Supplementary Information

Machine Learning Predicts New Anti-CRISPR Proteins

Simon Eitzinger[1†], Amina Asif[2,3,†], Kyle E. Watters[1,†], Anthony T. Iavarone[4], Gavin J. Knott[1], Jennifer A. Doudna[1,5-9,*], and Fayyaz ul Amir Afsar Minhas[2,10,*]

[1]Department of Molecular and Cell Biology, University of California Berkeley, Berkeley, CA 94720, USA

[2]Department of Computer and Information Sciences, Pakistan Institute of Engineering and Applied Sciences (PIEAS), PO Nilore, Islamabad, Pakistan.

[3]FAST School of Computing, National University of Computer and Emerging Sciences (NUCES), Islamabad, Pakistan

[4]QB3/Chemistry Mass Spectrometry Facility, University of California, Berkeley, Berkeley, CA, USA

[5]Department of Chemistry, University of California Berkeley, Berkeley, CA, 94720, USA

[6]Innovative Genomics Initiative, University of California Berkeley, Berkeley, CA, 94720, USA

[7]Center for RNA Systems Biology, University of California Berkeley, Berkeley, CA, 94720, USA

[8]Howard Hughes Medical Institute, University of California Berkeley, Berkeley, CA, 94720, USA

[9]Molecular Biophysics & Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA

[10]Department of Computer Science, University of Warwick, Coventry, CV4 7AL, UK

[†] These authors contributed equally to this work.

*To whom correspondence should be addressed: doudna@berkeley.edu, fayyaz.minhas14@alumni.colostate.edu

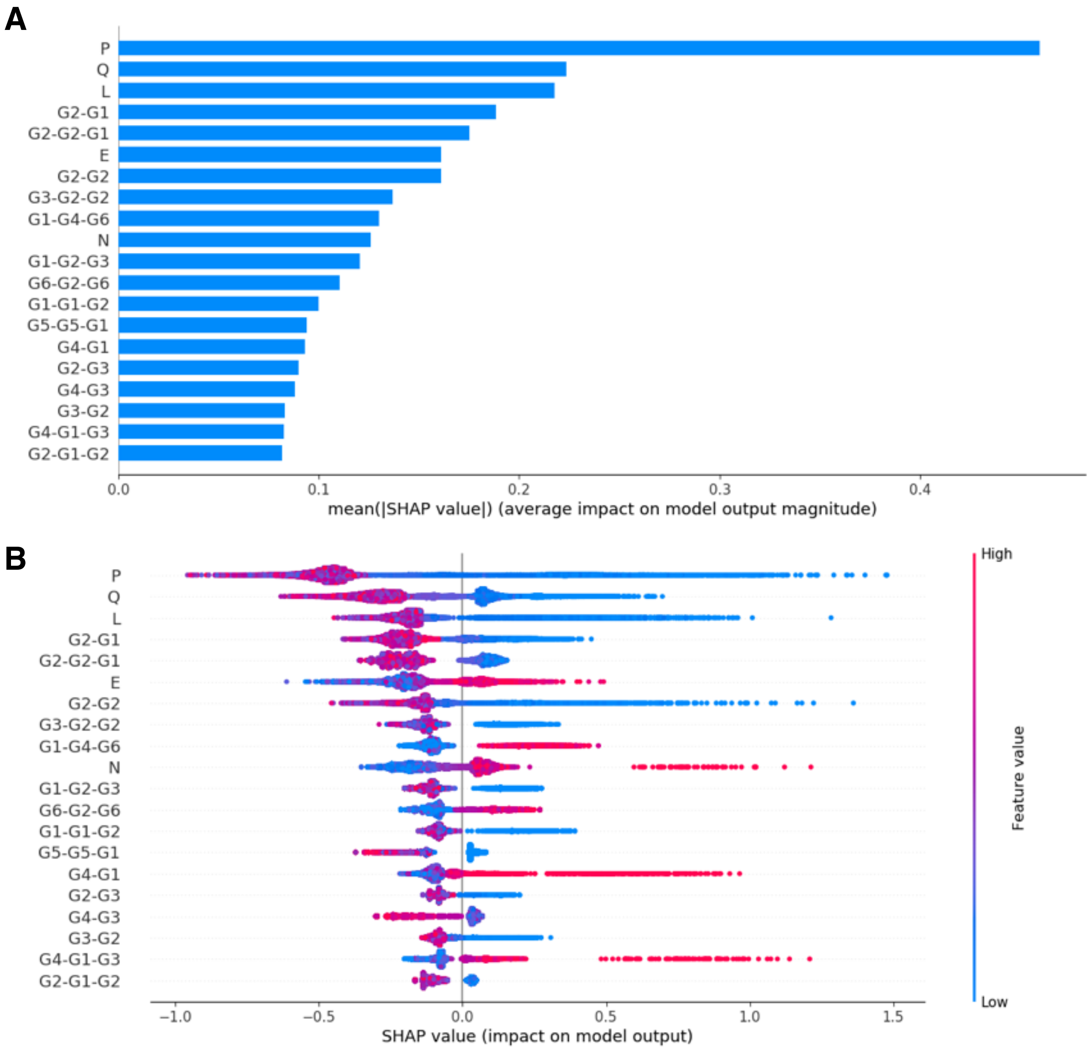

Figure S1. SHAP analysis of AcRanker features. (**A**) Absolute mean of the SHAP (SHapley Additive exPlanations) (1) values as measured for the 20 highest impact features in the AcRanker model. G1-G6 represent amino acid groupings used for computing dimeric and trimeric frequencies in AcRanker. Individual amino acids are grouped according to their side-chain volume and dipole moment (Table S3) (2). (**B**) Violin plots showing the SHAP value vs. the feature value for the 20 highest impact features in AcRanker. Higher feature values (red) with negative SHAP values indicate features that tend to be absent in the training set anti-CRISPRs, while high measured feature values with positive SHAP values suggest features that are more frequently found in the training set anti-CRISPRs. The data suggest that candidates with lower proline (P), glutamine (Q), and leucine (L) content will tend to have higher rankings.

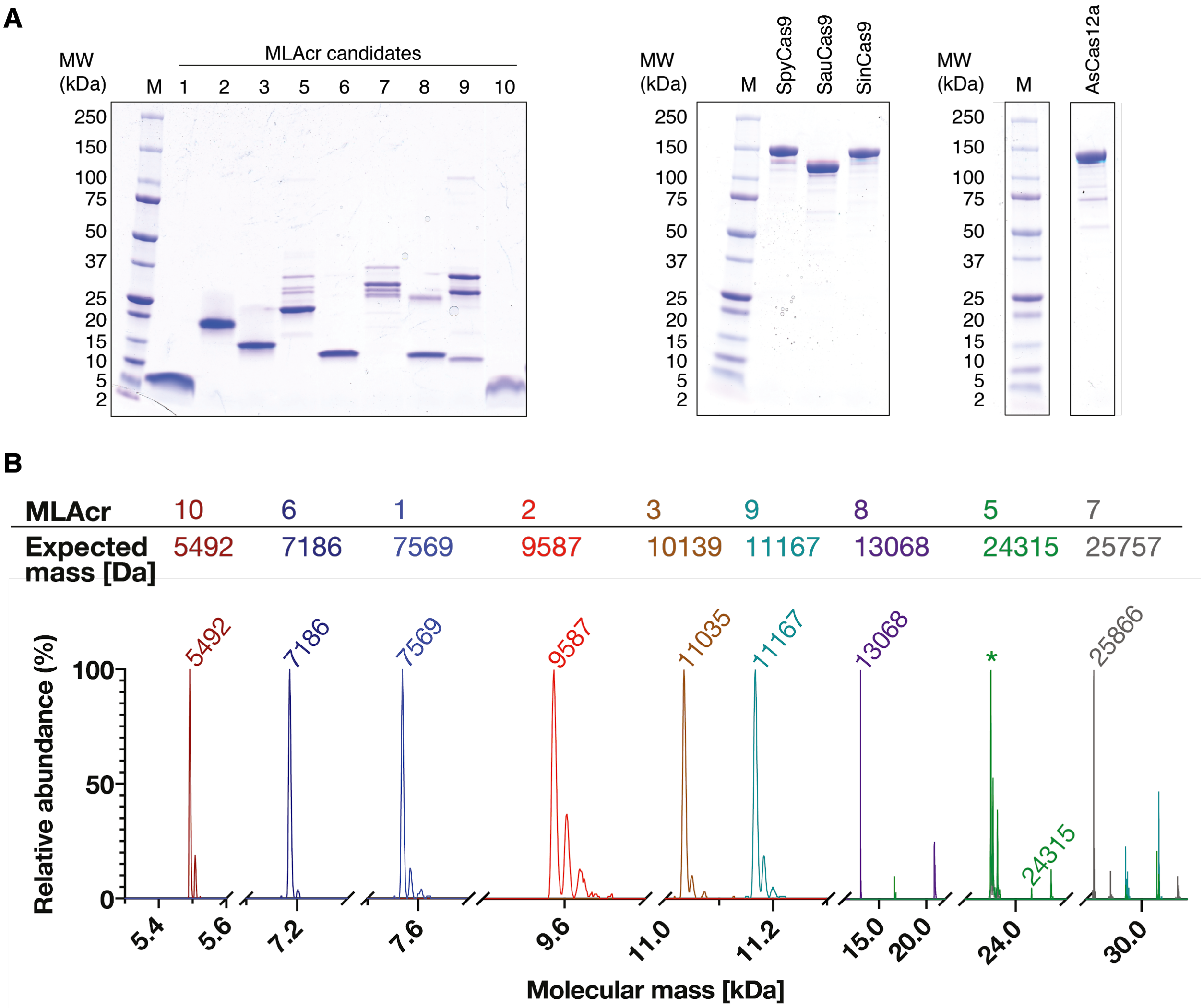

Figure S2. Purified Acr candidates and Cas effectors used in this study. (**A**) 4-20% gradient SDS-PAGE showing a size marker (M) and (left to right) purified machine learning Acr candidates, Cas9 effectors and AsCas12a used in this study. (**B**) Mass spectra of each purified Acr candidate used in this study. The measured mass of ML3 is 896 Da higher than the expected mass. We did not investigate the mass difference any further. ML5 contained a significant unidentified contaminant (*) of 23,510 Da in size.

