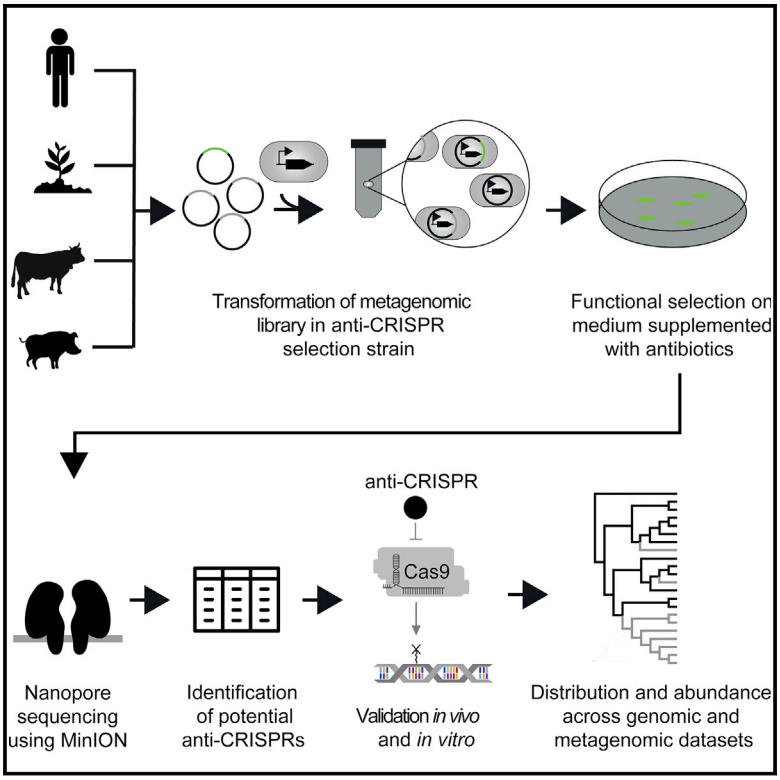
Short Article

Cell Host & Microbe

Discovery and Characterization of Cas9 Inhibitors Disseminated across Seven Bacterial Phyla

Graphical Abstract



Highlights

- A genetic circuit designed to functionally select anti-CRISPR genes from metagenomes
- Four protein families, AcrIIA7–10, that inhibit Cas9 *in vivo* and *in vitro* were identified
- AcrIIA7–10 are widely distributed across seven bacterial phyla
- Dissemination patterns of AcrIIA7–10 suggest interphylum horizontal gene transfer

Authors

Ruben V. Uribe, Eric van der Helm, Maria-Anna Misiakou, Sang-Woo Lee, Stefan Kol, Morten O.A. Sommer

Correspondence

msom@bio.dtu.dk

In Brief

Uribe et al. developed a high-throughput approach to screen for type II-A CRISPR-Cas inhibitors in metagenomic libraries based on their functional activity rather than sequence homology or genetic context. This approach led to the discovery of four novel anti-CRISPR proteins, homologs of which are widely distributed across seven bacterial phyla.

Uribe et al., 2019, Cell Host & Microbe *25*, 233–241
February 13, 2019
https://doi.org/10.1016/j.chom.2019.01.003

CellPress

Cell Host & Microbe
Short Article

Discovery and Characterization of Cas9 Inhibitors Disseminated across Seven Bacterial Phyla

Ruben V. Uribe,[1,2] Eric van der Helm,[1,2] Maria-Anna Misiakou,[1,2] Sang-Woo Lee,[1] Stefan Kol,[1] and Morten O.A. Sommer[1,3,*]
[1]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby 2800, Denmark
[2]These authors contributed equally
[3]Lead Contact
*Correspondence: msom@bio.dtu.dk
https://doi.org/10.1016/j.chom.2019.01.003

SUMMARY

CRISPR-Cas systems in bacteria and archaea provide immunity against bacteriophages and plasmids. To overcome CRISPR immunity, phages have acquired anti-CRISPR genes that reduce CRISPR-Cas activity. Using a synthetic genetic circuit, we developed a high-throughput approach to discover anti-CRISPR genes from metagenomic libraries based on their functional activity rather than sequence homology or genetic context. We identified 11 DNA fragments from soil, animal, and human metagenomes that circumvent *Streptococcus pyogenes* Cas9 activity in our selection strain. Further *in vivo* and *in vitro* characterization of a subset of these hits validated the activity of four anti-CRISPRs. Notably, homologs of some of these anti-CRISPRs were detected in seven different phyla, namely Firmicutes, Proteobacteria, Bacteroidetes, Actinobacteria, Cyanobacteria, Spirochaetes, and Balneolaeota, and have high sequence identity suggesting recent horizontal gene transfer. Thus, anti-CRISPRs against type II-A CRISPR-Cas systems are widely distributed across bacterial phyla, suggesting a more complex ecological role than previously appreciated.

INTRODUCTION

Bacteria in microbial communities are constantly exposed to several threats, from biochemical warfare among microbes to bacteriophage predation. Phages are often more abundant than bacteria (Suttle, 2007), playing a major role in bacterial population dynamics (Levin and Udekwu, 2010). In some environments, phages kill approximately 20% of the bacterial biomass per day (Suttle, 2007), leading to an evolutionary arms race between bacteria and phages. This ongoing arms race between bacteria and phages has resulted in the evolution of diverse mechanisms to avoid or promote infection (Stern and Sorek, 2011; Samson et al., 2013; Doron et al., 2018). Bacteria have evolved multiple mechanisms to avoid phage infection, including CRISPR-Cas systems that are adaptive immune mechanisms acting against foreign DNA or RNA elements in a sequence-specific manner (Makarova et al., 2015; Koonin et al., 2017). CRISPR-Cas systems are diverse, generally being divided in two classes that are further subdivided into six types and several subtypes based on the structure and organization of their effector module (Makarova et al., 2015; Koonin et al., 2017).

To overcome CRISPR-Cas systems, phages may mutate or delete their CRISPR target sites. In addition, phages have acquired inhibitory proteins termed anti-CRISPRs (ACRs) that interfere with CRISPR-Cas activity. ACRs were first identified for types I–F and I–E CRISPR-Cas systems (Bondy-Denomy et al., 2013; Pawluk et al., 2014). These ACRs are widespread across the phylum Proteobacteria, and they are present in the genome of multiple prophages, pathogenicity islands, and other mobile elements, suggesting that ACR genes are prone to horizontal gene transfer (HGT) (Bondy-Denomy et al., 2013; Pawluk et al., 2014, 2018). Subsequent computational mining identified ACR genes against type II-C Cas9 from *Neisseria meningitidis* (Pawluk et al., 2016) and type II-A Cas9 from *Streptococcus pyogenes* (SpCas9) (Rauch et al., 2017). Type II systems are of particular interest because nucleases such as SpCas9 have been exploited for precise and programmable gene editing with substantial impact on life sciences (Doudna and Charpentier, 2014). More recently two ACRs against SpCas9 were identified by cloning and testing multiple genes from mostly virulent phages that were able to escape CRISPR-based immunity from *Streptococcus thermophillus*, highlighting the importance of functional screening strategies for the identification of ACR genes (Hynes et al., 2017; Hynes et al., 2018). Given the abundance of CRISPR-Cas systems in bacteria (Burstein et al., 2016), as well as the abundance of uncharacterized phages (Simmonds et al., 2017) and other mobile elements, it is likely that we currently have elucidated only a minute proportion of ACR strategies in the environment. However, computational discovery guided by genomic context or homology is limited by the availability of reference ACR families.

To address this problem, we designed an *Escherichia coli* strain that harbors a genetic circuit for selection of genetically encoded ACR activity (Figures 1A and 1B). Using the synthetic genetic circuit, we developed a high-throughput approach to discover ACR genes from metagenomic libraries based on their functional activity rather than sequence homology or genetic context.

