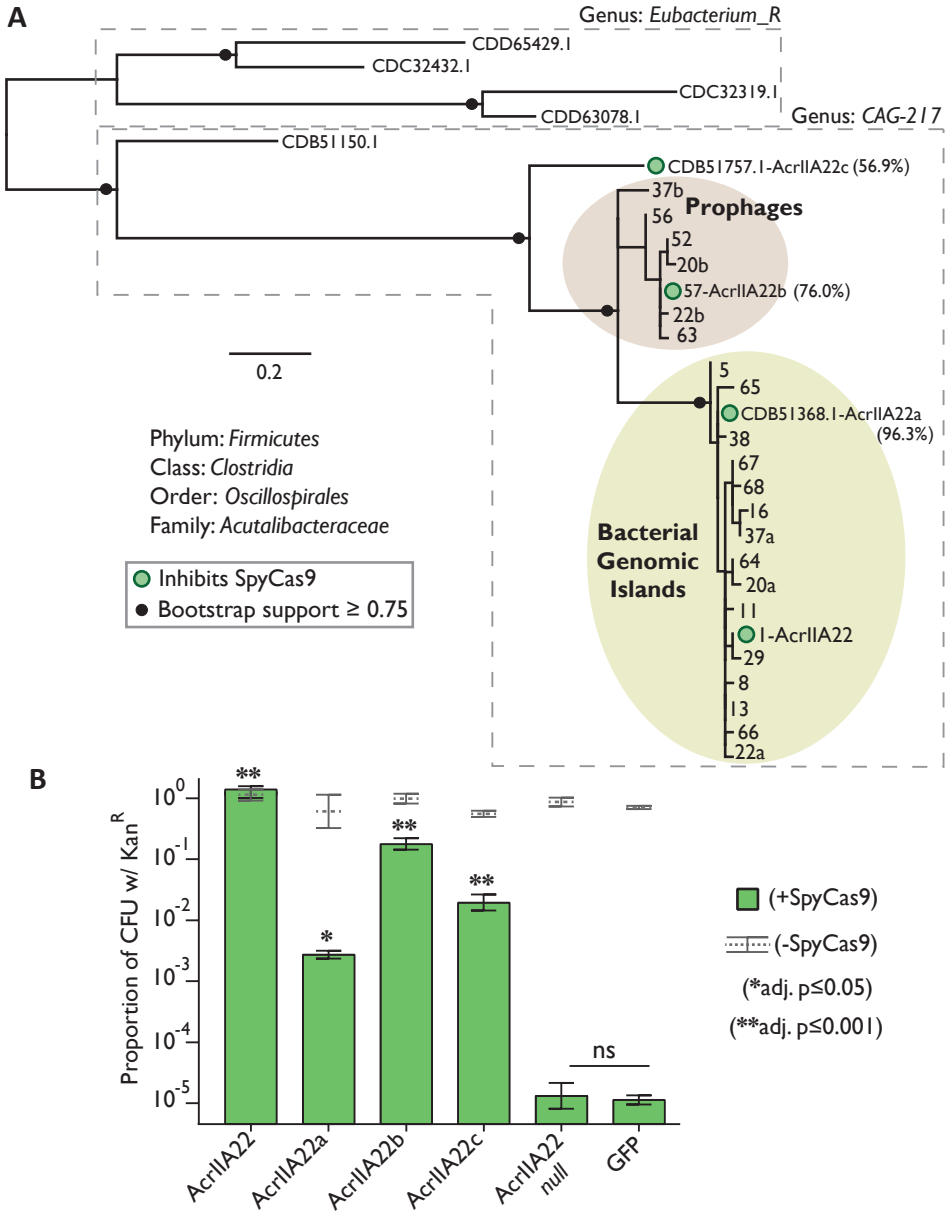


Figure 3. AcrIIA22 homologs capable of inhibiting SpyCas9 are common in the unnamed *Clostridial* genus, CAG-217. Phylogenetic classifications were assigned corresponding to the GTDB naming convention (Methods). (**A**) A phylogeny of all unique AcrIIA22 homologs identified from metagenomic and NCBI databases. Prophage sequences are shaded brown and homologs from hypervariable bacterial genomic islands are shaded yellow. Sequences obtained from NCBI are labeled with protein accession numbers. In other cases, AcrIIA22 homologs are numbered to match their contig-of-origin (Supplemental Table 3). In some cases, more than one AcrIIA22 homolog is found on the same contig ('a' or 'b' indicates its presence in a hypervariable genomic island or prophage genome, respectively). Circles at nodes indicate bootstrap support ≥ 0.75. Dashed boxes separate sequences identified from different bacterial genera. Filled green circles indicate homologs that were tested for their ability to inhibit SpyCas9 in the plasmid protection assay in panel B. (**B**) Homologs of AcrIIA22 in CAG-217 genomes inhibit SpyCas9. Asterisks depict statistically significant differences in plasmid retention under SpyCas9-inducing conditions between the indicated sample and a null mutant with an early stop codon in *acrIIA22*, per the legend at right (ns indicates no significance). All p-values were corrected for multiple hypotheses using Bonferroni's method (Student's t-test, n=3).

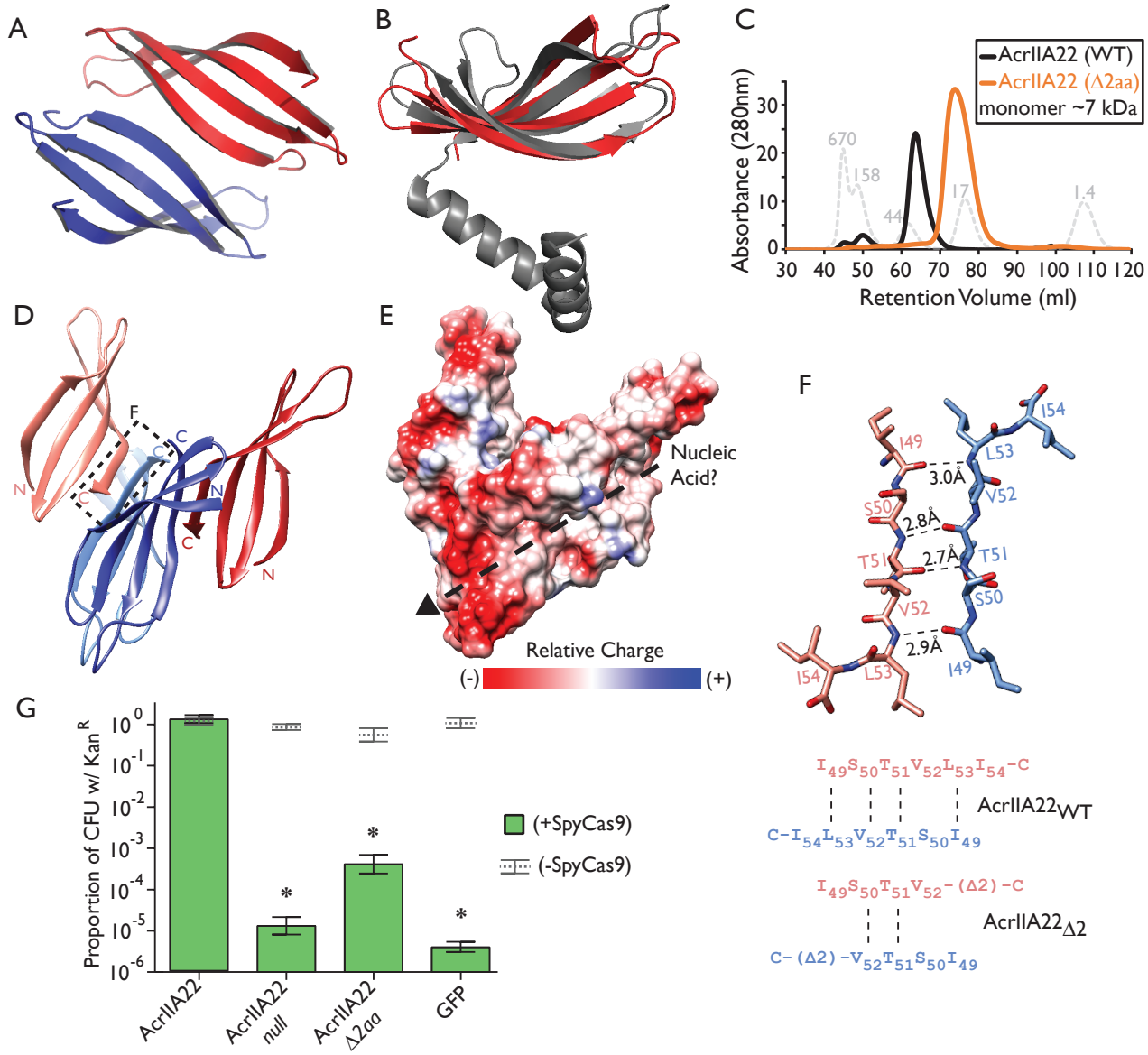


Figure 4. AcrIIA22 is a PC4-like protein that oligomerizes and inhibits SpyCas9. (**A**) AcrIIA22's crystal structure reveals a homodimer of two four-stranded β-sheets. (**B**) A monomer of AcrIIA22 (PDB:7JTA) is structurally similar to a predicted single-stranded DNA binding protein, which is proposed to promote recombination in phage T5 (PDB:4BG7, Z-score=6.2, matched residues 15%). (**C**) AcrIIA22 elutes as an oligomer that is 4-5 times the predicted molecular mass of its monomer. The gray, dashed trace depicts protein standards of the indicated molecular weight. The orange trace depicts the elution profile of a two-amino acid C-terminal AcrIIA22 truncation mutant. (**D**) Ribbon diagram of a proposed AcrIIA22 tetramer which requires binding between anti-parallel β-strands at the C-termini of AcrIIA22 monomers to form extended, concave β-sheets. This putative oligomerization interface is indicated by the dashed box and is detailed in panel F. (**E**) Space filling model of the tetrameric AcrIIA22 structure from panel D, with relative charge depicted, highlighting a groove (dashed line with arrowhead) that may accommodate nucleic acids. (**F**) A putative oligomerization interface between the C-termini of two AcrIIA22 monomers is shown with hydrogen bond distances between the polypeptide backbones indicated. The wild-type sequence and truncation mutant are indicated below. Dashed lines indicate potential hydrogen bonds. This interface occurs twice in the putative tetramer, between red-hued and blue-hued monomers in panel D. (**G**) The truncation mutant fails to protect a plasmid from SpyCas9 elimination, similar to an early stop codon mutant (*null*) and a *gfp* negative control. Asterisks depict statistically significant differences in plasmid retention under SpyCas9-inducing conditions between the indicated sample and the wild-type sequence (adj. $p < 0.002$, Student's t-test, n=3). All p-values were corrected for multiple hypotheses using Bonferroni's method.