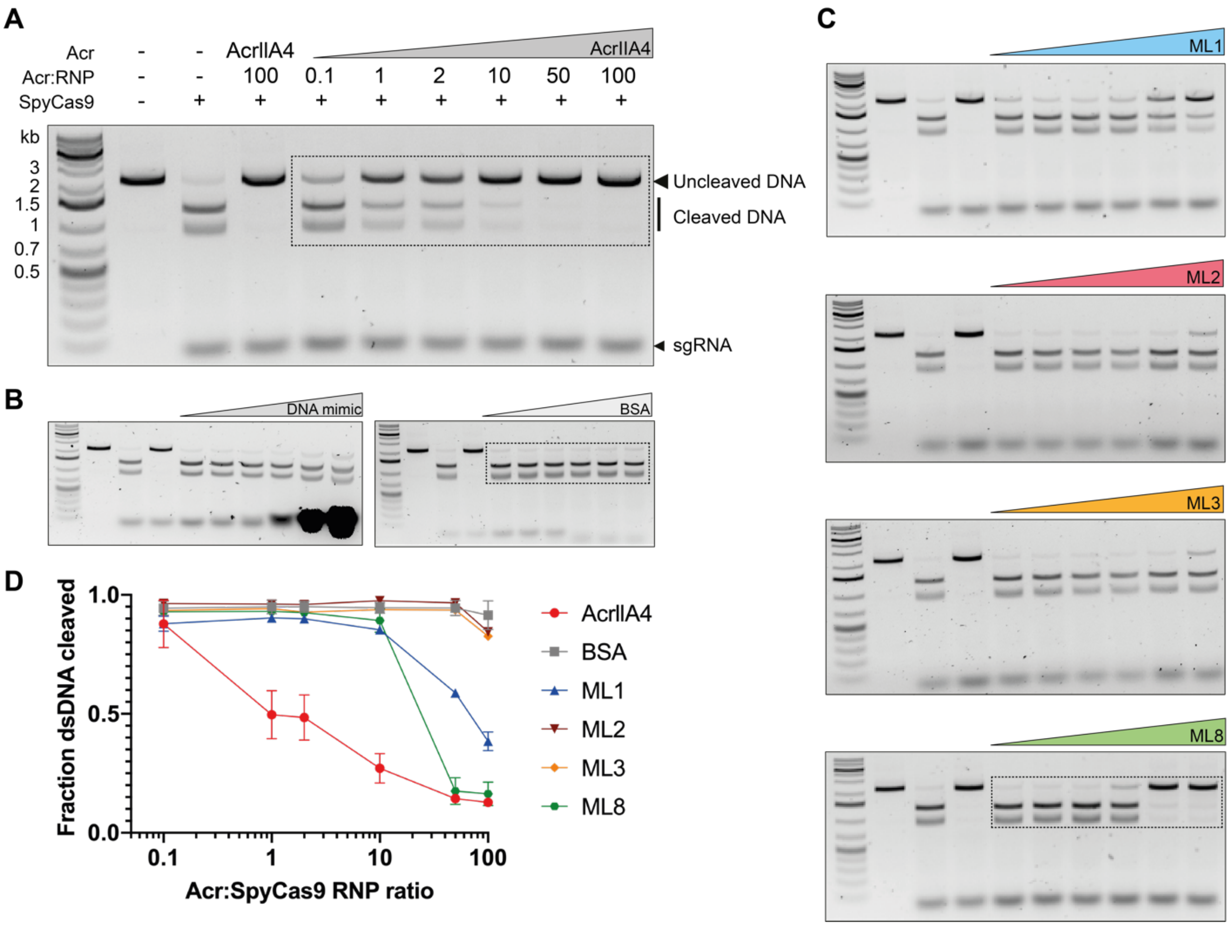

Figure S3. **Inhibition of SpyCas9 by newly discovered Acr candidates.** (**A**) *In vitro* cleavage of dsDNA by SpyCas9 in the presence of increasing concentrations of AcrIIA4 (positive control). (**B**) *In vitro* cleavage of dsDNA by SpyCas9 in the presence of increasing concentrations of (left) DNA mimic and (right) BSA (DNA or BSA:RNP 0.1-, 1-, 2- ,10-, 50- and 100-fold excess from left to right). (**C**) *In vitro* cleavage of dsDNA by SpyCas9 in the presence of increasing concentrations of ML1, ML2, ML3 and ML8 (Acr:RNP 0.1-, 1-, 2- ,10-, 50- and 100-fold excess from left to right). (**D**) Quantified band intensities of the *in vitro* cleavage assays. Fraction of dsDNA cleaved (y-axis) is plotted against the Acr to SpyCas9 RNP ratio (x-axis). AcrIIA4, BSA, ML1 and ML8 were run in triplicates.

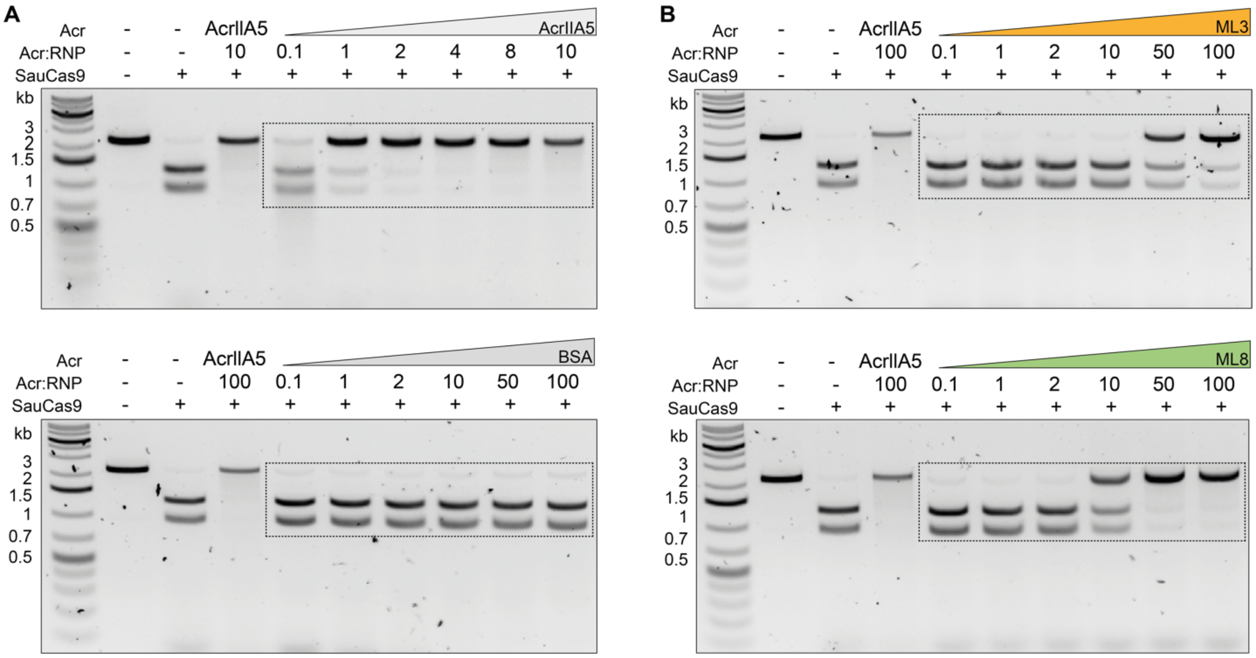

Figure S4. Inhibition of SauCas9 by newly discovered Acr candidates. (**A**) *In vitro* cleavage of dsDNA by SauCas9 in the presence of increasing concentrations of the positive control AcrIIA5 (top) or negative control BSA (bottom). (**B**) *In vitro* cleavage of dsDNA by SauCas9 in the presence of increasing concentrations of ML3 (top) and ML8 (bottom).

