Functional metagenomics-guided discovery of potent Cas9 inhibitors in the human microbiome

Kevin J. Forsberg[a,*], Ishan V. Bhatt[a], Danica T. Schmidtke[a],

Barry L. Stoddard[a], Brett K. Kaiser[a,c], and Harmit S. Malik[a,b]

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

Running head: A novel anti-CRISPR mechanism in the human microbiome

*Address correspondence to:

Kevin J. Forsberg,

1100 Fairview Ave, A2-025, Seattle, WA 98109; phone: +1-206-667-4512; fax: +1-206-667-6522; email: kforsber@fredhutch.org

1   **<u>Abstract</u>**

2   CRISPR-Cas systems protect bacteria and archaea from phages and other mobile genetic elements,

3   which use small anti-CRISPR (Acr) proteins to overcome CRISPR-Cas immunity. Because they are

4   difficult to identify, the natural diversity and impact of Acrs on microbial ecosystems is underappreciated.

5   To overcome this discovery bottleneck, we developed a high-throughput functional selection that isolates

6   *acr* genes based on their ability to inhibit CRISPR-Cas function. Using this selection, we discovered ten

7   DNA fragments from human oral and fecal metagenomes that antagonize *Streptococcus pyogenes* Cas9

8   (SpyCas9). The most potent *acr* discovered, *acrIIA11*, was recovered from a *Lachnospiraceae* phage

9   and is among the strongest known SpyCas9 inhibitors. *AcrIIA11* homologs are distributed across multiple

10 bacterial phyla and many divergent homologs inhibit SpyCas9. We show that AcrIIA11 antagonizes

11 SpyCas9 using a different mechanism than that of previously characterized inhibitors. Our study

12 highlights the power of functional selections to uncover widespread Cas9 inhibitors within diverse

13 microbiomes.

Introduction

CRISPR-Cas adaptive immune systems are present in ~50% of bacterial and ~90% of archaeal genomes (Makarova et al., 2015), where they protect their hosts from infection by phages (Barrangou et al., 2007), plasmids (Marraffini and Sontheimer, 2008), and other mobile genetic elements (MGEs) (Zhang et al., 2013). CRISPR-Cas systems mediate this defense by incorporating short (~30 bp) spacer sequences from invading genomes into an immunity locus in the host genome. These spacer sequences are then expressed and processed into CRISPR RNAs (crRNAs) that, together with various Cas nucleases, mediate homology-dependent restriction of invading genomes. CRISPR-Cas systems are classified into six types (I, II, III, etc.) and 25 subtypes (I-A, I-B, II-C, etc.) on the basis of functional differences, phylogenetic relatedness, and locus organization (Makarova et al., 2018).

In response to restriction, phages and other MGEs have evolved dedicated CRISPR-Cas antagonists, called anti-CRISPRs (Acrs) (Bondy-Denomy et al., 2013), which can promote phage infection, enable horizontal gene transfer (HGT), and thus shape microbial ecosystems (Borges et al., 2017; Pawluk et al., 2017). Acrs that inhibit the type II Cas9 (Pawluk et al., 2016a; Rauch et al., 2017) and type V Cas12 systems (Marino et al., 2018; Watters et al., 2018) have also garnered significant interest in biotechnology, with demonstrated utility for reducing off-target Cas9 lesions (Shin et al., 2017), suppressing gene drives (Basgall et al., 2018), and precisely controlling synthetic gene circuits (Nakamura et al., 2019).

Work over the past decade has revealed much about the activity (Hille et al., 2018), distribution (Makarova et al., 2015), and evolution (Koonin and Makarova, 2017) of CRISPR-Cas systems. In contrast, comparatively little is known about Acr diversity and function. Known Acrs inhibit only seven of the 25 CRISPR-Cas subtypes (Bondy-Denomy et al., 2018; Makarova et al., 2018), though it is quite likely that unidentified Acrs antagonize the remaining 18 groups (Pawluk et al., 2017). In cases where antagonists of CRISPR-Cas systems have been identified, many additional undiscovered Acrs almost certainly exist that antagonize those systems (Watters et al., 2018). These undiscovered Acrs likely act via a diversity of mechanisms to influence gene flow and phage dynamics in microbial communities (van Belkum et al., 2015; Westra et al., 2016; Borges et al., 2017), and may unlock new modes of manipulating phage- and CRISPR-Cas-enabled technologies (Sheth et al., 2016; Knott and Doudna, 2018).

Previous efforts to discover *acr*s have relied on phage genetics (Bondy-Denomy et al., 2013; Pawluk et al., 2014; Hynes et al., 2017; He et al., 2018), linkage to conserved genes (Pawluk et al., 2016a; Pawluk et al., 2016b; Hynes et al., 2018; Lee et al., 2018; Marino et al., 2018) or the presence of a self-targeting CRISPR spacer, which would create an unstable autoimmune state if not for the presence of

3

49 an Acr protein (Rauch et al., 2017; Watters et al., 2018). These clever strategies have revealed many
50 candidate *acr*s. But they have also highlighted the difficulty of finding new *acr* genes based on homology,
51 since *acr*s share little sequence conservation (Sontheimer and Davidson, 2017). As a result, most *acr*s
52 almost certainly lie unrecognized among the many genes of unknown function in phages, plasmids, and
53 other MGEs (Hatfull, 2015).

54

55   To overcome the challenges associated with anti-CRISPR discovery, we devised a functional
56 metagenomic selection that identifies *acr* genes from any cloned DNA, based on their ability to protect a
57 plasmid from CRISPR-Cas-mediated destruction. Because functional metagenomics selects for a
58 function of interest from large clone libraries (Handelsman, 2004), it is well-suited to identify individual
59 genes like *acr*s that have strong fitness impacts (Iqbal et al., 2014; Forsberg et al., 2015; Forsberg et al.,
60 2016; Genee et al., 2016). This approach may be particularly useful for Acr discovery because Acrs are
61 expressed from single genes and function readily in many genetic backgrounds (Pawluk et al., 2016a;
62 Rauch et al., 2017).

63

64   Using this functional selection, we find that many unrelated metagenomic clones from human oral
65 and gut microbiomes protect against *Streptococcus pyogenes* Cas9 (SpyCas9), the variant used most
66 commonly for gene editing applications (Knott and Doudna, 2018). We identify a broadly distributed but
67 previously undescribed Acr from the most potent Cas9-antagonizing clone in our libraries. This Acr,
68 named AcrIIA11, binds both SpyCas9 and double-stranded DNA (dsDNA) and exhibits a novel mode of
69 SpyCas9 antagonism, protecting both plasmids and phages from immune restriction. Thus, our functional
70 approach reveals not only new classes of Acrs, but also new mechanisms of CRISPR-Cas antagonism.

71

72 **Results**

73

74 **A functional metagenomic selection for type II-A anti-CRISPRs**

75   We designed a functional selection to isolate rare *acr*s from complex metagenomic libraries. We
76 based this selection on the ability of an *acr* gene product to protect a plasmid, which bears an antibiotic
77 resistance gene, from being cleaved by SpyCas9 (figures 1A, S1). By screening metagenomes, our
78 selection interrogates core bacterial genomes as well as DNA from the phages, plasmids, and other
79 mobile genetic elements that infect these bacteria, which must contend with CRISPR-Cas immunity.
80 Because most DNA inserts in large metagenomic libraries lack an *acr*, most clones in a library should be
81 susceptible to SpyCas9-mediated destruction. However, those few DNA inserts that encode a functional
82 Acr will resist SpyCas9 and can be recovered using the antibiotic resistance conferred by the plasmid
83 they protect.

84

85   In our *acr* selection scheme (figure S1), we targeted an inducible SpyCas9 nuclease to a kanamycin

86   resistance (*Kan^R*) gene on a plasmid used to construct the metagenomic libraries. Anticipating that

87   resistance to Cas9 cleavage could arise readily via point mutations in target sites, we targeted Cas9 to

88   two distinct loci within the *Kan^R* gene (figures 1B, S1) to reduce the frequency of these 'escape' plasmids.

89   We transformed metagenomic libraries into an *Escherichia coli* strain that contained SpyCas9 under an

90   arabinose-inducible promoter. After the cells were allowed to recover, we grew them overnight with

91   arabinose to induce SpyCas9 expression. During this phase, cells were not exposed to kanamycin and

92   thus were not under selection to maintain the metagenomic library bearing the *Kan^R* gene. We then plated

93   cells on solid medium with kanamycin, killing cells in which SpyCas9 cleaved the *Kan^R* gene and allowing

94   the recovery of metagenomic DNA from the few remaining Kan^R colonies.

95

96   Our analysis revealed that *Cas9* loss-of-function mutants (figure 1C) dominated the population in

97   early, single-iteration experiments, occurring in approximately 10^{-4} to 10^{-5} transformants (table S1, figure

98   S2). We therefore added a second iteration of SpyCas9 selection, reasoning that the Cas9 loss-of-

99   function rate would remain constant across iterations, whereas *acr*-encoding clones would be enriched

100   by a factor of $\sim10^4$ with each iteration. In theory, this enables us to select for an *acr* gene even if it is

101   present just once in a library of $\sim10^7$ clones. To accomplish these two rounds of Cas9 selection, we

102   purified total plasmid DNA from Kan^R colonies following one iteration, removed the original Cas9

103   expression plasmid via digestion, transformed the surviving plasmids into a fresh SpyCas9-expressing

104   strain, and exposed them to SpyCas9 selection a second time (figure S1). As expected, adding a second

105   iteration of Cas9 selection resulted in a significant enrichment for SpyCas9 antagonists above

106   background (figure 1D).

107

108   **SpyCas9 antagonism in human oral and fecal metagenomes**

109   We used our functional selection to search for *acr*s in five metagenomic libraries: two oral

110   metagenomes from Yanomami Amerindians (Clemente et al., 2015) and three fecal DNA metagenomes

111   from peri-urban residents of Lima, Peru (Pehrsson et al., 2016). We subjected each of these libraries,

112   with an estimated 1.3×10^6 - 3.4×10^6 unique clones per library, to SpyCas9 selection (table S2). For each

113   library, we observed a 10^4 to 10^5-fold reduction in the proportion of Kan^R colony forming units (CFU)

114   following one iteration of SpyCas9 selection. This value matches the reduction in Kan^R CFU seen for a

115   GFP control and the empirically determined frequency of Cas9 loss-of-function mutations (table S1,

116   figures 1D, S2). Since Kan^R is a measure of plasmid retention, this result indicates that most clones in

117   each library do not withstand SpyCas9. Following a second round of SpyCas9 selection for each library,

118   metagenomic inserts were amplified from pooled Kan^R colonies by PCR, deep-sequenced, and

119   assembled *de novo* using PARFuMS (Forsberg et al., 2012). Read coverage over each assembled contig

120   was used to estimate its abundance following selection. After quality-filtering (*e.g.* removal of low-

121   abundance contigs), we recovered a total of 51 contigs across all five libraries that putatively antagonize

122   SpyCas9 (tables S3, S4, figure S3).

123

124   After two rounds of SpyCas9 selection, two libraries (Oral_5, Fecal_01A) seemed more likely to

125   contain new *acr*s than the other three (Oral_3, Fecal_01E, Fecal_03G). The Oral_3 library poorly

126   withstood SpyCas9 restriction, so was not studied further (figure 1D). While the Fecal_03G library

127   completely resisted SpyCas9, subsequent analysis revealed that this was almost entirely due to a single

128   clone that acquired mutations in both Cas9 target sites (table S3, figures 1B, S3). Accordingly, just one

129   contig from this library passed quality filters. In contrast, the Fecal_01E library showed intermediate

130   SpyCas9 resistance and was largely devoid of 'escape' mutations (figure 1D). However, just one of the

131   18 contigs from this library was found to be phage-associated (table S4). Because *acr*s are expected to

132   originate in phages and MGEs (Borges et al., 2017; Pawluk et al., 2017; Sontheimer and Davidson,

133   2017), we predicted that contigs from Fecal_01E were unlikely to contain *bona fide acr*s and did not

134   prioritize them in this study. These contigs may nonetheless encode anti-Cas9 activity, perhaps via Cas9

135   regulatory factors employed by host bacteria rather than MGEs (Hoyland-Kroghsbo et al., 2017; Faure

136   et al., 2018), and therefore could represent a useful resource for probing host regulation of Cas9 activity.

137

138   Contigs from the Oral_5 and Fecal_01A libraries looked most promising for *acr* discovery. These

139   libraries conferred SpyCas9 resistance at levels 10-fold to 1,000-fold above background (figure 1D) and

140   had few escape mutations in Cas9 target sites, so likely withstood SpyCas9 due to functions encoded by

141   their DNA inserts (table S3). Many contigs from these libraries are phage-associated (30%; table S4,

142   figure 2A) and many encode genes of unknown function (figure S4); these are both hallmarks of known

143   *acr*s (Borges et al., 2017; Pawluk et al., 2017; Sontheimer and Davidson, 2017). Additionally, we

144   observed that Cas9 homologs are found in 15 of the 18 bacterial genera identified in these selections

145   (83%, table S4), a significant enrichment over the 12% of bacterial genera that harbor Cas9 (567/4822)

146   from a representative set of ~24,000 bacterial genomes (Mendler et al., 2018) ($p=4 \times 10^{-21}$, χ^2 test). This

147   enrichment strongly supports our hypothesis that many of the recovered contigs encode true *acr*s rather

148   than artifacts of functional selection, since phages and MGEs must encounter Cas9 to benefit from the

149   protective effect of *acr*s.

150

151   To confirm that the Oral_5 and Fecal_01A libraries encoded Acrs, we re-cloned the most abundant

152   contigs from these libraries into fresh plasmid and strain backgrounds and tested them individually for

153   SpyCas9 resistance. This step eliminated potential mutations to the plasmid backbone, the *cas9* gene,

154   or the host genome that may have accounted for Cas9 resistance in the original screen. For 10 of 14 re-

155   cloned contigs, the DNA insert still protected its parent plasmid from SpyCas9, confirming the power of

156   our selection to identify novel *acr*s from metagenomic DNA (figures 2A, 2B). Intriguingly, none of the

157   contigs recovered by functional selection encoded homologs of previously identified Acrs.

158

159   **Discovery of a widespread, potent Cas9 antagonist: AcrIIA11**

160   Among the 10 contigs we confirmed as having Cas9 antagonist activity, several features made us

161   focus on a single contig, F01A_2 (figure 2A). First, the F01A_2 contig was recovered from our iterative

162   selection of the Fecal_01A library, which conferred near-complete protection against SpyCas9 (figure

163   1D). Second, the F01A_2 contig was by far the most abundant contig from this library (with coverage

164   216-fold above the median coverage in the library), suggesting that it outperformed other contigs during

165   selection. In our re-testing, we confirmed that F01A_2 completely inhibited SpyCas9 activity (figure 2B).

166   Finally, F01A_2 shares near-perfect nucleotide identity (>99.9%) with a *Siphoviridae* phage. This phage

167   infects bacteria from the genus *Clostridium_Q* (figure 2C), within the family *Lachnospiraceae* (Parks et

168   al., 2018). As is typical for *Acr*-encoding loci, the five open reading frames (ORFs) on F01A_2 are small,

169   map to accessory regions of phage genomes, and appear to routinely undergo HGT (figure S5).

170

171   To identify the ORF(s) in F01A_2 responsible for SpyCas9 antagonism, we introduced an early stop

172   codon into each of the five predicted ORFs in the contig and re-tested this set of null mutants for anti-

173   Cas9 activity. *Orf_3* completely accounted for SpyCas9 inhibition: a null mutation in *orf_3* reduced the

174   frequency of Kan^R 10^5-fold, to the level of an empty-vector control (figure 3A). Furthermore, *orf_3* was

175   sufficient for SpyCas9 antagonism, protecting a target plasmid (figure 3B) from SpyCas9 approximately

176   as well as *acrIIA4*, the most potent inhibitor of SpyCas9 yet described (Rauch et al., 2017).

177

178   We also investigated the ability of *orf_3* to restore phage infection in the face of SpyCas9-mediated

179   immunity. We tested SpyCas9's ability to prevent Mu phage infection in the absence of any Acr or in the

180   presence of either *acrIIA4* or *orf_3*. Like with the plasmid protection assay, we found that *orf_3* provides

181   phage with nearly complete protection from SpyCas9, similar to *acrIIA4* (figures 3C, S6). Given its origin

182   from a phage contig, and its ability to potently inhibit SpyCas9 in both plasmid and phage protection

183   assays, we renamed *orf_3* from contig F01A_2 as *acrIIA11*, to distinguish it from *acrIIA4* and other

184   previously identified *acr*s (Borges et al., 2017; Pawluk et al., 2017; Lee et al., 2018; Uribe et al., 2019),

185   to which it bears no discernible homology.

186

187   Since *acrIIA11* is a newly discovered *acr*, we wished to determine its distribution in nature. We

188   therefore searched for homologs in NCBI and in IMG/VR, a curated database of cultured and uncultured

189   DNA viruses (Paez-Espino et al., 2017). We identified many proteins homologous to AcrIIA11 in both

190   phage and bacterial genomes and focused on a high-confidence set of homologs, those with ≥35% amino

191   acid identity over ≥75% of AcrIIA11's 182 amino acid sequence. AcrIIA11 homologs have a wider

192   phylogenetic distribution than most previously identified type II-A anti-CRISPRs (figures 3D, S7) and span

193   multiple bacterial phyla. We made a phylogenetic tree of AcrIIA11 homologs and found that they clustered

194   into three monophyletic clades (figure 4A), which correspond to the three bacterial taxonomic groups in

195   which they are found (figures 3D, S7). This concordance between gene and species clusters indicates

196   that, while HGT of *acrIIA11* may routinely occur across short phylogenetic distances (figure S5), intra-

197   class and intra-phylum gene flow is rare, in contrast to what has been observed for some other *acr*s

198   (Uribe et al., 2019).

199

200   AcrIIA11 appears restricted to specific bacterial taxa that are distantly related to *Streptococcus*.

201   Nevertheless, AcrIIA11 potently inhibits SpyCas9, even though SpyCas9 is highly diverged from the type

202   II-A Cas9 proteins found in AcrIIA11-containing taxa. For instance, the only Cas9 protein in *Clostridium_Q*

203   (NCBI accession CDD37961), AcrIIA11's genus-of-origin, shares just 32% amino acid identity with

204   SpyCas9. Intrigued by this observation, we tested divergent AcrIIA11 homologs from all three

205   phylogenetic groups against SpyCas9 and found that homologs from each group could inhibit its activity

206   (figure 4B). Type II-A CRISPR-Cas systems are enriched in the *Lachnospiraceae* relative to other

207   bacterial families (figure S7, p=5x10^{-27}, χ^2 test), which may explain why all tested AcrIIA11 homologs in

208   the 'a' group (figure 4A) inhibited SpyCas9 (figure 4B). Based on our findings, we hypothesize that

209   AcrIIA11 homologs may intrinsically possess the capacity to antagonize a diverse set of Cas9 orthologs.

210   Because homologs of AcrIIA11 are found in many genera prevalent within human gut microbiomes (figure

211   S7), its broad predicted range suggests that AcrIIA11 homologs will pose a meaningful barrier to Cas9

212   activity in this habitat – both in the context of natural phage infections as well as Cas9-based interventions

213   to manipulate microbiome composition (Sheth et al., 2016; Pursey et al., 2018).

214

215   **A novel mode of SpyCas9 antagonism by AcrIIA11**

216   Considering their relatively recent discovery, it is not surprising that very few Acrs have been

217   mechanistically characterized. Nonetheless, a common theme has emerged from biochemical studies of

218   AcrIIA2 and AcrIIA4, the only type II-A Acrs for which a mechanism of Cas9 inhibition has been

219   elucidated. These studies have shown that both AcrIIA2 and AcrIIA4 are dsDNA mimics; they inhibit

220   SpyCas9 by binding to its gRNA-loaded form and preventing association with a dsDNA target (Dong et

221   al., 2017; Shin et al., 2017; Yang and Patel, 2017; Jiang et al., 2019; Liu et al., 2019). Given that AcrIIA11

222   is an unrelated inhibitor, we sought to address whether it has the same mode of SpyCas9 antagonism

223   as these previously-studied type II-A Acr proteins.

8

224

225   We therefore purified recombinant AcrIIA11 from *E. coli* to analyze its biochemistry, interactions with

226   SpyCas9, and possible modes of inhibition. We first observed that untagged AcrIIA11 migrates at

227   approximately twice its predicted molecular weight by size exclusion chromatography (figure 5A),

228   suggesting that it may function naturally as a dimer. Validating our genetic findings, purified AcrIIA11

229   inhibited the dsDNA cleavage activity of SpyCas9 in a concentration-dependent manner (figures 5B and

230   5C). These data demonstrate that AcrIIA11 is sufficient for SpyCas9 inhibition and does not require

231   additional host factors. This autonomous activity is consistent with *acrIIA11*'s phylogenomic signature,

232   as it shows no linkage to any other gene across phage genomes (figure S5).

233

234   We considered several possibilities for AcrIIA11's mode of antagonism. We first tested whether

235   AcrIIA11 prevents SpyCas9 from binding gRNA. We found that AcrIIA11 has no clear effect on gRNA

236   binding, though it does bind gRNA at a large (160-fold) molar excess and may interact with the

237   SpyCas9/gRNA complex (figure S8). We therefore tested whether AcrIIA11 directly binds SpyCas9. We

238   found that AcrIIA11 was unable to bind the I-SmaMI meganuclease (used as a negative control) but

239   bound both the apo (without gRNA) and gRNA-loaded forms of SpyCas9. The addition of gRNA

240   enhanced AcrIA11 binding to SpyCas9 (figures 5D, S9) but not to the same extent as with AcrIIA4 binding

241   SpyCas9 (figure S9). Strong binding to the gRNA-loaded form of SpyCas9 has been previously

242   demonstrated for both AcrIIA2 and AcrIIA4 (Dong et al., 2017; Shin et al., 2017; Yang and Patel, 2017;

243   Jiang et al., 2019; Liu et al., 2019).

244

245   Given that AcrIIA11 has similar Cas9-binding properties as AcrIIA2 and AcrIIA4, we next tested

246   whether AcrIIA11 also acts as a DNA mimic to antagonize SpyCas9. Both AcrIIA2 and AcrIIA4 have the

247   key property of preventing gRNA-complexed SpyCas9 from binding target dsDNA (Dong et al., 2017;

248   Shin et al., 2017; Yang and Patel, 2017; Jiang et al., 2019; Liu et al., 2019). To test whether AcrIIA11

249   functions similarly, we performed an electrophoretic mobility shift assay (EMSA). Consistent with previous

250   work (Dong et al., 2017; Shin et al., 2017; Yang and Patel, 2017), we found that AcrIIA4 prevents the gel-

251   shift caused by SpyCas9/gRNA binding to its target dsDNA. In contrast, we observe that AcrIIA11 does

252   not prevent this gel-shift, even at molar ratios that abolish SpyCas9 cleavage activity (compare lanes 2-

253   5, figure 6A).

254

255   Instead of preventing a SpyCas9-induced gel-shift, we see that AcrIIA11 creates the opposite effect,

256   giving rise to a super-shifted SpyCas9/gRNA/dsDNA ternary complex (lane 4, figure 6A). This super-shift

257   may result from AcrIIA11 binding the ternary complex (as in figure 6B). Alternatively, it could result from

258   AcrIIA11 binding dsDNA overhangs unprotected by SpyCas9's footprint. Indeed, we find that AcrIIA11

259 weakly binds DNA (lanes 8 and 9, figure 6A) but that DNA-binding is greatly enhanced in the presence

260 of SpyCas9 (compare 'band #2' in lanes 4 and 12 with 8 and 11, figure 6A). Thus, AcrIIA11 and SpyCas9

261 could act together to bind DNA or one protein could stimulate the other's DNA-binding ability (figure 6B).