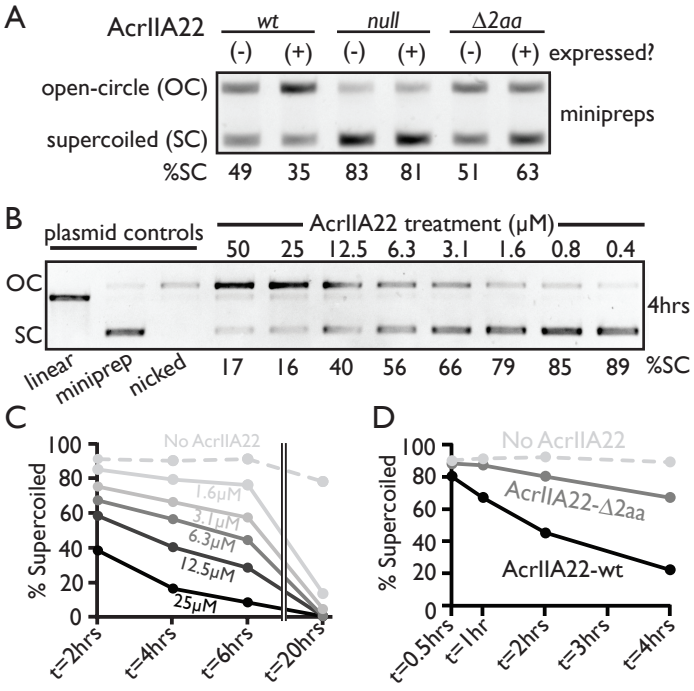


Figure 5. AcrIIA22 nicks supercoiled plasmids *in vivo* and *in vitro*. (**A**) Gel electrophoresis of plasmids purified from overnight *E. coli* cultures expressing the indicated genotypes. OC, open-circle plasmid; SC, supercoiled plasmid. %SC indicates the percentage of DNA in the supercoiled form for each sample. (**B**) AcrIIA22 nicks supercoiled plasmids *in vitro*. Supplemental Figure 7 depicts this experiment at additional time points. (**C**) Quantification of AcrIIA22-nicked plasmids in panel B and Supplemental Figure 7. AcrIIA22 nicks plasmids in a time and concentration-dependent manner. (**D**) A nickase assay as in panels B and C shows that the 2-aa truncation mutant is impaired for activity *in vitro*, relative to wild-type AcrIIA22. In both cases, 25 μM protein was used.

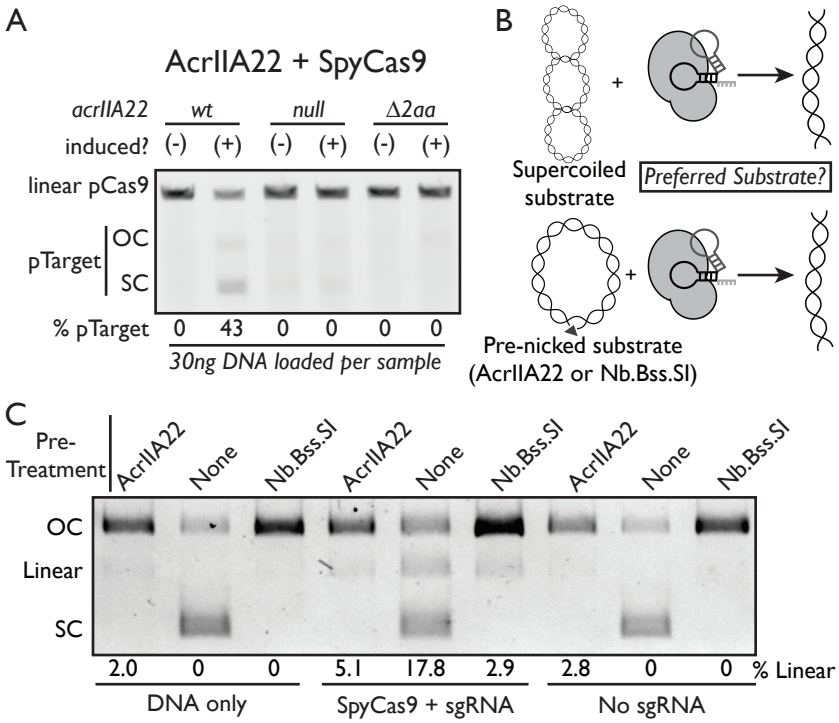
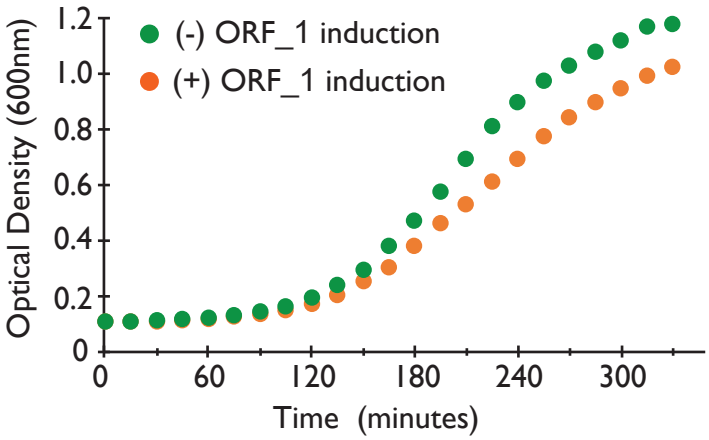
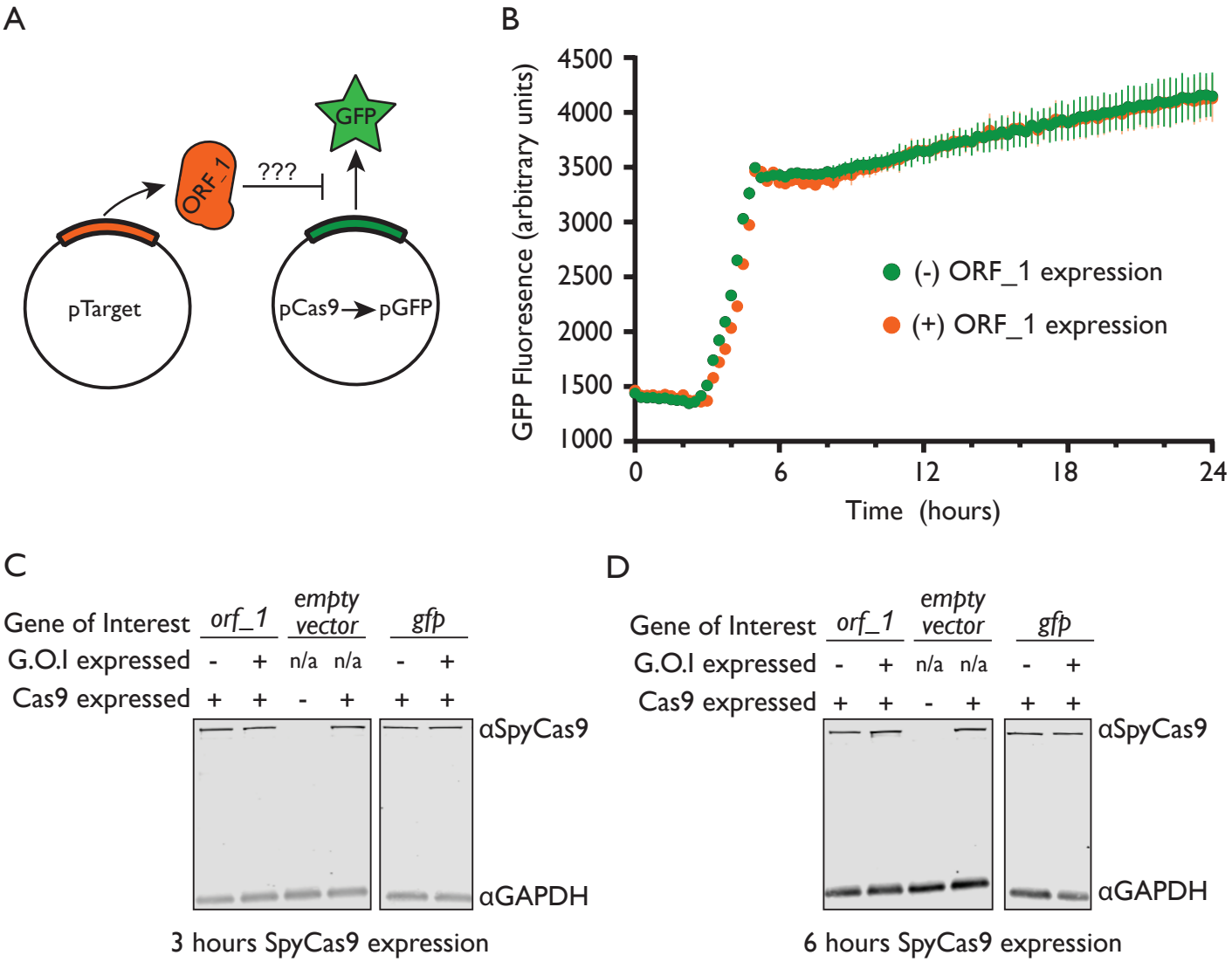