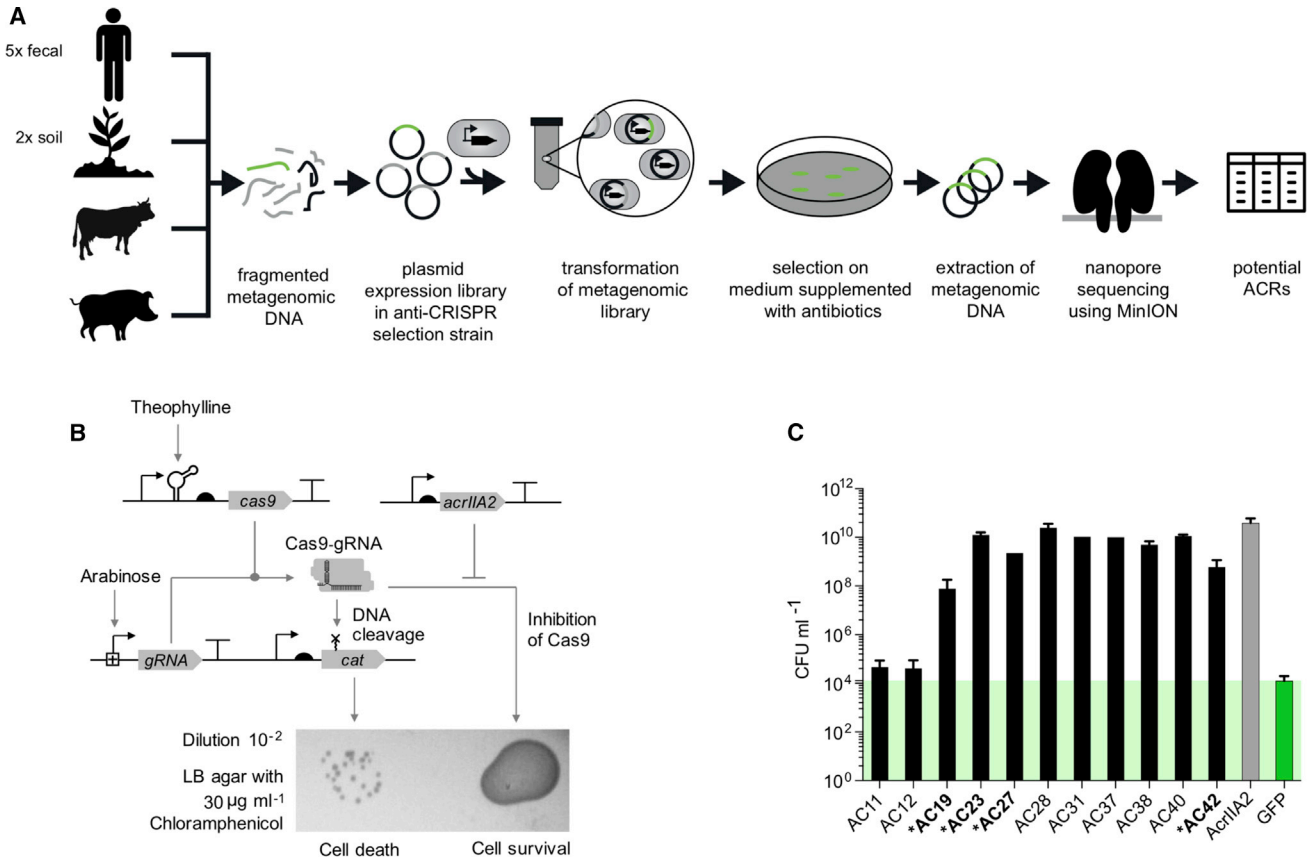


Figure 1. Anti-CRISPR Functional Selection System

(A) Workflow for identification of ACR proteins from metagenomic libraries.

(B) Schematic overview of the selection system; Cas9 and gRNA expression are induced by adding 2 mM theophylline and 1% arabinose, respectively.

(C) *In vivo* validation of metagenomic inserts containing ACR candidates retransformed into a fresh selection strain. Average and standard deviation of colony-forming units per milliliter (CFUs mL^{-1}) were calculated from biological triplicates. (*) Inserts that contain an open reading frame with putative ACR activity *in vitro*.

RESULTS

Anti-CRISPR Genes Identified Using a Functional Selection System

To identify ACRs from metagenomic libraries, we designed a genetic circuit that coupled ACR expression to survival under antibiotic selection. The genetic circuit was constructed by cloning SpCas9 with a cognate gRNA targeting a plasmid-borne chloramphenicol resistance gene. In the absence of ACR activity, this circuit yields a chloramphenicol-sensitive strain. However, introduction of a DNA fragment encoding and expressing a gene product that prevents SpCas9-mediated loss of chloramphenicol resistance would render the host strain resistant to chloramphenicol. In this way, metagenomic libraries can be rapidly selected for putative ACR genes (Figure 1A).

We first tested the genetic circuit using a previously discovered ACR, AcrIIA2 (Rauch et al., 2017), as a positive control and green fluorescent protein (GFP) as a negative control. Consistent with our expectation, bacterial cells equipped with the genetic circuit expressing AcrIIA2 were resistant toward 30 $\mu g/mL^{-1}$ of chloramphenicol, whereas cells expressing GFP were susceptible to the 30 $\mu g/mL^{-1}$ of chloramphenicol (Figures 1B and S1). When plating out cells harboring the genetic circuit and expressing GFP in high numbers, we observed escapers that evaded our selection system at a frequency of $\sim 10^{-4}$ colony-forming unit (CFU) mL^{-1} consistent with previous studies of CRISPR-Cas-based selection systems (Citorik et al., 2014; Caliando and Voigt, 2015; Cui and Bikard, 2016; Lauritsen et al., 2017).

To identify ACRs, we transformed nine different metagenomic libraries (Table S1) derived from soil, pig gut, cow gut, and human gut into our selection strain. We then selected the libraries in inhibitory concentrations of chloramphenicol (30 $\mu g/mL^{-1}$) and collected the clones that appeared on the selective plates. Metagenomic inserts were extracted from pooled clones on selection plates, barcoded per library, and sequenced using nanopore technology to obtain full inserts (van der Helm et al., 2017). The resulting contigs were annotated with BLASTx and manually curated because metagenomic selections can yield hits that are not directly relevant to the mechanism of investigation. For example, the current selection platform also resulted in genes encoding functionalities related to theophylline and arabinose degradation and export (Figure S1; Data S1). Therefore, we selected the top 5 inserts with the highest amount of mapped nanopore reads, most of which were annotated as hypothetical proteins. Additionally, we selected 34 inserts that had

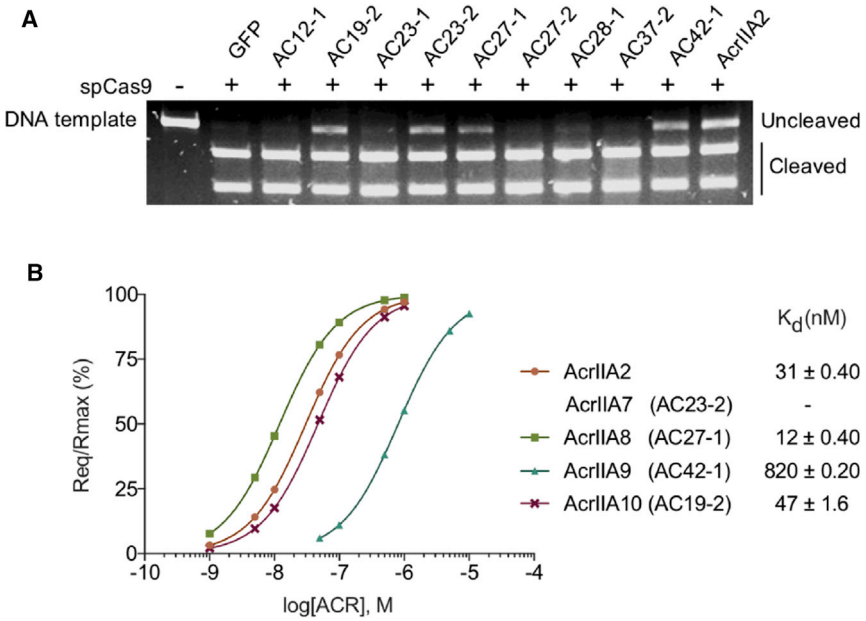


Figure 2. ACR Candidates' Activity and Binding against Cas9 *In Vitro*

(A) *In vitro* cleavage assay using linear double-stranded DNA template with the same 20 base pairs target sequence used in the *in vivo* assay. Presence of uncleaved DNA template indicates inhibition of SpCas9 activity mediated by an ACR. (B) Binding strength of ACRs to biotinylated SpCas9 showed as K_d determined using biolayer interferometry.

bacteriophage-related or putative mobile element annotations (Table S1) because it is expected that ACRs could be in the neighboring regions of such elements. The 39 inserts were re-cloned and transformed into the selection strain to individually re-test their ACR activity and discard the possibility of them being escapers or false positives. 11 of the 39 inserts showed ACR activity above background in our *in vivo* selection assay (Figure 1C).

Putative ACR Genes Inhibit Cas9 *In Vitro*

From the 11 inserts that retained ACR activity, we expressed 16 individual open reading frames (ORFs) in *E. coli* and purified the resulting proteins in order to directly test their effect on SpCas9 activity. We successfully expressed and purified 9 of the 16 cloned ORFs. Using an *in vitro* DNA cleavage assay (Figure 2) (Pawluk et al., 2016; Dong et al., 2017), we tested whether the individual proteins encoded by the ORFs could inhibit cleavage of target DNA by the SpCas9:gRNA complex. The four proteins AC19-2, AC23-2, AC27-1, and AC42-1 showed clear ACR activity. The putative ACR proteins preserved the DNA template intact, demonstrating that the activity of the proteins inhibit specifically SpCas9-mediated DNA cleavage (Figure 2A). The remaining expressed proteins were not able to inhibit SpCas9 activity *in vitro* using this assay. This could be due to difficulties in functionally expressing the proteins in our expression system, limitations of our assay, or that these proteins are derived from false positives of our selection system.