262 Although it remains unclear whether AcrIIA11's mode of antagonism is related to the 'super-shift' we

263 observe, our results nevertheless confirm that AcrIIA11 uses a different mechanism for SpyCas9

264 inhibition than AcrIIA2's or AcrIIA4's DNA mimicry.

265

266 **Discussion**

267   The identification and characterization of CRISPR-Cas systems has outpaced the discovery of anti-

268 CRISPRs, even though Acr diversity almost certainly exceeds CRISPR-Cas diversity in nature (Pawluk

269 et al., 2017; Watters et al., 2018). It seems likely that these undiscovered Acrs are important influences

270 on both phage infection outcomes and horizontal gene transfer in natural microbial communities (Borges

271 et al., 2017). Determining the impact of Acrs on these processes is a challenge, however, because *acr*s

272 lack shared sequence features so are difficult to detect. To overcome this obstacle and enable function-

273 based *acr* discovery, we developed a high-throughput selection that identifies *acr*s based solely on their

274 activity. Applying our functional selection to human fecal and oral metagenomes, we recovered 51 DNA

275 fragments that putatively antagonize SpyCas9, with 10 confirmed for activity and many additional host-

276 or phage-encoded inhibitors expected among the remaining 41 sequences.

277

278   In parallel work to ours, Uribe and colleagues recently described a functional metagenomic scheme

279 to identify *acr*s (Uribe et al., 2019). They used a single round of Cas9 selection and one targeting crRNA

280 to identify putative SpyCas9-antagonizing clones, though the authors suspected many of these to be

281 false positives. To find *acr*s, they selected 39 ideal candidates from their initial dataset for re-testing,

282 confirming 11 DNA fragments for anti-Cas9 activity, from which they identified four new *acr*s (*acrIIA7-*

283 *10*). Although the basic premise of both of our approaches is similar, our strategy uses two iterations of

284 Cas9 selection and two targeting crRNAs to identify *acr*s, which suppresses both Cas9 loss-of-function

285 mutations and escape mutations in the target plasmid. As a result, we confirmed that 10 of the 14 most

286 abundant clones following selection could antagonize SpyCas9 (figure 2B). More generally, we show that

287 iterative selection can enable high-confidence discoveries from high-diversity inputs, which will be critical

288 in the search for even rarer classes of *acr*s and other phage counter-defense strategies.

289

290   One of the major surprises from our work and that of Uribe *et al* is that most *acr*s in nature have not

291 been discovered. Neither study encountered a previously described *acr* and no *acr*s were shared across

292 datasets, though more than 10^7 unique clones were screened for antagonists of the same Cas9 allele

293 (SpyCas9) in the same selection host (*E. coli*). This lack of overlap emphasizes that the few *acr*s which

10

294 have been catalogued are a very minor subset of those that are found in nature. This tremendous diversity
295 exists because selection favors a highly-varied repertoire of *acr*s among even related MGEs (Bondy-
296 Denomy et al., 2013; Borges et al., 2017) and Acrs have evolved independently many times from a variety
297 of progenitor proteins (Rollins et al., 2018; Stone et al., 2018; Uribe et al., 2019). With extreme diversity
298 favored, and the relative ease by which new Acr function is born, it is almost certain that much more
299 mechanistic variety among Acrs remains to be discovered.

300

301 In this study, we have focused on the most potent SpyCas9 antagonist from our set, extensively
302 characterizing the phylogenetic distribution, functional range, and mechanism of inhibition for its causal
303 Acr, AcrIIA11. SpyCas9 is unlikely to be the natural target of AcrIIA11, as it is quite divergent from the
304 Cas9 homologs found in AcrIIA11-encoding bacteria (Chylinski et al., 2014). Yet, diverse AcrIIA11
305 homologs, which are distributed across multiple phyla, retain inhibitory activity against SpyCas9. This
306 suggests that AcrIIA11 is intrinsically predisposed to act against a wide variety of Cas9 sequences. Such
307 inhibitory breadth may be particularly useful to MGEs that infect bacteria with diverse CRISPR-Cas
308 systems (Makarova et al., 2015; Crawley et al., 2018) and could explain AcrIIA11's broad phylogenetic
309 distribution. This combination of prevalence and inhibitory breadth could also impact many Cas9-based
310 interventions (Sheth et al., 2016; Pursey et al., 2018), as a pre-existing set of Acrs could quickly overcome
311 diverse Cas9 homologs given the appropriate selective pressures. For instance, Cas9 inhibitors offer a
312 means to resist Cas9 gene drives, so will have high selection coefficients if encountered during a gene
313 drive (Bull and Malik, 2017). Gene drives, therefore, may inadvertently select for Acrs, especially those
314 already abundant in microbiomes and with a high proclivity for HGT (Maxwell, 2016).

315

316 Broad inhibitory activity implies that AcrIIA11 might interact with conserved residues on Cas9.
317 AcrIIC1, a broad-spectrum inhibitor of type II-C Cas9 enzymes, binds to conserved Cas9 residues to
318 prevent nuclease activity but not target DNA binding (Harrington et al., 2017). No type II-A Cas9 inhibitor
319 has been shown to act similarly, though AcrIIA11's behavior is consistent with such a mechanism; it binds
320 SpyCas9, inhibits DNA cleavage, but does not prevent target recognition. Alternatively, AcrIIA11's broad
321 predicted range may result from an indirect mode of inhibition that Cas9 cannot trivially escape. Such a
322 mechanism could be related to AcrIIA11's gRNA-binding ability, its DNA-binding capacity, or due to an
323 interaction with SpyCas9 that alters these properties in either protein. While AcrIIA11's exact mechanism
324 still awaits elucidation, its mode of antagonism is clearly distinct from the other two mechanistically-
325 characterized type II-A Acrs, AcrIIA2 and AcrIIA4, which act as dsDNA mimics to prevent target
326 recognition (Dong et al., 2017; Shin et al., 2017; Yang and Patel, 2017; Jiang et al., 2019; Liu et al.,
327 2019). A combination of Acrs that act via different mechanisms may be more effective at inhibiting
328 CRISPR-Cas activity than a combination of Acrs that act redundantly (Borges et al., 2017). Thus,

11

329   AcrIIA11's novel mode of antagonism could enable more potent SpyCas9 inhibition than can be achieved

330   using only dsDNA mimics.

331

332   Excitingly, AcrIIA11 represents just the tip of the iceberg. We have identified 50 additional

333   metagenomic clones that putatively antagonize SpyCas9 (with nine confirmed for activity); many of these

334   sequences are likely to contain new SpyCas9 inhibitors, *i.e.* ones not homologous to any known Acr.

335   Functional metagenomics not only offers a powerful means to identify new *acr* genes, but also enables

336   discovery of fundamentally new ways to inhibit Cas9. More broadly, our functional metagenomic

337   approach allows us to detect *acr*s beyond curated sequence databases and can enable the discovery of

338   Acrs against any CRISPR-Cas system of interest, in any microbial habitat, strain collection, or phage

339   bank. This information will improve our understanding of phage and MGE dynamics in microbial

340   ecosystems and holds significant promise for not only describing the longstanding evolutionary arms race

341   between phages and bacteria, but also for improving phage- and CRISPR-Cas-based technologies used

342   in therapeutic and engineering applications.

343 **Materials and Methods**

344

345 **Processing of metagenomic libraries**

346   For each metagenomic library (table S2, given generously by Gautam Dantas, Washington Univ. in

347 St. Louis), the original freezer stock (Clemente et al., 2015; Pehrsson et al., 2016) was inoculated into

348 52 ml and grown to an OD600 value of ~0.7 (300 µl/100 µl inoculum for oral/fecal libraries). All libraries

349 were previously constructed in a pZE21 MCS1 plasmid backbone (henceforth, pZE21) marked by

350 kanamycin resistance (Kan^R) (Lutz and Bujard, 1997; Clemente et al., 2015; Pehrsson et al., 2016). Each

351 library was then titered on agar plates containing lysogeny broth (LB; 10 g/L casein peptone, 10 g/L NaCl,

352 5 g/L ultra-filtered yeast powder) with 50 ug/ml Kanamycin and 10-12 ml was used to create replicate

353 freezer stocks. The remaining cells were pelleted by centrifugation at 4100 rcf for 6.5 minutes, purified

354 using Qiagen miniprep kits (four columns per library), and quantified using the Qubit BR dsDNA

355 quantification kit. The libraries Oral_3 and Oral_5 are combinations of two smaller libraries (Clemente et

356 al., 2015); each smaller library was processed independently and then combined in proportion to the

357 unique clones in each library before being transformed into *E. coli*.

358

359 **Cloning Cas9 expression constructs**

360   A series of slightly different plasmids were used to express *S. pyogenes* Cas9 during the development

361 of the Acr selection scheme and follow-up work (generically, pCas9). Plasmid composition, applications,

362 and resultant data are detailed in tables S1, S5, and figure S2. Early versions of pCas9 were built by

363 targeting a previously described Cas9 expression vector (addgene #48645) (Esvelt et al., 2013) to pZE21

364 with new crRNAs. The targeting crRNAs were engineered into the original locus (Esvelt et al., 2013) by

365 Gibson cloning with long-tailed primers. All Gibson assembly was performed using the NEBuilder HiFi

366 DNA assembly mastermix (NEB cat# E2621) per manufacturer's recommendations. The arabinose-

367 inducible *araC* and pBad regulon was amplified from pU2 (Lee et al., 2015) and, using Gibson assembly,

368 inserted into pCas9 in place of the constitutive proC promoter to regulate Cas9 expression. To create the

369 first 2-crRNA pCas9 construct, a gBlock was ordered from Integrated DNA Technologies (IDT) containing

370 an I-SceI cut site and second crRNA (sequence crRNA 'Z' in table S5) under the control of pJ107106 and

371 a T7TE-LuxIA terminator. This gBlock was cloned into pCas9 by Gibson assembly. Swapping crRNAs at

372 this locus was also performed by Gibson cloning with long-tailed primers. The pJ107106 promoter was

373 converted to pJ107111 (Zucca et al., 2015) using the Q5 site-directed mutagenesis kit (NEB), with

374 suggested protocols. All pCas9 constructs were transformed into *Escherichia coli* (NEB Turbo) by heat-

375 shock and made electrocompetent as described (protocols.io, 2015).

376

Metagenomic selection for Cas9 antagonists

In the first iteration of Cas9 selection, 360 ng of each metagenomic library was electroporated into a strain of ultra-competent *E. coli* (NEB Turbo) containing the spectinomycin-marked Cas9 expression plasmid (pCas9). As a control, 360 ng of pZE21 expressing GFP was also transformed. All electroporations occurred in a 1 mM cuvette using a Bio-Rad Gene Pulser Xcell with the following settings: 2.1kV, 100 Ω, 25μF. Following electroporation, cells recovered for three hours at 37°C in SOC media (NEB, cat. # B9020S). An aliquot was then taken to titer transformants and the remaining cells (860 µl) were inoculated into 25 ml of LB broth with 50 ug/ml spectinomycin (Spec) and 2 mg/ml arabinose. The GFP control was split across two 25 ml flasks, one with and one without arabinose (430 µl inoculum per flask). After 20 hours in selective conditions, all samples were titered on LB-Spec or LB-Kan/Spec plates. Figure 1D depicts the population proportion of Kan^R colony forming units (cfu) at 20 hours after Cas9 induction relative to their proportion before Cas9 induction.

To isolate plasmids surviving one round of selection, colonies from LB-Kan/Spec titer plates (table S2) were collected in 3 ml of LB-Spec after scraping colonies with an L-shaped cell scraper (Fisher Scientific cat. # 03-392-151) to gently remove them from the agar. Colonies were collected from all metagenomic libraries and the GFP control exposed to arabinose. One-third of the cells were used to make -80°C freezer stocks and plasmids from the remaining 2 ml were purified across two Qiagen miniprep columns into a combined 100 µl of nuclease-free H$_2$O. An I-SceI restriction site was engineered into pCas9 to enable its removal via treatment with the homing endonuclease. For all samples, 51 µl of miniprep eluate was combined with 6 µl NEB cutsmart buffer and 3 µl (15 units) I-SceI, incubated for 20 hours at 37°C, and the reaction heat-killed at 65°C for 20 minutes. After withholding 5 µl of each sample for gel electrophoresis, 2.98 µl of the *E. coli* RecBCD enzyme (149 units, given generously by Andrew Taylor), 4.9 µl 10mM ATP, and 1.62 µl NEB cutsmart buffer were added to the remaining 55 µl of each sample and incubated for one hour at 37°C. To kill the RecBCD reaction, EDTA was added to a final concentration of 20mM and the sample incubated at 70°C for 30 minutes. Linearization of pCas9 by I-SceI and the subsequent removal of linear DNA by RecBCD was confirmed by visualization with gel electrophoresis. Plasmid preparations were then purified through a Zymo Research DNA Clean & Concentrator column and once-selected metagenomic DNA libraries electroporated into the original Cas9 selection strain a second time (table S2). For the second iteration of Cas9 selection, electroporation, recovery, induction, titering, and outgrowth was performed exactly as described for the first iteration.

Purification and amplification of metagenomic DNA from Cas9 selections

From twice-selected metagenomic libraries, LB-Kan/Spec titer plates were used to collect Kan^R clones from the population. Two titer plates were processed independently for each library and were selected based on colony counts between (roughly) 10^2 and 10^4 (table S2). Only one titer plate was

14

412 collected for the GFP control. For each plate, colonies were collected, stored, and plasmid purified as
413 described above. No enzymatic removal of pCas9 was performed. Instead, purified plasmids were diluted
414 ten-fold and metagenomic DNA fragments amplified via PCR using 45µl Platinum HiFi polymerase mix
415 (Thermo Fischer cat. # 12532016), 1.67 µl plasmid template, and 5 µl of a custom primer mix. The custom
416 primer mix contained three forward and three reverse primers, each targeting the sequence immediately
417 adjacent the metagenomic clone site in pZE21, staggered by one base pair. The stagger enabled diverse
418 nucleotide composition during early Illumina sequencing cycles. A sample PCR contained the following
419 primer volumes, each from a 10µM stock: (primer F1, 5′-CCGAATTCATTAAAGAGGAGAAAG, 0.83µl); (primer
420 F2, 5′- CGAATTCATTAAAGAGGAGAAAGG, 0.83µl); (primer F3, 5′- GAATTCATTAAAGAGGAGAAAGGTAC, 0.83µl);
421 (primer R1, 5′- GATATCAAGCTTATCGATACCGTC, 0.36µl); (primer R2, 5′- CGATATCAAGCTTATCGATACCG,
422 0.71µl); (primer R3, 5′-TCGATATCAAGCTTATCGATACC, 1.43µl). PCRs were then executed using the
423 following conditions: 94°C for 2 minutes, 30 cycles of: 94°C for 10 seconds + 55°C for 30 seconds + 68°C
424 for 5.5min, and 68°C for 10 minutes. The amplified metagenomic fragments were cleaned using a Zymo
425 Research DNA Clean & Concentrator column and eluted in 80 µl of Qiagen elution buffer.

426

Illumina sample preparation and sequencing

428 For each sample, 50 µl of PCR amplicons were diluted to a final volume of 130 µl in Qiagen elution
429 buffer and sheared to a fragment size of approximately 500 bp using a Covaris LE220 sonicator. Sheared
430 DNA was purified and concentrated using a Zymo Research DNA Clean & Concentrator column and
431 eluted in 25 µl nuclease-free H_2O. DNA was then end-repaired for 30 minutes at 25°C using 250 ng
432 sample in a 25 µl reaction with 1.5 units T4 Polymerase (NEB cat. # M0203), 5 units T4 polynucleotide
433 kinase (NEB cat. # M0201), 2.5 units Taq DNA polymerase (NEB cat. # M0267), 40 µM dNTPs, and 1x
434 T4 DNA ligase buffer with 10 mM ATP (NEB cat. # B0202). Reactions were then heat-killed at 75°C for
435 20 minutes and 2.5 µl of 1 µM pre-annealed, barcoded sequencing adapters added with 0.8 µl of NEB
436 T4 DNA ligase (cat. # M0202T). This reaction was incubated at 16°C for 40 minutes and heat-killed at
437 65°C for 10 minutes. The barcoded adapters contained a 7bp sequence specific to each sample, which
438 allowed mixed-sample sequencing runs to be de-multiplexed. Forward and reverse sequencing adapters
439 were stored at 1 µM in TES buffer (10mM Tris, 1mM EDTA, 50 mM NaCl, pH 8.25) and annealed by
440 heating to 95°C and cooling at a rate of 0.1°C/sec to a final temperature of 4°C. Adapters were stored at
441 -20°C and thawed slowly on ice before use. After adapter-ligation, samples were purified through a Zymo
442 Research DNA Clean & Concentrator column and eluted in 12 µl of the supplied elution buffer.

443 Next, 6 µl from each sample was pooled together and the mixture size-selected to ~500-900 bp using
444 a 1.9% agarose gel visualized with SYBR-Safe dye (Thermo Fisher cat. # S33102). DNA was purified
445 from gel slices using a Zymo Research gel DNA recovery kit, eluting in 28µl nuclease-free H_2O. Purified
446 DNA was then enriched by PCR; a sample 25 µl reaction (in 1x Phusion HF buffer) contained 6 µl purified

447　DNA template, 0.5 µl dNTPs (10mM), 0.25 µl Phusion polymerase (2 units/µl, Thermo Fisher cat. # F530),

448　and 1.25 µl each of the following primers (10 µM stock): (primer F, 5' –

449　AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT); (primer R, 5' –

450　CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT). PCRs were amplified for

451　30 seconds at 98°C, subjected to 18 cycles of 98°C for 10 seconds, 65°C for 30 seconds, and 72°C for

452　30 seconds, and then incubated at 72°C for 5 minutes. Following enrichment PCR, samples were size-

453　selected a second time to ~500-900 bp as previously described and purified libraries quantified using the

454　Qubit fluorimeter (HS assay). Finally, 300 cycles of paired-end sequence data were generated at the

455　Fred Hutchinson Cancer Research Center Genomics Core using a 17 pM loading concentration on an

456　Illumina MiSeq machine with the v3 reagent kit.

457

458　**Assembly and annotation of functional metagenomic selections**

459　　Illumina paired-end reads were binned by barcode (perfect match required) such that data from each

460　titer plate was assembled and annotated independently. Metagenomic DNA fragments from each sample

461　were assembled using the first 93bp of each read with PARFuMS, a tool developed specifically for

462　assembling small-insert functional metagenomic selections and which was optimized for shorter read

463　lengths (Forsberg et al., 2012). The PARFuMS assembler uses three iterations of Velvet (Zerbino and

464　Birney, 2008) with variable job size, two iterations of PHRAP (de la Bastide and McCombie, 2007), and

465　custom scripts to clean reads, remove assembly chimeras, and link contigs by coverage and shared

466　annotation, as previously described (Forsberg et al., 2012). In sum, 206 contigs were assembled from all

467　ten titer plates (encompassing five selections, figure S3). We predicted open reading frames (ORFs) from

468　each contig with MetaGeneMark (Zhu et al., 2010) using default parameters. Proteins were annotated by

469　searching their amino acid sequences against the TIGRFAMs (Haft et al., 2001) and Pfam (Bateman et

470　al., 2000) profile HMM databases using HMMER3 (Finn et al., 2011). The highest-scoring profile was

471　used to annotate each ORF (minimum e-value $1e^{-5}$); these automated annotations were curated by hand

472　into the functional groups shown in figure S4. The NCBI taxonomies of these contigs were determined

473　using BLASTn against the 'nt' database; all contigs could confidently be assigned NCBI taxa at the

474　resolution of order (figure S4, table S4). Seven assembled contigs matched pCas9 (>99% nucleotide

475　identity) and were removed from the dataset. We also discarded contigs covered by less than 2

476　reads/base-pair/million reads (RPBM) to ensure contigs assembled from background DNA (*i.e.* plasmid

477　backbone, host genome) were not considered (the coverage maximum among pCas9-derived contigs

478　was 1.67 RPBM, which informed our 2 RPBM cutoff). Then, contigs which could be linked to a mutation

479　in either protospacer or protospacer adjacent motif (PAM) were eliminated (see methods section below).

480　Note that we use the term 'target site' to refer to a particular protospacer/PAM sequence. Subsequently,

481　redundant contigs assembled across both titer plates from a given library were removed with CD-HIT (Li

16

482 and Godzik, 2006), using the following parameters: -c 0.95, -aS 0.95, -g 1 (the shorter contigs among

483 clusters sharing 95% nucleotide identity over 95% the length of the shorter contig were removed). We

484 checked for cross-contamination among the non-redundant contigs with a BLASTn search against the

485 unfiltered contig set, detecting (and discarding) just one cross-contaminating contig (in library

486 Fecal_01E). Manual curation of this near-final dataset prompted removal of one chimeric contig and the

487 fusion of two incomplete assemblies. Note that the final dataset may exclude some metagenomic clones

488 that inhibit Cas9, specifically those with mild antagonism or high fitness costs (rare among surviving

489 colonies, with coverage < RPBM=2) and those which have acquired target site mutation yet also encode

490 antagonism (partial antagonism may potentiate target site mutation). Indeed, one contig (O5_7) contains

491 a mutation in the PAM of target site B yet still confers Cas9 protection when re-cloned into a wild-type

492 vector backbone (figure 2B). On the basis of empirical antagonism, we include this contig in our final

493 dataset (to give 51 final contigs across five selections, tables S3, S4), but exclude all untested contigs

494 linked to target site mutations. While these exclusions may omit *bona-fide* Cas9 antagonism, their

495 exclusion yields a dataset with higher-confidence enrichment for Cas9 antagonists. All final assembled

496 contigs have been submitted to GenBank.

497

498 **Determining coverage of assembled contigs**

499   Reads used as input for each assembly (i.e. those after adapter-trimming, vector-masking, and spike-

500 in removal by PARFuMs (Forsberg et al., 2012)) were also used to calculate coverage across each

501 contig. For each sample, these reads were mapped to assembled contigs with bowtie2 (local alignment,

502 unpaired reads, very-fast mode) and RPBM calculated for each contig.

503

504 **Determining the proportion of wild-type target sites**

505   To estimate the proportion of wild-type target sites in each library after two rounds of Cas9 selection,

506 titer-specific reads (before vector-cleaning) were concatenated by library and mapped to the pZE21

507 vector with bowtie2 (end-to-end alignment, unpaired reads, sensitive mode). For the GFP control, reads

508 were used from the only titer plate collected. Base frequencies at each protospacer and PAM position

509 were calculated using bam-readcount (github.com/genome/bam-readcount), requiring minimum mapping

510 and base quality scores of 25 and 30, respectively. The proportion of wild-type target sites was

511 conservatively calculated as the product of wild-type base calls at each protospacer/PAM position. In two

512 libraries, (Oral_3 and Fecal_01E), sanger sequencing of individually picked colonies revealed a

513 recombination event between a metagenomic DNA fragment and target site B, which would remove this

514 Cas9 target from these clones (but avoid frameshift within the Kan^R gene). For these two libraries, reads

515 were mapped to a variant of the pZE21 backbone corresponding to each recombination mutant. The

516 proportion of reads mapping across the recombination site was compared to the proportion mapping to

17

517   the wild-type vector backbone to estimate the frequency of recombination mutants in the population (0.4%

518   and 24.2% for Oral_3 and Fecal_01E, respectively). The overall proportion of wild-type target sites in

519   these libraries was then adjusted to reflect the proportion of these recombinants (table S3).