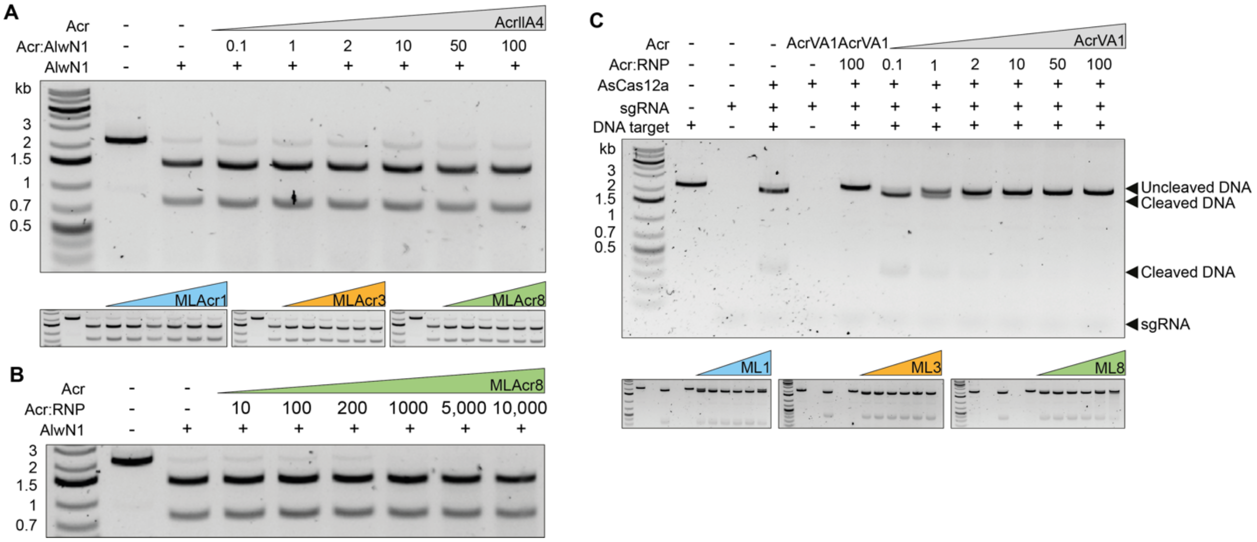

Figure S5. Control experiments for *in vitro* dsDNA cleavage assay. (**A**) *In vitro* cleavage of dsDNA by the restriction enzyme AlwN1 in the absence or presence of increasing concentrations of AcrllA4, ML1, ML3 and ML8 (Acr:AlwN1 0.1-, 1-, 2- ,10-, 50- and 100-fold excess from left to right). (**B**) *In vitro* cleavage of dsDNA by the restriction enzyme AlwN1 in the presence of increasing concentrations of ML8. (**C**) *In vitro* cleavage of dsDNA by AsCas12a in the absence or presence of increasing concentrations of AcrVA1, ML1, ML3 and ML8 (Acr:RNP 0.1-, 1-, 2- ,10-, 50- and 100-fold excess from left to right).

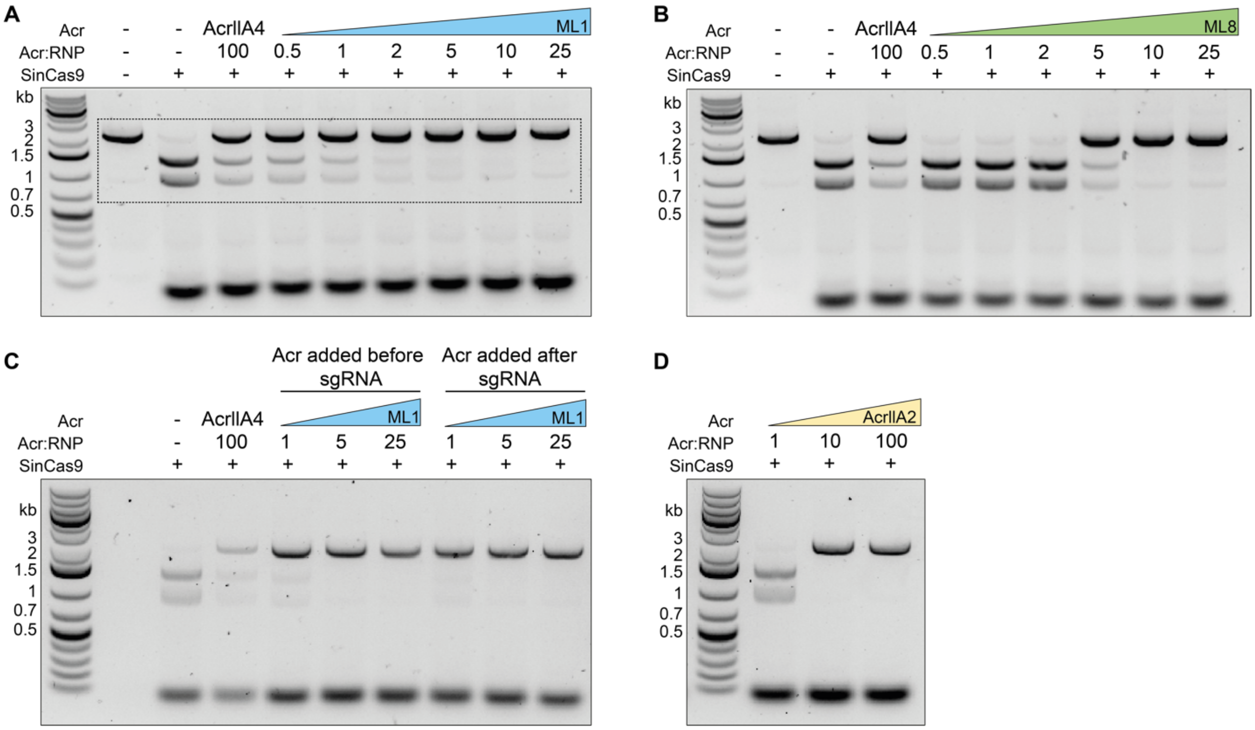

Figure S6. Inhibition of SinCas9 by ML1, ML8 and AcrIIA2. (**A**) *In vitro* cleavage of dsDNA by SinCas9 in the absence or presence of increasing concentrations of ML1. (**B**) *In vitro* cleavage of dsDNA by SinCas9 in the absence or presence of increasing concentrations of ML8. (**C**) *In vitro* cleavage assay where ML1 is incubated with SinCas9 before and after the incubation with sgRNA. (**D**) *In vitro* cleavage of dsDNA by SinCas9 in the presence of increasing concentrations of AcrIIA2. The same DNA target is used in all gels.

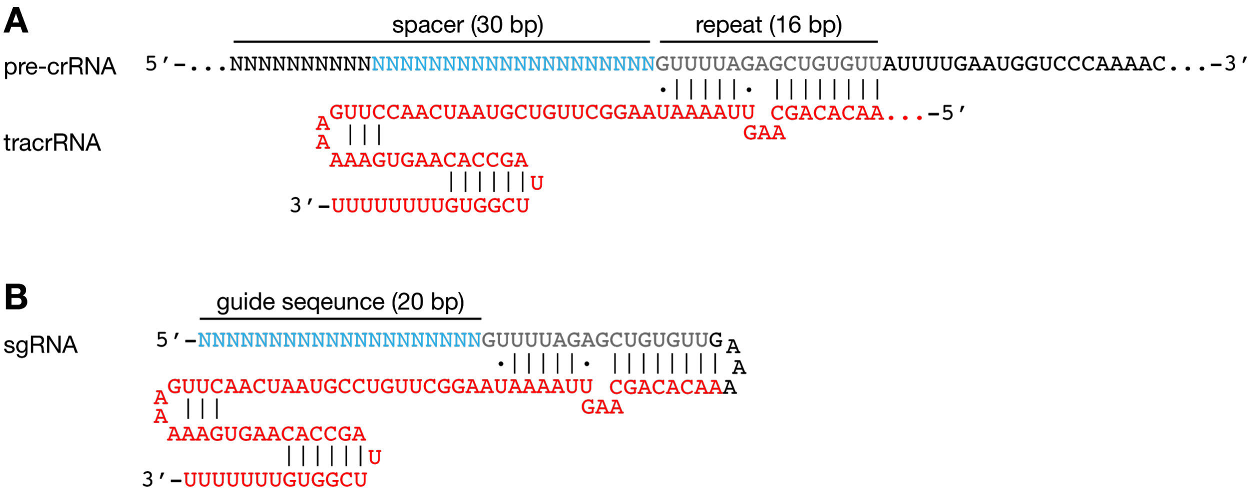

Figure S7. Design of the SinCas9 single guide RNA. (**A**) Predicted structure of the SinCas9 pre-crRNA:tracrRNA complex. (**B**) Schematic of the single-guide RNA designed for use in this study. The 20-bp spacer sequence is shown in blue, tracrRNA is shown in red and the direct repeat sequence in gray.