To further characterize proteins AC19-2, AC23-2, AC27-1, and AC42-1 that inhibit SpCas9 activity in our *in vitro* DNA cleavage assay, we examined their direct binding to SpCas9 using biolayer interferometry. AC19-2, AC27-1, and AC42-1 showed binding affinity to Cas9:gRNA complex when Mg^{2+} was added to the running buffer (Figure 2B) (Anders et al., 2015). The observed K_d for AC19-2, AC27-1, and AC42-1 was 47, 12, and 820 nM, respectively. We did not detect binding between AC23-2 and SpCas9 using biolayer interferometry. Taken together, the results from our combined *in vivo* and *in vitro* assays show that these 4 proteins display anti-SpCas9 activity, and accordingly we renamed them: AcrIIA7 (AC23-2), AcrIIA8 (AC27-1), AcrIIA9 (AC42-1), and AcrIIA10 (AC19-2) (Figure 2B). The nucleotide and amino acid (aa) sequences of the 4 proteins can be found in Table S2. AcrIIA7 (103 aa), AcrIIA8 (105 aa), and AcrIIA9 (141 aa) are derived from human gut metagenomic libraries, whereas AcrIIA10 (109 aa) originates from a soil metagenomic library.

AcrIIA7–10 Distribution across Genomic and Metagenomic Datasets

To determine the potential origin of the identified AcrIIAs and investigate how widespread these protein families are in nature, we examined their genetic context and diversity in comparison to previously known AcrIIAs (Rauch et al., 2017; Hynes et al., 2017) (Figure 3A). We assessed the distribution of all AcrIIA homologs across metagenomic datasets. Specifically, we interrogated publicly available viromes hosted by MetaVir (Roux et al., 2011) and whole metagenome datasets available at NCBI (Figure 3B). Of the 485 MetaVir virome datasets examined, more than 600 hits to these AcrIIAs were identified (Table S2). Across both databases, using Position-Specific Iterative Basic Local Alignment Search Tool (PSI-BLAST) with an e-value cutoff of 10^{-5}, only AcrIIA7 and AcrIIA9 homologs were detected in datasets derived from the human gastrointestinal (GI) tract, consistent with the isolation source of these ACRs. AcrIIA7 stood out for its dominant presence across diverse environments, ranging from freshwater and deep-sea sediments to hypersaline and insect samples (Figure 3B). As for the previously known ACRs, only AcrIIA5 and AcrIIA6 discovered in *S. thermophilus* bacteriophages (Hynes et al., 2017, 2018) had homologs in human metagenomic datasets (Table S2). The analysis of the distribution of AcrIIAs across different metagenomic datasets suggests that some of the AcrIIA families discovered in this study are much more abundant and diversely distributed across multiple environments compared to previously characterized AcrIIAs.

Computational analysis of the distribution of the distinct AcrIIA gene families in reference genomes revealed a varying host range and, in some cases, overlap in taxonomic identity (Table S2). The diverse phylogenetic distribution of homologs of the

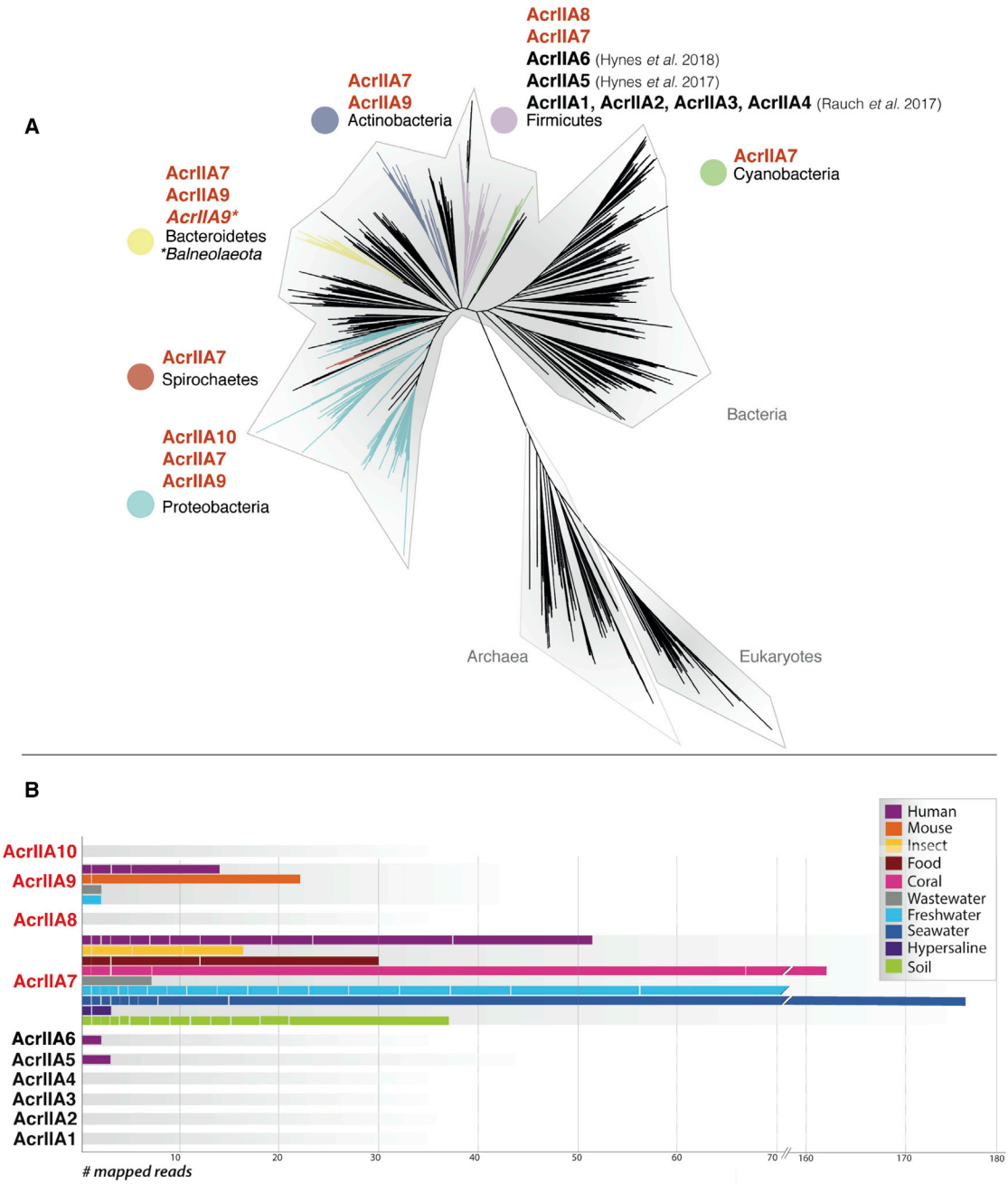


Figure 3. Distribution of AcrIIAs across Genomic and Metagenomic Datasets

(A) Hosts of all AcrIIAs known to date mapped to the tree of life (Hug et al., 2016). Gray shading refers to the domains of life (Bacteria, Archaea, and Eukaryotes). Colored tree branches correspond to all distinct bacterial phyla (NCBI Taxonomy) carrying AcrIIAs. The phylum Balneolaeota is marked in italics as it was recently separated from the Bacteroidetes phylum (Hug et al., 2016). Labels in bold above phyla names denote the AcrIIA's phylogenetic membership. Red-colored labels denote AcrIIAs identified by the current work; black-colored labels are previously known AcrIIAs.

(B) Quantification of AcrIIA homologs in publicly available viromes (MetaVir). The horizontal axis corresponds to the number of raw reads with homology to an AcrIIA and the vertical axis to the distinct virome datasets stacked by habitat and grouped by AcrIIA. Bars are assigned colors according to the type of habitat the datasets represent. For details on the origin of samples, accompanying publications, and MetaVir ID, see Table S2.

four AcrIIAs identified here substantially differs from the previously known AcrIIAs, which are all confined to Firmicutes genomes (Figure 3A). There were no homologs from the previously known ACRs or those identified here in the archaea domain (Figure 3A) as type II CRISPR-Cas systems have been exclusively found in bacteria (Shmakov et al., 2017). Furthermore, previously known AcrIIA1-6 only had homologs in a limited number of species in the retrieved reference genomes (Table S2). In the same

manner, a scarcity of homologs in reference genomes was observed for AcrIIA8 and AcrIIA10.

For AcrIIA8, we identified 52 homologs in 68 reference genomes (Table S2) of 38 unique species of Firmicutes, comprising about 1% of the 3,257 Firmicutes species (NCBI Genome List). One AcrIIA8 homolog was identified in a *Listeria monocytogenes* genome, which also harbors homologs of AcrIIA1, AcrIIA2, and AcrIIA4 (Rauch et al., 2017) (Table S2). AcrIIA8 homologs were absent in MetaVir datasets (Figure 3B); the flanking genes of AcrIIA8 homologs across diverse Firmicutes families are in conserved synteny containing genes for viral particle assembly (Figure S2), strongly suggesting a viral (prophage) origin for this protein. Localization near phage structural genes has been observed before for other ACRs (Bondy-Denomy et al., 2013; Rauch et al., 2017). Furthermore, conserved domains between ACRs and viral structural proteins suggest that in some cases, structural genes could be a potential evolutionary source for ACR genes (Stone et al., 2018). Interestingly, homology search of AcrIIA8 using PSI-BLAST and HHpred (Söding et al., 2005) indicates a phage head-tail adaptor function for this gene, which would suggest a common ancestry for these genes. The closest annotated homolog of AcrIIA8 is phage head-tail adapter protein from *Clostridium botulinum* (using blastp against the NCBI nr database). Although it is a statistically significant result (e-value: 4e−12), the aa sequence identity is 42%. Interestingly, the previously discovered AcrIIA6 has 86% aa identity to elongation factor G in *S. thermophilus* (Hynes et al., 2018). This suggests that some ACRs might have evolved from a protein family with a different initial function.