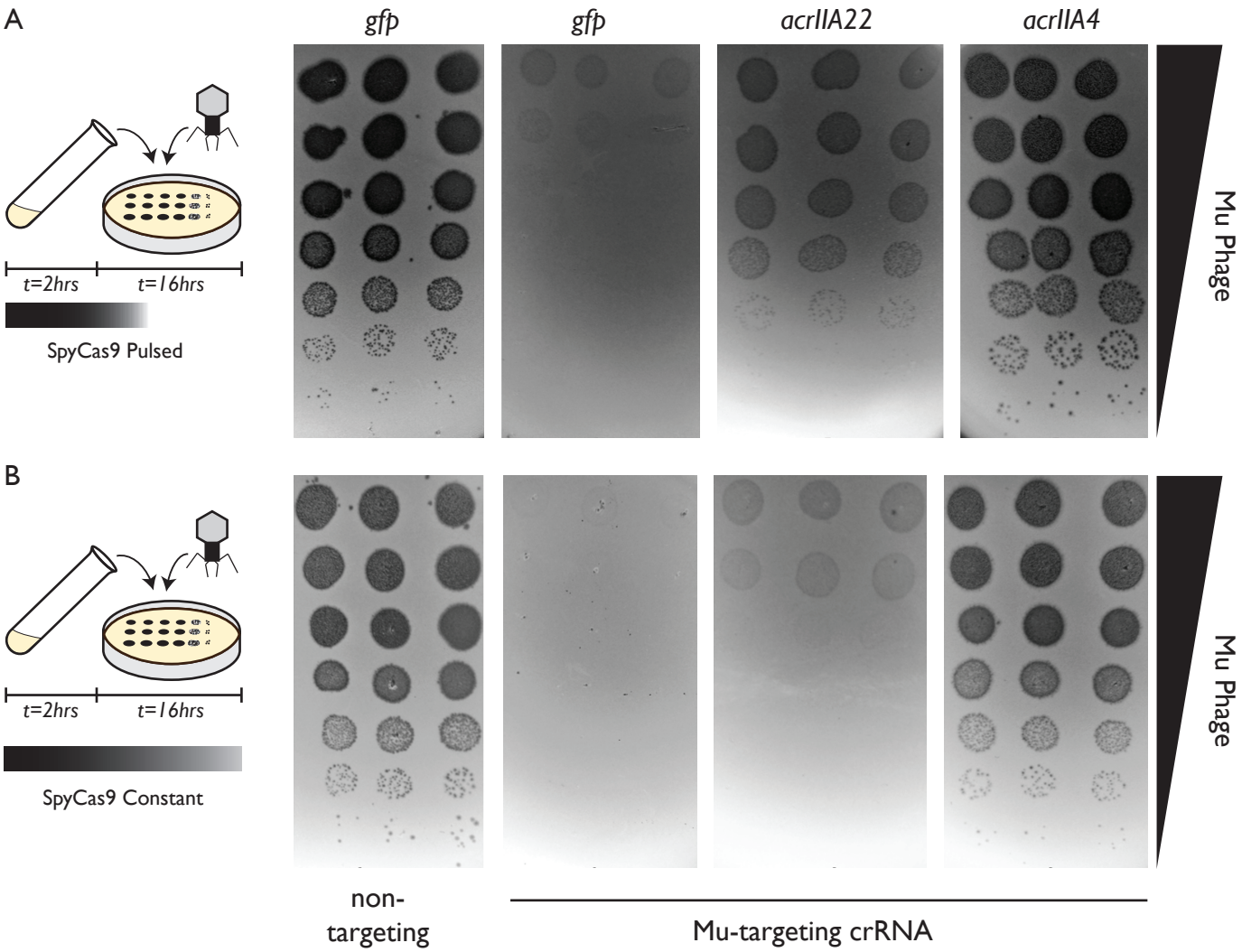
Figure 6. AcrIIA22 protects plasmids from SpyCas9 cleavage *in vivo* and *in vitro*. (**A**) Gel electrophoresis of plasmids purified from overnight *E. coli* cultures expressing the indicated *acrIIA22* genotypes and SpyCas9 from a second plasmid. The *acrIIA22*-encoding plasmids are indicated with the 'pTarget' label. OC, open-circle; SC, supercoiled. The SpyCas9 plasmid was linearized via a unique restriction site before electrophoresis. (**B**) Experimental design for the data depicted in panel C. The experiment tests whether SpyCas9 preferentially cleaves a supercoiled or open-circle plasmid target *in vitro*. (**C**) Nicked plasmids are less susceptible to linearization via SpyCas9 cleavage. Plasmid purifications from overnight cultures were either left unmodified or pre-treated with one of two nickase enzymes, AcrIIA22 or Nb.Bss.SI. Linear, open-circle (OC), and supercoiled (SC) plasmid forms are indicated. The percentage of DNA in the linear form is quantified below the gel, where the reaction components are also listed. See Supplemental Figure 9 for these data in different reaction conditions.



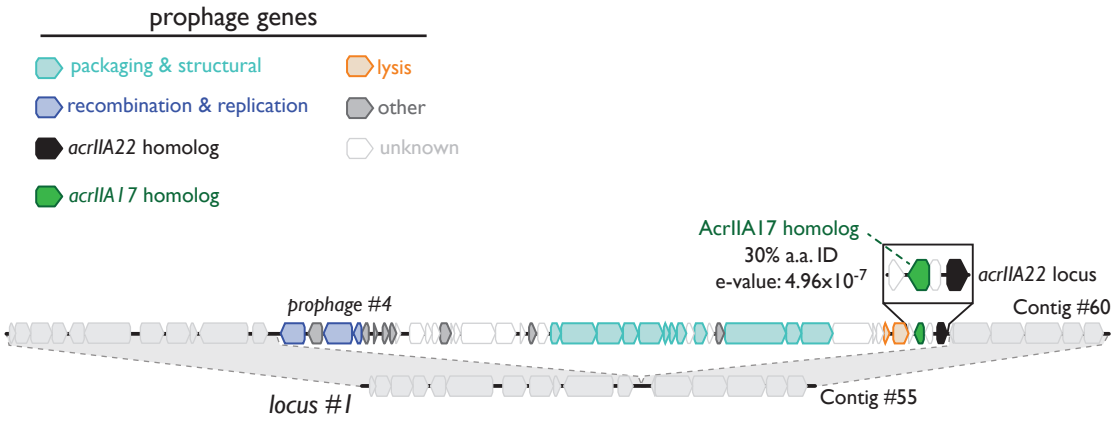
Supplemental Figure 1. *Orf_1* confers mild toxicity in *E. coli*. Growth curves with (orange) and without (green) *orf_1* induction. These growth data map directly to the cfu data in Figure 1C, demonstrating that anti-SpyCas9 activity occurs under conditions with minimal *orf_1* toxicity. Samples were removed after six hours of growth to plate for cfus. Growth curves are shown for samples without SpyCas9 induction to ensure that *orf_1* toxicity is not mitigated due to elimination of its plasmid. Points indicate averages from three replicates. Standard deviations at each timepoint are so small that the error bars do not exceed the bounds of the data point.



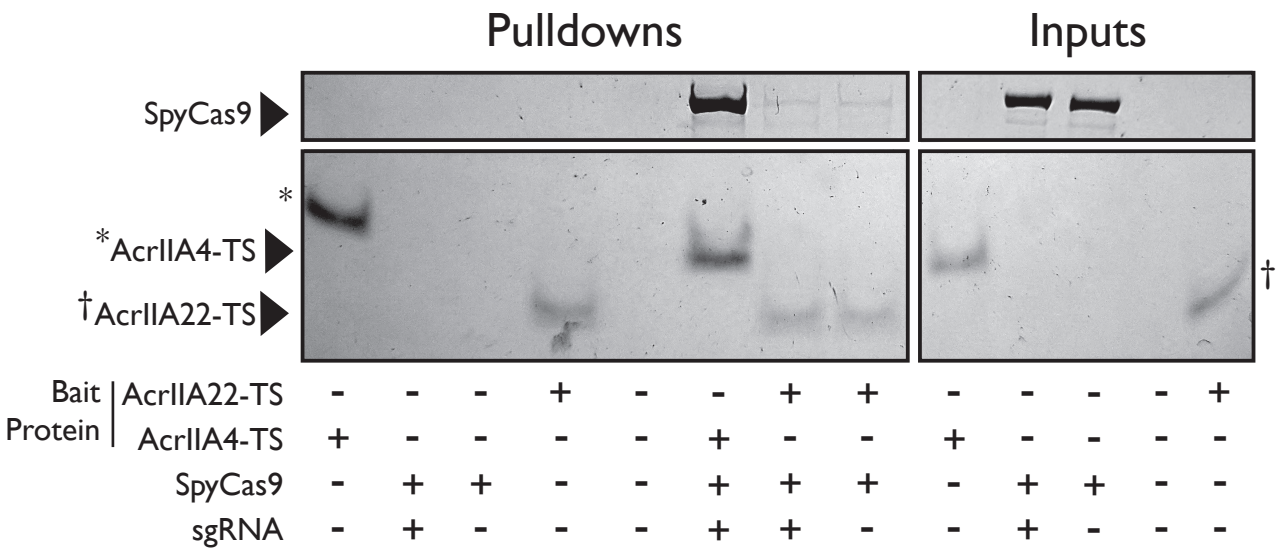
Supplemental Figure 2. *Orf_1* does not impact SpyCas9 expression. (**A**) Cartoon depicting the experimental design for (**B**). If ORF_1 prevented transcription from pCas9 or altered its copy number, we would expect expression of the *orf_1* gene to deplete the level of green fluorescence observed from a construct that replaces the *spycas9* gene with *gfp*. (**B**) Fluorescence measurements for the experiment depicted in panel A, throughout an *E. coli* growth curve. ORF_1 does not impact GFP expression. Points indicate averages from three replicates, error bars indicate standard deviation. (**C**) A western blot to detect SpyCas9 expression as a function of ORF_1 or GFP expression in growing *E. coli* cultures. As an internal control, GAPDH expression was also detected. (**D**) As in panel C, but samples were collected six hours after SpyCas9 induction, instead of three.