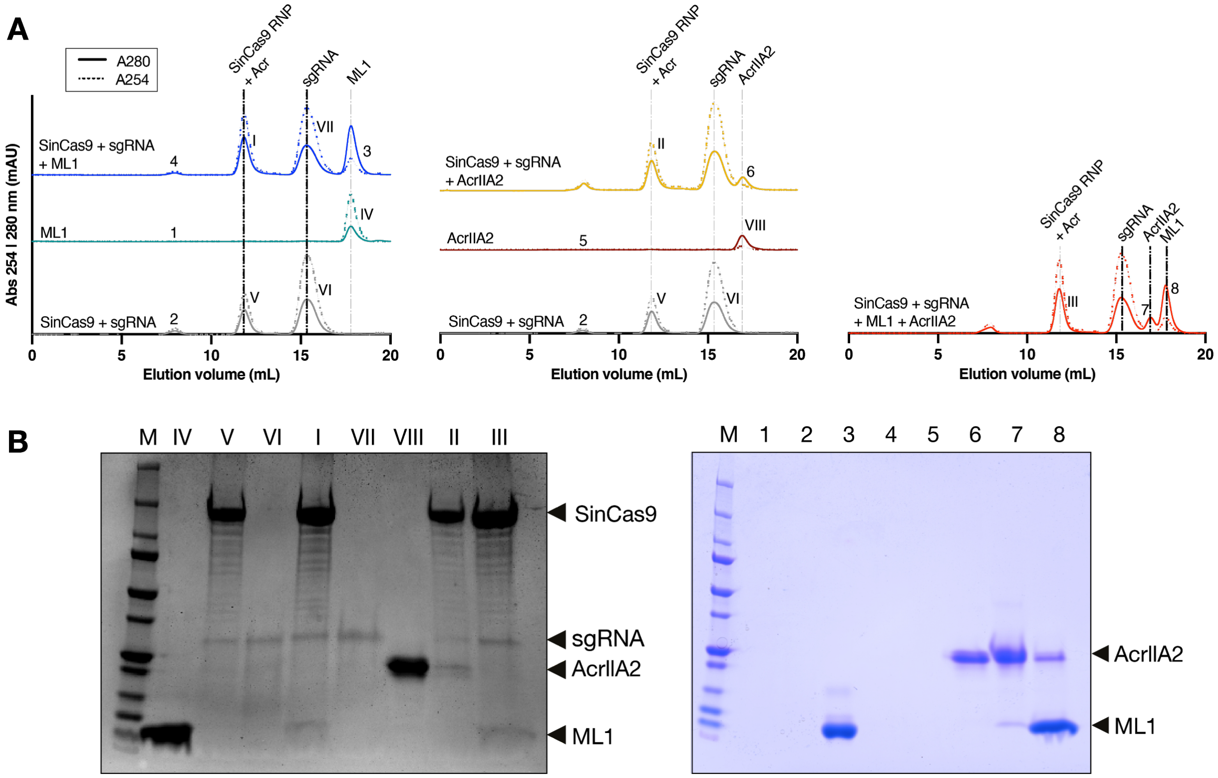

Figure S8. Competition binding experiment between ML1 and AcrIIA2. (**A**) Size-exclusion chromatogram of SinCas9-sgRNA in the absence or presence of ML1 (left), AcrIIA2 (middle) or both (right). (**B**) Coomassie-stained polyacrylamide gel illustrating the components of the fractions annotated with (I) to (VI) and 1 to 8 in panel (A).

Table S1. List of Acrs used for training and cross-validation of the AcRanker model.

Anti-CRISPRdb Name	**Acr Family**	**Protein Accession #**	**Species**	**Proteome Size**	**Ref**
anti_CRISPR0407	AcrIE1	YP_007392738.1	*Pseudomonas* phage JBD5	57	(3)
anti_CRISPR0408	AcrIE3	YP_950454.1	*Pseudomonas* phage DMS3	52	(3)
anti_CRISPR0409	AcrIE2	YP_007392439.1	*Pseudomonas* phage JBD88a	54	(3)
anti_CRISPR0410	AcrIE4	NP_938238.1	*Pseudomonas* phage D3112	54	(3)
anti_CRISPR0001	AcrIF1	YP_007392342.1	*Pseudomonas* phage JBD30	56	(4)
anti_CRISPR0007	AcrIF2	YP_002332454.1	*Pseudomonas* phage MP29	51	(4)
anti_CRISPR0003	AcrIF3	YP_007392440.1	*Pseudomonas* phage JBD88a	54	(4)
anti_CRISPR0002	AcrIF4	YP_007392799.1	*Pseudomonas* phage JBD24	57	(4)
anti_CRISPR0005	AcrIF5	YP_007392740.1	*Pseudomonas* phage JBD5	57	(4)
anti_CRISPR0008	AcrIF6	WP_043884810.1	*Pseudomonas aeruginosa*	6095	(5)
anti_CRISPR0011	AcrIF6	WP_019933870.1	*Oceanimonas smirnovii*	3045	(5)
anti_CRISPR0013	AcrIF6	WP_014702809.1	*Methylophaga frappieri*	2689	(5)
anti_CRISPR0022	AcrIF7	ACD38920.1	*Pseudomonas aeruginosa strain PACS458 clone fa1376 Pseudomonas aeruginosa*	57	(5)
anti_CRISPR0034	AcrIF8	AFC22483.1	*Pectobacterium* phage ZF40	68	(5)
anti_CRISPR0038	AcrIF9	WP_031500045.1	*Vibrio parahaemolyticus*	4928	(5)
anti_CRISPR0051	AcrIF10	KEK29119.1	*Shewanella xiamenensis*	3552	(5)
anti_CRISPR0134	AcrIIA1	AEO04364.1	*Listeria monocytogenes* J0161	2952	(6)
anti_CRISPR0246	AcrIIA2	AEO04363.1	*Listeria monocytogenes* J0161	2952	(6)
anti_CRISPR0384	AcrIIA4	AEO04689.1	*Listeria monocytogenes* J0161	2952	(6)
anti_CRISPR0433	AcrIIA5	D4276_028	*Streptococcus* phage D4276	54	(7)

Table S2. List of Acrs used for independent testing of AcRanker.

Acr Family	Protein Accession	Species	Proteome Size	Ref
AcrIE5	WP_074973300.1	*Pseudomonas otitidis* strain DSM 17224	5731	(8)
AcrIE6	WP_087937214.1	*Pseudomonas aeruginosa* strain S708_C14_RS	6794	(8)
AcrIE7	WP_087937215.1	*Pseudomonas aeruginosa* strain S708_C14_RS	6794	(8)
AcrIE4-F7	WP_064584002.1	*Pseudomonas citronellolis* strain SJTE-3	6260	(8)
AcrIF11	WP_038819808.1	*Pseudomonas aeruginosa* str. C1426	5888	(8)
AcrIF11.1	WP_033936089.1	*Pseudomonas aeruginosa* strain TRN6649	6373	(8)
AcrIF11.2	EGE18857.1	*Moraxella catarrhalis* BC8	1844	(8)
AcrIF12	ABR13388.1	*Pseudomonas aeruginosa* PAGI-5 genomic island sequence	121	(8)
AcrIF13	EGE18854.1	*Moraxella catarrhalis* BC8	1843	(8)
AcrIF14	AKI27193.1	*Moraxella* phage Mcat5	68	(8)
AcrIC1	WP_046701304.1	*Moraxella bovoculi strain 58069*	1944	(8)
AcrIIA3	WP_014930691.1	*Listeria monocytogenes serotype 7 str. SLCC2482*	2822	(6)
AcrIIA6	WP_149028791.1	*Streptococcus phage D1811*	40	(9)
AcrIIA7	AII65827.1	*Bacteroides dorei isolate HS1_L_1_B_010*	4519	(10)
AcrIIA9	WP_004289410.1	*Bacteroides fragilis strain DCMOUH0067B*	4286	(10)
AcrIIA13	AKS70260.1	*Staphylococcus schleiferi strain 5909-02*	2278	(11)
AcrIIC5	WP_002642161.1	*Simonsiella muelleri ATCC 29453*	2170	(12)
AcrIIIB1	NP_666582.1	*Sulfolobus islandicus rod-shaped virus 2*	54	(13)
AcrVA1	WP_046701302.1	*Moraxella bovoculi strain 58069*	1944	(8, 14)
AcrVA4	WP_046699156.1	*Moraxella bovoculi strain 22581*	2105	(14)

Table S3. Grouping of amino acids based on physiochemical properties. Groups of amino acids with similar side chains are grouped together to reduce the number of features to test in the machine learning model (2).