The soil-derived AcrIIA10 only had one close homolog (94% identity at nucleotide level), which was found in the genome of the soil isolate *Sinorhizobium sp.* GL28 (Proteobacteria) (Table S2).

Homologs to AcrIIA7 and AcrIIA9 were observed in substantially more reference genomes than the other ACRs. AcrIIA7 belonged to the most widespread protein family, with more than 1,000 homologs identified in more than 14,000 genomes of about 622 unique species across bacteria and bacteriophages (Table S2). Homologs of AcrIIA7 belong to the functionally uncharacterized DUF2829 superfamily, and distant homologs were distributed across 6 distinct phyla (Firmicutes, Proteobacteria, Bacteroidetes, Actinobacteria, Cyanobacteria, and Spirochaetes). The majority of homologs are found in Firmicutes, predominantly in *S. pneumoniae* strains (56% of all strains). The second most represented phylum is Proteobacteria, including Alpha-, Beta-, Gamma-, Delta-, and Epsilon-proteobacteria. Notably, AcrIIA7 was the only ACR with homologs in viral reference genomes, including all three tailed bacteriophage families (i.e., *Siphoviridae*, *Myoviridae*, and *Podoviridae*), further reflecting its ubiquitousness. On the protein tree with both metagenomic (MetaVir) and genomic homologs (Figure 4), AcrIIA7 clusters together with Bacteroidetes representatives, most likely *B. dorei* being its original host (NCBI EL88_22925), while its closest homolog is derived from a human gut virome sample. Evidence of transfer between different mobile genetic elements was observed in a homolog (WP_024086069) on a *Bacillus thuringiensis* prophage located on a plasmid. The diversity of the AcrIIA7 tree, mostly characterized by mixed-phylum clades, strongly supports the hypothesis that AcrIIA7 homologs have undergone several interphylum HGT events (Figure 4).

AcrIIA9 had over 600 homologs distributed across over 300 unique species (Table S2), mostly belonging to the phylum Bacteroidetes, distributed across two different families, and spanning several *Bacteroides* species and three *Parabacteroides* species with 100% sequence identity. Homologs of AcrIIA9 were also identified in genomes from the recently defined phylum Balneolaeota (Hahnke et al., 2016), as well as the phyla Firmicutes, Proteobacteria, and Actinobacteria (Figure S3). Most homologs of AcrIIA9 are annotated as members of the functionally uncharacterized PcfK superfamily that is found in bacteria and viruses according to PFAM. The flanking regions of these homologs often lacked annotation as is often the case with viral genomes (Roux et al., 2015; Krishnamurthy and Wang, 2017), making it difficult to determine if this ACR is located in a mobile element. However, we identified AcrIIA9 homologs in several viral metagenomic datasets (Figure 3B), indicating that AcrIIA9 homologs are likely viral-derived genes. Furthermore, out of 539 bacterial contigs carrying ACRIIA9 homologs, 174 were predicted as complete or partial phage sequence by VIRSorter—a tool that mines viral signal in bacterial genomes (Roux et al., 2015).

Abundance of Type II CRISPR-Cas Systems in Genomes Harboring AcrIIA7–10

Finally, since the ACRs were selected using the type II-A SpCas9 effector, we investigated the co-occurrence of AcrIIA7–10-carrying genomes with this CRISPR-Cas system. Accordingly, CRISPR-Cas systems' signature Cas genes (Makarova et al., 2015) were mined from the strains carrying each ACR homolog, and the frequencies of each CRISPR-Cas system were thus estimated (Table S3). Overall, we did not observe any correlation between the ACRs and their corresponding type II CRISPR-Cas system (Figure S4A). Notably, when examining the presence of different Cas types individually in AcrIIA-homolog-carrying reference genomes, there were multiple cases where Cas-II signature genes were completely absent (Figure S4B). An explanation for this could be that CRISPR-Cas systems are often themselves mobile (Makarova et al., 2015), and their presence in certain bacterial lineages may vary significantly depending on the selective pressure (Palmer, and Gilmore, 2010; Bikard et al., 2012).

DISCUSSION

Previous approaches for ACR protein discovery have relied mostly on cultivable hosts and phage genomes. However, genome databases are biased toward human pathogens and do not reflect the true species' diversity in nature, with some environments, such as soil, being underrepresented. Without relying on cultivated strains and specific sequence signatures, we identified four previously uncharacterized families of ACR proteins from functional metagenomic selections of DNA libraries derived from various environments. These ACRs inhibit the activity of SpCas9 *in vivo* (Figure 1C) and *in vitro* (Figure 2A) and share no homology with previously discovered ACRs and thus, substantially expand the known repertoire of ACRs against type II-A CRISPR-Cas system.

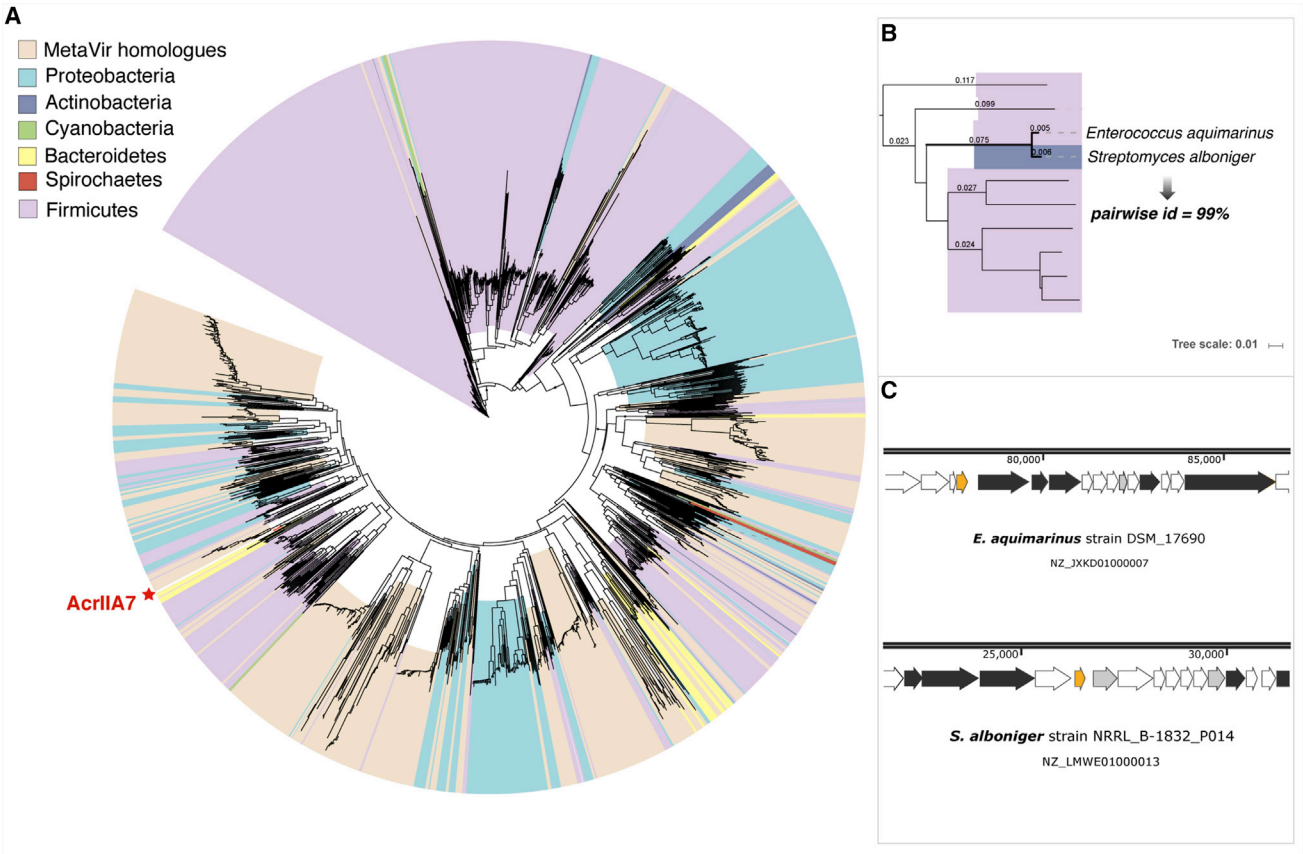


Figure 4. Phylogenetic Diversity of AcrIIA7

(A) The phylogenetic diversity of AcrIIA7 homologs represented by a protein tree, including proteins originating from both reference genomes and viral metagenomic (MetaVir) datasets. Clades of reference proteins are assigned colors according to phylum membership. Asterisk marks the placement of the original AcrIIA7 sequence on the tree, clustered within a *Bacteroides* clade.

(B) Zoomed-in region of a mixed-phylum clade with homologs of *E. aquamarinus* and *S. alboniger* sharing high amino acid identity. Numbers denote branch length.

(C) Corresponding genomic regions of the AcrIIA7 homolog in *E. aquamarinus* and *S. alboniger* strains reveal it is surrounded by phage hallmark genes. The sequence conservation of AcrIIA7 (90% and 99% identity at nucleotide and protein level, respectively) is not observed in the flanking regions. Orange, AcrIIA7 homolog; black, phage particle structural genes; gray, non-phage gene; white, hypothetical protein. The tree was based on multiple sequence alignment of AcrIIA7 homologs by Clustal Omega and constructed using Neighbor-Joining algorithm and visualized with iTOL.