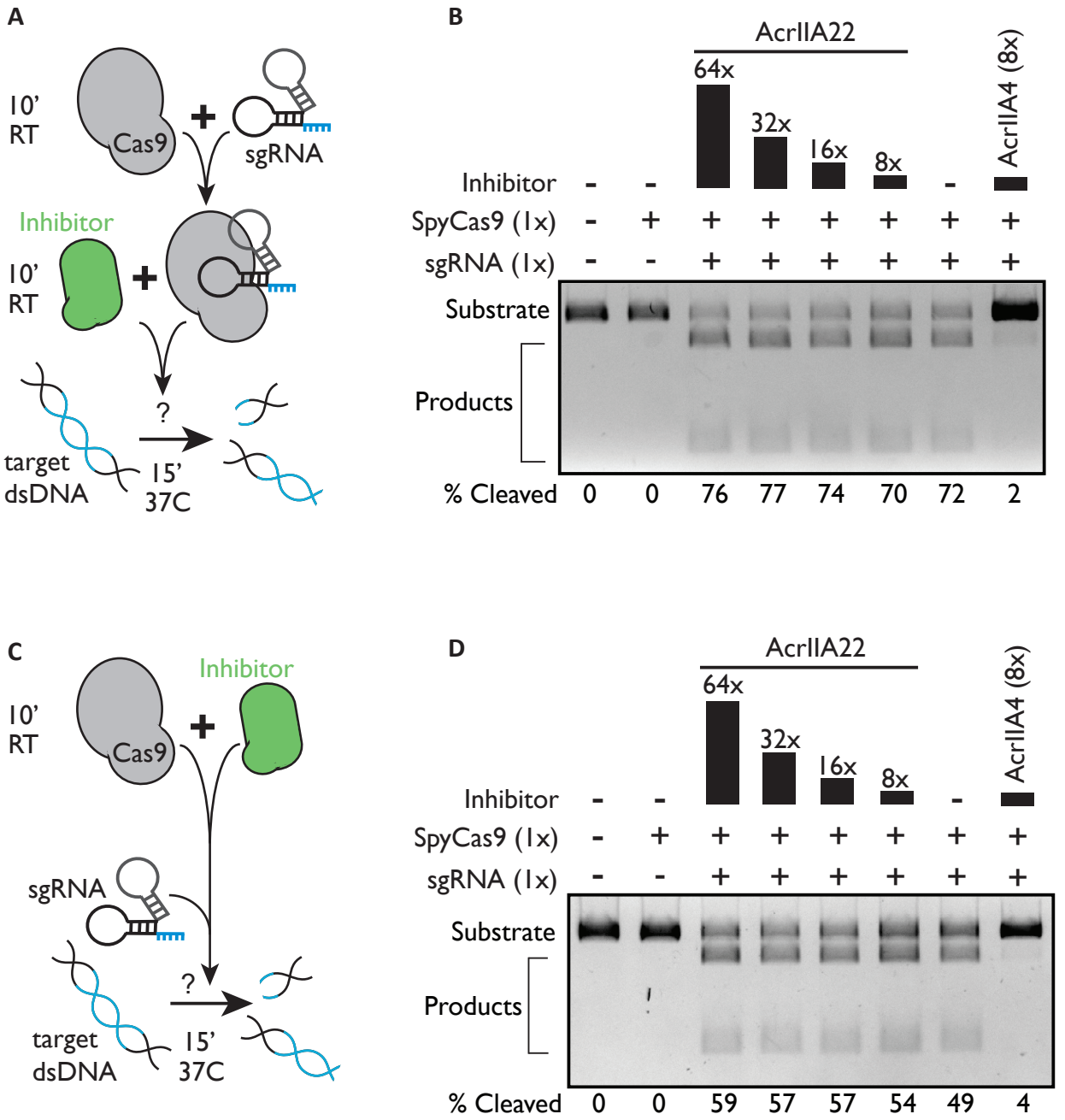
Supplemental Figure 3. Mu phage fitness, measured by plaquing on *E. coli*. Plaquing is measured in the presence of *gfp*, *acrIIA22*, or *acrIIA4* via serial ten-fold dilutions. Bacterial clearing (black) occurs when phage Mu overcomes SpyCas9 immunity and lyses *E. coli*. In (**A**) and in (**B**), SpyCas9 with a Mu-targeting crRNA confers substantial protection against phage Mu relative to a non-targeting (n.t.) control, in both conditions tested. These conditions are depicted at left, with the only difference being whether SpyCas9 was only expressed in liquid growth prior to phage infection (panel A) or expressed both in liquid media and in solid media throughout infection (panel B). The positive control *acrIIA4* significantly enhances Mu fitness by inhibiting SpyCas9 in all conditions. In contrast, *acrIIA22* confers milder protection against SpyCas9. The indicated *acr* gene or *gfp* control is expressed from a second plasmid, in trans.



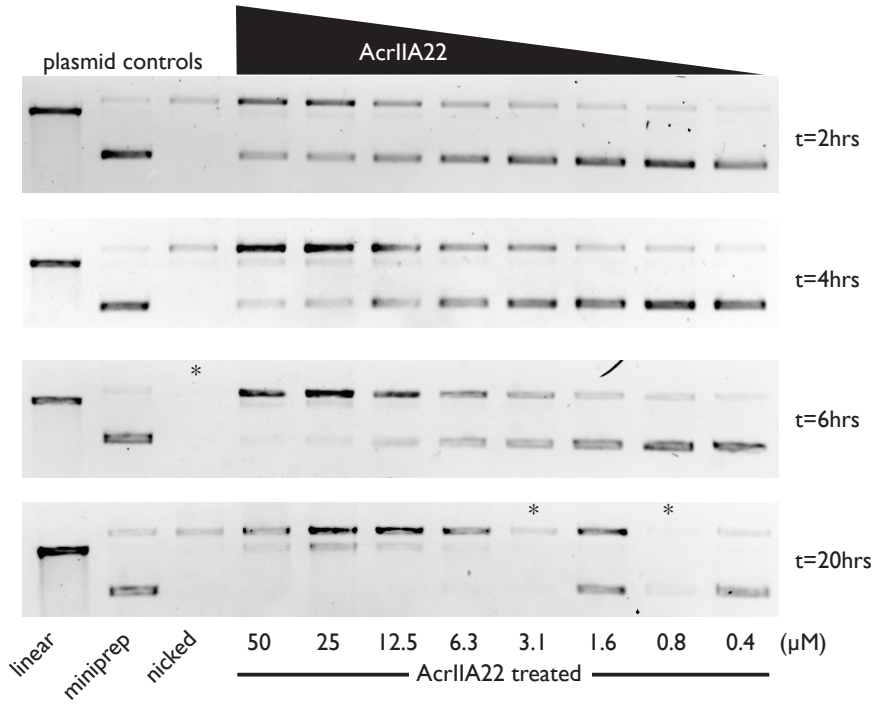
Supplemental Figure 4. An *acrIIA22*-encoding prophage similar to those depicted in Figure 2A. Prophage genes are colored by functional category, per the legend and as in Figure 2. This prophage encodes for a homolog of the previously described SpyCas9 inhibitor *acrIIA17* within one kilobase of an *acrIIA22* homolog. Sequence relatedness for the depicted *acrIIA17* gene and the original discovery by Mahendra *et al*. is shown. Because phages often encode multiple *acr*s in the same locus, the co-localization of *acrIIA17* with *acrIIA22* is consistent with the latter gene functioning natively to inhibit CRISPR-Cas activity. Contigs are numbered to indicate their descriptions in Supplemental Table 3, which contains their metadata, taxonomy, and sequence retrieval information. All sequences and annotations may also be found in Supplementary Datasets 1 and 2.



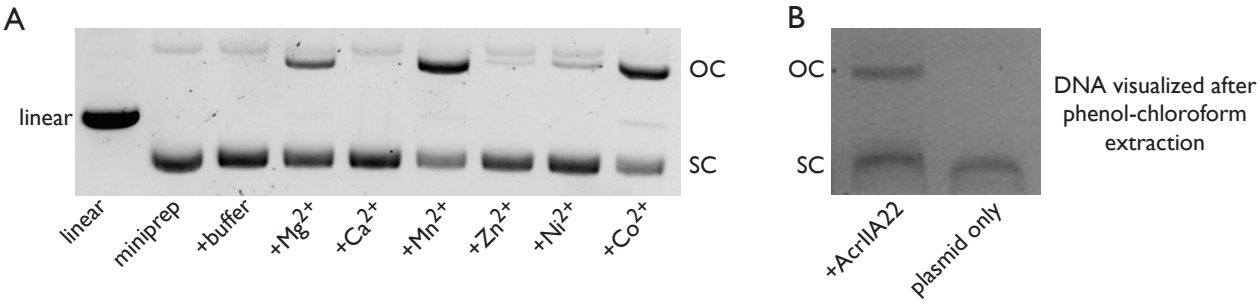
Supplemental Figure 5. AcrIIA22 does not strongly bind SpyCas9. SpyCas9 and sgRNA were pre-incubated before mixing with a twin-strep (TS) tagged AcrIIA22 or AcrIIA4. SpyCas9 without sgRNA was also used. (A) Streptactin pulldowns on AcrIIA4 also pulled down SpyCas9 pre-incubated with sgRNA, as previously reported. Similar pulldowns with AcrIIA22 indicate little to no interaction with SpyCas9, regardless of whether sgRNA was used. These images depict total protein content visualized by Coomassie stain. Reaction components are indicated below the gel image. Aterisks (*) and dagger (†) symbols indicate AcrIIA4 and AcrIIA22 protein bands that run at slightly different positions than expected due to gel smiling.



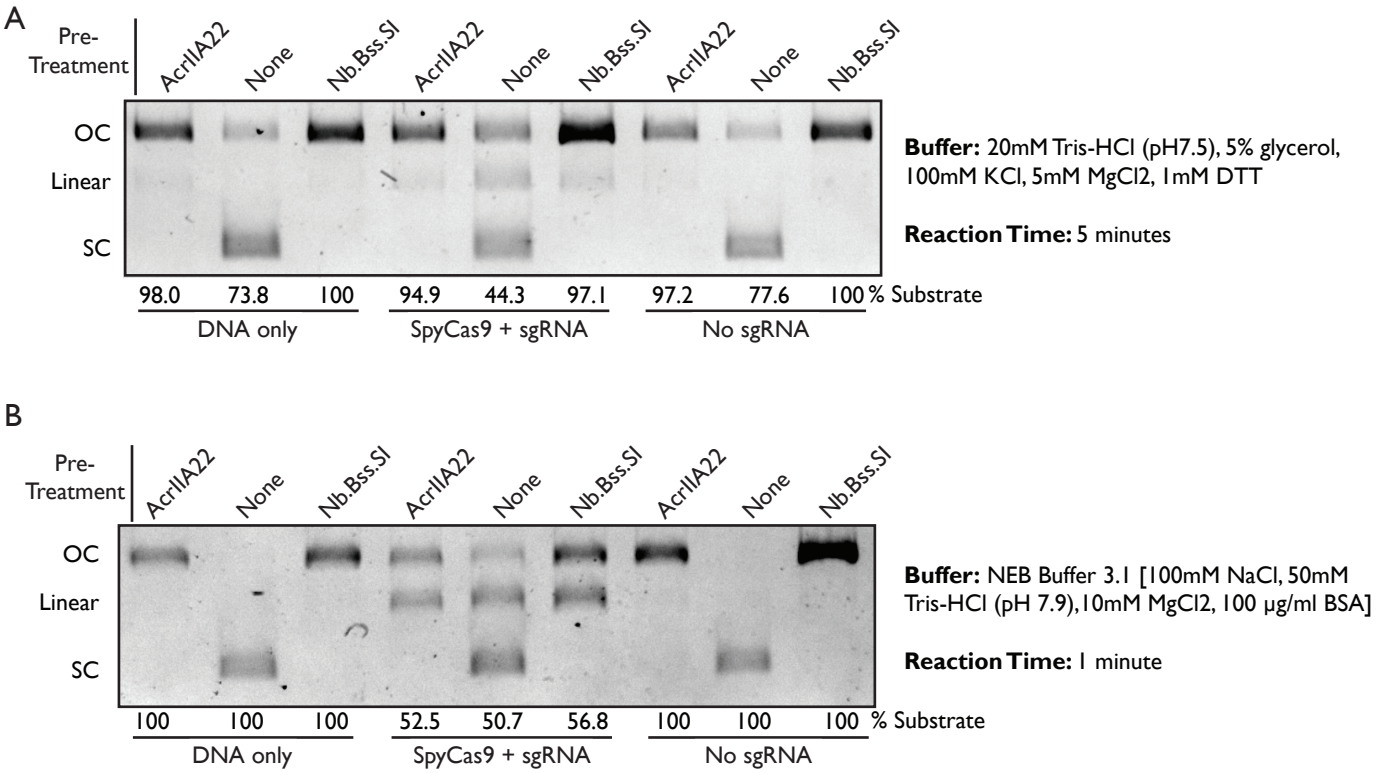
Supplemental Figure 6. AcrIIA22 does not inhibit SpyCas9 from cleaving linear DNA. (**A**) Cartoon depicting the experiment in (**B**). SpyCas9 was pre-incubated with sgRNA targeting linear DNA. Then, Acrs candidates were added. Subsequently, cleavage reactions were performed, and the DNA products visualized by gel electrophoresis in panel B. (**B**) Products of the reactions described in panel A for the inhibitors AcrIIA22 and AcrIIA4. Reaction components are depicted atop the gel image, with molar equivalents relative to SpyCas9 indicated. The percent of DNA substrate cleaved by SpyCas9 is quantified below each lane. (**C**) As in panel A, except candidate Acrs were incubated with SpyCas9 before sgRNA addition. Reactions were begun via the simultaneous addition of sgRNA and linear dsDNA instead of dsDNA in isolation. (**D**) The products of the reactions described in panel C for AcrIIA22 and AcrIIA4 inhibitors. To push these Cas9 digestion reactions toward completion, ratios of Cas9:DNA were ten-fold higher than those shown in Figure 6C and reactions were allowed to progress for three times as long.