Group #	Dipole Scale[a]	Volume Scale[b]	Amino Acids
1	-	-	A, G, V
2	-	+	I, L, F, P
3	+	+	Y, M, T, S
4	++	+	H, N, Q, W
5	+++	+	R, K
6	+'+'+'	+	D, E
7	+[c]	+	C

[a] Dipole scale (Debye): -, Dipole < 1.0; +, 1.0 < Dipole < 2.0; ++, 2.0 < Dipole < 3.0; +++, Dipole > 3.0; +'+'+', Dipole > 3.0 with opposite orientation

[b] Volume scale ($Å^3$): -, Volume < 50; +, Volume > 50

[c] Cysteine is separated from class 3 because of its ability to form disulfide bonds

Table S4. Comparison of XGBoost classification vs. pairwise ranking models during leave-one-out cross-validation. Each row of the table indicates which Acr was excluded from the training dataset and used as a test dataset, with the number indicating the rank obtained using either a blastp search against all other known Acrs in the training set (blastp), an XGBoost classification model (Class.), an XGBoost pairwise ranking model (Ranking). The best rank achieved by the XGBoost classification or pairwise ranking model within the complete or prophage proteome is marked with an asterisk. The best rank between blastp and either XGBoost model is bolded, and any method that produces the top rank is bolded with two asterisks. The pairwise ranking model performs better than the classification model, with the ranking model receiving a better rank 11 times vs. six times for the classification model in complete bacterial or phage proteomes. In the smaller prophage proteomes the ranking model is ranked higher five times vs. once for the classification model.

Protein	Acr Family	Complete Proteome				Prophage Proteome Subset			
		Size	blastp	AcRanker (XGBoost)		Size	blastp	AcRanker (XGBoost)	
				Class.	Ranking			Class.	Ranking
anti_CRISPR0407	AcrIE1	57	33	9	**1****				
anti_CRISPR0408	AcrIE3	52	17	**1****	**1****				
anti_CRISPR0409	AcrIE2	54	18	5	**2***				
anti_CRISPR0410	AcrIE4	54	11	2	**1****				
anti_CRISPR0001	AcrIF1	56	21	**4***	11				
anti_CRISPR0007	AcrIF2	51	34	**1****	**1****				
anti_CRISPR0003	AcrIF3	54	5	9	**1****				
anti_CRISPR0002	AcrIF4	57	36	**1****	3				
anti_CRISPR0005	AcrIF5	57	26	**19***	**19***				
anti_CRISPR0008	AcrIF6	6095	**1****	69*	80	361	**1****	17	15*
anti_CRISPR0011	AcrIF6	3045	**1****	25	13*	72	**1****	3	**1****
anti_CRISPR0013	AcrIF6	2689	**1****	541	130*	57	-	-	-
anti_CRISPR0022	AcrIF7	57	20	3	**1****				
anti_CRISPR0034	AcrIF8	68	30	3	**1****				
anti_CRISPR0038	AcrIF9	4928	198	**44***	333	37	-	-	-
anti_CRISPR0051	AcrIF10	3552	189	**2***	17	70	23	**1****	2
anti_CRISPR0134	AcrIIA1	2951	183	931	**770***	146	**60**	97	87*
anti_CRISPR0246	AcrIIA2	2952	210	15*	16	146	34	6	**3***
anti_CRISPR0384	AcrIIA4	2951	59	56	**21***	146	9	15	**4***
anti_CRISPR0433	AcrIIA5	54	**5**	12	**8***				

Table S5. Amino acid sequence and accession numbers of all the Acr candidates.

#ML cand.	Accession No.	Sequence
ML1	OHX26873.1	MKNYEVTNEVKNLNTQVETIGQAVDLYKEYGSNTIVWSIDKNEDLIDEVTELVAEYAEKGTVIK
ML2	WP_003731277.1	MGKTYWYNEGTDTLLTEKEYKELMEREAKALYEEVQEEEKDFESSEKTSFEEFLKTCYENESDFVLSDNEGNKLEEW
ML3	WP_003731276.1	MSKTMYKNDVIELIKNAKTNNEELLFTSVERNTREAATQYFRCPEKHVSDAGVYYGEDFEFDGFEIFEDDLIYTRSYDKEELN
ML4	WP_000946250.1	MLRRVNHVKNVLAHGEFAEWIENKIGIHYREANRMMTVAKQIPNVSTLKYLGATAKHVNGVAKRKQNFLSQISLIPTNPQLPHQTIINTYLYWQP
ML5	WP_001080841.1	MNRLKELRKEKKLTQEELAGEIGVSKITILRWENGERQIKPDKAKELAKYFNVSVGYLLGYAPNKKIDFQLNLDGTTLHLTKEQFLALENTSKSIKKIKNTINESVKQEEYIKNASKYYDFEKVSRRLTDRLFEIHTDLIELLMMLDHFPSGELSKSQQEAIFKFYKQLDYFVTDTPASFDYFKKNLESYGYKIYTEGDKIDFD
ML6	WP_000965633.1	MLYIDEFKEAIDKGYILGGTVAIVRKNGKIFDYVLPHEEVREEEVVTVERVEDVMRELE
ML7	WP_000591144.1	MIKIYFGKDAALNQAIQSRLDSYQIDYQAFSSKDIDAKTLMEWLFKSTDIFELLSTKMLKYKLNTQITLSQFVRKILKDVNSTLKLPIVVTDEVIYSNMSPDYVTVLLPKEYRKIKRIQLMRKMEQLDEGRLFWKNFELFRKQSELRWFELNELLFADVSDDLGEIKKAKDRFFSYKKNNQVPPNEIIERILKIFLVDREDFFKKSPSDLQNF
ML8	WP_000384271.1	MDYDNENYLIPKILLQDDFYSSLSAKDILVYAVLKDRQIEALEKGWIDTDGSIYLNFKLIELAKMFSCSRTTMIDVMQRLEEVNLIERERVDVFYGYSLPYKTYINEV
ML9	WP_000134666.1	MTEGFTIQLPKVTEKKLLARYDDMLQKAIEKALEDKELYKPIVRMAGLCRWLDVSTTTVVKWQKQGGMPHMVIDGVTLYDKHKVAQWLQQFER
ML10	WP_011058321.1	MNIEDIERIISEYLIFRSDIDGCAVIDIEDFLKHIRFSYERLK

Table S6. Amino acid sequence of all the Cas effectors used in this study.

Cas effector	Species	Sequence
Cas9	*Streptococcus pyogenes*	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD
Cas9	*Staphylococcus aureus*	MGKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINP

		YEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG
Cas9	*Streptococcus iniae*	MRKPYSIGLDIGTNSVGWAVITDDYKVPSKKMRIQGTTDRTSIKKNLIGALLFDNGETAEATRLKRTTRRRYTRRKYRIKELQKIFSSEMNELDIAFFPRLSESFLVSDDKEFENHPIFGNLKDEITYHNDYPTIYHLRQTLADRDQKADLRLIYLALAHIIKFRGHFLIEGNLDSENTDVHVLFLNLVNIYNNLFEEDIVETASIDAEKILTSKTSKSRRLENLIAEIPNQKRNMLFGNLVSLALGLTPNFKTNFELLEDAKLQISKDSYEEDLDNLLAQIGDQYADLFIAAKKLSDAILLSDIITVKGASTKAPLSASMVQRYEEHQQDLALLKNLVKKQIPEKYKEIFDNKEKNGYAGYIDGKTSQEEFYKYIKPILLKLNGTEKLISKLEREDFLRKQRTFDNGSIPHQIHLNELKAIIRRQEKFYPFLKENQKKIEKLFTFKIPYYVGPLANGQSSFAWLKRQSNESITPWNFEEVVDQEASARAFIERMTNFDTYLPEEKVLPKHSPLYEMFMVYNELTKVKYQTEGMKRPVFLSSEDKEEIVNLLFKKDRKVTVKQLKEEYFSKMKCFHTVTILGVEDRFNASLGTYHDLLKIFKDKAFLDDEANQDILEEIVWTLTLFEDQA

		MIERRLVKYADVFEKSVLKKLKKRHYTGWGRLSQKLINGIKDKQTGKTILGFLKDDGVANRNFMQLINDSSLDFAKIIKHEQEKTIKNESLEETIANLAGSPAIKKGILQSIKIVDEIVKIMGQNPDNIVIEMARENQSTMQGIKNSRQRLRKLEEVHKNTGSKILKEYNVSNTQLQSDRLYLYLLQDGKDMYTGKELDYDNLSQYDIDHIIPQSFIKDNSIDNIVLTTQASNRGKSDNVPNIEIVNKMKSFWYKQLKNGAISQRKFDHLTKAERGALSDFDKAGFIKRQLVETRQITKHVAQILDSRFNSNLTEDSKSNRNVKIITLKSKMVSDFRKDFGFYKLREVNDYHHAQDAYLNAVVGTALLKKYPKLEAEFVYGDYKHYDLAKLMIQPDSSLGKATTRMFFYSNLMNFFKKEIKLADDTIFTRPQIEVNTETGEIVWDKVKDMQTIRKVMSYPQVNIVMKTEVQTGGFSKESILPKGNSDKLIARKKSWDPKKYGGFDSPIIAYSVLVVAKIAKGKTQKLKTIKELVGIKIMEQDEFEKDPIAFLEKKGYQDIQTSSIIKLPKYSLFELENGRKRLLASAKELQKGNELALPNKYVKFLYLASHYTKFTGKEEDREKKRSYVESHLYYFDEIMQIIVEYSNRYILADSNLIKIQNLYKEKDNFSIEEQAINMLNLFTFTDLGAPAAFKFFNGDIDRKRYSSTNEIINSTLIYQSPTGLYETRIDLSKLGGK
Cas12a	*Acidaminococcus sp.*	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHE

		IIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRN

Table S7. sgRNAs used for the *in vitro* cleavage assay.