Except for the size of ACRs (<190 aa), there are no common features conserved among these protein families. ACRs have been found in most families of phages, conjugative elements, and pathogenicity islands (Pawluk et al., 2018). Likewise, we observed that multiple ACRs were present in the genome of some organisms with high variability in their location and arrangements of their neighboring regions, which could be the result of multiple HGT events (Figure 4).

In particular, AcrIIA7 appears to be more abundant in nature compared to other AcrIIAs, with homologs in seven phyla (Figure 3), five of which were previously not known to harbor AcrIIA homologs. Using our current setup, we identified only one representative of this protein family, which might be attributed to the intrinsic limitations of functional metagenomic screening (Gabor et al., 2004) and methodologies employed for sequencing and analysis (van der Helm et al., 2018). A detailed examination of this ACR family might help in understanding their distribution and abundance in multiple phyla and environments. Interestingly, AcrIIA7 was characteristic in our experiments for having no detectable binding to SpCas9 using biolayer interferometry (Figure 2B), even though AcrIIA7 was able to abolish SpCas9 activity in our *in vivo* and *in vitro* cleavage assays (Figures 1C and 2A). This observation suggests that the origin and mode of action of this ACR might be completely different from the previously characterized AcrIIA proteins; for instance, AcrIIA7 could be enzymatic in nature, or it may interfere with the formation of the gRNA:Cas9 complex, which would be consistent with our inability to demonstrate protein-protein interaction using biolayer interferometry (Figure 2B).

Little is known about the origin of ACRs or how they are related to other proteins; however, the evidence of conserved domains between viral structural proteins and ACRs suggests a possible mechanism for evolution of CRISPR-Cas inhibitors (Stone et al., 2018). The structural similarity of AcrIIA8 to head-tail adaptor proteins, in addition to the strong affinity to SpCas9 (Figure 2B) and its activity *in vivo* and *in vitro* (Figures 1C and 2A), suggest

a common evolutionary history for this ACR and viral structural proteins. The function of the evolutionary-related protein homologs remains to be experimentally tested, and they might have a stronger interaction with SpCas9 or a different function (i.e., phage head-tail adaptor). Subsequent experimental work is required to show if this Cas9 inhibitory effect is a property of several homologs of each ACR family when expressed heterologously and in the endogenous context of the host.

ACRs are likely to play a major role in the evolution, diversity, and distribution of CRISPR-Cas systems across different phyla. The cost to the cell associated with an active CRISPR-Cas system is still not clear. CRISPR-Cas immunity provides a clear advantage to bacteria against phage infection (Levin and Udekwu, 2010; Westra et al., 2014; van Houte et al., 2016); however, some potential disadvantages are associated with the toxicity of the nucleases expressed by these systems—potential self-targeting or limited transfer of novel genetic material (Palmer and Gilmore, 2010; Bikard et al., 2012; Vercoe et al., 2013; Cui and Bikard, 2016). In addition, it has been observed that some lineages of bacteria seem to completely lack CRISPR-Cas systems (Burstein et al., 2016), suggesting that the advantage of having these systems may be defined by additional factors. In that manner, the presence and diversity observed in CRISPR-Cas systems could be partially explained by the presence of equally diverse ACR strategies.

The approach presented in this manuscript explores the diversity of putative ACR genes that inhibit the activity of Cas9. These ACRs are located in chromosomal and extrachromosomal mobile elements disseminated across seven phyla. Our findings indicate that some ACRs are more widespread and abundant across bacterial phylogeny than previously believed, highlighting that we have only scratched the surface in terms of biological diversity of CRISPR-ACR interactions and suggesting that ACR biological functions may be more complex. Identification of more ACRs against diverse CRISPR-Cas systems and further characterization of their mechanisms in their endogenous context are needed to understand their evolutionary origin and their impact in bacterial population dynamics and transfer of genetic material.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Bacterial Strains and Growth Conditions
- METHOD DETAILS
  - Cloning of Selection System
  - *In Vivo* Assay for Screening of ACR Activity
  - Amplification of Positive Hits
  - Nanopore Sequence Library Preparation
  - Nanopore Sequencing
  - Nanopore Data Processing
  - Nanopore Sequence Analysis
  - Validation of Insert Activity
  - Individual ORF Identification
  - Protein Purification
  - *In Vitro* DNA Cleavage Assay *In Vitro*
  - Biolayer Interferometry for Binding Affinity
  - AcrII7-10 Homologue Retrieval
  - Phylogenetic Trees
  - AcrIIA8 Genomic Context Analysis
  - Identification of CRISPR-Cas System Types
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures, three tables, and one data file and can be found with this article online at https://doi.org/10.1016/j.chom.2019.01.003.

ACKNOWLEDGMENTS

The authors would like to thank Andreas Porse for discussion and comments on the manuscript, Lejla Imamovic for providing the metagenomic libraries, and Sara Petersen Bjørn for the plasmid pNIC28-Bsa4. Funding: This project has received funding from the Novo Nordisk Foundation under NFF grant number NNF10CC1016517, the Lundbeck Foundation under grant agreement R140-2013-13496, and the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 642738, MetaRNA.

AUTHOR CONTRIBUTIONS

R.V.U., E.v.d.H., and M.O.A.S. conceived and designed the project; R.V.U., E.v.d.H., S.-W.L., and S.K. conducted the experiments; E.v.d.H. performed nanopore sequencing and analysis of the results; M.-A.M. performed computational analysis of the ecological, phylogenetic, and genomic context; R.V.U., E.v.d.H., M.-A.M., and M.O.A.S. derived the theory and wrote the paper.

DECLARATION OF INTERESTS

Some of the authors have filed patents related to this work.

Received: March 13, 2018
Revised: August 24, 2018
Accepted: January 2, 2019
Published: February 5, 2019

REFERENCES

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Anders, C., Niewoehner, O., and Jinek, M. (2015). In vitro reconstitution and crystallization of Cas9 endonuclease bound to a guide RNA and a DNA target. Methods Enzymol. *558*, 515–537.

Bikard, D., Hatoum-Aslan, A., Mucida, D., and Marraffini, L.A. (2012). CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. Cell Host Microbe *12*, 177–186.

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.

Burstein, D., Sun, C.L., Brown, C.T., Sharon, I., Anantharaman, K., Probst, A.J., Thomas, B.C., and Banfield, J.F. (2016). Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. Nat. Commun. *7*, 10613.

Caliando, B.J., and Voigt, C.A. (2015). Targeted DNA degradation using a CRISPR device stably carried in the host genome. Nat. Commun. *6*, 6989.

Citorik, R.J., Mimee, M., and Lu, T.K. (2014). Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. Nat. Biotechnol. *32*, 1141–1145.

Cui, L., and Bikard, D. (2016). Consequences of Cas9 cleavage in the chromosome of *Escherichia coli*. Nucleic Acids Res. *44*, 4243–4251.

Dong, D., 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.

Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G., and Sorek, R. (2018). Systematic discovery of antiphage defense systems in the microbial pangenome. Science *359*, eaar4120.

Doudna, J.A., and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. Science *346*, 1258096.

Esvelt, K.M., Mali, P., Braff, J.L., Moosburner, M., Yaung, S.J., and Church, G.M. (2013). Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nat. Methods *10*, 1116–1123.

Gabor, E.M., Alkema, W.B.L., and Janssen, D.B. (2004). Quantifying the accessibility of the metagenome by random expression cloning techniques. Environ. Microbiol. *6*, 879–886.

Genee, H.J., Bali, A.P., Petersen, S.D., Siedler, S., Bonde, M.T., Gronenberg, L.S., Kristensen, M., Harrison, S.J., and Sommer, M.O.A. (2016). Functional mining of transporters using synthetic selections. Nat. Chem. Biol. *12*, 1015–1022.

Genee, H.J., Bonde, M.T., Bagger, F.O., Jespersen, J.B., Sommer, M.O., Wernersson, R., and Olsen, L.R. (2015). Software-supported USER cloning strategies for site-directed mutagenesis and DNA assembly. ACS Synth. Biol. *4*, 342–349.

Hahnke, R.L., Meier-Kolthoff, J.P., García-López, M., Mukherjee, S., Huntemann, M., Ivanova, N.N., Woyke, T., Kyrpides, N.C., Klenk, H.P., and Göker, M. (2016). Genome-based taxonomic classification of Bacteroidetes. Front. Microbiol. *7*, 2003.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., Rattei, T., Mende, D.R., Sunagawa, S., Kuhn, M., et al. (2016). EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. *44*, D286–D293.

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield, C.N., Hernsdorf, A.W., Amano, Y., Ise, K., et al. (2016). A new view of the tree of life. Nat. Microbiol. *1*, 16048.

Hynes, A.P., Rousseau, G.M., Agudelo, D., Goulet, A., Amigues, B., Loehr, J., Romero, D.A., Fremaux, C., Horvath, P., Doyon, Y., et al. (2018). Widespread anti-CRISPR proteins in virulent bacteriophages inhibit a range of Cas9 proteins. Nat. Commun. *9*, 2919.