520

521   **Linking contigs to target site genotype**

522   For some of the most abundant contigs in the final dataset, the genotype of the target site sequences

523   in the corresponding plasmid backbone could be determined by Sanger sequencing individual Kan^R

524   clones following two rounds of Cas9 selection. To link additional contigs with target site genotype in the

525   clones not sampled by our Sanger data, we examined paired-end reads that mapped to target site

526   sequences and an assembled contig. Because we used PCR to amplify metagenomic fragments before

527   Illumina library creation, only a small minority of reads mapped to the vector backbone. Additionally,

528   target sites A and B are 514 bp and 365 bp from the clone site of pZE21, respectively, so only read pairs

529   from long sequencing inserts could link target site and contig genotype. Accordingly, many assembled

530   contigs remain without links to target site genotype (figure S3), though the overall proportion of wild-type

531   target sites serves as a proxy for the number of Cas9 antagonists versus escape mutants in a given

532   library (table S3).

533   For each titer plate, paired-end reads were mapped to a reference database containing all assembled

534   contigs in both forward and reverse orientations within the pZE21 plasmid. We used bowtie2 in "no-

535   mixed" mode to consider only paired reads with an insert size ≥515 bp when mapping against this

536   reference database (additional mapping options: end-to-end alignment, paired reads, sensitive mode).

537   Only reads with a mapping quality score over ten were considered. We determined the true orientation

538   of a metagenomic DNA fragment by counting whether more reads mapped across the clone-site junctions

539   of the contig in the forward or reverse orientation. If more than one mapped read supported a particular

540   protospacer/PAM variant, the contig linked to that target site was removed from the final dataset.

541   Similarly, if fewer than ten reads mapped to a target site, a single read with any deviation from a wild-

542   type protospacer/PAM was sufficient to eliminate the corresponding contig from further consideration.

543

544   **Predicting phage-associated contigs**

545   The assembled metagenomic contigs (N50 length 2.2 Kb) were too short to provide sufficient

546   information content for phage prediction based on their nucleotide sequence (Roux et al., 2015). Instead,

547   protein sequences were used: those with close relatives in predicted phages were used to classify their

548   parent contig as phage-associated. Predicted proteins from each contig were queried against NCBI's NR

549   database using BLASTp (on September 12th, 2017) and the top five unique hits retained from each query.

550   For each BLAST hit, we downloaded sequence 25 Kb upstream and downstream of the corresponding

551   gene sequence and used these ~50KB sequences as input for phage prediction with VirSorter (Roux et

552 al., 2015). Sequences were compared against VirSorter's more encompassing 'virome' database (using

553 default parameters) and hits to bacteriophage of any confidence category (from 'possible' to 'most

554 confident') were considered phage-associated (Roux et al., 2015). The proteins that seeded these

555 bacteriophage predictions were used then to classify metagenomic DNA contigs. If any predicted protein

556 from our assembled contigs had a top-five BLAST hit that seeded a bacteriophage prediction, the whole

557 metagenomic contig was classified as being phage-associated. These classifications are referenced in

558 figure 2A and table S4.

559

560 **Validating contigs with Cas9 protection**

561 All contigs assembled from libraries Fecal_01A and Oral_5 with coverage above the sample median

562 (RPBM >3.6, Fecal_01A and RPBM > 8.2, Oral_5) were chosen for validation, totaling 15 contigs. Five

563 contigs were represented in a sparsely-sampled set of individually-picked and Sanger-sequenced

564 colonies. In these cases, the corresponding plasmids were re-transformed into NEB Turbo, target site

565 sequences confirmed as wild-type by additional Sanger Sequencing, and the plasmids co-transformed

566 with pCas9 into NEB Turbo for functional testing. The remaining contigs were amplified by PCR from the

567 plasmid preparations that followed two iterations of SpyCas9 selection (those used in PCRs for Illumina

568 library construction). PCR primers (table S6) targeting the junction between pZE21 and the assembled

569 contigs were used to facilitate re-cloning of the amplified fragments into pZE21 by Gibson assembly.

570 PCRs used the high-fidelity Q5 polymerase (NEB) per suggested protocols. Cycling conditions are as

571 follows: 98°C for 3 minutes, 35 cycles of 98°C for 15 seconds + 61°C or 65°C for 30 seconds + 72°C for

572 3 minutes, and 72°C for 10 minutes. Reaction-specific annealing temperatures are listed in table S6 and

573 PCR products of the expected size were purified from gel slices using a Zymo Research gel DNA

574 recovery kit. Nine of ten PCRs were successful [a contig from Fecal_01A (RPBM=3.66) did not amplify

575 well]. All Gibson assembly was performed using the NEBuilder HiFi DNA assembly mastermix per

576 manufacturer's recommendations. Following Gibson-assembly, sequence-verified clones were co-

577 transformed with pCas9 into NEB Turbo, resulting in 14 total strains re-tested for Cas9 antagonism.

578 These 14 strains and a control carrying an empty pZE21 vector were grown overnight in LB with 50

579 µg/ml kanamycin and spectinomycin (LB-Kan/Spec) to maintain the pZE21 variants and pCas9,

580 respectively. The next morning, cultures were diluted 30-fold into LB-Kan/Spec and grown to log phase

581 (absorbance readings at 600 nm ranged from 0.2 to 0.4). Mid-log cultures were then inoculated into LB

582 with 50 ug/ml spectinomycin and either 0.2 mg/ml arabinose (to induce Cas9) or no arabinose.

583 Kanamycin was omitted to allow for restriction of the pZE21 plasmid. Inocula were normalized by optical

584 density (a 1:40 inoculum was used for an absorbance reading of 0.4). Cultures were then arrayed in

585 triplicate across a 96-well plate (Greiner cat#655083, 200µl per well) and grown in a BioTek Cytation 3

586 plate reader at 37°C with linear shaking at 1096 cycles per minute (cpm). After six hours of growth (at

587 which time cultures had reached stationary phase), each well was serially-diluted ten-fold in peptone-

588 NaCl (1 g/L each, pH 7.0). Spots (5 µl) of each dilution were plated on agar plates with LB-Kan/Spec (to

589 count Kan^R colonies) and LB-Spec (to count total colonies). Figure 2B depicts the log-transformed

590 proportion of Kan^R/total cfu with and without Cas9 induction.

591

Identification of causal Acr proteins

593 Null alleles of each predicted ORF in F01A_2 were created by inserting an early stop codon into each

594 predicted protein. Stop codons were inserted by PCR using the high-fidelity Q5 polymerase (NEB) per

595 suggested protocols with the following cycling conditions: 98°C for 3 minutes, 35 cycles of 98°C for 15

596 seconds + [annealing temperature] for 30 seconds + 72°C for [extension time], and 72°C for 10 minutes.

597 Reaction-specific primers and conditions are listed in table S6. Constructs were then prepared using the

598 Q5 site-directed mutagenesis kit (NEB cat #E0554) or by Gibson assembly (using the NEBuilder HiFi

599 assembly mastermix, cat# E2621) per manufacturer's recommendations. Final constructs were

600 sequence-verified, co-transformed with pCas9 into NEB Turbo, and re-tested for Cas9 antagonism

601 alongside an empty-vector control and the parent pZE21-F01A_2. Plasmid protection assays were

602 performed exactly as described in the section titled 'Validating contigs with Cas9 protection', except that

603 Cas9 expression was induced using 2 mg/ml arabinose (rather than 0.2 mg/ml).

604

Cloning candidate Acrs

606 The third ORF (*acrIIA11*) from F01A_2 and *acrIIA4* were codon-optimized for *E. coli* and sub-cloned

607 into the KpnI (5') and HindIII (3') sites of a pZE21 containing the *tetR* gene to allow for doxycycline-

608 induced expression of the candidate Acrs from the pLtetO-1 promoter. To generate this plasmid, *tetR*

609 (and its pLac promoter + rrnB terminator) was amplified from pCRT7 (addgene #52053) with NEB's Q5

610 polymerase per suggested protocols and using the conditions listed in table S6. This amplicon was then

611 cloned by Gibson assembly into the pZE21 backbone. To clone candidate Acrs, KpnI and HindIII sites

612 were added to each gene by PCR with the Q5 polymerase (see table S6), vectors and genes digested

613 using these enzymes, and sticky-end ligations performed using the Fast-Link ligation kit (epicentre cat#

614 LK0750H) per suggested protocols. The *acrIIA11 and acrIIA4* gene sequences were synthesized as

615 gBlocks from IDT (the AcrIIA4 amino acid sequence is identical to NCBI accession AEO04689.1).

616 AcrIIA11 homologs were codon-optimized for *E. coli*, synthesized by GenScript, and cloned into this

617 identical vector context. All constructs were co-transformed with pCas9 into NEB Turbo prior to

618 performing plasmid protection assays. Because NEB Turbo contains the *lacI^q* mutation, 0.5 mM IPTG

619 was used throughout cloning and handling of the pZE21_tetR plasmid in this strain. IPTG was used

620 because *TetR* is driven by the pLac promoter in this construct. Including IPTG ensured that TetR was at

621   sufficient intracellular concentrations to maintain repressed transcription from the pLtetO-1 promoter prior

622   to Acr induction.

623

Testing candidate Acrs for plasmid protection

625   Overnight cultures of candidate Acr constructs were grown in LB-Kan/Spec + 0.5mM IPTG (LB-

626   Kan/Spec/IPTG) to maintain pCas9 and the Acr variants in an uninduced state. The next morning,

627   cultures were diluted 1:50 into LB-Kan/Spec/IPTG and grown at 37°C for approximately two hours to mid-

628   log with or without 100 ng/ml doxycycline to control Acr expression (final absorbance readings ranged

629   from 0.2 to 0.6). Mid-log cultures were then inoculated into LB with 50 ug/ml spectinomycin, 0.5 mM

630   IPTG, and either 2 mg/ml arabinose (to induce Cas9) or no arabinose. Doxycycline was included in this

631   medium at 100 ng/ml to maintain Acr expression only for previously induced cultures. Kanamycin was

632   omitted to allow for elimination of the pZE21 target plasmid and inocula were normalized by optical

633   density (a 1:40 inoculum was used for an absorbance reading of 0.4). Cultures were grown in a 96-well

634   plate reader and the proportion of Kan^R cfu determined exactly as described in the section 'Validating

635   contigs with Cas9 protection'. Figures 3 and 4 depicts the log-transformed proportion of Kan^R/total cfu

636   with and without Cas9 induction for each Acr.

637

Phage plaquing assay

639   Overnight cultures with pCas9 and pZE21_tetR expressing either GFP or an Acr were diluted 1:50 in

640   LB-Kan/Spec/IPTG supplemented with 5 mM MgSO$_4$ and grown shaking at 37C for three hours until late-

641   log phase (OD600 range 0.5-0.8, see table S5 for exact crRNA sequences). GFP and the candidate Acrs

642   were induced by adding doxycycline to a final concentration of 100 ng/ul and cultures grown for another

643   two hours before SpyCas9 was induced by adding 0.2 mg/ml arabinose. After three additional hours of

644   growth in the presence of both inducers, 200μl of culture was used in a top-agar overlay, allowed to

645   harden, and ten-fold serial dilutions of phage Mu spotted on top. The top and bottom agar media were

646   made with LB-Kan/Spec/IPTG supplemented with 5mM MgSO4, 0.02 mg/ml arabinose, and 100 ng/ul

647   doxycycline and contained 0.5% and 1% Difco agar, respectively. Plates were incubated at 37°C

648   overnight and plaques imaged the following day.

649

AcrIIA11 homolog discovery and phylogenetic placement

651   To determine the distribution AcrIIA1, AcrIIA2, AcrIIA3, AcrIIA4, AcrIIA5, AcrIIA6, and AcrIIA11 across

652   bacteria (figures 3 and S7), homologs of each Acr were identified via a BLASTp search against NCBI nr;

653   all sequences with ≥35% amino acid identity over ≥75% of the query length were retrieved. Many AcrIIA11

654   homologs were derived from bacteria with phylogenetically-ambiguous taxonomic labels in NCBI. For

655   instance, the NCBI genus 'Clostridium' is extremely polyphyletic, appearing in 121 genera and 29 families

21

656 in a rigorous re-assessment of bacterial phylogeny (work via the genome taxonomy database, or GTDB

657 (Parks et al., 2018)). Importantly, the GTDB taxonomy ensures that taxonomic labels are linked to

658 monophyletic groups, applies taxonomic ranks (phylum, class, order, etc.) at even phylogenetic depths,

659 and substantially improves problematic taxonomies (Parks et al., 2018). To make phylogenetically

660 meaningful inference with respect to AcrIIA11, we adopted the GTDB taxonomic scheme which, crucially,

661 is largely congruent with NCBI taxonomy outside a few key problem areas (e.g. Clostridiales, see table

662 S4). The taxonomic assignments for AcrIIA1-6 homologs are unchanged between NCBI and GTDB

663 schema.

664   To map protein homologs onto a bacterial genus tree, the NCBI genome assemblies matching each

665 homolog were downloaded and classified according to the GTDB scheme using the 'classify_wf' workflow

666 within GTDB toolkit (v0.1.3) (Parks et al., 2018); default parameters were used. To visualize these

667 classifications on a tree of life, the minimal phylogenetic tree encompassing all lineages encoding

668 AcrIIA1-6 or AcrIIA11 homologs was downloaded from AnnoTree (v1.0; node ID 31285). The AnnoTree

669 phylogeny additionally includes KEGG and PFAM annotations for nearly 24,000 bacterial genome

670 assemblies and is classified according to GTDB taxonomy (Mendler et al., 2018). Node 31285 includes

671 five bacterial phyla which are identified with letters on figure S7 (a: Firmicutes, b: Firmicutes_D, c:

672 Firmicutes_A, d: Fusobacteria, and e: Firmicutes_F). The KEGG identifier K09952 was used to determine

673 which GTDB genera encoded Cas9 while the PFAM IDs PF09711 and PF16813 were used to identify

674 taxa which encode Csn2, the signature protein of type II-A CRISPR-Cas systems (Makarova et al., 2015).

675 Enrichment for type II-A CRISPR-Cas systems was determined via a chi-squared test using the number

676 of Csn2-encoding genomes within each GTDB family, with p-values corrected for multiple hypotheses

677 using the Bonferroni correction.

678

679 **Phylogenetic tree of AcrIIA11 homologs**

680   The gene tree in figure S4 was built using the homologs in NCBI (see above section) and additionally

681 a set of homologs identified via a BLASTP search of IMG/VR (the January 1, 2018 release), a curated

682 database of cultured and uncultured DNA viruses (Paez-Espino et al., 2017). An e-value cutoff of 1×10^{-10}

683 was used in the IMG/VR homolog search. A preliminary phylogeny of all AcrIIA11 homologs was used

684 to select genes that optimally sample AcrIIA11 diversity for gene synthesis and anti-SpyCas9 activity

685 screening. The final phylogeny in figure S4 contains all NCBI homologs and the viral homologs which

686 were selected for gene synthesis (see table S7 for sequences and accession numbers). Alignments were

687 performed using the Geneious (v8) alignment tool and a maximum-likelihood tree was generated with

688 PhyML using the LG substitution model and 100 bootstraps.

689

690

Protein expression and purification

Codon-optimized AcrIIA4 and AcrIIA11 were cloned into pET15b to contain thrombin-cleavable 6XHis N-terminal tags, with and without C-terminal 2xStrep2 tags. All plasmids were transformed into *E. coli* BL21(DE3) RIL cells except the C-terminally-tagged AcrIIA11 variant, which was transformed into a BL21(DE3) pLysS strain. Overnight cultures were grown in LB/Ampicillin (100 µg/mL), diluted 100-fold into 1 L pre-warmed LB/Amp media, grown until an OD600 of 0.6-0.8, then incubated on ice for 30 minutes. IPTG was added to 0.2 mM, and the cultures where shaken for 18-20 hours at 18°C. The cells were pelleted and stored at -20°C. Cell pellets were resuspended in lysis/wash buffer (500 mM NaCl, 25 mM Tris, pH 7.5, 20 mM Imidazole), lysed by sonication, and centrifuged for 25 minutes in an SS34 rotor at 18,000 rpm for 30 minutes. The soluble fraction was filtered through a 5 µm filter and incubated in batch with Ni-NTA resin (Invitrogen, Cat# R90115) at 4°C for 1 hour. The resin was transferred to a gravity filtration column and washed with at least 50 volumes of wash buffer, followed by elution in 200 mM NaCl, 25 mM Tris, pH 7.5, 200 mM Imidazole. The buffer was exchanged into 200 mM NaCl/25 mM Tris, pH 7.5 by concentrating and diluting using an Amicon filter (EMD Millipore, 10,000 MWCO). Biotinylated thrombin (EMD Millipore) was added (1 U per mg of protein) and incubated for 16 hours at 18°C. Streptavidin-agarose was then used to remove thrombin according to the manufacturer's instructions (EMD Millipore). The proteins were diluted to 150 mM NaCl, 25 mM Tris pH 7.5, loaded onto a 1 mL HiTrapQ column (GE Life Sciences), and eluted by a sodium chloride gradient (150 mM to 1 M over 20 mL). Peak fractions were pooled and concentrated, bound to Ni-NTA to remove any uncleaved proteins, and the flow-through was purified via size exclusion chromatography (using Superdex75 16/60 (GE HealthCare) or SEC650 (BioRad) columns) in 200 mM NaCl, 25 mM Tris, pH 7.5, 5% glycerol. Figure 5A depicts size exclusion chromatography data for thrombin-cleaved AcrIIA11 lacking a 2xStrep2 tag. Peak fractions were pooled, concentrated, flash frozen as single-use aliquots in liquid nitrogen, and stored at -80°C.

Plasmid pMJ806 (addgene #39312) was used to express SpyCas9 containing an N-terminal 6XHis-MBP tag followed by a TEV protease cleavage site. The protein was expressed and purified over Ni-NTA as described above. The eluate from the Ni-NTA resin was dialyzed into 25 mM Tris, pH 7.5, with 300 mM NaCl, 1 mM DTT, and 5% glycerol. It was simultaneously cleaved overnight with homemade TEV protease at 4°C. The cleaved protein was loaded onto a 1 mL heparin HiTrap column (GE) in 300 mM NaCl, 25 mM Tris, 7.5 and eluted in a gradient extending to 1 M NaCl/ 25 mM Tris, 7.5. Pooled fractions were concentrated and buffer-exchanged into with 200 mM NaCl, 25 mM Tris (pH 7.5), and 20 mM imidazole, and bound to Ni-NTA to remove uncleaved fusion proteins. The flow-through was purified over a Superdex 200 16/60 column (GE Healthcare) equilibrated in buffer containing 200 mM NaCl, 25 mM Tris (pH 7.5), 5% glycerol, and 2 mM DTT. Peak fractions were pooled, concentrated to 3-6 mg/mL, flash frozen as single-use aliquots in liquid nitrogen, and stored at -80°C. A second SpyCas9 variant was also

23

726 expressed and purified from a pET28a backbone (addgene #53261) as described above except that no

727 protease cleavage site was present, so tags were not removed. This second SpyCas9 variant includes

728 an N-terminal 6xHis tag, a C-terminal HA-tag, and a C-terminal NLS sequence. DNA cleavage assays

729 and Acr pulldowns were performed using tagged SpyCas9, whereas EMSAs used the untagged version.

730 EMSAs were also performed using the tagged SpyCas9 variant and the results did not differ between

731 SpyCas9 purifications.

732

733 **gRNA Generation**

734 gRNA was generated by T7 RNA polymerase using Megashortscript Kit (Thermo Fisher #AM1354).

735 Double-stranded DNA template was generated by a single round of thermal cycling (98°C for 90 seconds,

736 55°C for 15 seconds, 72°C for 60 seconds) in 50 µl reactions using Phusion PCR polymerase mix (NEB)

737 containing 25 pmol each of the following ultramers (the protospacer-matching sequence is underlined):

738 GAAATTAATACGACTCACTATAGG<u>TAATGAAATAAGATCACTAC</u>GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTA

739 GTCCG and AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGC.

740 The dsDNA templates were purified using an Oligo Clean and Concentrator Kit (ZymoResearch) and

741 quantified by Nanodrop. Transcription reactions were digested with DNAse, extracted with phenol-

742 chloroform followed by chloroform, ethanol precipitated, resuspended in RNase free water and stored at

743 -20°C. RNA was quantified by Nanodrop and analyzed on 15% acrylamide/TBE/UREA gels.

744

745 **DNA Cleavage Assay**

746 The buffer used in DNA cleavage reactions was NEB buffer 3.1 (100 mM NaCl, 50 mM Tris-HCl, pH

747 7.9, 10 mM MgCl2, 100 µg/mL BSA); proteins were diluted in 130 mM NaCl, 25 mM Tris, pH 7.4, 2.7mM

748 KCl. SpyCas9 (0.4 µM) and AcrIIA11 (0.4 – 12.8 µM) were incubated for 10 minutes at room temperature

749 before the reaction was started by simultaneously adding 0.4 µM gRNA and 4 nM linearized plasmid (2.6

750 Kb) and transferring reactions to a 37°C water bath. After 10 minutes at 37°C, the reaction was stopped

751 by adding 0.1% SDS and 50 mM EDTA. Reactions were then run on a 1.25% agarose gel containing

752 ethidium bromide at 115V for 2 hours at room temperature. Gels were imaged using the ethidium bromide

753 detection protocol on a BioRad Chemidoc gel imager.

754

755 **Pull-down assays using Strep-tagged AcrIIA4 and AcrIIA11**

756 The binding buffer for pull-down assays was 200 mM NaCl, 25 mM Tris (pH 7.5); protein dilutions

757 were made in the same buffer. In 20 µl binding reactions, 90 pmol of SpyCas9 and gRNA were incubated

758 for 20 minutes at room temperature, followed by incubation with 150 pmol of strep-tagged Acr for an

759 additional 20 minutes at room temperature. 50 µl of a 10% slurry of Streptactin Resin (IBA biosciences

760 #2-1201-002) equilibrated in binding buffer was added to the binding reactions and incubated at 4°C on

24

a nutator. Thereafter all incubations and washes were carried out at 4°C or on ice. The beads were washed a total of four times, including one tube transfer, by centrifuging 1 minute at 2000 rpm, carefully aspirating the supernatant with a 25 gauge needle and resuspending the beads in 100 µl binding buffer. After the final bead aspiration, Strep-tagged proteins were eluted by resuspending 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 minutes at room temperature. The beads were spun and 30 µl of the supernatant was carefully removed and mixed with 2X SDS Sample Buffer (Novex). Proteins were then separated by SDS PAGE on BOLT 4-12% gels in MES buffer (Invitrogen), followed by Coomassie staining. In instances where DNA was also visualized, the gel was imaged using the DyLight 488 detection protocol on a BioRad Chemidoc gel imager (Coomassie staining followed DNA visualization using the same gel).