Cas effector	**Species**	**sgRNA sequence***	**Main Text Fig.**	**Suppl. Fig.**
Cas9	*Streptococcus pyogenes*	**ATACGGGAGGGCTTACCATC**GTTTTAGAGCTATGCTGTTTTGGAAACAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT	2A-B	3A-D
Cas9	*Staphylococcus aureus*	**TATCGTAGTTATCTACACGACG**GTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGATTTTT	2C-D	4A-B
Cas9	*Streptococcus iniae*	**ATACGGGAGGGCTTACCATC**GTTTTAGAGCTGTGTTGAAAAACACAGCAAGTTAAAATAAGGCTTGTCCGTAATCAACTTGAAAAAGTGAACACCGATTCGGTGTTTTTTT	3A-B	6A-D
Cas12a	*Acidaminococcus sp.*	AAUUUCUACUCUUGUAGAU**AAAGUGCUCAUCAUUGGAAAACGU**	-	5C

* Spacer sequences are shown in bold

Table S8. DNA target used for the *in vitro* cleavage assay.

Cas effector	Species	DNA target sequence*	Main Text Fig.	Suppl. Fig.
Cas9	*Streptococcus pyogenes*	AATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCG**TATCGTAGTTATCTACACG**ACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTG	2A-B	3A-D
Cas9	***Staphylococcus aureus***		2C-D	4A-B
Cas9	*Streptococcus iniae* UEL-Si1		3A-B	6A-D
Cas12a	*Acidaminococcus sp.*		-	5C

		CCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTGTGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCAGAGTTTGTAGAAACGCAAAAAGGCCATCCGTCAGGATGGCCTTCTGCTTAATTTGATGCCTGGCAGTTTATGGCGGGCGTCCTGCCCGCCACCCTCCGGGCCGTTGCTTCGCAACGTTCAAATCCGCTCCCGGCGGTTGAGAAGAGAAAAGAAAACCGCCGATCCTGTCCACCGCATTACTGCAAGGTAGTGGACAAGACCGGCGGTCTTAAGTTTTTTGGCTGAAATGCCTGGCAGTTCCCTACTCTCGCATGGGGCTCGCGGTTAACTGATTATTTTATTTATCTAGGCTACTTACGAACG		
		DNA mimic** GCTGACAATGATACGAACGAGACACACGCTCACGACTCAG	-	3B

* Target sequences are shown in blue (*Streptococcus pyogenes, Streptococcus iniae*), green (*Acidaminococcus sp.*) or bold (*Staphylococcus aureus*); **DNA mimic used for control experiments

Table S9. Independent testing set validation results. 20 proteomes containing non-redundant (<40% sequence identity) Acrs from bacterial and phage sources were ranked using AcRanker and blastp. Bacterial proteomes that had Acrs within PHASTER-predicted prophages were also tested with a subset of the proteome containing only the prophage proteins. Cases where the top rank is returned are in boldface.

		Complete Proteome			**Prophage Subset**		
Acr Accession #	**Acr Family**	**Proteome size**	**AcRanker rank**	**Blastp rank**	**Proteome size**	**AcRanker rank**	**Blastp rank**
WP_064584002.1	AcrIE4-F7	6260	68	**1**	111	4	**1**
WP_074973300.1	AcrIE5	5731	10	63	-	-	-
WP_087937214.1	AcrIE6	6794	80	4383	-	-	-
WP_087937215.1	AcrIE7	6794	742	6546	-	-	-
WP_038819808.1	AcrIF11	5888	138	2995	64	3	38
WP_033936089.1	AcrIF11.1	6373	38	2293	92	**1**	38
EGE18857.1	AcrIF11.2	1844	412	90	59	30	**1**
ABR13388.1	AcrIF12	121	7	10	-	-	-
EGE18854.1	AcrIF13	1844	187	755	-	-	-
AKI27193.1	AcrIF14	68	14	3	68	14	3
WP_046701304.1	AcrIC1	1944	6	313	72	**1**	15
WP_014930691.1	AcrIIA3	2822	10	1184	74	2	40
WP_149028791.1	AcrIIA6	40	23	21	40	23	21
AII65827.1	AcrIIA7	4519	179	2208	-	-	-
WP_004289410.1	AcrIIA9	4286	53	930	-	-	-
AKS70260.1	AcrIIA13	2278	22	355	145	3	29
WP_002642161.1	AcrIIC5	2170	10	1954	367	6	237
NP_666582.1	AcrIIIB1	54	44	25	54	44	25
WP_046701302.1	AcrVA1	1944	114	376	72	10	18
WP_046699156.1	AcrVA4	2105	1100	1405	293	220	81

Table S10. List of expected lethal self-targeting *Streptococcus* genomes obtained with Self-Target Spacer Searcher (STSS). Searching *Streptococcus* assemblies from NCBI with STSS returned 385 cases of self-targeting derived from type II-A arrays representing 241 individual genomes. Of those genomes, 20 contained at least one spacer with the characteristic NRG 3′ PAM for SpyCas9, shown in the table below. Only *Streptococcus iniae* strain UEL-Si1 contains a previously discovered anti-CRISPR (AcrIIA3). Also shown in the table are the self-targeting spacers for *Listeria monocytogenes* strain R2-502, which was also ranked with AcRanker.

Target Accession#	**Locus Accession#**	**Species/Strain**	**Self-Targeting Spacer Sequence(s)**	**3′ PAM Region**	**Anti-CRISPRs Present**
NZ_MNAC01000031.1	NZ_MNAC01000010.1	*Streptococcus iniae* strain UEL-Si1	TTGATAAGTATAATTTCCTGTCTTTGTTTT	AGGAGTTTT	AcrIIA3 (WP_071127625.1)
NZ_MNAC01000046.1			TAAGGAATTTGAAGCAATACGTCTTAATTT	AGCAATGAC	
NZ_MNAC01000023.1			CAAAAAAGTTCGGTAACTTACGGTAACTTA	CGGTAACTT	
			TCTAAAAAATCAAAAGTTACCGTGTTACCG	TAGTTTTGA	
			AATATGACTTTTGGGAAATTAACTAATCAA	TGGCTGAAA	
			TTTTTGAGTGTACTGATGTTGCTTTTGAGC	TGGCCACTT	
NZ_MNAC01000021.1			ATAATCAATCACATTAATGCTGACATCAAC	TGGAGCAGA	
			GAGTTTAATTAAGTGACATAATATCTTCAT	CGGTTATAG	
NZ_JRLL01000002.1	NZ_JRLL01000058.1	*Streptococcus pyogenes* SS1447	TCGTCAGATTTGTCAGTATAGTAATCATCA	CGATATAAA	None
NZ_JRLL01000072.1			CTATATTGTTGAGCTGTGGGCTTTGCATAA	AGGTTTAAA	

NZ_JRLL01000026.1			GTAATAATAGCATTGCCTGTTCTATCCTGT	CGGTAGAAC	
NZ_CQAV01000003.1	NZ_CQAV01000001.1	*Streptococcus agalactiae* strain DE-NI-032	TATTTGATAGCGGTAACGGGTCATATACAA	AGGCATCTA	None
			TGGTGGTATTTATAATGTACGAGCAAATCG	AGGCGCTCC	
			ACCTTGCTCCGATGACACCATCGCGAACCT	TGGTCTAAT	
NZ_CP010449.1	NZ_CP010449.1	*Streptococcus pyogenes* strain NGAS322	ATCGTAAGGCAACAGATTATCGTAAGATCT	AGGTGTATA	None
NZ_ALQN01000014.1	NZ_ALQN01000018.1	Streptococcus agalactiae CCUG 37430	ATTTGCAACTTTCTCAAGTGTTGCGAGAGA	TGGAGAATT	None
NZ_ALQN01000018.1			GCAAGCACTAAATGAAGCTACTAGACTTAA	AGGTCGCAG	
			TAATGACATGTGGATTGATATCTCAGAGAA	CGGCGATTA	
			TGTCATTGTTAAAATCATTTGCATATTTTT	TGGATATAA	
			TACTTGACGAATTGAAGATGACGGAATTTA	TTGCTCCAC	
NZ_CPVL01000019.1	NZ_CPVL01000003.1	*Streptococcus agalactiae* strain DE-NI-007	AAGGCACGCGCAAGATGAATTCATTTCTAA	TGGCTACAC	None
			TGATGTTCTTTATCAAACATTCTAAATACT	TGGAAGCCC	
			GAGCCTTGCTTGAGTTTGTGGAGCTTTATA	GGGATGGAA	
			GTATAATTTAGTTAAGCTTAAATTTAACCA	AGGAGACGT	