Supplemental Figure 7. AcrIIA22 nicks supercoiled plasmids *in vitro*. Plasmid controls are in the leftmost three lanes. Reaction times are indicated to the right of each gel. AcrIIA22 nicks supercoiled plasmids in a concentration and time dependent manner. Extended incubations at high concentrations produce a linearized product. Asterisks (*) indicate loading errors, where less sample was loaded than other lanes.



Supplemental Figure 8. AcrIIA22 is a nickase. (**A**) The impact of different divalent cations on AcrIIA22's nickase activity. OC, open-circle plasmid form. SC, supercoiled plasmid. (**B**) The open-circle plasmid product persists through phenol-chloroform extraction following AcrIIA22 treatment.



Supplemental Figure 9. Nicked plasmids are less susceptible to linearization via SpyCas9 cleavage. (**A**) As in Figure 6C, Plasmid purifications from overnight cultures were either left unmodified or pre-treated with one of two nickase enzymes, AcrIIA22 or Nb.Bss.SI. Linear, open-circle (OC), and supercoiled (SC) plasmid forms are indicated. The % substrate value indicates the percentage of DNA in the nicked form for AcrIIA22 or Nb.Bss.SI-treated plasmids or in the supercoiled form for the untreated miniprep. Reaction components are listed below each lane. Buffer conditions and reaction time is listed at right. (**B**) As in (**A**), but with different reaction conditions, listed at right. In these conditions, the reaction proceeded too quickly to detect SpyCas9's substrate preference (all substrates were rapidly linearized).

Supplemental Table 1. Whether anti-CRISPRs are known to bind Cas proteins and inhibit their cleavage activity as purified proteins.

Acr	Binds cognate Cas protein?	Inhibit as pure proteins?	References
AcrIIA1	Yes	No	(Osuna et al., 2020)
AcrIIA2	Yes	Yes	(Jiang et al., 2019; Liu et al., 2019)
AcrIIA3	unknown	unknown	(Rauch et al., 2017)
AcrIIA4	Yes	Yes	(Dong et al., 2017; Shin et al., 2017; Yang and Patel, 2017)
AcrIIA5	Yes	Yes	(An et al., 2020; Garcia et al., 2019; Song et al., 2019)
AcrIIA6	Yes	Yes	(Fuchsbauer et al., 2019)
AcrIIA7	No	Yes	(Uribe et al., 2019)
AcrIIA8	Yes	Yes	(Uribe et al., 2019)
AcrIIA9	Yes	Yes	(Uribe et al., 2019)
AcrIIA10	Yes	Yes	(Uribe et al., 2019)
AcrIIA11	Yes	Yes	(Forsberg et al., 2019)
AcrIIA12	probable	Yes	(Eitzinger et al., 2020; Osuna et al., 2020)
AcrIIA13	unknown	Yes	(Watters et al., 2020)
AcrIIA14	unknown	Yes	(Watters et al., 2020)
AcrIIA15	unknown	Yes	(Watters et al., 2020)
AcrIIA16	Yes	Yes	(Mahendra et al., 2020)
AcrIIA17	Yes	No	(Mahendra et al., 2020)
AcrIIA18	Yes	No	(Mahendra et al., 2020)
AcrIIA19	Yes	No	(Mahendra et al., 2020)
AcrIIA20	unknown	Yes	(Eitzinger et al., 2020)
AcrIIA21	unknown	Yes	(Eitzinger et al., 2020)
AcrIIA22	**No**	**No**	**This study**
AcrIIC1	Yes	Yes	(Pawluk et al., 2016)
AcrIIC2	Yes	Yes	(Pawluk et al., 2016)
AcrIIC3	Yes	Yes	(Pawluk et al., 2016)
AcrIIC4	Yes	Yes	(Lee et al., 2018)
AcrIIC5	Yes	Yes	(Lee et al., 2018)
AcrVA1	Yes	Yes	(Knott et al., 2019b; Watters et al., 2018; Zhang et al., 2019)
AcrVA2	unknown	unknown	(Marino et al., 2018)
AcrVA3	unknown	unknown	(Marino et al., 2018)
AcrVA4	Yes	Yes	(Knott et al., 2019a; Knott et al., 2019b; Watters et al., 2018; Zhang et al., 2019)

AcrVA5	transiently	Yes	(Knott et al., 2019b; Watters et al., 2018; Zhang et al., 2019)
AcrVIA1(Lse)	Yes	Yes	(Meeske et al., 2020)
AcrVIA1(Lwa)	Yes	unknown	(Lin et al., 2020)
AcrVIA2	Yes	unknown	(Lin et al., 2020)
AcrVIA3	Yes	unknown	(Lin et al., 2020)
AcrVIA4	Yes	unknown	(Lin et al., 2020)
AcrVIA5	Yes	unknown	(Lin et al., 2020)
AcrVIA6	Yes	unknown	(Lin et al., 2020)
AcrVIA7	unknown	unknown	(Lin et al., 2020)
AcrIB1	unknown	unknown	(Lin et al., 2020)
AcrIC1	unknown	unknown	(Leon et al., 2020)
AcrIC2	probable	unknown	(Leon et al., 2020)
AcrIC3	unknown	unknown	(Leon et al., 2020)
AcrIC4	probable	unknown	(Leon et al., 2020)
AcrIC5	probable	unknown	(Leon et al., 2020)
AcrIC6	unknown	unknown	(Leon et al., 2020)
AcrIC7	probable	unknown	(Leon et al., 2020)
AcrIC8	probable	unknown	(Leon et al., 2020)
AcrID1	Yes	unknown	(He et al., 2018)
AcrIE1	Yes	unknown	(Pawluk et al., 2017)
AcrIE2	unknown	unknown	(Pawluk et al., 2014)
AcrIE3	probable	unknown	(Stanley, 2018)
AcrIE4	unknown	unknown	(Pawluk et al., 2014)
AcrIE5	unknown	unknown	(Pawluk et al., 2014)
AcrIE6	unknown	unknown	(Pawluk et al., 2014)
AcrIE7	unknown	unknown	(Pawluk et al., 2014)
AcrIE4-IF7	unknown	unknown	(Marino et al., 2018)
AcrIE8	unknown	unknown	(Pinilla-Redondo et al., 2020)
AcrIF1	Yes	unknown	(Bondy-Denomy et al., 2015; Chowdhury et al., 2017; Guo et al., 2017)
AcrIF2	Yes	unknown	(Bondy-Denomy et al., 2015; Chowdhury et al., 2017; Guo et al., 2017)
AcrIF3	Yes	unknown	(Bondy-Denomy et al., 2015; Wang et al., 2016a; Wang et al., 2016b)
AcrIF4	Yes	unknown	(Bondy-Denomy et al., 2015)
AcrIF5	unknown	unknown	(Bondy-Denomy et al., 2013)
AcrIF6	Yes	Yes	(Zhang et al., 2020)

AcrIF7	Yes	unknown	(Hirschi et al., 2020)
AcrIF8	Yes	Yes	(Zhang et al., 2020)
AcrIF9	Yes	Yes	(Hirschi et al., 2020; Zhang et al., 2020)
AcrIF10	Yes	unknown	(Guo et al., 2017)
AcrIF11	unknown	unknown	(Marino et al., 2018)
AcrIF12	unknown	unknown	(Marino et al., 2018)
AcrIF13	unknown	unknown	(Marino et al., 2018)
AcrIF14	unknown	unknown	(Marino et al., 2018)
AcrIF15	probable	unknown	(Pinilla-Redondo et al., 2020)
AcrIF16	unknown	unknown	(Pinilla-Redondo et al., 2020)
AcrIF17	unknown	unknown	(Pinilla-Redondo et al., 2020)
AcrIF18	probable	unknown	(Pinilla-Redondo et al., 2020)
AcrIF19	unknown	unknown	(Pinilla-Redondo et al., 2020)
AcrIF20	unknown	unknown	(Pinilla-Redondo et al., 2020)
AcrIF21	unknown	unknown	(Pinilla-Redondo et al., 2020)
AcrIF22	unknown	unknown	(Pinilla-Redondo et al., 2020)
AcrIF23	unknown	unknown	(Pinilla-Redondo et al., 2020)
AcrIF24	unknown	unknown	(Pinilla-Redondo et al., 2020)
AcrIII-1	No (degrades CA4 second messenger)	No	(Athukoralage et al., 2020)
AcrIIIB1	Yes	unknown	(Bhoobalan-Chitty et al., 2019)

Supplemental Table 2. PC4-like proteins with structural homology to AcrIIA22

Structural Homolog		Function	Similarity to AcrIIA22			
PDBID	Name	DNA/RNA Binding*	Zscore	r.m.s.d.	n-align	% A.A. ID
4bg7	PC4 putative transcriptional coactivator p15	DNA	6.2	2.5	54	15
3k44	D. melanogaster Pur-α	DNA/RNA	5.9	2.6	47	9
5fgp	Pur-α repeat I and II from D. melanogaster	DNA/RNA	5.6	2.1	48	8
3n8b	Pur-α from B. burgdorferi	DNA/RNA	5	2.8	48	6
2gje	Mitochondrial RNA Binding Protein (Trypanosoma brucei)	RNA	4.9	2.5	52	8
5zkl	Protein of unknown function SP_0782, S. pneumoniae	DNA	4.7	3.6	52	12
5fgo	D. melanogaster Pur-α repeat III	No info	4.5	2.7	44	14
1pcf	Replication & transcription cofactor PC4 CTD	DNA	4.5	2.5	45	7
2ltt	Putative Uncharacterized Protein YDBC	DNA	4.5	2.8	50	12
4bhm	MoSub1-DNA PC4 transcription cofactor	DNA	3.9	2.8	45	4
3cm1	SSGA-like sporulation specific cell division protein	No info	2.8	3.7	47	13
1l3a	Transcription factor PBF-2 (P24, WHY1)	DNA	2.8	5	48	8
4ntq	Anti-toxin CdiI, E. cloacae	No info	2.7	3	49	12
3n1k	WHY2 transcription factor, S. tuberosum	DNA	2.6	2.8	52	4

*RNA/DNA binding data from (Janowski and Niessing, 2020).