Electrophoretic Mobility Shift Assays

Reactions were carried out in EMSA binding buffer (56 mM NaCl, 10 mM Tris, pH 7.4, 1.2 mM KCl, 5% glycerol, 1 mM DTT, 2 mM EDTA, 50 µg/ml heparin, 100 µg/ml BSA, 0.01% Tween-20); proteins were diluted in 130 mM NaCl, 25 mM Tris, pH 7.4, 2.7mM KCl. Omitting MgCl2 ensured that Cas9 did not cleave target DNA, as previously described (Lee et al., 2018). DNA gel shifts used the 6FAM-labeled 60-mer target dsDNA (see table S6 for sequence) as template and was visualized on a BioRad Chemidoc gel imager as described for the pulldowns. All incubations were carried out at room temperature. Cas9 and gRNA (each at 2 µM) were incubated for 25 minutes, followed by addition of Acrs (2-16 µM) for 20 minutes, followed by addition of 20 nM dsDNA template and incubation for 20 minutes. Samples were loaded on an 8% acrylamide/0.5X TBE gel that was pre-run (30 minutes, 90 V, 4°C), and resolved for 160 minutes, 4°C, 90 V in 0.5X TBE buffer. For gRNA EMSA experiments, AcrIIA11 (at 4-32 µM) was incubated with 2 µM Cas9 for 20 minutes, followed by incubation with 0.2 µM gRNA for 20 minutes. The same buffers were used as in DNA EMSAs, except that 3 mM MgCl2 was included in the reactions. Samples were run for 180 minutes under the gel conditions described above. The gels were post-stained with a 1:10,000 dilution of SYBR-Gold (Invitrogen) in 0.5X TBE to visualize RNA.

797   interpretation, or the decision to publish this study. HSM is an Investigator of the Howard Hughes Medical

798   Institute.

799

800   **Competing Interests**

801   All authors declare no significant competing financial, professional, or personal interests that might

802   have influenced the performance or presentation of the work described in this manuscript.

References

803 Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P.
804 (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709-1712.
805 Basgall, E.M., Goetting, S.C., Goeckel, M.E., Giersch, R.M., Roggenkamp, E., Schrock, M.N., Halloran, M., and
806 Finnigan, G.C. (2018). Gene drive inhibition by the anti-CRISPR proteins AcrIIA2 and AcrIIA4 in Saccharomyces
807 cerevisiae. Microbiology *164*, 464-474.
808 Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L., and Sonnhammer, E.L. (2000). The Pfam protein families
809 database. Nucleic Acids Res *28*, 263-266.
810 Bondy-Denomy, J., Davidson, A.R., Doudna, J.A., Fineran, P.C., Maxwell, K., Moineau, S., Peng, X., Sontheimer, E.J.,
811 and Wiedenheft, B. (2018). A Unified Resource for Tracking Anti-CRISPR Names. The CRISPR Journal *1*, 304-305.
812 Bondy-Denomy, J., Pawluk, A., Maxwell, K.L., and Davidson, A.R. (2013). Bacteriophage genes that inactivate the
813 CRISPR/Cas bacterial immune system. Nature *493*, 429-432.
814 Borges, A.L., Davidson, A.R., and Bondy-Denomy, J. (2017). The Discovery, Mechanisms, and Evolutionary Impact
815 of Anti-CRISPRs. Annu Rev Virol *4*, 37-59.
816 Bull, J.J., and Malik, H.S. (2017). The gene drive bubble: New realities. PLoS Genet *13*, e1006850.
817 Chylinski, K., Makarova, K.S., Charpentier, E., and Koonin, E.V. (2014). Classification and evolution of type II CRISPR-
818 Cas systems. Nucleic Acids Res *42*, 6091-6105.
819 Clemente, J.C., Pehrsson, E.C., Blaser, M.J., Sandhu, K., Gao, Z., Wang, B., Magris, M., Hidalgo, G., Contreras, M.,
820 Noya-Alarcon, O.*, et al.* (2015). The microbiome of uncontacted Amerindians. Sci Adv *1*.
821 Crawley, A.B., Henriksen, J.R., and Barrangou, R. (2018). CRISPRdisco: An Automated Pipeline for the Discovery
822 and Analysis of CRISPR-Cas Systems. The CRISPR Journal *1*, 171-181.
823 de la Bastide, M., and McCombie, W.R. (2007). Assembling genomic DNA sequences with PHRAP. Curr Protoc
824 Bioinformatics *Chapter 11*, Unit11 14.
825 Dong, Guo, M., Wang, S., Zhu, Y., Wang, S., Xiong, Z., Yang, J., Xu, Z., and Huang, Z. (2017). Structural basis of
826 CRISPR-SpyCas9 inhibition by an anti-CRISPR protein. Nature *546*, 436-439.
827 Esvelt, K.M., Mali, P., Braff, J.L., Moosburner, M., Yaung, S.J., and Church, G.M. (2013). Orthogonal Cas9 proteins
828 for RNA-guided gene regulation and editing. Nat Methods *10*, 1116-1121.
829 Faure, G., Makarova, K.S., and Koonin, E.V. (2018). CRISPR-Cas: Complex Functional Networks and Multiple Roles
830 beyond Adaptive Immunity. J Mol Biol.
831 Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching.
832 Nucleic Acids Res *39*, W29-37.
833 Forsberg, K.J., Patel, S., Wencewicz, T.A., and Dantas, G. (2015). The Tetracycline Destructases: A Novel Family of
834 Tetracycline-Inactivating Enzymes. Chem Biol *22*, 888-897.
835 Forsberg, K.J., Patel, S., Witt, E., Wang, B., Ellison, T.D., and Dantas, G. (2016). Identification of Genes Conferring
836 Tolerance to Lignocellulose-Derived Inhibitors by Functional Selections in Soil Metagenomes. Appl Environ
837 Microbiol *82*, 528-537.
838 Forsberg, K.J., Reyes, A., Wang, B., Selleck, E.M., Sommer, M.O., and Dantas, G. (2012). The shared antibiotic
839 resistome of soil bacteria and human pathogens. Science *337*, 1107-1111.
840 Genee, H.J., Bali, A.P., Petersen, S.D., Siedler, S., Bonde, M.T., Gronenberg, L.S., Kristensen, M., Harrison, S.J., and
841 Sommer, M.O. (2016). Functional mining of transporters using synthetic selections. Nat Chem Biol *12*, 1015-1022.
842 Haft, D.H., Loftus, B.J., Richardson, D.L., Yang, F., Eisen, J.A., Paulsen, I.T., and White, O. (2001). TIGRFAMs: a
843 protein family resource for the functional identification of proteins. Nucleic Acids Res *29*, 41-43.
844 Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. Microbiol Mol Biol
845 Rev *68*, 669-685.
846 Harrington, L.B., Doxzen, K.W., Ma, E., Liu, J.J., Knott, G.J., Edraki, A., Garcia, B., Amrani, N., Chen, J.S., Cofsky, J.C.*,
847 et al.* (2017). A Broad-Spectrum Inhibitor of CRISPR-Cas9. Cell.
848 Hatfull, G.F. (2015). Dark Matter of the Biosphere: the Amazing World of Bacteriophage Diversity. J Virol *89*, 8107-
849 8110.

27

850 He, F., Bhoobalan-Chitty, Y., Van, L.B., Kjeldsen, A.L., Dedola, M., Makarova, K.S., Koonin, E.V., Brodersen, D.E.,
851 and Peng, X. (2018). Anti-CRISPR proteins encoded by archaeal lytic viruses inhibit subtype I-D immunity. Nat
852 Microbiol.
853 Hille, F., Richter, H., Wong, S.P., Bratovic, M., Ressel, S., and Charpentier, E. (2018). The Biology of CRISPR-Cas:
854 Backward and Forward. Cell *172*, 1239-1259.
855 Hoyland-Kroghsbo, N.M., Paczkowski, J., Mukherjee, S., Broniewski, J., Westra, E., Bondy-Denomy, J., and Bassler,
856 B.L. (2017). Quorum sensing controls the Pseudomonas aeruginosa CRISPR-Cas adaptive immune system. Proc
857 Natl Acad Sci U S A *114*, 131-135.
858 Hynes, A.P., Rousseau, G.M., Agudelo, D., Goulet, A., Amigues, B., Loehr, J., Romero, D.A., Fremaux, C., Horvath,
859 P., Doyon, Y*., et al.* (2018). Widespread anti-CRISPR proteins in virulent bacteriophages inhibit a range of Cas9
860 proteins. Nat Commun *9*, 2919.
861 Hynes, A.P., Rousseau, G.M., Lemay, M.L., Horvath, P., Romero, D.A., Fremaux, C., and Moineau, S. (2017). An
862 anti-CRISPR from a virulent streptococcal phage inhibits Streptococcus pyogenes Cas9. Nat Microbiol.
863 Iqbal, H.A., Craig, J.W., and Brady, S.F. (2014). Antibacterial enzymes from the functional screening of
864 metagenomic libraries hosted in Ralstonia metallidurans. FEMS Microbiol Lett *354*, 19-26.
865 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.
866 (2019). Temperature-Responsive Competitive Inhibition of CRISPR-Cas9. Molecular Cell *73*, 601-610.e605.
867 Knott, G.J., and Doudna, J.A. (2018). CRISPR-Cas guides the future of genetic engineering. Science *361*, 866-869.
868 Koonin, E.V., and Makarova, K.S. (2017). Mobile Genetic Elements and Evolution of CRISPR-Cas Systems: All the
869 Way There and Back. Genome Biol Evol *9*, 2812-2825.
870 Lee, J., Mir, A., Edraki, A., Garcia, B., Amrani, N., Lou, H.E., Gainetdinov, I., Pawluk, A., Ibraheim, R., Gao, X.D*., et*
871 *al.* (2018). Potent Cas9 Inhibition in Bacterial and Human Cells by AcrIIC4 and AcrIIC5 Anti-CRISPR Proteins. mBio
872 *9*, e02321-02318.
873 Lee, S.A., Gallagher, L.A., Thongdee, M., Staudinger, B.J., Lippman, S., Singh, P.K., and Manoil, C. (2015). General
874 and condition-specific essential functions of Pseudomonas aeruginosa. Proc Natl Acad Sci U S A *112*, 5189-5194.
875 Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide
876 sequences. Bioinformatics *22*, 1658-1659.
877 Liu, L., Yin, M., Wang, M., and Wang, Y. (2019). Phage AcrIIA2 DNA Mimicry: Structural Basis of the CRISPR and
878 Anti-CRISPR Arms Race. Molecular Cell *73*, 611-620.e613.
879 Lutz, R., and Bujard, H. (1997). Independent and tight regulation of transcriptional units in Escherichia coli via the
880 LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Res *25*, 1203-1210.
881 Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J.,
882 Charpentier, E., Haft, D.H*., et al.* (2015). An updated evolutionary classification of CRISPR-Cas systems. Nat Rev
883 Microbiol *13*, 722-736.
884 Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2018). Classification and Nomenclature of CRISPR-Cas Systems: Where
885 from Here? The CRISPR Journal *1*, 325-336.
886 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.,
887 Kleinstiver, B.P*., et al.* (2018). Discovery of widespread Type I and Type V CRISPR-Cas inhibitors. Science.
888 Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci
889 by targeting DNA. Science *322*, 1843-1845.
890 Maxwell, K.L. (2016). Phages Fight Back: Inactivation of the CRISPR-Cas Bacterial Immune System by Anti-CRISPR
891 Proteins. PLoS Pathog *12*, e1005282.
892 Mendler, K., Chen, H., Parks, D.H., Hug, L.A., and Doxey, A.C. (2018). AnnoTree: visualization and exploration of a
893 functionally annotated microbial tree of life. bioRxiv, 463455.
894 Nakamura, M., Srinivasan, P., Chavez, M., Carter, M.A., Dominguez, A.A., La Russa, M., Lau, M.B., Abbott, T.R., Xu,
895 X., Zhao, D*., et al.* (2019). Anti-CRISPR-mediated control of gene editing and synthetic circuits in eukaryotic cells.
896 Nature Communications *10*, 194.
897 Paez-Espino, D., Chen, I.A., Palaniappan, K., Ratner, A., Chu, K., Szeto, E., Pillay, M., Huang, J., Markowitz, V.M.,
898 Nielsen, T*., et al.* (2017). IMG/VR: a database of cultured and uncultured DNA Viruses and retroviruses. Nucleic
899 Acids Res *45*, D457-D465.

900 Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., and Hugenholtz, P. (2018). A
901 standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nature
902 Biotechnology *36*, 996.
903 Pawluk, A., Amrani, N., Zhang, Y., Garcia, B., Hidalgo-Reyes, Y., Lee, J., Edraki, A., Shah, M., Sontheimer, E.J.,
904 Maxwell, K.L.*, et al.* (2016a). Naturally Occurring Off-Switches for CRISPR-Cas9. Cell *167*, 1829-1838 e1829.
905 Pawluk, A., Bondy-Denomy, J., Cheung, V.H., Maxwell, K.L., and Davidson, A.R. (2014). A new group of phage anti-
906 CRISPR genes inhibits the type I-E CRISPR-Cas system of Pseudomonas aeruginosa. MBio *5*, e00896.
907 Pawluk, A., Davidson, A.R., and Maxwell, K.L. (2017). Anti-CRISPR: discovery, mechanism and function. Nat Rev
908 Microbiol.
909 Pawluk, A., Staals, R.H., Taylor, C., Watson, B.N., Saha, S., Fineran, P.C., Maxwell, K.L., and Davidson, A.R. (2016b).
910 Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. Nat Microbiol *1*, 16085.
911 Pehrsson, E.C., Tsukayama, P., Patel, S., Mejia-Bautista, M., Sosa-Soto, G., Navarrete, K.M., Calderon, M., Cabrera,
912 L., Hoyos-Arango, W., Bertoli, M.T.*, et al.* (2016). Interconnected microbiomes and resistomes in low-income
913 human habitats. Nature *533*, 212-216.
914 protocols.io (2015). NEB: Making your own electrocompetent cells (NEB).
915 Pursey, E., Sunderhauf, D., Gaze, W.H., Westra, E.R., and van Houte, S. (2018). CRISPR-Cas antimicrobials:
916 Challenges and future prospects. PLoS Pathog *14*, e1006990.
917 Rauch, B.J., Silvis, M.R., Hultquist, J.F., Waters, C.S., McGregor, M.J., Krogan, N.J., and Bondy-Denomy, J. (2017).
918 Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. Cell *168*, 150-158 e110.
919 Rollins, M.F., Chowdhury, S., Carter, J., Golden, S.M., Miettinen, H.M., Santiago-Frangos, A., Faith, D., Lawrence,
920 C.M., Lander, G.C., and Wiedenheft, B. (2018). Structure reveals mechanism of CRISPR RNA-guided nuclease
921 recruitment and anti-CRISPR viral mimicry. bioRxiv, 453720.
922 Roux, S., Enault, F., Hurwitz, B.L., and Sullivan, M.B. (2015). VirSorter: mining viral signal from microbial genomic
923 data. PeerJ *3*, e985.
924 Sheth, R.U., Cabral, V., Chen, S.P., and Wang, H.H. (2016). Manipulating Bacterial Communities by in situ
925 Microbiome Engineering. Trends Genet *32*, 189-200.
926 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,
927 J.A. (2017). Disabling Cas9 by an anti-CRISPR DNA mimic. Science Advances *3*.
928 Sontheimer, E.J., and Davidson, A.R. (2017). Inhibition of CRISPR-Cas systems by mobile genetic elements. Curr
929 Opin Microbiol *37*, 120-127.
930 Stone, N.P., Hilbert, B.J., Hidalgo, D., Halloran, K.T., Lee, J., Sontheimer, E.J., and Kelch, B.A. (2018). A
931 Hyperthermophilic Phage Decoration Protein Suggests Common Evolutionary Origin with Herpesvirus Triplex
932 Proteins and an Anti-CRISPR Protein. Structure *26*, 936-947 e933.
933 Uribe, R.V., van der Helm, E., Misiakou, M.-A., Lee, S.-W., Kol, S., and Sommer, M.O.A. (2019). Discovery and
934 Characterization of Cas9 Inhibitors Disseminated across Seven Bacterial Phyla. Cell Host & Microbe.
935 van Belkum, A., Soriaga, L.B., LaFave, M.C., Akella, S., Veyrieras, J.B., Barbu, E.M., Shortridge, D., Blanc, B.,
936 Hannum, G., Zambardi, G.*, et al.* (2015). Phylogenetic Distribution of CRISPR-Cas Systems in Antibiotic-Resistant
937 Pseudomonas aeruginosa. MBio *6*, e01796-01715.
938 Watters, K.E., Fellmann, C., Bai, H.B., Ren, S.M., and Doudna, J.A. (2018). Systematic discovery of natural CRISPR-
939 Cas12a inhibitors. Science.
940 Westra, E.R., Dowling, A.J., Broniewski, J.M., and Houte, S.v. (2016). Evolution and Ecology of CRISPR. Annual
941 Review of Ecology, Evolution, and Systematics *47*, 307-331.
942 Yang, H., and Patel, D.J. (2017). Inhibition Mechanism of an Anti-CRISPR Suppressor AcrIIA4 Targeting SpyCas9.
943 Mol Cell *67*, 117-127 e115.
944 Zerbino, D.R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs.
945 Genome Res *18*, 821-829.
946 Zhang, Q., Rho, M., Tang, H., Doak, T.G., and Ye, Y. (2013). CRISPR-Cas systems target a diverse collection of
947 invasive mobile genetic elements in human microbiomes. Genome Biol *14*, R40.
948 Zhu, W., Lomsadze, A., and Borodovsky, M. (2010). Ab initio gene identification in metagenomic sequences.
949 Nucleic Acids Res *38*, e132.

950 Zucca, S., Pasotti, L., Politi, N., Casanova, M., Mazzini, G., Cusella De Angelis, M.G., and Magni, P. (2015). Multi-
951 Faceted Characterization of a Novel LuxR-Repressible Promoter Library for Escherichia coli. PLoS One *10*,
952 e0126264.