Hynes, A.P., Rousseau, G.M., Lemay, M.L., Horvath, P., Romero, D.A., Fremaux, C., and Moineau, S. (2017). An anti-CRISPR from a virulent streptococcal phage inhibits *Streptococcus pyogenes* Cas9. Nat. Microbiol. *2*, 1374–1380.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacetrial immunity. Science *337*, 816–821.

Koonin, E.V., Makarova, K.S., and Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. Curr. Opin. Microbiol. *37*, 67–78.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. *27*, 722–736.

Krishnamurthy, S.R., and Wang, D. (2017). Origins and challenges of viral dark matter. Virus Res. *239*, 136–142.

Lauritsen, I., Porse, A., Sommer, M.O.A., and Nørholm, M.H.H. (2017). A versatile one-step CRISPR-Cas9 based approach to plasmid-curing. Microb. Cell Fact. *16*, 135.

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. *44*, W242–W245.

Levin, B.R., and Udekwu, K.I. (2010). Population dynamics of antibiotic treatment: a mathematical model and hypotheses for time-kill and continuous-culture experiments. Antimicrob. Agents Chemother. *54*, 3414–3426.

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

Loman, N.J., and Quinlan, A.R. (2014). Poretools: a toolkit for analyzing nanopore sequence data. Bioinformatics *30*, 3399–3401.

Loman, N.J., Quick, J., and Simpson, J.T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat. Methods *12*, 733–735.

Lutz, R., and Bujard, H. (1997). Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2regulatory elements. Nucleic Acids Res. *25*, 1203–1210.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J.J., Charpentier, E., Haft, D.H., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. Nat. Rev. Microbiol. *13*, 722–736.

Martínez-Garćía, E., Aparicio, T., Goñi-Moreno, A., Fraile, S., and De Lorenzo, V. (2015). SEVA 2.0: an update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities. Nucleic Acids Res. *43*, D1183–D1189.

Medema, M.H., Takano, E., and Breitling, R. (2013). Detecting sequence homology at the gene cluster level with multigeneblast. Mol. Biol. Evol. *30*, 1218–1223.

Palmer, K.L., and Gilmore, M.S. (2010). Multidrug-resistant enterococci lack CRISPR-cas. MBio *1*, e00227–10.

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.e9.

Pawluk, A., Bondy-Denomy, J., Cheung, V.H.W., 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., Davidson, A.R., and Maxwell, K.L. (2018). Anti-CRISPR: discovery, mechanism and function. Nat. Rev. Microbiol. *16*, 12–17.

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.e10.

Roux, S., Faubladier, M., Mahul, A., Paulhe, N., Bernard, A., Debroas, D., and Enault, F. (2011). Metavir: a web server dedicated to virome analysis. Bioinformatics *27*, 3074–3075.

Roux, S., Enault, F., Hurwitz, B.L., and Sullivan, M.B. (2015). VirSorter: mining viral signal from microbial genomic data. PeerJ *3*, e985.

Samson, J.E., Magadán, A.H., Sabri, M., and Moineau, S. (2013). Revenge of the phages: defeating bacterial defences. Nat. Rev. Microbiol. *11*, 675–687.

Savitsky, P., Bray, J., Cooper, C.D.O., Marsden, B.D., Mahajan, P., Burgess-Brown, N.A., and Gileadi, O. (2010). High-throughput production of human proteins for crystallization: the SGC experience. J. Struct. Biol. *172*, 3–13.

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I., et al. (2017). Diversity and evolution of class 2 CRISPR–Cas systems. Nat. Rev. Microbiol. *15*, 169–182.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539.

Simmonds, P., Adams, M.J., Benkő, M., Breitbart, M., Brister, J.R., Carstens, E.B., Davison, A.J., Delwart, E., Gorbalenya, A.E., Harrach, B., et al. (2017). Consensus statement: virus taxonomy in the age of metagenomics. Nat. Rev. Microbiol. *15*, 161–168.

Smith, T.F., and Waterman, M.S. (1981). Identification of common molecular subsequences. J. Mol. Biol. *147*, 195–197.

Söding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. *33*, W244–W248.

Sommer, M.O.A., Dantas, G., and Church, G.M. (2009). Functional characterization of the antibiotic resistance reservoir in the human microflora. Science *325*, 1128–1131.

Stern, A., and Sorek, R. (2011). The phage-host arms race: shaping the evolution of microbes. Bioessays *33*, 43–51.

Stone, N.P., Hilbert, B.J., Hidalgo, D., Halloran, K.T., Lee, J., Sontheimer, E.J., and Kelch, B.A. (2018). A hyperthermophilic phage decoration protein suggests common evolutionary origin with herpesvirus triplex proteins and an anti-CRISPR protein. Structure *26*, 936–947.e3.

Suttle, C.A. (2007). Marine viruses — major players in the global ecosystem. Nat. Rev. Microbiol. *5*, 801–812.

van der Helm, E., Genee, H.J., and Sommer, M.O.A. (2018). The evolving interface between synthetic biology and functional metagenomics. Nat. Chem. Biol. *14*, 752–759.

van der Helm, E., Imamovic, L., Hashim Ellabaan, M.M.H., Van Schaik, W., Koza, A., and Sommer, M.O.A. (2017). Rapid resistome mapping using nanopore sequencing. Nucleic Acids Res. *45*, gkw1328.

van Houte, S., Ekroth, A.K.E., Broniewski, J.M., Chabas, H., Ashby, B., Bondy-Denomy, J., Gandon, S., Boots, M., Paterson, S., Buckling, A., et al. (2016). The diversity-generating benefits of a prokaryotic adaptive immune system. Nature *532*, 385–388.

Vercoe, R.B., Chang, J.T., Dy, R.L., Taylor, C., Gristwood, T., Clulow, J.S., Richter, C., Przybilski, R., Pitman, A.R., and Fineran, P.C. (2013). Cytotoxic chromosomal targeting by CRISPR/cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. PLoS Genet. *9*, e1003454.

Westra, E.R., Buckling, A., and Fineran, P.C. (2014). CRISPR–Cas systems: beyond adaptive immunity. Nat. Rev. Microbiol. *12*, 317–326.

Yamada, K.D., Tomii, K., and Katoh, K. (2016). Application of the MAFFT sequence alignment program to large data-reexamination of the usefulness of chained guide trees. Bioinformatics *32*, 3246–3251.

Zhu, W., Lomsadze, A., and Borodovsky, M. (2010). Ab initio gene identification in metagenomic sequences. Nucleic Acids Res. *38*, e132.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
One ShotTOP10 Electrocomp *E. Coli*	Invitrogen	Cat#C404052
BL21-AI One Shot Chemically Competent *E. coli*	Invitrogen	Cat#C607003
TOP10 (pCasens3+pDual3)	This paper	bRU001
TOP10 (pDual3, pCasens3, pZE21-AcrIIA2)	This paper	bRU002
TOP10 (pDual3, pCasens3, pZE21-GFP)	This paper	bRU003
Chemicals, Peptides, and Recombinant Proteins		
Cas9 Nuclease, *S. pyogenes*	NEB	Cat#M0386M
Proteinase K from *Tritirachium album*	Sigma-Aldrich	Cat#P6556
AcTEV Protease	Thermo Scientific	Cat#12575015
USER Enzyme	NEB	Cat#M5505L
AcTEV protease	Thermo Scientific	Cat#12575015
Phusion high fidelity PCR Master Mix	Thermo Scientific	Cat#F531S
Critical Commercial Assays		
TranscriptAid T7 High Yield Transcription Kit	Thermo Scientific	Cat#K0441
EZ-Link Sulfo-NHS-LC-Biotinylation Kit	Thermo Scientific	Cat#21435
Nanopore Genomic Sequencing Kit SQK-LSK108	Oxford Nanopore, UK	SQK-LSK108
NEBNext Ultra II End Repair/dA Tailing	NEB, USA	E7546S
Blunt/TA Ligase Master Mix	NEB, USA	M0367S
HisTRAP	GE Healthcare, USA	GE17-5247-01
MBPTrap HP column	GE Healthcare, USA	28-9187-79
HiTrap Q Fast Flow	GE Healthcare, USA	GE17-5156-01
Deposited Data		
Nucleotide sequences of ACR coding sequences	This paper	ENA: PRJEB29470
Nucleotide sequences of the 39 functionally selected inserts	This paper	Biostudies: S-BSST226
Oligonucleotides		
Primers to amplify contig from metagenomic libraries, cloning the plasmids pCasens3, pDual3 and the contigs from the libraries into the expression vector pZE21 see Table S1	This paper	N/A
Recombinant DNA		
pZE21 plasmid	Lutz and Bujard, 1997	pZE21
pZE21-AcrIIA2	This paper	N/A
pZE21-GFP	This paper	N/A
pMJ806	Addgene	Cat#39312
pNIC28-Bsa4	Addgene	Cat#26103
Metagenomic libraries from human feces	van der Helm et al., 2017	120A, 120B, 120C, 120D and 120E
Metagenomic libraries from pig feces, cow feces and soil.	Genee et al., 2016	GranjaPig, GranjaCow and GranjaWorker
Metagenomic libraries from soil	Sommer et al., 2009	AB95D01
pCasens3	This paper	N/A
pDual3	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
Poretools v0.6	Loman and Quinlan, 2014	https://github.com/arq5x/poretools/releases/tag/v0.6.0
poreFUME v1.1	van der Helm et al., 2017	https://github.com/EvdH0/poreFUME/releases/tag/v1.1
Canu v1.5	Koren et al., 2017	https://github.com/marbl/canu
cd-hit V4.6	Li and Godzik, 2006	https://github.com/weizhongli/cdhit/releases/tag/V4.6.8
MAFFT v7.310	Yamada et al., 2016	https://mafft.cbrc.jp/alignment/software/
ETE3	Huerta-Cepas et al., 2016, Mol. Bio. Evol.	http://etetoolkit.org/download/
Nanopolish (commit 04fd9aecbb4ab266350476b957f4abb8ed994d8d)	Loman et al., 2015	https://github.com/jts/nanopolish
Albacore	Albacore basecaller from Oxford Nanopore	https://github.com/dvera/albacore
MetaGeneMark v. 3.26	Zhu et al., 2010	http://exon.gatech.edu/GeneMark/license_download.cgi
PSI-BLAST v. 2.6.0+	Altschul et al., 1997	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/
ClustalOmega v. 1.2.2	Sievers et al., 2011	http://www.clustal.org/omega/#Download
iTOL v4.0.2	Letunic and Bork, 2016	https://itol.embl.de
MultiGeneBlast v1.1.14	Medema et al., 2013	https://sourceforge.net/projects/multigeneblast/files/
Other		
MetaVir virome database	Roux et al., 2011	http://metavir-meb.univ-bpclermont.fr
Newick file of phylogenetic tree of life	Hug et al., 2016	https://media.nature.com/original/nature-assets/nmicrobiol/2016/nmicrobiol201648/extref/nmicrobiol201648-s6.txt
Cas profiles	Makarova et al., 2015	ftp://ftp.ncbi.nih.gov/pub/wolf/_suppl/CRISPR2015/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Morten O.A. Sommer (msom@bio.dtu.dk).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Bacterial Strains and Growth Conditions