Supplemental Table 3. All sequences used in this study. Sequence names and databases are indicated. All sequences and annotations are also available as supplemental data. Sequences retrieved from Pasolli *et al.* refer to the following study: (Pasolli et al., 2019).

Contig No.	Sequence Name	How Used?	NCBI Nuc. ID	Pasolli et al SGB?	Pasolli et al Raw Assembly?	Pasolli Reconstructed Genome Name	SGB id	GTDB Taxonomy	External Data Available at:
1	4303_LiJ_2014__V1.UC63-0__bin.67_NODE_112_length_95405_cov_4.60675	Figures 2, find gene functions from 54 unique genomic loci; Figure 3 Acr Seq	n/a	Yes	No	LiJ_2014__V1.UC63-0__bin.67	4303	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
2	Bengtsson-PalmeJ_2015__TRAVELRES9_NODE_4_length_538580_cov_9.43148	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
3	ChengpingW_2017__AS9raw_NODE_922_length_27664_cov_3.49089	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
4	CosteaPI_2017__SID713B025-11-0-0_NODE_4_length_351620_cov_7.46108	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
5	BritoIL_2016__M1.64.ST_NODE_47_length_140472_cov_9.49805	find gene functions from 54 unique genomic loci; Figure 3 Acr Seq	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
6	BritoIL_2016__M2.57.ST_NODE_3_length_405636_cov_14.0428	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
7	BritoIL_2016__WL.14.ST_NODE_13_length_259523_cov_10.8408	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
8	ChengpingW_2017__AS67raw_NODE_2_length_439177_cov_9.00174	Figure 2, find gene functions from 54 unique genomic loci; Figure 3 Acr Seq	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
9	CM_madagascar__A90_04_1FE_NODE_125_length_81453_cov_9.00904	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
10	CM_madagascar__V12_01_2FE_NODE_5_length_202628_cov_9.50435	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
11	CosteaPI_2017__SID713A046-11-0-0_NODE_322_length_69000_cov_4.32987	find gene functions from 54 unique genomic loci; Figure 3 Acr Seq	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
12	CosteaPI_2017__SID713A045-11-0-0_NODE_78_length_64886_cov_3.97493	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
13	CosteaPI_2017__SID713A004-11-0-0_NODE_1_length_647860_cov_14.1013	find gene functions from 54 unique genomic loci; Figure 3 Acr Seq	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
14	CosteaPI_2017__peacemaker-11-60-	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html

	0_NODE_48_length_49378_cov_15.5445								
15	CosteaPI_2017__SID713A063-11-90-0_NODE_2082_length_16960_cov_2.98527	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
16	CosteaPI_2017__SID713A088-11-0-0_NODE_89_length_113297_cov_5.77445	find gene functions from 54 unique genomic loci; Figure 3 Acr Seq	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
17	CosteaPI_2017__SID713A062-11-0-0_NODE_38_length_192196_cov_4.03099	Figure 2, find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
18	CosteaPI_2017__SID713B051-11-0-0_NODE_14_length_298619_cov_7.20988	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
19	FengQ_2015__SID31872_NODE_2_length_392843_cov_5.93617	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
20	FengQ_2015__SID530258_NODE_5_length_350476_cov_17.595	find gene functions from 54 unique genomic loci; Figure 3 Acr Seq	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
21	FengQ_2015__SID530373_NODE_21_length_272157_cov_9.73468	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
22	HeQ_2017__SZAXPI029561-52_NODE_1_length_502752_cov_8.09488	find gene functions from 54 unique genomic loci; Figure 3 Acr Seq	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
23	HeQ_2017__SZAXPI029575-90_NODE_229_length_94918_cov_3.79903	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
24	KarlssonFH_2013__S463_NODE_1_length_570037_cov_16.3973	Figure 2, find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
25	LiJ_2014__O2.UC12-1_NODE_323_length_49995_cov_5.04395	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
26	LiJ_2014__V1.FI02_NODE_274_length_84286_cov_3.49253	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
27	LiJ_2017__H1M413815_NODE_71_length_81514_cov_18.301	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
28	LiJ_2017__H2M514909_NODE_68_length_69076_cov_10.283	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
29	LiuW_2016__SRR3992969_NODE_1149_length_18999_cov_8.45033	find gene functions from 54 unique genomic loci; Figure 3 Acr Seq	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html

30	LiuW_2016__SRR3992984_NODE_127_length_61384_cov_18.0593	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
31	LiuW_2016__SRR3993014_NODE_8_length_143441_cov_89.3981	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
32	QinJ_2012__NOM001_NODE_179_length_28679_cov_2.87521	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
33	QinJ_2012__T2D-050_NODE_25_length_192521_cov_10.1129	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
34	VatanenT_2016__G78791_NODE_43_length_22491_cov_6.98654	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
35	XieH_2016__YSZC12003_35392_NODE_87_length_196476_cov_13.3023	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
36	XieH_2016__YSZC12003_35563_NODE_11_length_353850_cov_18.1068	Figure 2, find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
37	XieH_2016__YSZC12003_36005_NODE_238_length_99923_cov_5.91259	Figure 2, find gene functions from 54 unique genomic loci, source of orf1-encoding phage genome; Figure 3 Acr Seq	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
38	XieH_2016__YSZC12003_36794_NODE_1_length_781521_cov_10.2961	find gene functions from 54 unique genomic loci, source of orf1-encoding phage genome; Figure 3 Acr Seq	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
39	XieH_2016__YSZC12003_37133_NODE_3_length_676817_cov_24.9073	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
40	XieH_2016__YSZC12003_37322_NODE_5_length_601737_cov_115.712	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
41	XieH_2016__YSZC12003_37399_NODE_3_length_598430_cov_49.9887	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
42	XieH_2016__YSZC12003_37878_NODE_8_length_402183_cov_76.149	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
43	YuJ_2015__SZAXPI003435-11_NODE_1_length_772218_cov_11.5924	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
44	YuJ_2015__SZAXPI015230-16_NODE_32_length_174349_cov_17.3543	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
45	ZeeviD_2015__PNP_DietIntervention_11_NODE_16_le	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html

	ngth_97163_cov_10.0000_ID_22997								
46	ZeeviD_2015__PNP_Main_234_NODE_10_length_202229_cov_13.9987_ID_180729	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
47	ZeeviD_2015__PNP_Main_294_NODE_20_length_208110_cov_20.9981_ID_106095	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
48	ZeeviD_2015__PNP_Main_390_NODE_33_length_137723_cov_10.9985_ID_46475	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
49	ZeeviD_2015__PNP_Main_578_NODE_20_length_138741_cov_8.9988_ID_132563	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
50	ZeeviD_2015__PNP_Main_741_NODE_13_length_214417_cov_12.0572_ID_91679	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
51	ZellerG_2014__CCIS03857607ST-4-0_NODE_542_length_35291_cov_2.7674	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
52	ZellerG_2014__CCIS22958137ST-20-0_NODE_40_length_181493_cov_7.91373	find gene functions from 54 unique genomic loci, source of orf1-encoding phage genome; Figure 3 Acr Seq	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
53	XieH_2016__YSZC12003_35635_NODE_109_length_156568_cov_5.12141	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
54	ZeeviD_2015__PNP_Main_85_NODE_182_length_52997_cov_7.0000_ID_133080	find gene functions from 54 unique genomic loci	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
55	4303_HeQ_2017__SZAXPI029570-85__bin.1_NODE_2_length_608092_cov_26.3259	Figure 2	n/a	Yes	No	HeQ_2017__SZAXPI029570-85__bin.1	4303	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
56	4303_CosteaPI_2017__SID713B074-11-90-0__bin.57_NODE_18_length_238289_cov_5.37382	Figure 2, source of orf1-encoding phage genome; Figure 3 Acr Seq	n/a	Yes	No	CosteaPI_2017__SID713B074-11-90-0__bin.57	4303	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
57	Clostridiales_bacterium_isolate_CIM:MAG_317_1 contig_8085	Figure 2, source of orf1-encoding phage genome	QALM01000002.1	No	No	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	NCBI Genbank
58	TPA_asm:_Ruminococcaceae_bacterium_isolate_UBA8277_contig_226	Figure 2; Figure 3 Acr Seq	DPDR010001 0.1	No	No	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	NCBI Genbank
59	KarlssonFH_2013__S424_NODE_2_length_526279_cov_9.22761	source of orf1-encoding phage genome	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html

60	XieH_2016__YSZC12003_36696_NODE_1_length_776477_cov_39.8546	source of orf1-encoding phage genome, AcrIIA17 encoding phage (figure S8)	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
61	XieH_2016__YSZC12003_37308R1_NODE_3_length_717276_cov_26.9646	source of orf1-encoding phage genome	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
62	ZellerG_2014__CCIS88007743ST-4-0_NODE_31_length_210910_cov_8.07406	source of orf1-encoding phage genome	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
63	4303_QinN_2014__LD-22__bin.75_NODE_22_length_329763_cov_10.7401	Figure 3 Acr Seq	n/a	Yes	No	QinN_2014__LD-22__bin.75	4303	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
64	ZellerG_2014__CCMD25963797ST-21-0_NODE_9_length_356111_cov_10.1715	Figure 3 Acr Seq	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
65	ZellerG_2014__CCIS41222843ST-4-0_NODE_17_length_267133_cov_14.7383	Figure 3 Acr Seq	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
66	FengQ_2015__SID530168_NODE_20_length_224404_cov_6.02914	Figure 3 Acr Seq	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
67	FengQ_2015__SID530041_NODE_7_length_421742_cov_9.32571	Figure 3 Acr Seq	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html
68	FengQ_2015__SID31223_NODE_13_length_228767_cov_7.50553	Figure 3 Acr Seq	n/a	No	Yes	n/a	n/a	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__CAG-217 (inferred)	http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html

Supplemental Table 4. Plasmids used in this study. Supplemental Table S5 indicates genes expressed from pZE21_tetR.