953
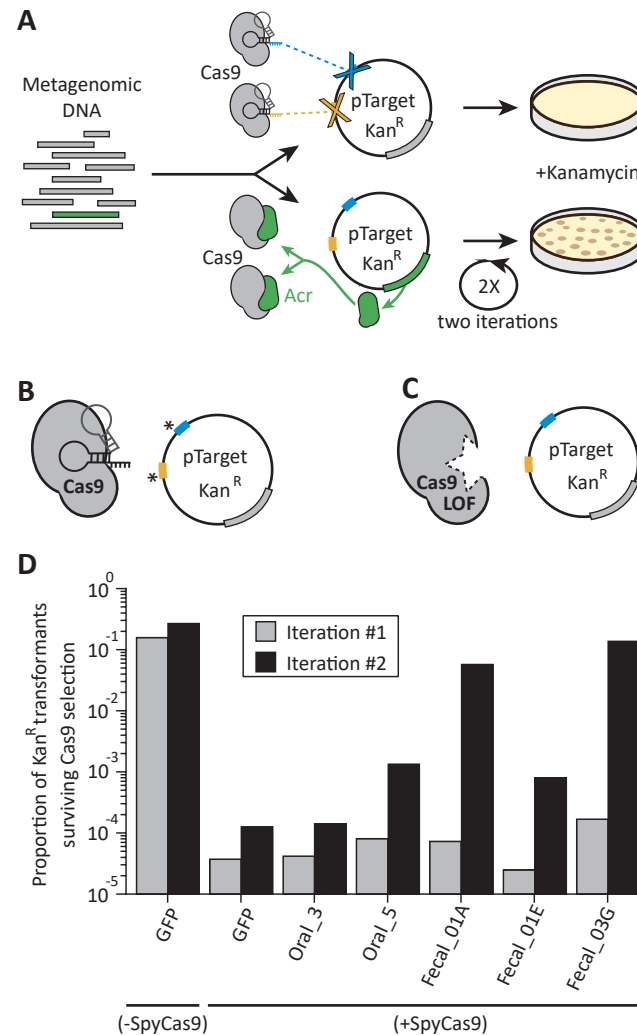
Figure 1. A functional metagenomic selection for type II-A anti-CRISPRs. (A) Plasmids in a metagenomic library without *acr*s are targeted by Cas9 and eliminated. Those few with *acr*s (green) withstand Cas9 and can be recovered via kanamycin selection. However, individual target site mutations, indicated by an asterisk **(B)** or Cas9 loss of function (LOF) mutations **(C)** allow plasmids to evade Cas9 independent of the metagenomic DNA insert that they carry. To reduce these major sources of false positives, we employ two target sites (blue, yellow) and two rounds of selection. **(D)** Relative to a GFP control, metagenomic libraries protect plasmids after two iterations through Cas9-inducing conditions. Each bar represents a single experiment.
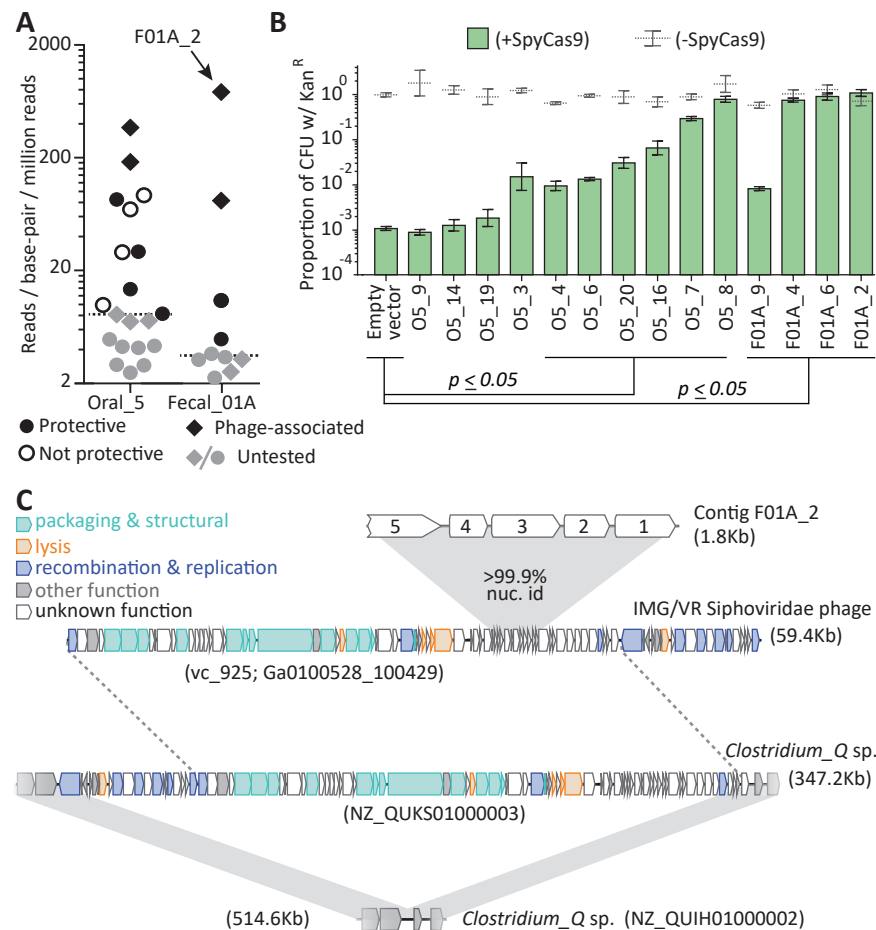
**Figure 2. Metagenomic DNA inserts antagonize SpyCas9. (A)** The read-coverage of all contigs from libraries Oral_5 and Fecal_01A is shown. Dashed lines depict median coverage for each library. Relative coverage can be used as a proxy for contig abundance within a library but cannot be used for meaningful comparisons across libraries. **(B)** Upon individual re-testing, six oral and four fecal inserts protect a plasmid from SpyCas9 relative to an empty vector (student's t-test); +/- refers to SpyCas9 induction. Error bars depict standard error of the mean. **(C)** Contig F01A_2 is from a Siphoviridae phage which lysogenizes *Clostridium_Q* sp. Gray, shaded regions depict key regions of near perfect (>99%) nucleotide identity. Homology between the *Clostridium_Q* prophage and the IMG/VR phage is not depicted as these phages are nearly identical (except for one translocation or assembly-artefact involving the recombination and replication machinery). All accession numbers denote NCBI Genbank IDs except for the Siphoviridae phage, for which we use IMG/VR convention and indicate viral_cluster with the scaffold_id in parentheses. Contig lengths are depicted next to each sequence.
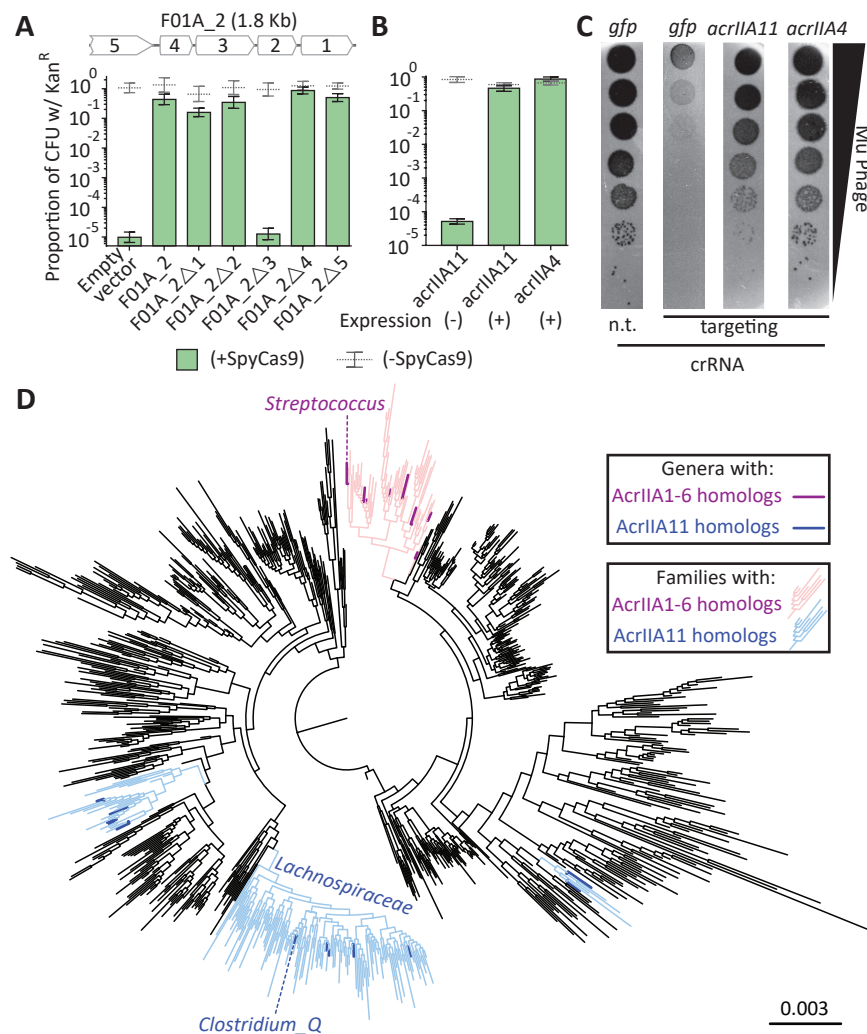
Figure 3. AcrIIA11 is a widely disseminated anti-CRISPR that protects plasmid and phage from SpyCas9. (A) The F01A_2 contig is depicted above the bar chart. Delta symbols ($\triangle$) indicate early stop codons in each gene of the contig. Only the third gene on contig F01A_2 is necessary for SpyCas9 antagonism. **(B)** Induction of the third gene, named *acrIIA11*, is sufficient for SpyCas9 antagonism, protecting a plasmid as well as *acrIIA4*. Error bars depict standard error of the mean. **(C)** Mu phage fitness, measured by plaquing on *E. coli* expressing Mu-targeting SpyCas9, is measured in the presence of *gfp*, *acrIIA11*, or *acrIIA4* via serial ten-fold dilutions (also see replicated in Figure S6). Based on a non-targeting (n.t.) crRNA control, we conclude that SpyCas9 confers 10^5-fold protection against phage Mu in these conditions. Both *acrIIA11* and *acrIIA4* significantly enhance Mu fitness by inhibiting SpyCas9. **(D)** A phylogenetic tree of Firmicutes and related phyla that shows the widespread dissemination of *acrIIA11* relative to other type II-A *acr*s. Each node represents a genus; families colored light blue contain *acrIIA11* homologs while families colored pink contain homologs of *acrIIA1 - acrIIA6*. Genera are colored dark blue or magenta if *acr* homologs could confidently be assigned to this phylogenetic resolution. The phylogenetic placement of *Clostridium_Q* and *Streptococcus* is shown. See figure S7 for a completely annotated version of this phylogeny.
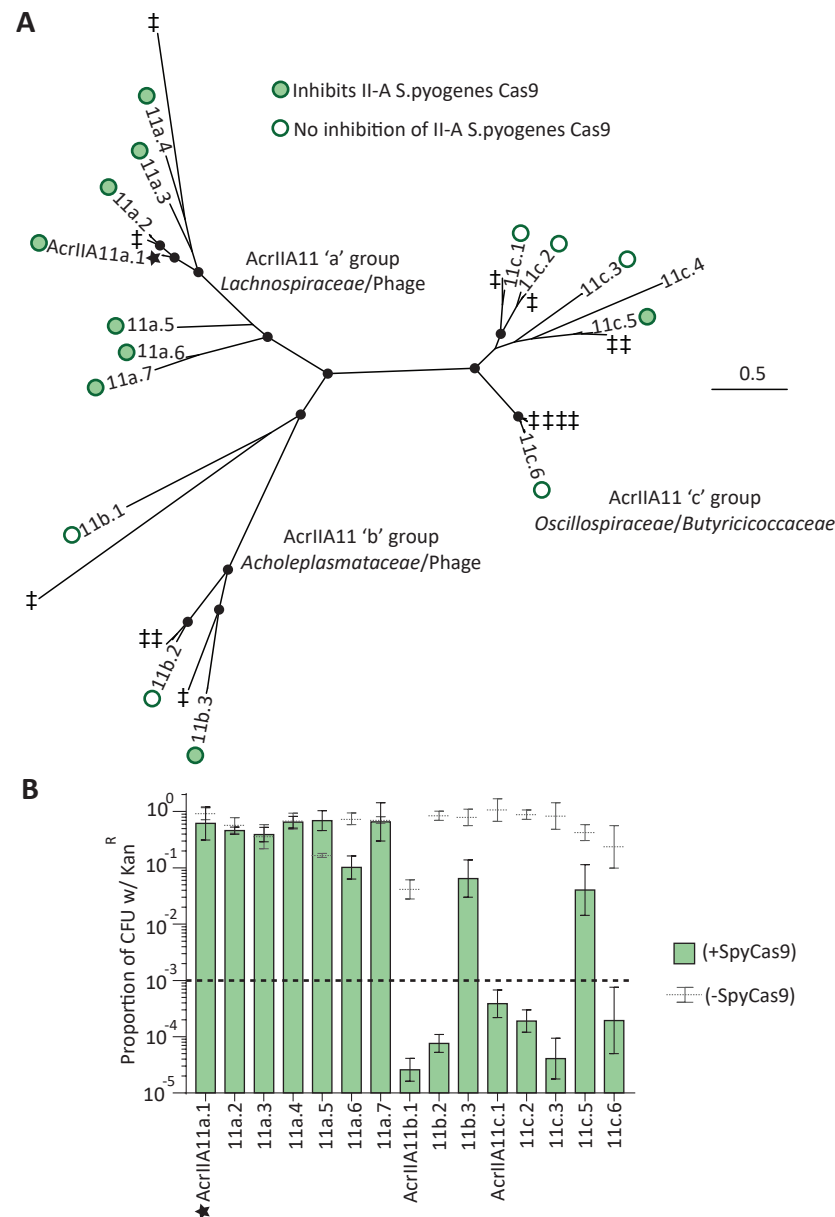
Figure 4. Diverse AcrIIA11 homologs inhibit SpyCas9. (A) An unrooted phylogenetic tree of predicted AcrIIA11 homologs clusters into three major clades that correspond to the bacterial families depicted in figures 3D and S7. Circles at nodes indicate bootstrap support >0.75. The star indicates the original AcrIIA11 sequence isolated via functional selection. Open and filled circles indicate whether a given homolog inhibits SpyCas9 in a plasmid protection assay. Untested homologs are indicated with a diesis (‡); toxicity with AcrIIA11c.4 prevented it from yielding meaningful data **(B)** AcrIIA11 homologs antagonize SpyCas9. The dashed lined indicates the threshold used to assign inhibitory labels in (A). Error bars depict standard error of the mean.
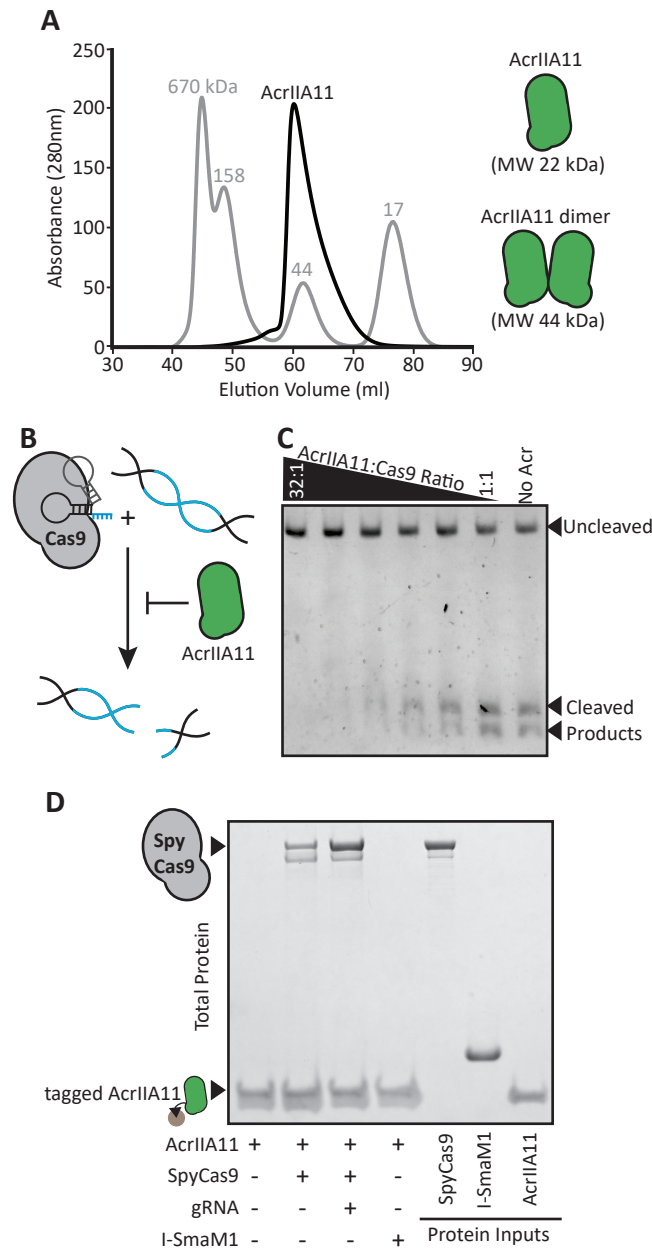
Figure 5. AcrIIA11 binds SpyCas9 to inhibit dsDNA cleavage. (A) AcrIIA11 elutes as a dimer on size exclusion chromatography. The gray trace depicts protein standards of the indicated molecular weight. The black trace shows AcrIIA11 elution. The predicted molecular weights of its monomeric and dimeric forms are depicted to the right. **(B)** Schematic of dsDNA cleavage assay in (C). **(C)** AcrIIA11 inhibits the ability of SpyCas9 (0.4μM) to cleave a linear 2.6 Kb dsDNA substrate in a concentration-dependent manner. **(D)** AcrIIA11 binds SpyCas9, with a moderate preference for the gRNA-loaded form. A Coomassie stain of total protein following pulldown of a 2x-strep-tagged AcrIIA11 incubated with either purified SpyCas9 (+/- gRNA) or the meganuclease I-SmaMI (as a negative control). SpyCas9 often runs as a doublet (see the protein marker), likely due to partial protein degradation.
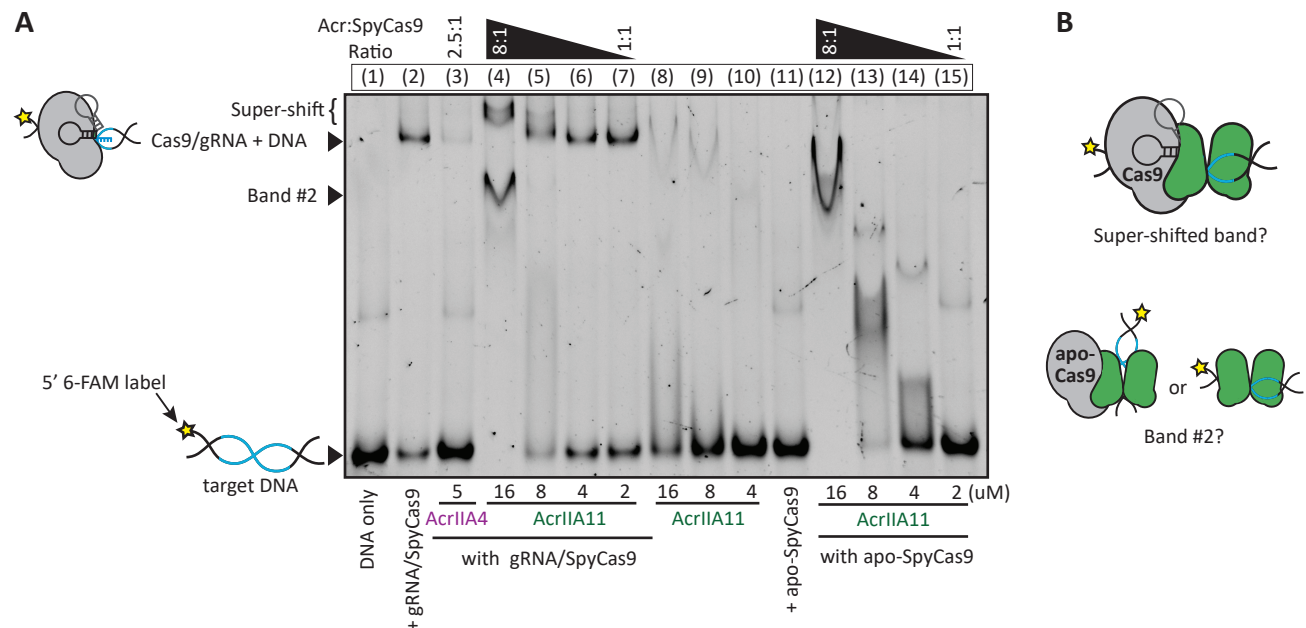
Figure 6. AcrIIA11 inhibits SpyCas9 via a novel mechanism. (A) Using an EMSA, we examine the gel shift experienced by target DNA (fluorescently labeled with a 6-FAM marker) incubated with either purified SpyCas9, gRNA, and/or purified Acrs. Divalent cations were omitted from reactions such that SpyCas9 could bind, but not cleave, target DNA, as previously shown (Lee et al., 2018). Lane numbers and Acr:SpyCas9 ratios are depicted above the native gel. Lane contents and Acr concentrations are depicted below the gel; 2uM SpyCas9 was used. Key bands are annotated to the left of the gel. **(B)** Schematics of possible explanations for the super-shifted band and band #2 in (A) are depicted.
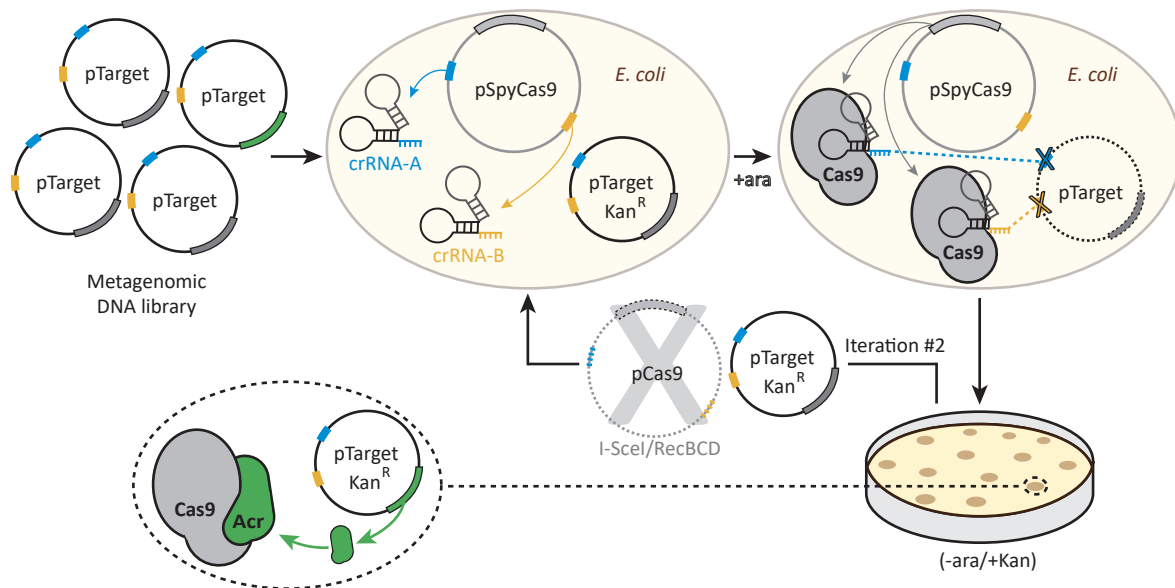
Figure S1. **A functional selection for type II-A anti-CRISPRs.** A library of plasmids each bearing a different metagenomic DNA fragment is transformed into an *E. coli* strain expressing SpyCas9 and two crRNAs, which target the library for destruction. Two crRNAs targeting two sites in the plasmid backbone reduce the number of target-site escape mutations, mitigating this source of false-positives. After transformed cells are allowed to recover, SpyCas9 is induced with arabinose and the library is subjected to SpyCas9 selection for twenty hours. Plasmids that survive this first round of selection are purified from Kan^R clones and the pCas9 plasmid is removed via digestion with I-SceI and RecBCD treatment. The metagenomic library is then subjected to SpyCas9 exposure a second time, which enriches for plasmid-intrinsic SpyCas9 resistance (*i.e.* what may be encoded by the metagenomic DNA inserts). The second iteration allows *acr*-encoding clones to be enriched above background, which is set by the frequency of Cas9 loss-of-function mutations. Kan^R clones following two rounds of selection were then harvested and their metagenomic DNA inserts sequenced to identify putative *acr*-encoding metagenomic DNA fragments.