Cloning and functional screening of the metagenomic libraries was carried out using *E.coli* TOP10. For expression and purification of recombinant ACR proteins the strain *E.coli* BL21(AI) was used (Key Resources Table). Both strains were routinely growth and maintained at 37°C in 2xYT broth at 250 RPM.

METHOD DETAILS

Cloning of Selection System

Construction of pCasens3 plasmid containing SpCas9 was performed in a single step using USER cloning (Genee et al., 2015). The fragment containing SpCas9 was amplified from DS-SPcas addgene ID48645 (Esvelt et al., 2013) and clone into the backbone of pSEVA47 that contains the low copy number origin of replication pSC101 and the antibiotic resistance gene *aadA* that confers

resistance against spectinomycin (Martínez-Garćia et al., 2015). The theophylline riboswitch was placed in front of Cas9 using a long forward primer from IDT (5'-AAGTCTAGCGAACCGCACTTAATACGACTCACTATAGGTACCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCTGCTAAGGTAACAACAAGATGATGGATAAGAAATACTCAATAGGCTTAGATATCGGCAC-3'). Aditionally, a sigma70 constitutive promoter was also introduced using a reverse primer in order to introduce a different promoter for Cas9 (5'-ctctagTagctagcactgtacctaggactgagctagccgtcaaGTTAGCTGTGCTCTAGAAGCTAGCAG-3').

Construction of pDual plasmid containing the arabinose inducible gRNA was constructed using USER cloning (Genee et al., 2015). The chimeric gRNA under a pBAD inducible system and terminator were synthesized from IDT (5'-**CTATAACCAGACCGTTCAGC**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT-3'). The 20bp target of the chloramphenicol gene is marked it underlined in bold letter. The backbone of the plasmid was from pSEVA3610 a plasmid that contains a chloramphenicol resistance gene (*cat*), an arabinose inducible expression system and low copy number origin of replication p15A (Martínez-Garćia et al., 2015).

In Vivo Assay for Screening of ACR Activity

The strain for selection of metagenomic libraries consisted of an *E.coli* TOP10 strain harbouring the plasmids pCasens3 and pDual3 (Figure S1). An additional plasmid was added as positive control (pZE21-AcrIIA2), negative control pZE21-GFP or metagenomic library that was previously cloned in the multi-cloning site of pZE21 (Lutz and Bujard, 1997).

An overnight culture of the selection strain harboring pZE21-AcrIIA2, pZE21-GFP or the metagenomic library was prepared in 2xYT media supplemented with 50 μg/ml spectinomycin, 30 μg/ml chloramphenicol and 50 μg/ml of kanamycin and incubated at 37°C; 250RPM. A 1:100 subculture of the overnight culture was prepared in 2xYT media supplemented with 50 μg/ml spectinomycin, 50 μg/ml of kanamycin with, 2 mM theophylline and 1% arabinose, in order to activate the selection system. The culture was incubated for 8 hrs at 37°C; 250RPM. hen serial dilutions from 10^{-1} to 10^{-8} of the culture were prepared in LB media and plated on LB-agar supplemented with 50 μg/ml spectinomycin, 50 μg/ml of kanamycin and 30 μg/ml chloramphenicol.

Amplification of Positive Hits

Clones appearing on 10^{-2}-10^{-6} dilutions were collected as previously described (van der Helm et al., 2017). Briefly, the clones from plates were collected by adding 5 ml of H2O, after which the colonies were scraped off the plate with cell scraper. The bacterial cells were then pelleted by centrifugation and the pellet was resuspended in 10 ml of H2O. Two ml of the collected bacterial cells was used for plasmid extractions with the Plasmid Mini Kit (Invitrogen, USA). Primers were synthesized that amplify the common region on pZE21-MCS- together with the specific barcodes (van der Helm et al., 2017). One ng of DNA was amplified by PCR using Phusion Mastermix.

Nanopore Sequence Library Preparation

Nanopore sequencing library was prepared as briefly described below. DNA QC was performed using Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific, USA). Sequencing library preparation was carried out with Nanopore Genomic Sequencing Kit SQK-LSK108 (Oxford Nanopore, UK) using 1D - R9.4 chemistry. The NEBNext Ultra II End Repair/dA Tailing module (E7546S, NEB, USA) was used to prepare 1000 ng of the functionally selected DNA. Next, 350 ng (with an average size of 2.5 kb) of End-prepared DNA was used to fulfill the 0.2 pmol requirement of the ligation protocol. The End-prepared DNA was ligated with 1D adapters (AMX1D) using Blunt/TA Ligase Master Mix (M0367S, NEB, USA and purified with AMPure XP beads.

Nanopore Sequencing

The MinION was initially primed for 10 minutes with 800 μl priming solution (520 μl nuclease free water, 480 μl Running Buffer with Fuel mix) though the priming port and finally with 200 μl priming solution immediately before loading the sequencing library.

For sequencing, 12 μl library was mixed with 2.5 μl nuclease free water, 35 μl Running Buffer with Fuel mix and 25.5 μl Library Loading Beads (LLB) and immediately loaded to the SpotON sample port of the MinION. MinKNOW software was used to sequence the library without live basecalling.

Nanopore Data Processing

The sequencing data produced by MinKNOW was basecalled using Albacore (read_fast5_basecaller.exe) with the flags '–recursive -c FLO-MIN106_LSK108_linear.cfg'. Poretools (Loman and Quinlan, 2014) was used to extract 1D FASTQ reads using the 'poretools fastq' command.

Nanopore Sequence Analysis

The obtained reads were demultiplexed on barcode using the Smith-Waterman algorithm (Smith and Waterman, 1981) of the poreFUME package (van der Helm et al., 2017). User defined barcodes are detected within 60 basepairs of the read ends. Barcode alignment was scored using +2.7 for match, -4.5 for mismatch, -4.7 gap opening and -1.6 for gap extension. A score threshold of >58 was used for the combined score of the asymmetric barcodes. Second, the demultiplexed reads were error corrected using the error correction module of Canu v1.5 (Koren et al., 2017) and subsequently the corrected reads were assembled by Canu with the flags '*genomeSize=11m correctedErrorRate=0.075 minReadLength=300 minOverlapLength=300 corOutCoverage=10000 -nanopore-raw contigFilter="2 300 1.0 1.0 2"*'. Both the assembled and unassembled contigs from Canu were first clustered with cd-hit V4.6

(Li and Godzik, 2006) with the flags '*cd-hit-est -c 0.8 -n 4 -d 0 -M 4000 -p 1 -r 1 -g 1*' and subsequently with MAFFT (Yamada et al., 2016) v 7.310 '*–reorder –adjustdirectionaccurately –maxiterate 0*'. Next, an average linkage tree was constructed using the UPGMA (Unweighted Pair Group Method with Arithmetic Mean) method. The constructed tree was analyzed using ETE3 [Huerta-Cepas et al., 2016, Mol bio Evol] and nodes were collapsed with a distance shorter than 1.0 propagating the node with the most mapped nanopore reads as representative.