NZ_ANCM01000101.1	NZ_ANCM01000101.1	*Streptococcus agalactiae* FSL S3-586	GAAAAAGGCGATGTAGCTTAGAAAGGAGAA	GGGATGGAA	None
NZ_ANCM01000006.1			GAAAAAGGCGATGTAGCTTAGAAAGGAGAA	CACCATGAA	
NZ_ANCM01000028.1			TACGAAAAGGTTGTGATAAAAGCCATATCA	TCGAGTTTG	
NZ_ALTM01000012.1	NZ_ALTM01000016.1	*Streptococcus agalactiae* GB00548	AACAACTTTCTTACAAAAGGTTCTAGTTTTCTT	TCGCAAAAC	
NZ_ALTM01000013.1			ACGCTCTGAGGCAGATGAGGAACAGGCGCA	TAGGCACCC	
NZ_ALUZ01000056.1	NZ_ALUZ01000054.1	*Streptococcus agalactiae* GB00984	TGAAAACAAGCGCAAAGCTGTCAGAAAACA	CGGAACTAA	None
			TACTTGACGAATTGAAGATGACGGAATTTA	TGGCTCCAC	
NZ_ALRF01000019.1	NZ_ALRF01000066.1	*Streptococcus agalactiae* BSU188	GAAACTTCGATTAGTTTGCGTACTCGCTCA	CGGCAAAAC	None
NZ_ANEM01000019.1	NZ_ANEM01000012.1	*Streptococcus agalactiae* MRI Z1-022	TTGCTGCTAGACCCAAACAGTTTATTTTTAG	GGCCAAAAA	None
NZ_ANEM01000074.1			TATTTCATCATAGAAAATCCTGCTAGTGGT	CGGTTATGG	
NZ_CQEL01000006.1	NZ_CQEL01000002.1	*Streptococcus agalactiae* strain DK-NI-014	ACACCTAGTTTCAAGTTTTTAGCAGATTTTTT	GGTTACATT	None
NZ_CQEL01000008.1			ACGCTCTGAGGCAGATGAGGAACAGGCGCA	TAGGCACCC	
NZ_MAWX01000026.1	NZ_MAWX01000055.1		ATTGACTGTTTACGATTTCCTTCCACCGTT	GGGTACAAA	None

		Streptococcus agalactiae strain DK-PW-096	TGATGAGATTTTTAAAAGACTCACTGATAT	AGGATTGAC	
			CGCTTAGATGAAGTACAGATTGTAACAAGT	TCGGAAGTA	
NZ_CTJD01000013.1	NZ_CTJD01000001.1	*Streptococcus agalactiae* strain GB-NI-015	TGAAAACAAGCGCAAAGCTGTCAGAAAACA	CGGAACTAA	None
			TACTTGACGAATTGAAGATGACGGAATTTA	TGGCTCCAC	
NZ_CPZS01000003.1	NZ_CPZS01000001.1	*Streptococcus agalactiae* strain IT-NI-009	TATTTGATAGCGGTAACGGGTCATATACAA	AGGCATCTA	None
			ACCTTGCTCCGATGACACCATCGCGAACCT	TGGTCTAAT	
NZ_CPVQ01000026.1	NZ_CPVQ01000002.1	*Streptococcus agalactiae* strain RBH12	AACACAGCTTCCTCGAAAGGGATATATCTA	CGGACAACT	None
NDGB01000049.1	NDGB01000023.1	*Streptococcus agalactiae* strain ST 618	ATTAAGTTGCTTAGTGCTTTCATAATCATC	TGGAATAAC	None
NDGB01000030.1			ATTAAGTTGCTTAGTGCTTTCATAATCATC	TGGAATAAC	
NZ_KQ969340.1	NZ_KQ969342.1	*Streptococcus oralis* strain DD14	TTCCATTTCTGATTTGATTCAACAGCAGCA	GGAAATCCT	None
			TACAGCGGATACAACCCCACCAATAGCCTC	AGGAATTGC	
NZ_KQ961462.1	NZ_KQ961485.1	*Streptococcus pasteurianus* strain GED7275A	TTTATTCGGCATCGGCTGGTGTTATGGACT	TGGCTGCGG	None

NZ_AWTL01000007.1	NZ_AWTL01000011.1	*Streptococcus pyogenes* GA03805	TAGAGTAAACCGAATCTTTGCCATCTCTGG	CAGTTTGAC	None
NZ_LRGN01000012.1	NZ_LRGN01000001.1	*Streptococcus pyogenes* strain SST2091-1	TAGAGTAAACCGAATCTTTGCCATCTCTGG	CAGTTTGAC	None
LRGT01000330.1	LRGT01000062.1	*Streptococcus pyogenes* strain SST2097-1	TGGTCTAACTGCGTCTGGTCTGTGAATGA	TAGGTACAA	None
NC_021838.1	NC_021838.1	*Listeria monocytogenes* strain R2-502	GGTAAAACAAGCATCGGCGAAGCAGTAACA	TGGCTTCTT	AcrIIA3 (WP_023553812.1), AcrIIA2 (WP_023553814.1), AcrIIA1 (WP_003722518.1), AcrIIA1 (WP_012581438.1)
			GGTAAAACAAGCATCGGCGAAGCAGTAACA	TGGCTACTC	
			TAGGTTTAGGGAGTAAATTAGCTCCTTTGG	CAGCTGGGT	
			TAACTTTAGATACTGCTAAAGAATTAGCAA	TGGTGCAAA	
			TTGGGCAAAATGACCGTAATAAATCCATTC	CGGTTCATC	
			TAGGTTTAGGGAGTAAATTAGCTCCTTTGG	CGGCTGGAT	

Table S11. Top Acr gene candidates within each genome ranked by AcRanker. The proteins found within the prophages of 20 *Streptococcus* genomes were ranked using AcRanker; up to the top 10 highest ranking genes are listed in ascending order. Known Acr genes and the 10 genes synthesized for biochemical testing are indicated in the rightmost column. Genomes with fewer than 10 listed have very few annotated proteins found within predicted prophages.

Organism	**Source Contig**	**Protein**	**Rank**	**Candidate # or Acr**
Streptococcus iniae strain UEL-Si1	NZ_MNAC01000021.1	WP_071127623.1	1	ML1
	NZ_MNAC01000023.1	WP_071127667.1	2	
		WP_071127683.1	3	
		WP_071127693.1	4	
	NZ_MNAC01000021.1	WP_071127625.1	5	AcrIIA3
		WP_071127624.1	6	
	NZ_MNAC01000023.1	WP_071127689.1	7	
	NZ_MNAC01000021.1	WP_071127610.1	8	
	NZ_MNAC01000023.1	WP_071127674.1	9	
	NZ_MNAC01000021.1	WP_071127619.1	10	
Streptococcus pyogenes strain SS1447	NZ_JRLL01000026.1	WP_032460883.1	1	
		WP_029713970.1	2	
		WP_003057301.1	3	
	NZ_JRLL01000072.1	WP_032461152.1	4	
		WP_076634198.1	5	
	NZ_JRLL01000026.1	WP_032460878.1	6	
	NZ_JRLL01000072.1	WP_002986828.1	7	
	NZ_JRLL01000026.1	WP_080286986.1	8	
		WP_012678849.1	9	
		WP_032460877.1	10	
Streptococcus agalactiae strain DE-NI-032	NZ_CQAV01000003.1	WP_000640620.1	1	
		WP_000164461.1	2	
		WP_025194532.1	3	
		WP_017827941.1	4	
		WP_050201842.1	5	
		WP_050305756.1	6	
		WP_001162136.1	7	

		WP_000431575.1	8	
		WP_000138374.1	9	
		WP_001872365.1	10	
Streptococcus pyogenes strain NGAS322	NZ_CP010449.1	WP_002983328.1	1	
		WP_080370149.1	2	
		WP_002983750.1	3	
		WP_002984315.1	4	
		WP_032465789.1	5	
		WP_002982773.1	6	
		WP_011054546.1	7	
		WP_010921912.1	8	
		WP_080370134.1	9	
		WP_053308468.1	10	
Streptococcus agalactiae strain CCUG 37430	NZ_ALQN01000018.1	WP_000649300.1	1	
		WP_079261174.1	2	
		WP_000660740.1	3	
		WP_000076700.1	4	
		WP_000033707.1	5	
		WP_000343312.1	6	
		WP_000130090.1	7	
		WP_000582684.1	8	
		WP_000431581.1	9	
		WP_000323860.1	10	
Streptococcus agalactiae strain DE-NI-007	NZ_CPVL01000019.1	WP_000694571.1	1	
		WP_001166092.1	2	
		WP_000359663.1	3	
		WP_000141918.1	4	
		WP_000648623.1	5	
		WP_079260963.1	6	
		WP_000205000.1	7	
		WP_000130289.1	8	
		WP_000946250.1	9	ML4
		WP_001021397.1	10	
		WP_001080841.1	12	ML5
Streptococcus agalactiae FSL S3-586	NZ_ANCM01000028.1	WP_017643458.1	1	
		WP_000134940.1	2	