GFP False Positive Rate
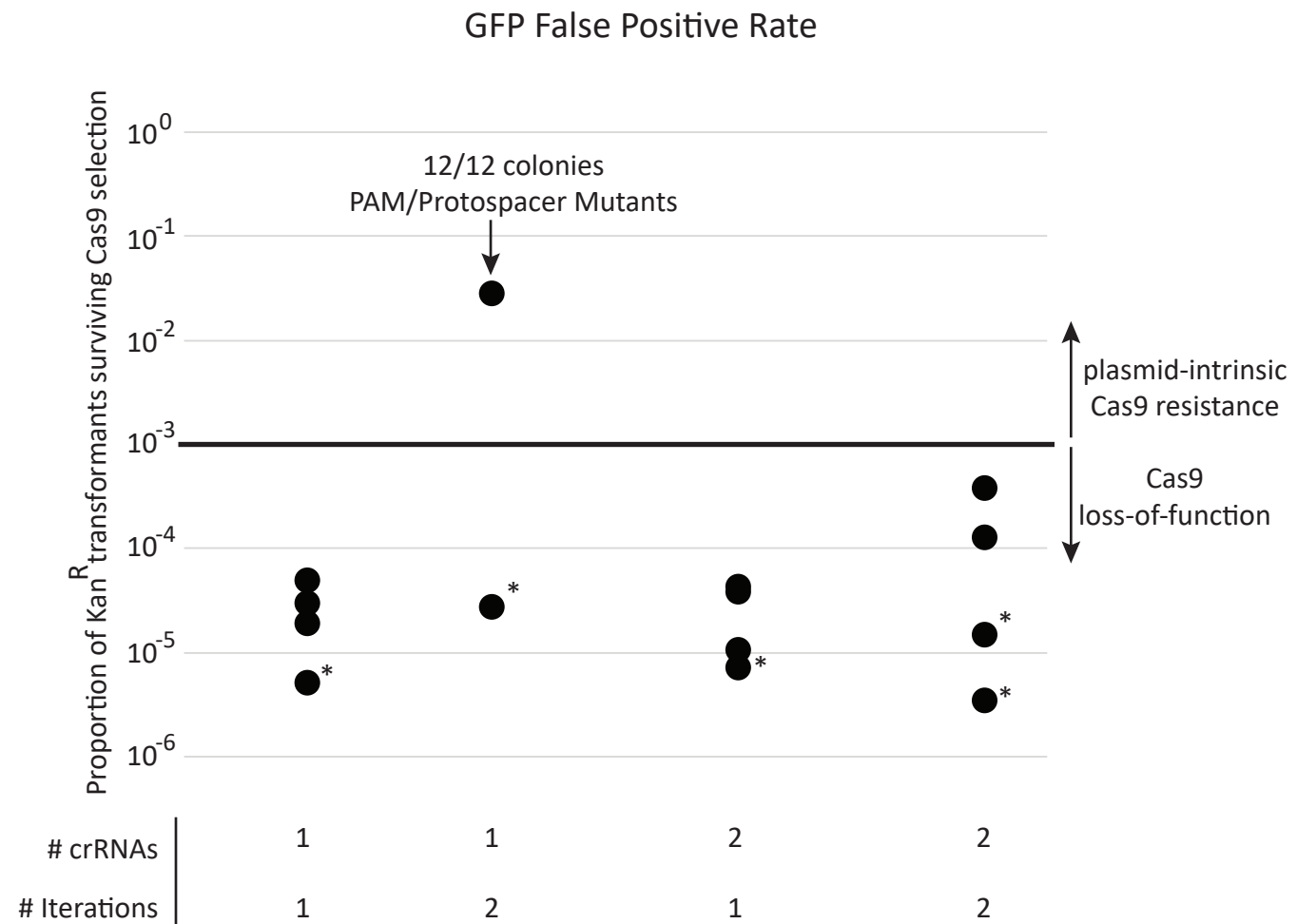
Figure S2. **Cas9 loss-of-function mutations are the major source of false-positives during *acr* selection.** Each data point represents a separate experiment toward developing the final selection for SpyCas9 antagonists. All transformations use the pZE21-GFP control target plasmid and approximate a metagenomic library expressing only neutral functions. Surviving colonies therefore represent sources of false positives. A single iteration of SpyCas9 exposure reduces Kan^R transformants by a factor of 10^4 to 10^5. This false-positive rate remained constant across all experiments except for one experiment that used a single target site on pZE21. In this experiment, mutations to the protospacer or PAM region of the Cas9 target site in pZE21 dominated, prompting two target loci to be used thereafter. All other colonies genotyped (those in asterisked experiments) escaped selection due to inactivating mutations in pCas9. This loss-of-function rate, importantly, remained constant across rounds of selection, allowing for plasmid-intrinsic Cas9 resistance to be identified via two iterations through SpyCas9 selection. When libraries were subject to two rounds of selection, this plasmid-intrinsic resistance was predominantly due to genes encoded by its metagenomic DNA fragment (fig. 1D).
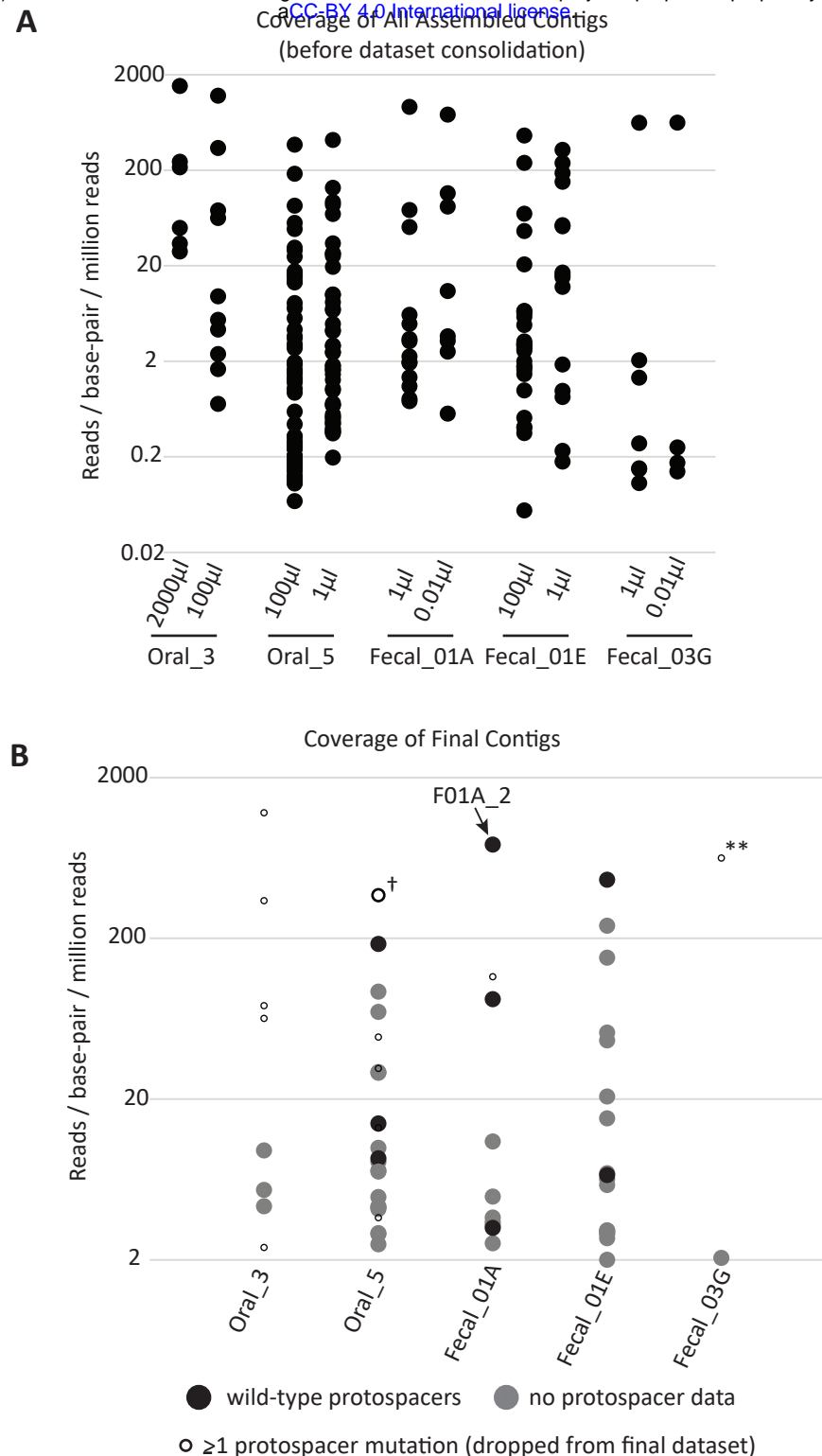
Figure S3. **Coverage of assembled contigs by library. (A)** Coverage of each assembled contig before dataset processing. The contigs from each titer plate were sequenced separately and assembled independently. Redundant contigs were removed during data processing to give the final contigs depicted in (B). Relative coverage can be used as a proxy for contig abundance within a library but cannot be used for meaningful comparisons across libraries. **(B)** Coverage of contigs in the final dataset. Filled circles depict contigs reported in the final dataset (n=51); grey indicates contigs without genotypic data for Cas9 target sites and black indicates contigs from plasmids with confirmed wild-type target sites. Small, empty circles depict assembled contigs linked to at least one target site mutation and were not included in the final dataset, with the cross (†) indicating the lone exception. This contig was found in a plasmid containing a single target site mutation but nonetheless was confirmed for anti-SpyCas9 activity when cloned into a fresh plasmid background, so was included in the final dataset. The double asterisks (**) highlight a clone with escape mutations in both target sites that dominated the selection using library Fecal_03G.
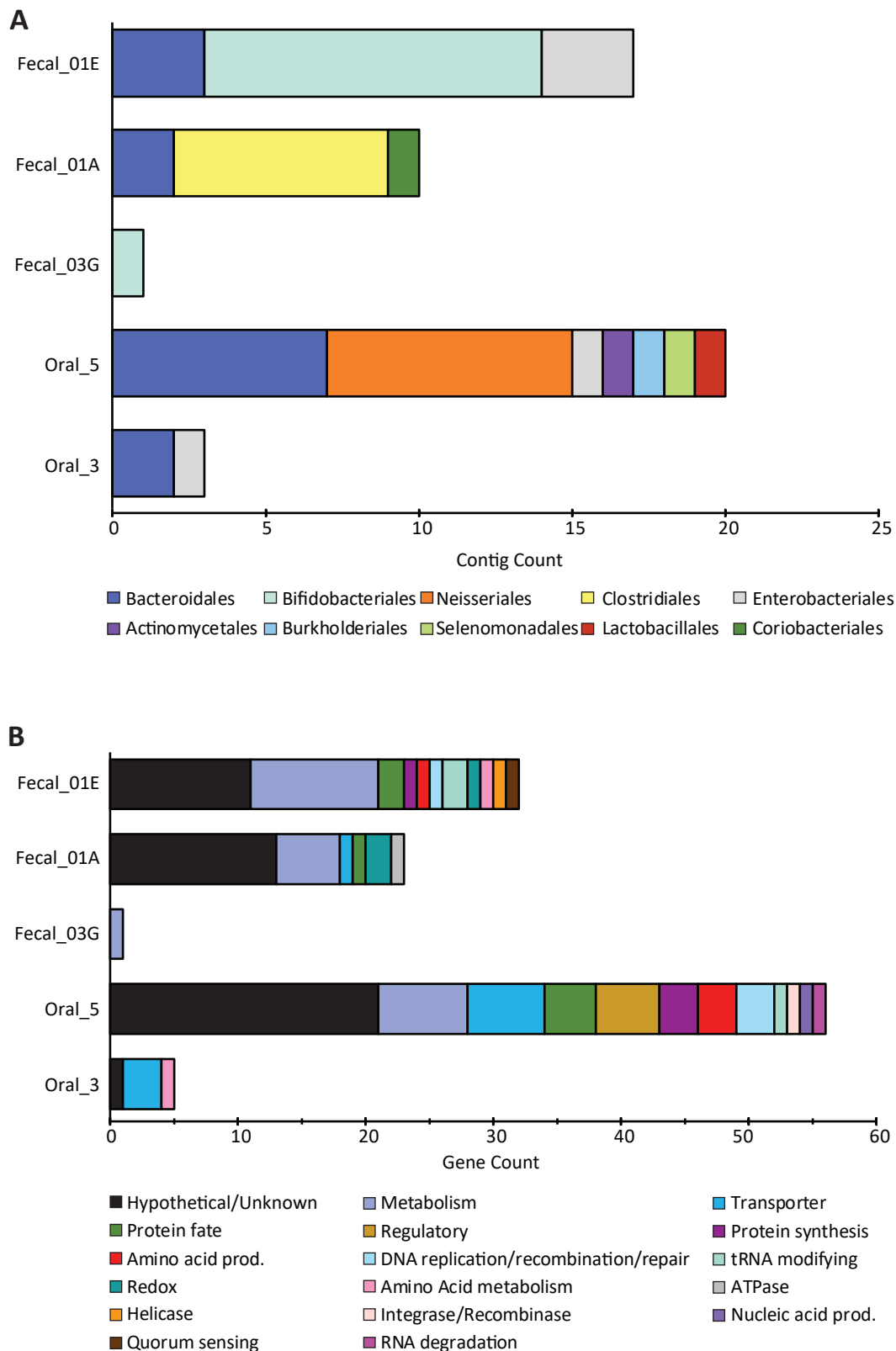
Figure S4. **NCBI taxonomies and gene functions recovered from all SpyCas9-antagonizing contigs. (A)** Taxonomic orders for the contigs surviving SpyCas9 targeting, grouped by library. **(B)** The encoded gene functions on these contigs, hand-curated into broad categories.
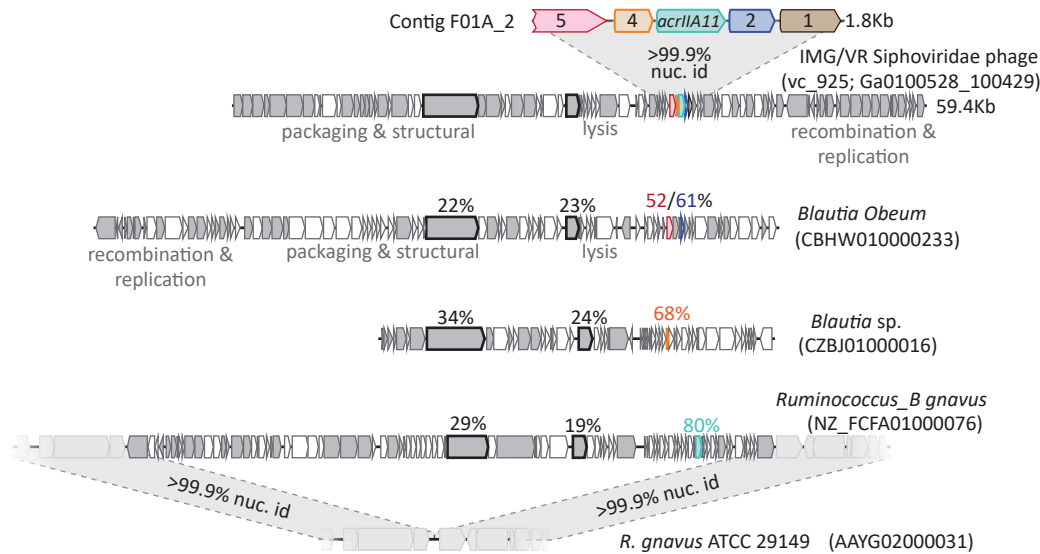
Figure S5. **The genes on contig F01A_2 are small accessory proteins of *Lachnospiraceae* phages.** These genes move readily by horizontal gene transfer and are typically unlinked from one another. Gray genes denote homologs shared between the AcrIIA11-encoding Siphoviridae phage and the three other related phages. The amino acid identities (relative to the Siphoviridae phage) for two genes present in all genomes, a tape measure protein and reverse transcriptase, are depicted in bold outline to illustrate phage relatedness. Homologs of F01A_2 genes are indicated by common color and the amino acid identities for each gene product are depicted above each gene. The depicted homolog of *acrIIA11* (named *acrIIA11a.2*) is found in an actively circulating temperate phage of *Ruminococcus gnavus*. All accession numbers denote NCBI Genbank IDs except for the Siphoviridae phage (for this phage, we use IMG/VR convention and indicate viral_cluster; scaffold_id in parentheses). Contig lengths are depicted next to each sequence.
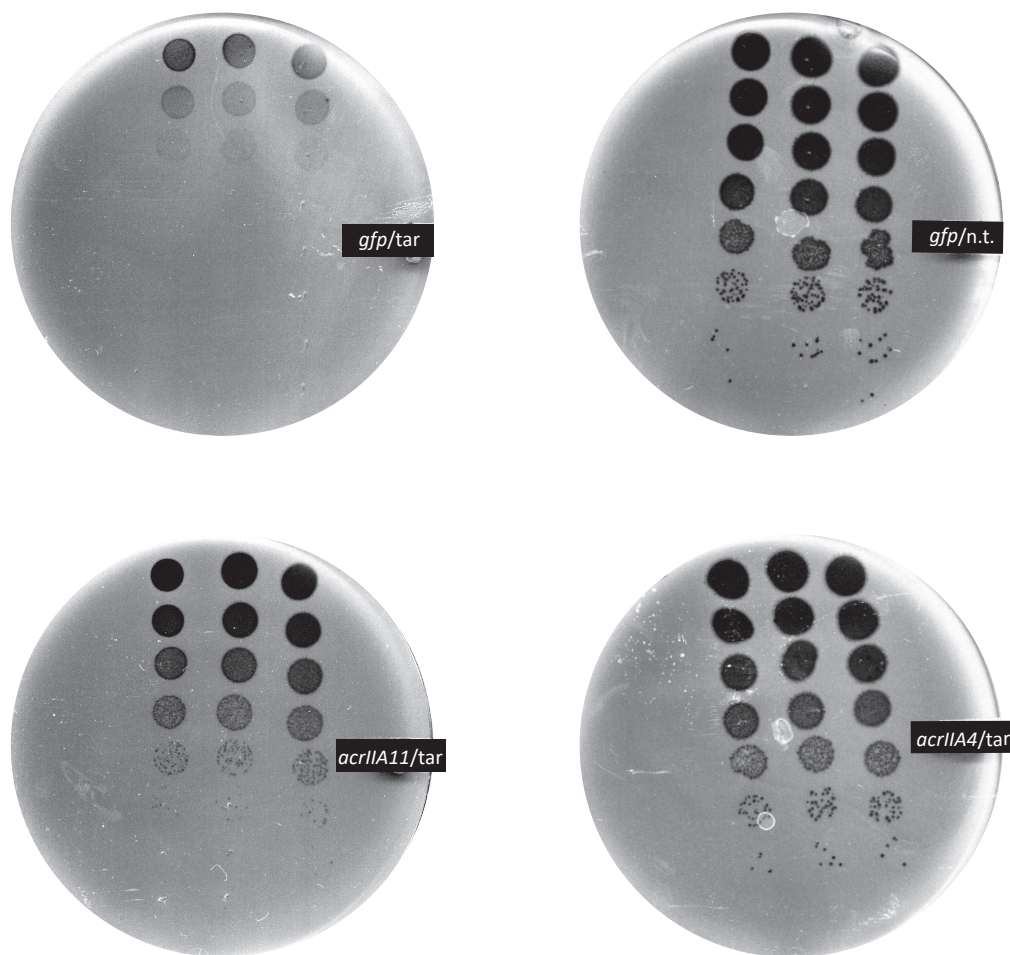
Figure S6. *AcrIIA11 protects phage from SpyCas9.* Mu phage fitness, measured by plaquing on *E. coli* expressing Mu-targeting SpyCas9, is measured in the presence of *gfp*, *acrIIA11*, or *acrIIA4* via serial ten-fold dilutions. Bacterial clearing (black) occurs when phage Mu overcomes Cas9 immunity and lyses *E. coli*. Based on a non-targeting (n.t.) crRNA control, we conclude that SpyCas9 confers 10^5-fold protection against phage Mu in these conditions. Both *acrIIA11* and *acrIIA4* significantly enhance Mu fitness by inhibiting SpyCas9. The indicated anti-CRISPR gene or *gfp* control is expressed from a second plasmid, in trans.
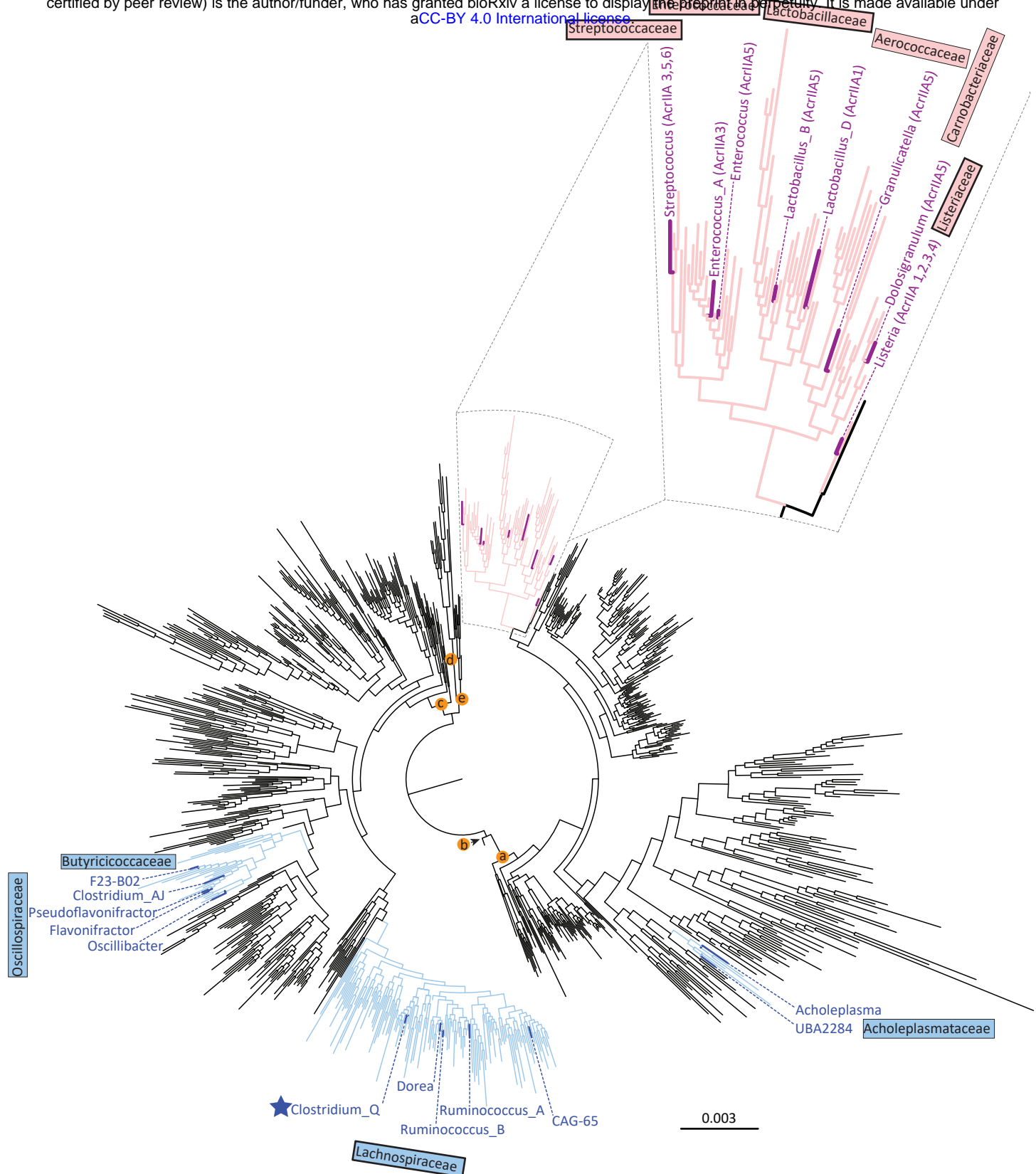
Figure S7. *AcrIIA11* **is widely distributed across Firmicutes and related phyla compared to** *acrIIA1-acrIIA6.* As in figure 3D, each node represents a genus and families colored light blue contain *acrIIA11* homologs while families colored pink contain homologs of *acrIIA1 - acrIIA6*. Genera are colored dark blue or magenta if *acr* homologs could confidently be assigned to this phylogenetic resolution. The star indicates the *Clostridium_Q* genus depicted in figure 2C. A thick border around a family indicates that is enriched for type II-A CRISPR-Cas systems relative to all bacterial families (chi-squared test, $p<1\times10^{-4}$, family classification used the GTDB scheme with functional attributions provided by AnnoTree). We used the GTDB taxonomy because it helps to resolve a well-recognized polyphyly among *Clostridia* and ensures that taxonomic labels are linked to monophyletic groups while applying taxonomic ranks (phylum, class, order, etc.) at even phylogenetic depths. Lettered nodes indicate GTDB phyla (a: Firmicutes, b: Firmicutes_D, c: Firmicutes_A, d: Fusobacteria, e: Firmicutes_F).
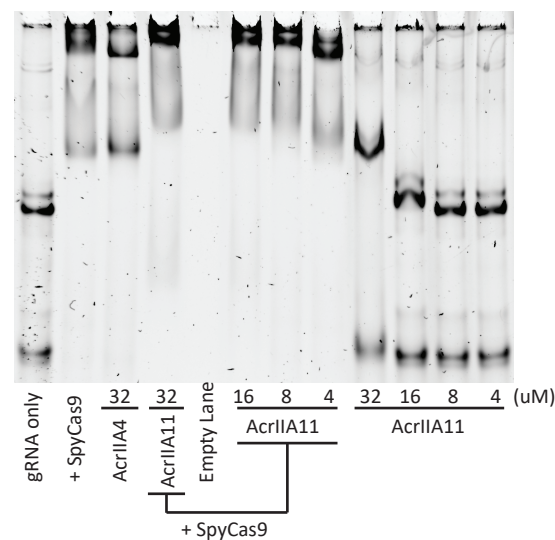
Figure S8. **AcrIIA11 does not prevent a gRNA gel-shift, indicating that the RNA is protein-bound.** An EMSA examining the relative mobility of *S. pyogenes* gRNA (0.2µM) through an 8% acrylamide gel in the presence of SpyCas9 and/or various Acrs. Neither AcrIIA4 nor AcrIIA11 prevent a gel-shift shift upon SpyCas9 addition, though the nature of the shift is different between Acrs. AcrIIA11 also appears to super-shift the SpyCas9/gRNA complex, which may represent AcrIIA11 bound to this complex. At high concentrations, AcrIIA11 binds gRNA.