Plasmid	crRNA promoter, sequence (5'-3')	Notes	Refs	Purpose
pZE21_tetR	n/a	Contains tetR behind pLac promoter for inducible expression of candidate Acrs. Targeted by crRNA_A; PAM = AGG.	(Forsberg et al., 2019)	Expressing genes to test *in vivo* anti-CRISPR activity
pSpyCas9 _crA	pJ23100, GTTCATTCAGGGCACCGGAC	Arabinose-inducible SpyCas9 with pZE21 targeting pZE21_tetR	(Forsberg et al., 2019)	Target pZE21_tetR for elimination with SpyCas9
pSpyCas9 _crMu	pJ23100, GTAATACTTGTCCCGCAAAG	Mu-targeting spacer for phage Mu immunity testing. Otherwise identical to pSpyCas9_crA	(Forsberg et al., 2019)	Phage Mu immunity testing
pSpyCa9_crNT	pJ23100, GAACGAAAAGCTGCGCCGGG	non-targeting spacer used as control. Otherwise identical to pSpyCas9_crA	(Forsberg et al., 2019)	Phage Mu immunity testing, Western blots
pCloDF13_GFP	pJ23100, GAACGAAAAGCTGCGCCGGG	*eGFP* gene replaces *spyCas9* in pSpyCas9_crA		Measure generic protein expression from pSpyCas9 expression vector
pIDTsmart	n/a	Plasmid used for *in-vitro* SpyCas9 digestion. Sequence available at: https://www.idtdna.com/pages/products/genes-and-gene-fragments/custom-gene-synthesis		Plasmid template for *in-vitro* nuclease reactions

Supplemental Table 5. Gene sequences used in this study.

Gene Name	Sequence	Notes
acrIIA22wt	ATGGTAGTAGAAGAGACGCGGGATTTAGCCGAAACTGCGGATTGTGTAGTGATCGAAGCCATTTTAGTGGATGACGG ATTGCGTTACAGACAGCTTTCTGTCGGCATCAAAGACGAAAACGGCGACATTATTCGTATCGTCCCTATTTCAACCGTT *CTGATC*TAG	The italicized six base pairs were deleted in the ∆2aa truncation mutant
acrIIA22-null	ATGGTAGTAGAAGAGACGCGGGATTTAGCCGAAACTGCGGATTGTGTAGTGATCGAAGCCATTT**A**AGTGGATGACGG ATTGCGTTACAGACAGCTTTCTGTCGGCATCAAAGACGAAAACGGCGACATTATTCGTATCGTCCCTATTTCAACCGTT CTGATCTAG	Mutation to introduce early stop codon in bold, underline
acrIIA22a	ATGGTCATAGAAGAGACGCGGGATTTAGCTGAAACTGCGGATTGTGTAGTGATCGAAGCCATTTTAGTGGATGACGGA TTGCGTTACAAACAGCTTTCCGTCGGCATCAAAGACGAAAACGGTGACATTATTCGTATCGTCCCTATTTCAACCGTTC TGATCTAG	Same amino acid sequence as NCBI protein CDB51368.1
acrIIA22b	ATGATTGTGGAAGATACCAAAGATTTGGTTGAAACTGCGGACTATGTGATCATCGAAGCTGTTTTAGTGGATGATGGAT TGCGTTACAAACAACTTTCTGTTGGCATTAAAGCCAAAAATGGTGACATTATCCGCATAATTCCAATATCGACAATGCT GATGTAA	
acrIIA22c	ATGAAAATGATTGTGGAAGATACGAAAGATCTGGTAGAAACGGACGATTATGTAATCATTGAAGCGACTTTGTCAGAG GGCGATTTGTTGTTTGTGCAAATTGCCGTGGGCATTCGCAACGAAGTGGGCGACATTGTTCGTATTATTCCCATTTCC ACCAACCCAATCTAA	Same amino acid sequence as NCBI protein CDB51757.1
purF	ATGTTCGATAGTTTGCACGAGGAATGCGGTGTTTTCGGCGTATTTGAAAATCAGACCACTACGGTGGCCCAGACGGC GTATCTGGCTCTGTTTGCCTTGCAGCACAGAGGGCAGGAGAGTTGCGGCATTGCCGTGAATGACGACGGCGTGTTTC GCCACCATCGGGGCGACGGACTGGTGCCGGATGTGTTTAGCAAGGAGCAGCTGGCTGCCCTGGGTACAGGTAATAT GGCCATCGGTCATGTGCGCTACTCCACCACCGGCGGCAAAAACGCCAACAATATTCAGCCCCTGGTCATTCGCCATA TTAAGGGTAATTTGGCGGTGGCACATAACGGCAATTTGGTAAACGCCCCGGAGCTGCGCCGCCAGTTTGAGCTGAAG GGCGCCATTTTTCACGGCACATCGGACACCGAGTCCATTGCCTATTCTATTGTAGAGGAGCGCCTGCACAGTAAGAG CACGGAAGAGGCCATCGAAAAAATCATGCCCCGGCTGCAAGGGGCATTCTCTTGCGTGGTGATGACTGCCACCAAAC TCATTGCGTTTCGTGACCCCAACGGCTTTCGGCCTCTTTGCCTGGGTAAGACTGCGGACGATGCTTATGTGGTGGCG TCGGAGAGTTGTGCGCTGGATTCCATCGGCGCCCACTTTGTGCGGAATATTGCTCCCGGCGAGATCGTTGTGATCAG CAAGGATGGCGTGCGCTCTATTACCACCCATTGCGGCGGACTACGCCACATTTGTGTGTTTGAGTACATCTATTTTGC TCGGCCGGACAGTGTGATTGAGGGCGTGTCTGTGCAGCACGCCAGAATGCGGGCCGGTGCGTACCTGGCGAAGGA ACACCCGGTAGACGCGGATATTGTCATCGGCGTGCCGGACAGCGGCCTGGACGCCGCCTTGGGCTATGCCCAGGAG AGCGGCATTCCTTACGGTATTGGATTTATCAAGAACCGCTACATCGGCCGCAGCTTTATTCAGCCTACCCAAGGTCAG CGTGAGGACGCGGTGAAGATCAAGCTGAATGTACTGCGAGAGAATATCAAGGGCAAGCGGGTGGTGATGATCGATG ACTCTATTGTTCGCGGCACCACCAGCGCTCGGATTGTCAGCCTGCTGCGAGAGGCCGGCGCCACCGAGGTGCATAT GCGGGTTTCTGCCCCTCCGTTTCGGCATCCTTGCTTCTTTGGAACGGACATTGATAGCGAAGAAAACCTGATTGCATG CAAATTTCACGAAATTTCTGAAATTTCTTGCCAATTAGGGGTTGACAGCTTGGGGTATCTTAGTGTAAAATCTACTCAC GAACTTGCGAAGGAGTCCGGATTCGATTTTTGCGACGGTTGCTTCACCGGCCATTATCCCATCCCCACCCCGAAGCA ACAGTCCAAGGATAAGTTTGAGGAAAAGCTGAATCAGTTCTCCTCTTACTACCAGGTCTTGGATTAA	Flanks *acrIIA22*-encoding bacterial genomic islands. Used as bait to retrieve additional examples of this locus for genomic and evolutionary analyses.
radC	ATGCGTGCCGCTTATCTGCAAGGCGGCGGCGACGCTATGCCGGACCACCAGTTGCTGGAATTGCTGCTGTCCATCAG CATTCCCCGCAGAGATGTAAAGCCCATTGCCTATGCGCTCATTAACCGCTTCGGCTCGCTGGAGCAGGTGTTTGCCG CCGGCGCAGCAGATCTGCAACAAGTGCCGGGCGTCGGCGAACAGACCGCCGTACAGATTCTGCTGGTACGGGATCT GAACCGGCGGATCCATCAAAATCAAAACAAACCGGTCAAGCACCTGACAGATGCCACCCAGTCCTGCGCCTACTTTTC CAATCTGTTACGGGACAAAACCGCCGAGCAGGTGTACTTGGTCACCCTGGACGGCAGTGCCAAAATCCTGCAAACCC ACGCCGTAGGCAGCGGCAGCGTCAACCTGGCCTCTGTGGATCAGCGCACTTTGATGGAACATATTCTGCGAGACAAC GCCAACGCTGTTATGCTGGCACACAACCATCCCGGCGGCAAGGCCCAGCCCTCTGCGCAGGATCTGGAATTCACCAT TCGTCTGCTTTCCATTCTGCGTTCCATTCATGTGCAGCTGCTGGATCATATTATCGTCAGTCCTACCGACACCTACTCC ATGCGCAGCGACCCGGAGTACGGCAGCTTCTTCACCGTCAAATAA	Flanks *acrIIA22*-encoding bacterial genomic islands. Used as bait to retrieve additional examples of this locus for genomic and evolutionary analyses.
acrIIA4	ATGAATATTAACGATTTGATCCGTGAGATTAAGAATAAGGATTATACTGTCAAATTGTCCGGGACAGATTCCAATTCTAT TACACAATTAATCATCCGTGTGAATAACGATGGTAATGAGTATGTCATCTCTGAATCAGAAAACGAGAGCATCGTAGAA AAGTTCATCAGTGCCTTCAAGAACGGGTGGAACCAAGAGTATGAAGATGAGGAGGAATTTTACAATGATATGCAGACA ATTACGCTTAAATCAGAATTGAATTAA	Discovered by (Rauch et al., 2017)

References for Supplemental Tables.

An, S.Y., Ka, D., Kim, I., Kim, E.-H., Kim, N.-K., Bae, E., and Suh, J.-Y. (2020). Intrinsic disorder is essential for Cas9 inhibition of anti-CRISPR AcrIIA5. Nucleic Acids Research *48*, 7584-7594.

Athukoralage, J.S., McMahon, S.A., Zhang, C., Gruschow, S., Graham, S., Krupovic, M., Whitaker, R.J., Gloster, T.M., and White, M.F. (2020). An anti-CRISPR viral ring nuclease subverts type III CRISPR immunity. Nature *577*, 572-575.

Bhoobalan-Chitty, Y., Johansen, T.B., Di Cianni, N., and Peng, X. (2019). Inhibition of Type III CRISPR-Cas Immunity by an Archaeal Virus-Encoded Anti-CRISPR Protein. Cell *179*, 448-458 e411.

Bondy-Denomy, J., Garcia, B., Strum, S., Du, M., Rollins, M.F., Hidalgo-Reyes, Y., Wiedenheft, B., Maxwell, K.L., and Davidson, A.R. (2015). Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins. Nature *526*, 136-139.

Bondy-Denomy, J., Pawluk, A., Maxwell, K.L., and Davidson, A.R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature *493*, 429-432.

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.* (2017). Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex. Cell *169*, 47-57 e11.