The resulting contigs were annotated with blastx using the NT database at the 29th of April 2017. The annotations were manually curated and categorized into 'non-relevant enzyme activity (ie. orotidine 5'-phosphate decarboxylase)', 'DNA binding/modifying enzyme (ie. transposon)', 'RNA binding/modifying enzyme', 'phage component', 'transporter/membrane protein (ie. ABC transporter)', 'no annotation available, 'hypothetical protein', 'system interference (ie araC)'. 34 inserts representing interesting biology were selected and combined with the top 5 inserts that had the most nanopore reads mapped but displayed no relevant biological annotation (ie. hypothetical protein or empty read). The accuracy of the resulting 39 contigs was increased using nanopolish (Loman et al., 2015) with the flags '*variants –consensus –min-candidate-frequency 0.1*'.

Validation of Insert Activity

Primers for the 39 inserts were designed for USER cloning (Genee et al., 2015). The fragments were amplified from their corresponding metagenomic library and cloned back into the pZE21 vector in order to validate the activity. The fragments were PCR amplified from the barcoded library using Phusion high fidelity PCR Master Mix. 5 uL of crude PCR mixture was combined with 0.5 uL of linearized pZE21 with USER overlaps for the inserts and 1 unit of USER enzyme, incubated for 30 minutes at 37°C and after 30 minutes at room temperature. Then 6 uL of the assembly reaction was chemically transformed in the selection strain harboring pDual3 and pCasens3. Colony PCR was used to validate the correct insertion of the 39 fragments. The selection strain harboring each of the 39 inserts were tested individually for anti-CRISPR activity.

Individual ORF Identification

Identification of putative ORFs from inserts with ACR activity were detected using MetaGeneMark v3.25 (Zhu et al., 2010) with the flags '*gmhmmp -m MetaGeneMark_v1.mod -f G*'. In parallel the inserts were re-annotated using blastx using the NT database accessed at the 17th of May 2017. Based on the combined annotation, potentially biologically active ORFs were manually identified (ie. >80% of the subject gene present, no missing N-terminus) and USER primers were designed to clone the ORFs into the pNIC28-Bsa4 plasmid (Savitsky et al., 2010).

Protein Purification

Proteins were expressed in *E. coli* strain BL21 (AI) grown in the 2xYT medium at 18°C for 16 hours following induction with 1% arabinose. Proteins were purified by a combination of affinity, ion exchange, and size exclusion chromatography steps. Briefly, cells were lysed by three passes through an EmulsiFlex-C5 homogenizer (Avestin, Mannheim, Germany) at 10 000–15 000 psi, any debris and unbroken cells were removed by centrifuging at 18 000 *g* at 4°C for 30 minutes. The supernatant was loaded onto nickel-nitrilotriacetic acid (Ni^{2+}-NTA) resin columns (HisTRAP, GE Healthcare, Chicago, IL, USA) on an Äkta Pure system connected to an F9-C fraction collector (GE Healthcare).

Cas9 inhibitors were eluted by increasing the imidazole concentration in a stepwise manner to 25 mM, 50 mM, 75 mM and finally 500 mM. After pooling and concentration, protein samples were buffer exchanged into IEX start buffer. Anion exchange was performed on a HiTrap Q FF column (GE Healthcare) in 20 mM phosphate buffer pH 7.0 (AC12-1, AC19-2, AC28-1, AC42-1, AcrIIA2 and GFP) or 20 mM TRIS-HCl pH 8.0 (AC23-2 and AC27-1). Cation exchange was performed on a HiTrap SP FF column (GE Healthcare) in 20 mM phosphate buffer pH 7.0 (AC23-1 and AC27-2). The fractions containing the protein of interest were pooled, concentrated, flash frozen, and stored at -80°C. Purity analysis was performed using a Coomassie-stained SDS-PAGE gel analysed by ImageQuant TL software (GE Healthcare).

MBP-Cas9 was expressed from plasmid pMJ806 and was essentially purified as described (Jinek et al., 2012) with some modifications. After performing expression and His-tag affinity purification as described above, MBP-Cas9 was further purified using an MBPTrap HP column (GE Healthcare). After cleavage with AcTEV protease (ThermoFisher Scientific, Waltham, MA, USA) during overnight dialysis and negative His-tag affinity purification, the sample was loaded onto a Superdex 200 Increase 10/300 GL column (GE Healthcare) equilibrated with 50 mM Tris-HCl pH 7.5, 150 mM NaCl. The fractions containing Cas9 were pooled, concentrated and biotinylated using EZ-Link Sulfo-NHS-LC-Biotinylation Kit (ThermoFisher Scientific). After buffer exchange, samples were flash frozen and stored at -80°C.

In Vitro DNA Cleavage Assay *In Vitro*

DNA cleavage assay was carried out as described previously (Pawluk et al., 2016; Dong et al., 2017) with some modifications. SpyCas9 (New England Biolabs) (100nM), gRNA (*in vitro* transcribed) (100nM), and purified anti-CRISPR protein were mixed together in cleavage buffer (20 mM HEPES-KOH (pH 7.5), 75 mM KCl, 10% glycerol, 1 mM DTT, and 10 mM MgCl2) and incubated for 30 min. Then, PCR amplified DNA target (10nM) was added and the mixture was incubated for 10 min for cleavage. The reaction was stopped by adding proteinase K and incubating at 60°C for 15 min. The cleaved and un-cleaved fraction of DNA target were visualized in 1% agarose gel.

Biolayer Interferometry for Binding Affinity

Equimolar gRNA was mixed together with biotinylated Cas9 and incubated at 25°C for 15min to form a biotinylated Cas9:gRNA complex. Streptavidin biosensors (Pall FortéBio) were pre-equilibrated in PBS buffer for 600s, loaded with a biotinylated Cas9:gRNA complex at optimal concentrations and times, and brought to baseline in kinetics buffer (1X PBS, 0.02% Tween-20, 0.1% BSA, 75mM KCl, 10mM MgCl) for 300s. Association with anti-CRISPR proteins was measured in the same kinetics buffer for 600s, and then dissociation was measured in the kinetics buffer without anti-CRISPR proteins for 1000s. All biolayer interferometry experiments were performed on Octet RED96 system (Pall FortéBio) in 96-well microplates at 30°C with 200 μl volume. Binding kinetics were calculated using the FortéBio Data Analysis v7.1 software by fitting the association and dissociation data to a 1:1 model.

AcrII7-10 Homologue Retrieval

MetaVir-harboured Data Sets

485 viromes for which metadata were available were downloaded from MetaVir (Table S2). ORFs were predicted *de novo* using MetaGeneMark (v. 3.26) and a database of 81,706,359 predicted ORfFs was constructed. PSI-BLAST (v. 2.6.0+) with 4 iterations was used to search for all AcrIIA1-10 homologues with an e-value cut off of 10^{-5}.

NCBI-harboured Data Sets

Web-based PSI-BLAST was run for each ACR until convergence (e-value $<10^{-5}$) against the reference proteins (refseq_protein) and metagenomic protein databases (env_nr) of NCBI. Genomes carrying the respective protein hits were retrieved from the NCBI RefSeq database (last updated 24 Jan. 2017).

Phylogenetic Trees

In order to construct the protein trees of AcrIIA7 and AcrIIA9, respective homologues were retrieved from predicted proteins of reference genomes and metagenomic data sets (MetaVir). Multiple Sequence Alignment was performed with ClustalOmega (v. 1.2.2) and phylogenetic trees were constructed using the Neighbour-Joining algorithm (no distance corrections). Trees were visualized and graphically customized in iTOL (v4.0.2). For the tree of life (Figure 3A), iTOL was used to visualize the raw newick file provided by the work of Hug et al. on ribosomal protein alignment and phylogenetic tree construction (Hug et al., 2016).

AcrIIA8 Genomic Context Analysis

MultiGeneBlast (v1.1.14) (Medema et al., 2013) was used for inspection of the genomic neighbourhood of AcrIIA8 homologues. Initially, individual entries for contigs carrying the homologue were retrieved for each reference genome's WGS GenBank file and used to construct a custom MultiGeneBlast database. Then the contig carrying the closest homologue (belonging to *Erysipelotrichaceae bacterium* 21_3, NZ_JH590843) was selected as a query for a MultiGeneBlast homology search with default parameters (30% identity and 25% coverage) against this database, by specifying an approx. 10 kb genomic region around the AcrIIA8 homologue. All hits had at least 10^{-6} evalue. The graphical output was edited in Adobe Illustrator CS6 to mark the position of the ACR homologue (MultiGeneBlast is based on blastp while the more sensitive PSI-BLAST algorithm was used originally to recover these homologues) and to align contigs to the protein tree of AcrIIA8.

Identification of CRISPR-Cas System Types

Cas profiles originating from CDD, COG and PFAM databases as well as custom profiles, generated by the work of (Makarova et al., 2015, last updated June 02, 2015) were used to generate individual Position-Specific Scoring Matrices (PSSMs). These were subsequently provided to PSI-BLAST (v. 2.6.0+) for individual searches against the protein collection of each reference genome, for all genomes carrying the respective AcrIIA7-10. For each genome, significant hits (e-value $< 10^{-5}$) were then searched for the presence of signature CAS genes for each CRISPR-Cas system and the frequency of each type was estimated accordingly (Table S3).

QUANTIFICATION AND STATISTICAL ANALYSIS

Average +/- standard deviation of biological triplicates is shown throughout unless stated otherwise.

DATA AND SOFTWARE AVAILABILITY

The accession number for the four acr coding sequences reported in this paper is ENA: PRJEB29470. The accession number for the 39 functionally selected metagenomic inserts is Biostudies: S-BSST226.