A

AcrIIA11	-	-	-	+	+	+	+	SpyCas9	I-SmaM1	AcrIIA11
SpyCas9	+	+	-	-	+	+	-			
gRNA	-	+	-	-	-	+	-			
I-SmaM1	-	-	+	-	-	-	+			

Pulldown   Protein Inputs

B

AcrIIA4	+	+	+	+	SpyCas9	I-SmaM1	AcrIIA4
SpyCas9	+	+	-	-			
gRNA	-	+	-	-			
I-SmaM1	-	-	+	-			

Pulldown   Protein Inputs
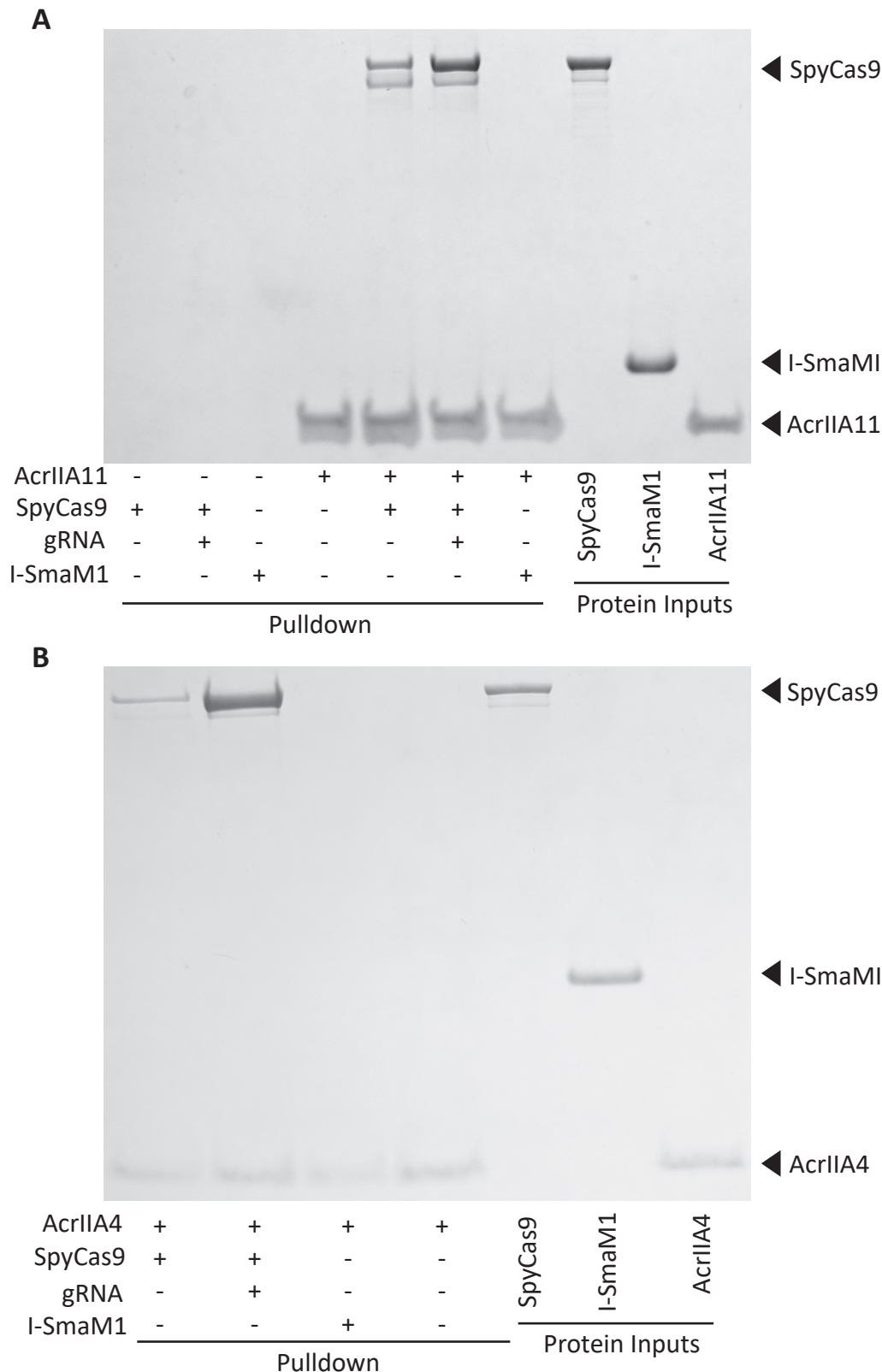
Figure S9. **AcrIIA11 binds SpyCas9 with a moderate preference for the gRNA-loaded form. (A)** AcrIIA11 binds SpyCas9. **(B)** AcrIIA4 binds SpyCas9. SpyCas9 and gRNA were pre-incubated before mixing with a 2x-strep-tagged AcrIIA11 (A) or AcrIIA4 (B). SpyCas9 without gRNA and the meganuclease I-SmaMI were also used. (A) Pulldowns on AcrIIA11 brought with them SpyCas9 but not I-SmaMI, and the presence of gRNA improved the strength of this interaction, but not to the degree seen with AcrIIA4 in (B). These images depict total protein content visualized by Coomassie stain. The gel in (A) is identical to that depicted in fig. 5D except that the three leftmost control lanes have not been cropped from this image.

Table S1. False positive genotypes and frequencies encountered during the development of the Cas9 Acr selection. Asterisks (*) indicate genotyped experiments that are referenced in figure S2. Parentheses indicate the number of colonies queried and the number with the indicated genotype.

Experiment	Target Plasmid	pCas9 used (table S5)	Cas9 expression	No. of target sites	Iteration No.	pCas9 genotype	Target site genotype	Surviving Proportion of Kan^R Transformants	Notes
A	pZE21-GFP	pCas9_crA	Constitutive	1	1	loss-of-function mutant (8/8)	wild-type (8/8)	not determined	Constitutive Cas9 restricted transformation
B	pZE21-GFP	pCas9_crZ	Constitutive	1	1	loss-of-function mutant (6/6)	wild-type (6/6)	not determined	Constitutive Cas9 restricted transformation
C*	pZE21-GFP	pCas9_in_crA	Inducible	1	1	loss-of-function mutant (6/6)	wild-type (6/6)	5.05E-06	In figure S2
D	pZE21-GFP	pCas9_in_crA	Inducible	1	1	no colonies tested	no colonies tested	1.87E-05	In figure S2
E*	pZE21-GFP	pCas9_in_crA	Inducible	1	2	wild-type (6/6)	PAM or protospacer mutant (12/12)	0.03	In figure S2
F	pZE21-GFP	pCas9_in_crA	Inducible	1	1	no colonies tested	no colonies tested	4.82E-05	In figure S2
G*	pZE21-GFP	pCas9_in_crAcrZ	Inducible	2	1	loss-of-function mutant (11/11)	wild-type (10/10)	7.03E-06	In figure S2
H	pZE21-GFP	pCas9_in_crA	Inducible	1	1	no colonies tested	no colonies tested	2.98E-05	In figure S2
I	pZE21-GFP	pCas9_in_crAcrZ	Inducible	2	1	no colonies tested	no colonies tested	1.06E-05	In figure S2
J*	pZE21-GFP	pCas9_in_crA	Inducible	1	2	loss-of-function mutant (4/4)	wild-type (8/8)	2.72E-05	In figure S2
K*	pZE21-GFP	pCas9_in_crAcrZ	Inducible	2	2	loss-of-function mutant (6/6)	wild-type (9/9)	3.44E-06	In figure S2
L*	pZE21-GFP	pCas9_in_crAcrZ	Inducible	2	2	loss-of-function mutant (6/6)	wild-type (20/20)	1.46E-05	In figure S2
M	pZE21-GFP	pCas9_in_crAcrZ	Inducible	2	1	no colonies tested	no colonies tested	4.20E-05	In figure S2
N	pZE21-GFP	pCas9_in_crAcrZ	Inducible	2	2	no colonies tested	no colonies tested	3.77E-04	In figure S2
O	pZE21-GFP	pCas9_in_crAcrB	Inducible	2	1	no colonies tested	no colonies tested	3.75E-05	In figure S2
P	pZE21-GFP	pCas9_in_crAcrB	Inducible	2	2	no colonies tested	no colonies tested	1.27E-04	In figure S2

Table S2. Metagenomic library information. Each library has an average DNA insert size of ~2 Kb (Clemente et al., 2015; Pehrsson et al., 2016). In calculating the number of clones subjected to Cas9 selection, we assumed that transformation is a Poisson process.

Library Information		Iteration #1			Iteration #2		
Library	Unique Clones (x10^6)	Transformants (x10^6)	Unique Clones Subjected to Cas9 Selection (x10^6)	KanR Colonies Collected	Amt. Transformed (ng)	Transformants (x10^6)	Titer Volumes Collected
Oral_3	3.65	5	2.72	2850	272	56	2ml / 100µl
Oral_5	3.88	7.85	3.37	3580	396	71.8	100µl / 1µl
Fecal_01A	1.65	21.4	1.65	3110	261	32.9	1µl / 0.01µl
Fecal_01E	1.35	4.88	1.31	520	236	29.8	100µl / 1µl
Fecal_03G	2.7	4.36	2.16	2910	216	46.2	1µl / 0.01µl
GFP	n/a	42.8	n/a	5650	308	3.86	100µl

Table S3. Summary of twice-iterated functional selections for Cas9 Acrs. Wild-type target site statistics were determined using bulk plasmids surviving two rounds of SpyCas9 selection. The 'ts' acronym stands for 'target site', which encompasses both the protospacer and PAM sequences required for SpyCas9 restriction.

Sample	GFP +ara	Oral_3	Oral_5	Fecal_01A	Fecal_01E	Fecal_03G
% wild-type target sites (tsA/tsB)	99.4/99.0	98.5/86.2	98.8/82.2	95.9/95.9	99.6/54.9	32.2/0.4
Contigs assembled (titer plate 1 / 2)	n/a	6/10	69/45	14/8	28/15	7/4
Contigs with RPBM > 2 (titer plate 1 / 2)	n/a	3/6	21/19	7/6	16/8	2/1
Final unique contig count	n/a	3	20	10	17	1

Table S4. All contigs in the final dataset. Contigs are organized by library. Reads per base-pair per million reads (RPBM) indicate coverage for a given contig, which is a proxy for that sequence's abundance following SpyCas9 selection. Proteins with close homologs in a predicted phage were used to classify a given contig as phage-associated (see methods). NCBI taxa were determined by blastn and blastp searches of contig and protein sequences; GTDB taxa were determined as described in the methods and were used throughout most of the manuscript to resolve a well-recognized polyphyly among the Clostridiales.

Library	Contig	RPBM	Phage-Associated?	NCBI Order	NCBI Genus	GTDB Genus	Cas9 in GTDB Genus?
Oral_3	O3_1	5.45	No	Enterobacteriales	Escherichia	Escherichia	No
Oral_3	O3_2	4.31	No	Bacteroidales	Porphyromonas	Porphyromonas	**Yes**
Oral_3	O3_3	9.60	No	Bacteroidales	Porphyromonas	Porphyromonas	**Yes**
Oral_5	O5_1	8.15	**Yes**	Actinomycetales	Actinomyces	Actinomyces	**Yes**
Oral_5	O5_2	7.21	**Yes**	Enterobacteriales	Escherichia	Escherichia	No
Oral_5	O5_3	29.20	No	Neisseriales	Neisseria	Neisseria	**Yes**
Oral_5	O5_4	13.74	No	Bacteroidales	Porphyromonas	Porphyromonas	**Yes**
Oral_5	O5_5	4.31	No	Neisseriales	Neisseria	Neisseria	**Yes**
Oral_5	O5_6	184.49	**Yes**	Bacteroidales	Porphyromonas	Porphyromonas	**Yes**
Oral_5	O5_7	371.30	**Yes**	Neisseriales	Neisseria	Neisseria	**Yes**
Oral_5	O5_8	85.44	No	Neisseriales	Neisseria	Neisseria	**Yes**
Oral_5	O5_9	69.77	No	Bacteroidales	Prevotella	Prevotella	**Yes**
Oral_5	O5_10	4.15	No	Bacteroidales	Prevotella	Prevotella	**Yes**
Oral_5	O5_11	7.09	**Yes**	Burkholderiales	Unknown	Unknown	n/a
Oral_5	O5_12	4.25	No	Neisseriales	Neisseria	Neisseria	**Yes**
Oral_5	O5_13	2.94	No	Selenomonadales	Selenomonas	Centipeda	No
Oral_5	O5_14	9.97	No	Neisseriales	Neisseria	Neisseria	**Yes**
Oral_5	O5_15	2.90	No	Bacteroidales	Alloprevotella	Alloprevotella	**Yes**
Oral_5	O5_16	8.32	No	Lactobacillales	Streptococcus	Streptococcus	**Yes**
Oral_5	O5_17	4.93	No	Neisseriales	Kingella	Kingella	**Yes**
Oral_5	O5_18	2.50	No	Neisseriales	Neisseria	Neisseria	**Yes**
Oral_5	O5_19	93.26	No	Bacteroidales	Alloprevotella	Alloprevotella	**Yes**
Oral_5	O5_20	29.47	No	Bacteroidales	Alloprevotella	Alloprevotella	**Yes**

Fecal_01A	F01A_1	3.28	**Yes**	Clostridiales	Ruminococcus	Ruminococcus_E	**Yes**
Fecal_01A	F01A_2	766.28	**Yes**	Clostridiales	Clostridium	Clostridium_Q	**Yes**
Fecal_01A	F01A_3	2.54	**Yes**	Clostridiales	Coprococcus	Coprococcus	**Yes**
Fecal_01A	F01A_4	83.66	**Yes**	Clostridiales	Clostridium	CAG-217	No
Fecal_01A	F01A_5	3.67	No	Bacteroidales	Bacteroides	Bacteroides	**Yes**
Fecal_01A	F01A_6	10.91	No	Clostridiales	Roseburia	Roseburia	**Yes**
Fecal_01A	F01A_7	2.25	No	Coriobacteriales	Collinsella	Collinsella	**Yes**
Fecal_01A	F01A_8	3.43	No	Clostridiales	Ruminococcus	Ruminococcus_B	**Yes**
Fecal_01A	F01A_9	4.95	No	Clostridiales	Unknown	KLE1615	No
Fecal_01A	F01A_10	3.26	No	Bacteroidales	Bacteroides	Bacteroides	**Yes**
Fecal_01E	F01E_1	6.72	No	Bifidobacteriales	Bifidobacterium	Bifidobacterium	**Yes**
Fecal_01E	F01E_2	20.76	No	Bifidobacteriales	Bifidobacterium	Bifidobacterium	**Yes**
Fecal_01E	F01E_3	6.36	No	Bifidobacteriales	Bifidobacterium	Bifidobacterium	**Yes**
Fecal_01E	F01E_4	5.85	No	Bacteroidales	Bacteroides	Bacteroides	**Yes**
Fecal_01E	F01E_5	2.89	No	Bifidobacteriales	Bifidobacterium	Bifidobacterium	**Yes**
Fecal_01E	F01E_6	3.05	No	Bifidobacteriales	Bifidobacterium	Bifidobacterium	**Yes**
Fecal_01E	F01E_7	2.73	No	Enterobacteriales	Escherichia	Escherichia	No
Fecal_01E	F01E_8	239.49	No	Bifidobacteriales	Bifidobacterium	Bifidobacterium	**Yes**
Fecal_01E	F01E_9	3.04	No	Enterobacteriales	Escherichia	Escherichia	No
Fecal_01E	F01E_10	46.35	No	Enterobacteriales	Escherichia	Escherichia	No
Fecal_01E	F01E_11	6.75	No	Bacteroidales	Bacteroides	Bacteroides	**Yes**
Fecal_01E	F01E_12	463.50	No	Bifidobacteriales	Bifidobacterium	Bifidobacterium	**Yes**
Fecal_01E	F01E_13	2.01	No	Bifidobacteriales	Bifidobacterium	Bifidobacterium	**Yes**
Fecal_01E	F01E_14	2.96	No	Bifidobacteriales	Bifidobacterium	Bifidobacterium	**Yes**
Fecal_01E	F01E_15	151.77	No	Bifidobacteriales	Bifidobacterium	Bifidobacterium	**Yes**
Fecal_01E	F01E_16	51.89	**Yes**	Bifidobacteriales	Bifidobacterium	Bifidobacterium	**Yes**
Fecal_01E	F01E_17	15.19	No	Bacteroidales	Bacteroides	Bacteroides	**Yes**
Fecal_03G	F03G_1	2.06	No	Bifidobacteriales	Bifidobacterium	Bifidobacterium	**Yes**

Table S5. Plasmids used in this study.

Plasmid	crRNA #1 promoter, sequence (5'-3')	crRNA #2 promoter, sequence (5'-3')	Notes	Refs	Purpose	Figures/Tables
pZE21_MCS1	n/a	n/a	PAMs for protospacers A,Z,B: AGG,TGG,AGG		Library construction, library screening, contig re-testing	Figs 1, S1, S2 / Tables S1, S2, S3
pZE21_RBS	n/a	n/a	Deletion of multiple cloning site moves RBS near clone site	(Forsberg et al., 2015)	Express GFP in pZE21	Figs 1, S1 / Tables S1, S2, S3
pZE21_tetR	n/a	n/a	Contains tetR behind pLac promoter for inducible expression of candidate Acrs			Figs 3, 4, S6
pCas9_crA	pJ23100, GTTCATTCAGGGC ACCGGAC	n/a	Addgene #48645 with crRNA_A	(Esvelt et al., 2013)	Acr selection development	Fig S1 / Table S1
pCas9_crZ	pJ23100, GCCCTGCAAAGTA AACTGGA	n/a	Addgene #48645 with crRNA_Z	(Esvelt et al., 2013)	Acr selection development	Fig S1 / Table S1
pCas9_in_crA	pJ23100, GTTCATTCAGGGC ACCGGAC	n/a	pCas9_crA with AraC + pBad for inducible Cas9	(Lee et al., 2015)	Acr selection development, contig/ORF re-testing	Figs 1, 2, 3, 4, S1, S2 / Table S1
pCas9_in_crA_crZ	pJ23100, GTTCATTCAGGGC ACCGGAC	pJ107106, GCCCTGCAAAGTA AACTGGA	pCas9_in_crA with I-SceI site and crRNA_Z		Acr selection development	Fig S1 / Table S1
pCas9_in_crA_crB	pJ23100, GTTCATTCAGGGC ACCGGAC	pJ107111, GAACAAGATGGAT TGCACGC	pCas9_in_crA with I-SceI site and crRNA_B		Acr selection development, library screening	Figs 1, S1, S2 / Tables S1,S2,S3
pCas9_in_crMu	pJ23100, GTAATACTTGTCC CGCAAAG	n/a	Mu-targeting spacer for phage Mu immunity testing		Phage Mu immunity testing	Figs 3, S6
pCa9_in_crNT	pJ23100, GAACGAAAAGCTG CGCCGGG	n/a	non-targeting spacer used in phage Mu immunity testing		Phage Mu immunity testing	Figs 3, S6

Table S6. Oligos used in this study.

Oligo_Num	Oligo Name	Sequence (5'->3')	Purpose	TM (C)	Ext. Time
1	O5_3_F	CCGGGCCCCCCCTCGAGGTCAGTTATCGGCGAAGCTTCAGA	Amplify O5_3 for cloning into pZE21_MCS1	65	3.5 min
2	O5_3_R	TCAAGCTTATCGATACCGTCGTATCGATGTGCCGGAAAGC	Amplify O5_3 for cloning into pZE21_MCS1	65	3.5 min
3	O5_8_F	CCGGGCCCCCCCTCGAGGTCCTCATCCAATGGACTTTGCAC	Amplify O5_8 for cloning into pZE21_MCS1	61	3.5 min
4	O5_8_R	TCAAGCTTATCGATACCGTCTGCTCAAAAATTCAATTTACAGGG	Amplify O5_8 for cloning into pZE21_MCS1	61	3.5 min
5	O5_9_F	CCGGGCCCCCCCTCGAGGTCGGTCTAGACCTTCACTTCCATTGGATT	Amplify O5_9 for cloning into pZE21_MCS1	65	3.5 min
6	O5_9_R	TCAAGCTTATCGATACCGTCACCGTCGGCTTGCATTGAG	Amplify O5_9 for cloning into pZE21_MCS1	65	3.5 min
7	O5_14_F	CCGGGCCCCCCCTCGAGGTCGGTTCAAGCCGTTTACGGTCG	Amplify O5_14 for cloning into pZE21_MCS1	65	3.5 min
8	O5_14_R	TCAAGCTTATCGATACCGTCCGATACCGTGCGGCGTT	Amplify O5_14 for cloning into pZE21_MCS1	65	3.5 min
9	O5_16_F	CCGGGCCCCCCCTCGAGGTCCGAGGTCACCTTGGGGCAT	Amplify O5_16 for cloning into pZE21_MCS1	65	3.5 min
10	O5_16_R	TCAAGCTTATCGATACCGTCATACCGTCGTTTTTACCAGTTGATG	Amplify O5_16 for cloning into pZE21_MCS1	65	3.5 min
11	O5_19_F	CCGGGCCCCCCCTCGAGGTCCGCGTGCCTCATAGAGTTG	Amplify O5_19 for cloning into pZE21_MCS1	65	3.5 min
12	O5_19_R	TCAAGCTTATCGATACCGTCACCGTCGCAGTTGTCTTCA	Amplify O5_19 for cloning into pZE21_MCS1	65	3.5 min
13	O5_20_F	CCGGGCCCCCCCTCGAGGTCGTCAACTGTAAAGTGATGCGTGC	Amplify O5_20 for cloning into pZE21_MCS1	65	3.5 min
14	O5_20_R	TCAAGCTTATCGATACCGTCATACCGTCCTACGCAGATGCT	Amplify O5_20 for cloning into pZE21_MCS1	65	3.5 min
15	F01A_6_F	CCGGGCCCCCCCTCGAGGTCCCTGCACCTTTTCTGCAC	Amplify F01A_6 for cloning into pZE21_MCS1	61	3.5 min
16	F01A_6_R	TCAAGCTTATCGATACCGTCATACCGTCAGAATAAATGTATCGGTTG	Amplify F01A_6 for cloning into pZE21_MCS1	61	3.5 min
17	F01A_9_F	CCGGGCCCCCCCTCGAGGTCAGCATGCTGGACTGTATCATC	Amplify F01A_9 for cloning into pZE21_MCS1	61	3.5 min
18	F01A_9_R	TCAAGCTTATCGATACCGTCCCGAAACGGAGCTGGAAA	Amplify F01A_9 for cloning into pZE21_MCS1	61	3.5 min
19	F01A_2Δ1	TTATAATACTTTACGATTTATGCTGTTATCATTACGCACATTC	Generate early STOP codon in F01A_2_1 by Gibson Assembly	61	7 min
20	F01A_2Δ1	GTGCGTAATGATAACAGCATAAATCGTAAAGTATTATAACGATTATAACCAGA	Generate early STOP codon in F01A_2_1 by Gibson Assembly	61	7 min

21	F01A_2Δ2	CGCCTGCGCTCCCAGTATTAGTCCTCATCCGGTATATTCCAC	Generate early STOP codon in F01A_2_2 by Gibson Assembly	65	7 min
22	F01A_2Δ2	GAATATACCGGATGAGGACTAATACTGGGAGCGCAGG	Generate early STOP codon in F01A_2_2 by Gibson Assembly	65	7 min
23	F01A_2Δ3	CAAAACTGTCATAATGTCATATCTGCC	Generate early STOP codon in F01A_2_3 by site directed mutagenesis	63	7 min
24	F01A_2Δ3	CGAGAGATACCGCAAGGG	Generate early STOP codon in F01A_2_3 by site directed mutagenesis	63	7 min
25	F01A_2Δ4	TCTCTTTCTCAGTATGCGTGC	Generate early STOP codon in F01A_2_4 by site directed mutagenesis	63	7 min
26	F01A_2Δ4	AAGCGGAGGGAGATACAT	Generate early STOP codon in F01A_2_4 by site directed mutagenesis	63	7 min
27	F01A_2Δ5	CGTTTTCCTCAGCCAATCCTGC	Generate early STOP codon in F01A_2_5 by site directed mutagenesis	65	7 min
28	F01A_2Δ5	CCTGGAGCTGACACAGCC	Generate early STOP codon in F01A_2_5 by site directed mutagenesis	65	7 min
29	pCRT7_F	CAATTCCGACGTCTAAGAAACTGGATAACCGTATTACCGCCTTTG	Amplify tetR expression regulon from pCRT7 for Gibson Assembly with pZE21_MCS1	64	3.5 min
30	pCRT7_R	CTTTGAGTGAGCTGATACCGCTCATTCAGGCTGCGCAACT	Amplify tetR expression regulon from pCRT7 for Gibson Assembly with pZE21_MCS1	64	3.5 min
31	pZE21_tetR_F	AGCGGTATCAGCTCACTCAAAG	Amplify pZE21_MCS1 backbone for Gibson Assembly with pCRT7 tetR regulon	64	3.5 min
32	pZE21_tetR_R	GTTTCTTAGACGTCGGAATTGC	Amplify pZE21_MCS1 backbone for Gibson Assembly with pCRT7 tetR regulon	64	3.5 min
33	AcrIIA11_KpnI	CATTAAAGAGGAGAAAGGTACCATGGCAGATATGACATTAAGACAG	Amplify AcrIIA11 with KpnI and HindII sites for cloning into pZE21_tetR	63	1 min
34	AcrIIA11_HindIII	GCAGCCGGATCAAGCTTTCATTAGTCCTCCTTGAGATTCCACA	Amplify AcrIIA11 with KpnI and HindII sites for cloning into pZE21_tetR	63	1 min

35	AcrIIA4_KpnI	CATTAAAGAGGAGAAAGGTACCATGAATATTAATGACTTAATTAGAGAAATC	Amplify AcrIIA4 with KpnI and HindII sites for cloning into pZE21_tetR	58	1 min
36	AcrIIA4_HindIII	GCAGCCGGATCAAGCTTTCATTAGTTCAACTCACTTTTTAAGGTG	Amplify AcrIIA4 with KpnI and HindII sites for cloning into pZE21_tetR	58	1 min
37	SpyCas9_targetDNA	ACGTTCCAACTTTCACCATAATGAAATAAGATCACTACAGGGCGTATTTTTTGAGTTATC	Double-stranded SpyCas9 target DNA olligo, ordered with 5' 6-FAM conjugate from IDT	n/a	n/a

Table S7. Acr sequences used to create the phylogenetic tree in figure 4A.