Dong, Guo, M., Wang, S., Zhu, Y., Wang, S., Xiong, Z., Yang, J., Xu, Z., and Huang, Z. (2017). Structural basis of CRISPR-SpyCas9 inhibition by an anti-CRISPR protein. Nature *546*, 436-439.

Eitzinger, S., Asif, A., Watters, K.E., Iavarone, A.T., Knott, G.J., Doudna, J.A., and Minhas, F. (2020). Machine learning predicts new anti-CRISPR proteins. Nucleic Acids Res *48*, 4698-4708.

Forsberg, K.J., Bhatt, I.V., Schmidtke, D.T., Javanmardi, K., Dillard, K.E., Stoddard, B.L., Finkelstein, I.J., Kaiser, B.K., and Malik, H.S. (2019). Functional metagenomics-guided discovery of potent Cas9 inhibitors in the human microbiome. Elife *8*.

Fuchsbauer, O., Swuec, P., Zimberger, C., Amigues, B., Levesque, S., Agudelo, D., Duringer, A., Chaves-Sanjuan, A., Spinelli, S., Rousseau, G.M.*, et al.* (2019). Cas9 Allosteric Inhibition by the Anti-CRISPR Protein AcrIIA6. Mol Cell *76*, 922-937 e927.

Garcia, B., Lee, J., Edraki, A., Hidalgo-Reyes, Y., Erwood, S., Mir, A., Trost, C.N., Seroussi, U., Stanley, S.Y., Cohn, R.D.*, et al.* (2019). Anti-CRISPR AcrIIA5 Potently Inhibits All Cas9 Homologs Used for Genome Editing. Cell Rep *29*, 1739-1746 e1735.

Guo, T.W., Bartesaghi, A., Yang, H., Falconieri, V., Rao, P., Merk, A., Eng, E.T., Raczkowski, A.M., Fox, T., Earl, L.A.*, et al.* (2017). Cryo-EM Structures Reveal Mechanism and Inhibition of DNA Targeting by a CRISPR-Cas Surveillance Complex. Cell *171*, 414-426 e412.

He, F., Bhoobalan-Chitty, Y., Van, L.B., Kjeldsen, A.L., Dedola, M., Makarova, K.S., Koonin, E.V., Brodersen, D.E., and Peng, X. (2018). Anti-CRISPR proteins encoded by archaeal lytic viruses inhibit subtype I-D immunity. Nat Microbiol *3*, 461-469.

Hirschi, M., Lu, W.T., Santiago-Frangos, A., Wilkinson, R., Golden, S.M., Davidson, A.R., Lander, G.C., and Wiedenheft, B. (2020). AcrIF9 tethers non-sequence specific dsDNA to the CRISPR RNA-guided surveillance complex. Nat Commun *11*, 2730.

Janowski, R., and Niessing, D. (2020). The large family of PC4-like domains - similar folds and functions throughout all kingdoms of life. RNA Biol *17*, 1228-1238.

Jiang, F., Liu, J.J., Osuna, B.A., Xu, M., Berry, J.D., Rauch, B.J., Nogales, E., Bondy-Denomy, J., and Doudna, J.A. (2019). Temperature-Responsive Competitive Inhibition of CRISPR-Cas9. Mol Cell *73*, 601-610 e605.

Knott, G.J., Cress, B.F., Liu, J.J., Thornton, B.W., Lew, R.J., Al-Shayeb, B., Rosenberg, D.J., Hammel, M., Adler, B.A., Lobba, M.J.*, et al.* (2019a). Structural basis for AcrVA4 inhibition of specific CRISPR-Cas12a. Elife *8*.

Knott, G.J., Thornton, B.W., Lobba, M.J., Liu, J.J., Al-Shayeb, B., Watters, K.E., and Doudna, J.A. (2019b). Broad-spectrum enzymatic inhibition of CRISPR-Cas12a. Nat Struct Mol Biol *26*, 315-321.

Lee, J., Mir, A., Edraki, A., Garcia, B., Amrani, N., Lou, H.E., Gainetdinov, I., Pawluk, A., Ibraheim, R., Gao, X.D.*, et al.* (2018). Potent Cas9 Inhibition in Bacterial and Human Cells by AcrIIC4 and AcrIIC5 Anti-CRISPR Proteins. mBio *9*.

Leon, L.M., Park, A.E., Borges, A.L., Zhang, J.Y., and Bondy-Denomy, J. (2020). Mobile element warfare via CRISPR and anti-CRISPR in <em>Pseudomonas aeruginosa</em&gt. bioRxiv, 2020.2006.2015.151498.

Lin, P., Qin, S., Pu, Q., Wang, Z., Wu, Q., Gao, P., Schettler, J., Guo, K., Li, R., Li, G.*, et al.* (2020). CRISPR-Cas13 Inhibitors Block RNA Editing in Bacteria and Mammalian Cells. Mol Cell *78*, 850-861 e855.

Liu, L., Yin, M., Wang, M., and Wang, Y. (2019). Phage AcrIIA2 DNA Mimicry: Structural Basis of the CRISPR and Anti-CRISPR Arms Race. Mol Cell *73*, 611-620 e613.

Mahendra, C., Christie, K.A., Osuna, B.A., Pinilla-Redondo, R., Kleinstiver, B.P., and Bondy-Denomy, J. (2020). Broad-spectrum anti-CRISPR proteins facilitate horizontal gene transfer. Nat Microbiol *5*, 620-629.

Marino, N.D., Zhang, J.Y., Borges, A.L., Sousa, A.A., Leon, L.M., Rauch, B.J., Walton, R.T., Berry, J.D., Joung, J.K., Kleinstiver, B.P.*, et al.* (2018). Discovery of widespread Type I and Type V CRISPR-Cas inhibitors. Science.

Meeske, A.J., Jia, N., Cassel, A.K., Kozlova, A., Liao, J., Wiedmann, M., Patel, D.J., and Marraffini, L.A. (2020). A phage-encoded anti-CRISPR enables complete evasion of type VI-A CRISPR-Cas immunity. Science *369*, 54-59.

Osuna, B.A., Karambelkar, S., Mahendra, C., Christie, K.A., Garcia, B., Davidson, A.R., Kleinstiver, B.P., Kilcher, S., and Bondy-Denomy, J. (2020). Listeria Phages Induce Cas9 Degradation to Protect Lysogenic Genomes. Cell Host Microbe *28*, 31-40 e39.

Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P.*, et al.* (2019). Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. Cell *176*, 649-662 e620.

Pawluk, A., Amrani, N., Zhang, Y., Garcia, B., Hidalgo-Reyes, Y., Lee, J., Edraki, A., Shah, M., Sontheimer, E.J., Maxwell, K.L.*, et al.* (2016). Naturally Occurring Off-Switches for CRISPR-Cas9. Cell *167*, 1829-1838 e1829.

Pawluk, A., Bondy-Denomy, J., Cheung, V.H., Maxwell, K.L., and Davidson, A.R. (2014). A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of Pseudomonas aeruginosa. MBio *5*, e00896.

Pawluk, A., Shah, M., Mejdani, M., Calmettes, C., Moraes, T.F., Davidson, A.R., and Maxwell, K.L. (2017). Disabling a Type I-E CRISPR-Cas Nuclease with a Bacteriophage-Encoded Anti-CRISPR Protein. MBio *8*.

Pinilla-Redondo, R., Shehreen, S., Marino, N.D., Fagerlund, R.D., Brown, C.M., Sørensen, S.J., Fineran, P.C., and Bondy-Denomy, J. (2020). Discovery of multiple anti-CRISPRs uncovers anti-defense gene clustering in mobile genetic elements. bioRxiv, 2020.2005.2022.110304.

Rauch, B.J., Silvis, M.R., Hultquist, J.F., Waters, C.S., McGregor, M.J., Krogan, N.J., and Bondy-Denomy, J. (2017). Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. Cell *168*, 150-158 e110.

Shin, J., Jiang, F., Liu, J.-J., Bray, N.L., Rauch, B.J., Baik, S.H., Nogales, E., Bondy-Denomy, J., Corn, J.E., and Doudna, J.A. (2017). Disabling Cas9 by an anti-CRISPR DNA mimic. Science Advances *3*.

Song, G., Zhang, F., Zhang, X., Gao, X., Zhu, X., Fan, D., and Tian, Y. (2019). AcrIIA5 Inhibits a Broad Range of Cas9 Orthologs by Preventing DNA Target Cleavage. Cell Rep *29*, 2579-2589 e2574.

Stanley, S.Y. (2018). An Investigation of Bacteriophage Anti-CRISPR and Anti-CRISPR Associated Proteins. In Department of Molecular Genetics (http://hdl.handle.net/1807/97883: University of Toronto), pp. 120.

Uribe, R.V., van der Helm, E., Misiakou, M.A., Lee, S.W., Kol, S., and Sommer, M.O.A. (2019). Discovery and Characterization of Cas9 Inhibitors Disseminated across Seven Bacterial Phyla. Cell Host Microbe *25*, 233-241 e235.

Wang, J., Ma, J., Cheng, Z., Meng, X., You, L., Wang, M., Zhang, X., and Wang, Y. (2016a). A CRISPR evolutionary arms race: structural insights into viral anti-CRISPR/Cas responses. Cell Res *26*, 1165-1168.

Wang, X., Yao, D., Xu, J.G., Li, A.R., Xu, J., Fu, P., Zhou, Y., and Zhu, Y. (2016b). Structural basis of Cas3 inhibition by the bacteriophage protein AcrF3. Nat Struct Mol Biol *23*, 868-870.

Watters, K.E., Fellmann, C., Bai, H.B., Ren, S.M., and Doudna, J.A. (2018). Systematic discovery of natural CRISPR-Cas12a inhibitors. Science.

Watters, K.E., Shivram, H., Fellmann, C., Lew, R.J., McMahon, B., and Doudna, J.A. (2020). Potent CRISPR-Cas9 inhibitors from Staphylococcus genomes. Proc Natl Acad Sci U S A *117*, 6531-6539.

Yang, H., and Patel, D.J. (2017). Inhibition Mechanism of an Anti-CRISPR Suppressor AcrIIA4 Targeting SpyCas9. Mol Cell *67*, 117-127 e115.

Zhang, H., Li, Z., Daczkowski, C.M., Gabel, C., Mesecar, A.D., and Chang, L. (2019). Structural Basis for the Inhibition of CRISPR-Cas12a by Anti-CRISPR Proteins. Cell Host Microbe *25*, 815-826 e814.

Zhang, K., Wang, S., Li, S., Zhu, Y., Pintilie, G.D., Mou, T.C., Schmid, M.F., Huang, Z., and Chiu, W. (2020). Inhibition mechanisms of AcrF9, AcrF8, and AcrF6 against type I-F CRISPR-Cas complex revealed by cryo-EM. Proc Natl Acad Sci U S A *117*, 7176-7182.