Protein Name	NCBI ID	IMG/VR ID	Used in	Sequence
AcrIIA11a.1 (AcrIIA11)	--	--	functional tests, phylogenetic trees	MADMTLRQFCERYRKGDFLAKDRETQIEAGWYDWFCDDKALAGRLA KIWGILKGITSDYILDNYRVWFKNNCPMVGPLYDDVRFEPLDEEQRDE LYFGVAIDDKRREKKYVIFTARNDYENECGFNNVREVRQFINGWEDE LKNEEFYKAREKKRQEMEEANNKFAEIMQRADEILWNLKED
AcrIIA11a.2	WP_064786071.1	--	functional tests, phylogenetic trees	MADMTLREFCERYRKGDFLAKDRNTQIEAGWYDWFCSDKALAGRLA KIWSILKGVTSNYILDNYRVWFKNNCPMVGPLYDDVRFEPLDEEKRD ELYFGVAIDDERRDNKYIIFTARNDYEDECGFNNVREVRQFINGWEEE LKNEEFYKERERKKEELKKENDRCLALLRKADEVLGKHEE
AcrIIA11a.3	--	7000000582_____SRS019787_WUGC _scaffold_26282_____SRS019787_WU GC_scaffold_26282__gene_57340	functional tests, phylogenetic trees	MGNEMTLRQFCERYRRGDFLSKDRDVQIEAGWYDWFCSDDALAGR LAKIWNILKGIDSDYVLDNYRVWFKNNCPCEGPLYDDVRFEPIDEDKR DELYFGVAIDCVWHDSEYAVFTARNGYETEREFSNIREVRAFINGWE DALKDEEFYQKRAEKDAAMKRLSEEADRLIKMGEDILKGYQQDSEKT Q
AcrIIA11a.4	--	3300014553_____Ga0134451_100255 _____Ga0134451_10025556	functional tests, phylogenetic trees	MTVREFCERYRNGDFLIKDRNVQIEAGWYDWFCSDTALAGRLAKIWH ILKGIDSDFILDNFRVWFKNNCPIEDPLYDDVRFEPLDESKRDKMYFGI AIDDKRNAHKYTVFSARADYKNEAGFDKVKDVRAFINGWEDALKDPL FYVRKAAIHKAQEQATKEATSGLVKCLTELNAVLEEGQNDEAGK
AcrIIA11a.5	--	3300019372_____Ga0187904_100010 9_____Ga0187904_10001093	functional tests, phylogenetic trees	MKQMTLREFQERYRNGDFLAKDFDTQVRAGWYDWFCHTSSLSNRL KKIFKIIDGITDDWMLDNFRVWFKNNCPCEGPLYDDVRFEPLDEKLRIF QYFVVAIDDKRNDSKYVIYTARMNYEKEAGFGNVREVRAWINDWHKE FDKEGVTWLN
AcrIIA11a.6	--	3300019371_____Ga0187896_100196 1_____Ga0187896_100196114	functional tests, phylogenetic trees	MADMTIREWQDEYRRGMFEDGDFDTQVKAGWYDWFCNDSSLKNKT EKIAKIITGITNDWMLDNFRVWFKNNCPCAGPLYDDVRFEPMDESKR DELYFVVSIDDKREDDKYCVFTARNGYEKEAGFKSIKELKEWINGWE QEVSA
AcrIIA11a.7	--	3300010270_____Ga0129306_100331 0_____Ga0129306_100331014	functional tests, phylogenetic trees	MAEMTIRQWQDEYRRGMFDDGDFDTQVKAGWYDWFCKDESLKRK TDKFAKIINGITNDWMLDNFRIWFKNNCPCARPLYDDIRFEPMDESKR DQLYFVISIDDKREDDKYCVFTARKGYEKEAGLGNVNDVIAWINAWE QEVA
AcrIIA11b.1	--	7000000628_____SRS015782_WUGC _scaffold_33785_____SRS015782_WU GC_scaffold_33785__gene_75489	functional tests, phylogenetic trees	MITVREWIQKFNAGDFDKPNSGAGWYDWFCDDSQLPKRLKAMGNVI KDIKNDFILDNYYVWFKNNCPCCYPLYDDFRLEPITREGDDEKTRVER RNALYFGVDCGHPFGSEFKYEIFTARNGYEVEFKCKNKWEVLRRINE LATDFEKLEI
AcrIIA11b.2	OHE28210.1	--	functional tests, phylogenetic trees	MSKELSLRTWIEKFNQGDFDSKDLETQIKAGWYDWFCKDDSLGNKT KRMGSIVKQFKDCGKLNLDNMYVWFKNNCPLAGPLYDDFRIADIESG DTLFTIQINCFREENRYTVYGKRNDFDTPLFGTESIKELVAWMNEGWA DNV
AcrIIA11b.3	OHE43765.1	--	functional tests, phylogenetic trees	MMEKQQPLKDWIQAFNSGSFESSDVKVQIKAGWYDWFCKDSSLKNK TKRMGNIIKQIKPGGKVDLDNSYVWFKNNCPLQGSLYDDFRIADLESD VTLIVVQLNSPWHDKTYTVYERLTHYEKVVFSTDSVKELVKWLNEGW DTHV
AcrIIA11c.1	WP_006572312.1	--	functional tests, phylogenetic trees	MSVRQWQERFRAGDFSSRDRAVQCEAGWYDWFCRDEALAGHLKRI SKVVLGINAPFILDNYYVWFKNNCPMAGPLYDDVRFEPLSGERGGKH FVVSLDCPHELAKWVLYTERYGYDAPEFCSGNVRNMVQYINSMAAE LEQGIQPAFILEKQAVSDYVLRHEGTCGIPVYRNGEHEFSYISRKDRQ LRKVTVSGGLEALPPGYEAGQAEQHGDLYVFGTEPPILETPPGQSKA AQRKGRER

AcrIIA11c.2	WP_009258904.1	--	functional tests, phylogenetic trees	MNEMSVRTWQERFRAGDFSSRDRAVQCEAGWYDWFCRDDALAGR LKKISSVVLGITDPFILDNYYVWFKNNCPLEGPLYDDVRFEPLTGERD GKYFLVALDSHHELIKWTLYTERYGYDAPEFCCGNVREMTAYINAMA PELAQGIQPRFVLEKAAVGEYVRQHEGKAAYSIRREGDHLFAYQSSR DWKYRTVAVSDSPENVPQGFPAERAEQHGMLYVFPSKAPALDRADY VVRRAQRRKEQTR
AcrIIA11c.3	WP_054338718.1	--	functional tests, phylogenetic trees	MSEERISVRQWQERFRAGAYESKDLKVQCAAGWYDWFCQDQALAG RLKKISRVVMGITDPYILDNYYVWFKNNCPVVGGLYDDVRFEPLSGD RDGKYFVVSLDSLHESLRWTLFTEQYDFGAPEYGCENVRDMVKYINS MAHELEEGVIPAFTVERWAVLKYMQRHDKTVNRPVHRNGEHRFSYT TFPGKQTKNIMVVSDLKDAPPDFVPDHAEQYGSFYLYCPEDTERPAP EQKRSKPARKKEAER
AcrIIA11c.4	WP_016321673.1	--	functional tests, phylogenetic trees	MKEEISIRQWQKQFKAGFYDSPDIHTQCGAGWYDWFCQDRALAGRL KKIAKVVMGVTNPFILDHYYIWFKNNALVSGPMYDDVRFEPLSGKRD GKYFLVRLDCAGRKKWSLFSERYGFFAPEFECGNVRGMAKYIDGIGR QFAQEIQPVFLLEKRAVEHFITQQDGLCDSIVYRAGEHCYHYKSSKNP KLRTAIAASASGPPPDGFPADQAKEFRGILVWSPDGMERDMKKEADA QKKPNLKKKEGTER
AcrIIA11c.5	WP_023346767.1	--	functional tests, phylogenetic trees	MPDELSVRQWQEQFQAGAFERSDYATQCAAGWYDWFCQDSALAG RLKKIGRVVMGITDPFILDNYYVWFKNNCPLNGPLYDDARFEPLSGER GGKYFVVSLDSPHERMKWALVTERYGFDAPEFDCRNIRDMIRYVNSI GPELRQGIIPPFIAEKDAVTAYAQRRGEPEGLHIYRDGEHCYSYTSRQ DRRKRTVLAAASLEDAPPGFVSEQAHSIKGMYVYCPEDAGIPLPDLAP QDTAKSQKRKEPER
AcrIIA11c.6	WP_055271317.1	--	functional tests, phylogenetic trees	MTEMSVRQWQERFRAGDFSSKDRAVQCEAGWYDWFCQDDALAGR LQKLSKVVMGITDPYILDHYYVWFKNNCPLSDPLYDDIRFEPLHGDRS GKYFVVIRDSPHEAHKWTLYTGRHGFEQPEFTCGNVRDMLRHINSM APESWRGNPPPEKAMHPPQKKRKEAER
--	WP_118651841.1		phylogenetic trees	MADMTLRQFCERYRKGDFLAKDRETQIEAGWYDWFCDDKALAGRLA KIWGILKGITSDYILDNYRVWFKNNCPMVGPLYDDVRFEPLDEEQRDE LYFGVAIDDKRREKKYVIFTARNDYENECGFNNVREVRQFINGWEDE LKNEEFYKAREKKRQEMEEANNKFAEIMQRADEILGNLKED
--	WP_118617142.1		phylogenetic trees	MADMTLREFCERYRKGDFLAKDRNTQIEAGWYDWFCSDKALAGRLA KIWSILKGVTSNYILDNYRVWFKNNCPMVGPLYDDVRFEPLDEEKRD ELYFGVAIDDERRDNKYIIFTARNDYEDECGFNNVREVRQFINGWEEE LKNEEFYKERERKKEELKKENDRCLELLRKADEVLGKHEE
--	WP_118315543.1		phylogenetic trees	MADMTLREFCERYRKGDFLAKDRNTQIEAGWYDWFCSDKALAGRLA KIWSILKGVTSNYILDNYRVWFKNNCQMVGPLYDDVRFEPLDEEKRD ELYFGVAIDDERRDNKYIIFTARNDYEDECGFNNVREVRQFINGWEEE LKNEEFYKERERKKEELKKENDRYLELLRKADEVLGKHEE
--	WP_101872220.1		phylogenetic trees	MADMTLREFCERYRKGDFLAKDRNTQIEAGWYDWFCSDKALAGRLA KIWSILKGVTSNYILDNYRVWFKNNCPMVGPLYDDVRFEPLDEEKMD ELYFGVAIDDERRDNKYIIFTARNDYEDECGFNNVREVRQFINGWEEE LKNEEFYKERERKKEELKKENDRCLALLRKADEVLGKHEE
--	WP_118005709.1		phylogenetic trees	MADMTLREFCERYRKGDFLAKDRKTQIEAGWYDWFCDEKALAGRLA KIWSILKGITSDYILDNYRVWFKNNCPMVGPLYDDVRFEPLDEERRDE LYFGVAIDDKRREKKYIIFTARNDYEDECGFNNVRELRQFINGWEDEL KNEEFYKEKERKEEELKKENDKYLALLKEAEGILKKHEE
--	WP_103240931.1		phylogenetic trees	MSCPFQAMEGGNGMERKMALREFCGRYRKGDFKGTERAVQIEAGW YDWFCEDGELEARLEKIWEILRGIRNDYVLDNYRVWFKNNCPVEEPL YDDVRFEPLDEGRRDELYFGVSIHHGKLASGYQVFTARSGYGIEAEF GSVEDVWDFINNWEGNWDDAMLHERKEAEDGRPREMLEGILQVLN EAADMLAGHLD

--	WP_118642345.1		phylogenetic trees	MTEMSVRQWQERFRAGDFSSKDRAVQCEAGWYDWFCQDDALAGR LQKLSKVVMGITDPYILDHYYVWFKNNCPLSGPLYDDVRFEPLHGDR NGRYFVVIRDSPHETHKWTIYTERHGFEQPEFTCANVRDMLRHINTM APETWRDDPQPAKTPRSPQKKRKEAER
--	WP_087378810.1		phylogenetic trees	MTEMSVRQWQERFRAGDFSSKDRAAQCEAGWYDWFCQDDALAGR LQKLSKVVMGITDPYILDNYYVWFKNNCPLSGPLYDDVRFEPLHGDR NGRYFVVIRDSPHETHKWTIYTERHGFEQPEFTCANVRDMLRHINSM APETWRGDPQPAKAPRSPQKKRKEAER
--	WP_009260171.1		phylogenetic trees	MTEMSVRQWQERFRAGDFSSKDRAVQCEAGWYDWFCQDDALAGR LQKLSKVVMGITDPYILDHYYVWFKNNCPLSGPLYDDIRFEPLHGDRS GKYFVVIRDSPHEAHKWTLYTGRHGFEQPEFTCGNVRDMLRHINSM APESWRGNPPPEKAMHPPQKKRKEAER
--	WP_024723214.1		phylogenetic trees	MTEMSVRQWQERFRAGDFSSKDRAVQCEAGWYDWFCQDVALAGR LQKLSKVVMGITDPYILDHYYVWFKNNCPLSGPLYDDIRFEPLHGDRS GKYFVVIRDSPHEAHKWTLYTGRHGFEQPEFTCGNVRDMLRHINSM APESWRGNPPPEKAMHPPQKKRKEAER
--	WP_008980755.1		phylogenetic trees	MTEMSVRQWQERFRAGDFSSKDRAVQCEAGWYDWFCQDDALAGR LQKLSKVVMGITDPYILDNYYVWFKNNCPLSGPLYDDVRFEPLHGDR NGRYFVVIRDSPHETHKWTLYTERRGFEQPEFTCANVRDMLRHINSM APESWRGNPPPEKAMRPPQKKRKEAER
--	WP_044943868.1		phylogenetic trees	MTEMSVRQWQERFRAGDFSSKDRAVQCEAGWYDWFCQDDALAGR LQKLSKVVMGITDPYILDHYYVWFKNNCPLSGPLYDDIRFEPLHGDRS GKYFVVIRDSPHEAHKWTLYTERHGFEQPEFTCGNVRDMLRHINSMA PESWRGNPPPEKAMHPPQKKRKEAER
--	WP_021633244.1		phylogenetic trees	MTEMSVRQWQERFRAGDFSSKDRAVQCEAGWYDWFCQDVALAGR LQKLSKVVMGITDPYILDHYYVWFKNNCPLSGPLYDDIRFEPLHGDRS GKYFVVIRDSPHEAHKWTLYTERHGFEQPEFTCGNVRDMLRHINSMA PESWRGNPPPEKAMHPPQKKRKEAER
--	WP_122789648.1		phylogenetic trees	MNEMSVREWQERFRAGDFSSRDRAVQCEAGWYDWFCRDDALAGR LKKISGVVLGITDPFILDNYYVWFKNNCPLDGPLYDDVRFEPLTGERD GKYFVVSLDSPHEHMKWALVTERYGYDAPEFECGNVRDMVKYINAIA PELARGIQPRFVQEKAAVGEYVRQHEGKSSYSIRRAGDHLFAYQSPR DWKYRTVAVSDSLENVPQGFPAEQSEQHGMLYVFPSEAPALDRADM VQRAQRRKEQTR
--	WP_087265981.1		phylogenetic trees	MNEMSVRTWQERFRAGDFSSRDRAVQCEAGWYDWFCRDDALAGR LKKISGVVLGITDPFILDNYYVWFKNNCPVNGPLYDDVRFEPLTGERD GKYFVVSLDSPHERMKWALVTERYGYDAPEFECGNVRDMVKYINAIA PELAQGIQPRFVLEKAAVGEYVRQHEGKSSYSIRRAGDHLFAYQSPR DWKYRTVAVSDSLENVPQGFPAEQAEQHGMLYVFPSEAPALDRADM VQRAQRRKEQTR
--	WP_006574413.1		phylogenetic trees	MTEMSVRQWQERFRSGDFSSKDRAVQCEAGWYDWFCQDDALAGR LQKLSKVVMGITDPYILDNYYVWFKNNCPLSGPLYDDVRFEPLRGDR NGKYFVVIRDSPHETHKWTLYTERHGFEQPEFTCANVRDMLRHINSM APETWRGDPQPAKAPRSPQNKRKEAER
--	OLA43941.1		phylogenetic trees	MNEMSVRTWQERFRAGDFSSRNRAVQCEAGWYDWFCRDDALAGR LKKISGVVLGITDSFILDNYYVWFKNNCPLDGPLYDDVRFEPLTGERD GKYFVVSLDSPHEHMKWALVTERYGYDAPEFECGNVRDMVKYINAIA PELARGIQPRFVQEKAAVGEYVRQHEGKSSYSIRRAGDHLFAYQSPR DWKYRTVAVSDSPENVPQGFPAELAEQHCMLYVFPSEAPALDRADV LQRAQRRKEQTR
--	WP_087265909.1		phylogenetic trees	MTEMSVRQWQERFRAGDFSSKDRAVQCEAGWYDWFCQDDALAGR LQKLSKVVMGITDPYILDHYYVWFKNNCPLSGPLYDDVRFEPLHGDR

				NGRYFVVIRDSPHETHKWTIYTERHGFEQPEFTCANVWDMLRHINTMAPETWRGDPQPAKAPHSPQKKRKEAER
--	WP_021749148.1		phylogenetic trees	MTEMSVRQWQERFRAGDFNSRDLSVQCEAGWFDWFCRNDALAGRLKKLSSAVLGIKAPFILDNYYVWFKNNCPMAGPLYDDVRFEPLSGEREGKYFVITLDCPHELAKWVLYTERYGYDAPEFCSGNVRDMGQYINSMAAELEQGIQPAFLLEKRAVSKYIFRHEGEHGIPVYRDREHEFSYISRKDRQLRKVMVTDSMEALPPGYGAEQAERHGKLYVFGVEPPIPEGPTVRPKTVQRGGQER
--	WP_054338718.1		phylogenetic trees	MSVRQWQERFRAGAYESKDLKVQCAAGWYDWFCQDQALAGRLKKISRVVMGITDPYILDNYYVWFKNNCPVVGGLYDDVRFEPLSGDRDGKYFVVSLDSLHESLRWTLFTEQYDFGAPEYGCENVRDMVKYINSMAHELEEGVIPAFTVERWAVLKYMQRHDKTVNRPVHRNGEHRFSYTTFPGKQTKNIMVVSDLKDAPPDFVPDHAEQYGSFYLYCPEDTERPAPEQKRSKPARKKEAER
--	CRY95966.1		phylogenetic trees	MPDELSVRQWQERFRAGAFNVQDRYTQCKAGWYDWFCQDHALAGRLKKIGRVVMGITDPFILDNYYVWFKNNCPLNGPLYDDVRFEPLSGERGGKYFVVSLDSPHEREKWALVTERYGFDAPEFDCRDIREMVKYINSIGPELQQGIVPPFIAEKDAVTAYTRLRGEPEGIHIYRDGNHRFSYAPRRGRQMRTVLTAASLEDAPPGFVSEQAHSIKGMYVYCPEDAGLPLPDLVPQDTTKSQKRKEPER
--	WP_023347629.1		phylogenetic trees	MPDELSVRQWQEQFRAGAFNLQDRYTQCRAGWYDWFCQDHALAGRLKKIGRVVMGITDPFILDNYYVWFKNNCPLNGPLYDDARFEPLSGNRDGKYFVISLDSPHERMKWALVTERYGFDAPEFDCRDIREMIRYVNSIGPELQKGVIPPFIAEKDAVTAYARQRGEPEGLHIYRDGDHQYSYTSRRDRRKRTVLAAASLENAPAGFVSEQAHAVKGMYLYCPEDVGIPLPEHPENIKKSREKKGVER
--	WP_106720848.1		phylogenetic trees	MSKEQSLRTWIESFNKGEFDSKDLQTQIKAGWYDWFCKDESLANKTKRMGNIVKQFKDGGKVNLDKIYVWFKNNCPLAGPLYDDFRIADIESGDTLFTIQINCFREEKRYTVYGKKNDFDTSLLNTDSIKELVSWFNEGWQDNV
--	WP_118225168.1		phylogenetic trees	MQKSAKYMRSIFKGGKSKMRSITVREWISKFNNGEFGAGDFDTQTKAGWNDWFCPDRELASRLKKMGNLIKNIENDYILDNFFLVFYNICSIDYPLFDQARFTPFRRIADKDFEFVLSFDCPYNGYKYEICTARSDYHTEFQCNTVEELFDYLKQLTEDYKKEEMNMDCQHQK
--	WP_106699806.1		phylogenetic trees	MSKEQSLRTWIESFNKGEFDSKDLQTQIKAGWYDWFCKDESLVNKTKRMGNIVKQFKDGGKVNLETMYVWFKNNCPLAGSLYDDFRIADIESGDTLFTIQINCFREEKRYTIYGKKNNFDTPLYNTDSIKELVNWFNEGWQDNV
--	WP_012242545.1		phylogenetic trees	MEKQQLLKDLIQAFNSGSFESSDVKVQIKAGWYDWFCKDSSLKNKTKRMGNIIKQIKPGGKVDLDNSYVWFKNNCPLQGSLYDDFRIADLESDVTLIVVQLNSPWHDKTYTVYERLTHYEKVVFSTDSVKELVKWLNEGWETHV
--	PKK98686.1		phylogenetic trees	MEKQIKLSEWIERFKSGEFDRPDTTTQINAGWFDWFCRDTSLANKTKKMGNIIKQIKAGGKVDLETSYVWFKNNCPLNGPLYDDFRIADIETNNNLIVIQIDCVWNDSKYTVYERLDGFDKPAYKTNSSRELVKWLNKGWNE