	NZ_ANCM01000006.1	WP_001875290.1	3	
	NZ_ANCM01000028.1	WP_000789102.1	4	
		WP_003051787.1	5	
		WP_000032136.1	6	
		WP_000342242.1	7	
		WP_000686776.1	8	
		WP_000988928.1	9	
	NZ_ANCM01000101.1	WP_017643459.1	10	
Streptococcus agalactiae strain GB00548	NZ_ALTM01000002.1	WP_000331953.1	1	
		WP_000259017.1	2	
		WP_000793595.1	3	
		WP_079254676.1	4	
		WP_000384271.1	5	ML8
		WP_001018249.1	6	
		WP_000568029.1	7	
		WP_001097380.1	8	
		WP_001867157.1	9	
		WP_000656477.1	10	
		WP_000134666.1	12	ML9
		WP_000591144.1	29	ML7
Streptococcus agalactiae strain GB00984	NZ_ALUZ01000056.1	WP_000660738.1	1	
		WP_000164461.1	2	
		WP_017827941.1	3	
		WP_000965653.1	4	
		WP_000431574.1	5	
		WP_000138374.1	6	
		WP_000614971.1	7	
		WP_000258802.1	8	
		WP_000763911.1	9	
		WP_000118546.1	10	
Streptococcus agalactiae strain BSU188	NZ_ALRF01000068.1	WP_001042289.1	1	
		WP_000965633.1	2	ML6
		WP_025194532.1	3	
		WP_000660741.1	4	
		WP_001162136.1	5	
		WP_000274022.1	6	

		WP_000076712.1	7	
		WP_001183891.1	8	
		WP_000431576.1	9	
		WP_000763914.1	10	
Streptococcus agalactiae strain MRI Z1-022	NZ_ANEM01000074.1	WP_017648179.1	1	
		WP_079265830.1	2	
		WP_000033707.1	3	
		WP_000582684.1	4	
		WP_017648175.1	5	
		WP_017648177.1	6	
		WP_000802599.1	7	
		WP_000343901.1	8	
		WP_025195242.1	9	
		WP_000142566.1	10	
Streptococcus agalactiae strain DK-NI-014	NZ_CQEL01000002.1	WP_000421991.1	1	
		WP_000640620.1	2	
		WP_011058321.1	3	ML10
		WP_000965642.1	4	
		WP_000660741.1	5	
		WP_000906736.1	6	
		WP_001162136.1	7	
		WP_000076715.1	8	
		WP_000027835.1	9	
		WP_001872365.1	10	
Streptococcus agalactiae strain DK-PW-096	NZ_MAWX01000026.1	WP_000258802.1	1	
		WP_001229661.1	2	
		WP_001921522.1	3	
		WP_000774601.1	4	
		WP_011324937.1	5	
		WP_000218309.1	6	
		WP_079261306.1	7	
		WP_000411527.1	8	
		WP_001270064.1	9	
		WP_000659174.1	10	
Streptococcus agalactiae strain GB-NI-015	NZ_CTJD01000013.1	WP_000640620.1	1	
		WP_000660738.1	2	

		WP_000164461.1	3	
		WP_017827941.1	4	
		WP_000965655.1	5	
		WP_000431574.1	6	
		WP_000138374.1	7	
		WP_001872365.1	8	
		WP_000614971.1	9	
		WP_000258802.1	10	
Streptococcus agalactiae strain IT-NI-009	NZ_CPZS01000003.1	WP_000640620.1	1	
		WP_000164461.1	2	
		WP_079261174.1	3	
		WP_050201842.1	4	
		WP_001162136.1	5	
		WP_000431575.1	6	
		WP_000138374.1	7	
		WP_001872365.1	8	
		WP_000474006.1	9	
		WP_000258802.1	10	
Streptococcus agalactiae strain RBH12	NZ_CPVQ01000026.1	WP_000650503.1	1	
		WP_000164461.1	2	
		WP_079261174.1	3	
		WP_050198474.1	4	
		WP_001058281.1	5	
		WP_079454162.1	6	
		WP_050199334.1	7	
		WP_000612386.1	8	
		WP_000963485.1	9	
		WP_000206191.1	10	
Streptococcus agalactiae strain ST 618	NDGB01000030.1	OTG45472.1	1	
		OTG45475.1	2	
		OTG45496.1	3	
		OTG45484.1	4	
		OTG45499.1	5	
		OTG45477.1	6	
		OTG45479.1	7	
		OTG45483.1	8	

		OTG45481.1	9	
		OTG45480.1	10	
Streptococcus oralis strain DD14	NZ_KQ969340.1	WP_061420077.1	1	
		WP_061420097.1	2	
		WP_061420111.1	3	
		WP_061420115.1	4	
		WP_061420334.1	5	
		WP_061420080.1	6	
		WP_061420123.1	7	
		WP_061420133.1	8	
		WP_061420073.1	9	
		WP_061420062.1	10	
Streptococcus pasteurianus strain GED7275A	NZ_KQ961462.1	WP_061100257.1	1	
		WP_061100237.1	2	
		WP_061100224.1	3	
		WP_061100243.1	4	
		WP_061100244.1	5	
		WP_061100249.1	6	
		WP_082731474.1	7	
		WP_061100238.1	8	
		WP_061100250.1	9	
		WP_061100233.1	10	
Streptococcus pyogenes GA03805	NZ_AWTL01000007.1	WP_011528797.1	1	
		WP_011888786.1	2	
		WP_023079933.1	3	
		WP_023079900.1	4	
		WP_023079918.1	5	
		WP_002985387.1	6	
		WP_011528776.1	7	
		WP_023079897.1	8	
		WP_011017565.1	9	
		WP_023079923.1	10	
Streptococcus pyogenes strain SST2091-1	NZ_LRGN01000012.1	WP_011889039.1	1	
		WP_011285632.1	2	
		WP_010922455.1	3	
		WP_010922464.1	4	

		WP_002994106.1	5	
		WP_002994744.1	6	
		WP_063629031.1	7	
		WP_080464960.1	8	
		WP_063629030.1	9	
		WP_063629029.1	10	
Streptococcus pyogenes strain SST2097-1	LRGT01000330.1	OAC70929.1	1	
		OAC70939.1	2	
		OAC70918.1	3	
		OAC70933.1	4	
		OAC70928.1	5	
		OAC70915.1	6	
		OAC70921.1	7	
		OAC70941.1	8	
		OAC70938.1	9	
		OAC70937.1	10	
Listeria monocytogenes strain R2-502	NC_021838.1	WP_003731672.1	1	
		WP_003733710.1	2	
		WP_003731277.1	3	ML2
		WP_003731276.1	4	ML3
		WP_023553812.1	5	AcrIIA3
		WP_014601509.1	6	
		WP_003733721.1	7	
		WP_003731655.1	8	
		WP_003725074.1	9	
		WP_014601388.1	10	
		WP_023553814.1	34	AcrIIA2
		WP_003722518.1	71	AcrIIA1
		WP_012581438.1	95	AcrIIA1

Table S12. BLAST vs. AcRanker rankings for the selection candidates ML1-ML10. After selecting the 10 candidate proteins for biochemical investigation, we performed a blastp ranking to determine the ability of BLAST to predict new Acr proteins. The three validated anti-CRISPRs are indicated with tan shading and in all three cases, AcRanker gives a much higher ranking than BLAST.

Candidate	Prophage proteome size	Blastp rank (e-value)	AcRanker rank
ML1 (AcrIIA20)	56	12 (0.38)	1
ML2	190	155 (4.85)	1
ML3 (AcrIIA12)	190	132 (2.48)	2
ML4	26	4 (0.16)	9
ML5	26	16 (0.7)	12
ML6	75	37 (1.3)	2
ML7	32	29 (5.84)	29
ML8 (AcrIIA21)	32	27 (3.24)	5
ML9	11	11 (0.5)	12
ML10	74	74 (4.5)	3

References

1. Lundberg,S.M. and Lee,S.-I. (2017) A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems.*
2. Shen,J., Zhang,J., Luo,X., Zhu,W., Yu,K., Chen,K., Li,Y. and Jiang,H. (2007) Predicting protein-protein interactions based only on sequences information. *Proc. Natl. Acad. Sci. USA*, **104**, 4337–4341.
3. Pawluk,A., Bondy-Denomy,J., Cheung,V.H.W., Maxwell,K.L. and Davidson,A.R. (2014) A New Group of Phage Anti-CRISPR Genes Inhibits the Type I-E CRISPR-Cas System of Pseudomonas aeruginosa. *mBio*, **5**, e00896-14.
4. Bondy-Denomy,J., Pawluk,A., Maxwell,K.L. and Davidson,A.R. (2013) Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*, **493**, 429–432.
5. Pawluk,A., Staals,R.H.J., Taylor,C., Watson,B.N.J., Saha,S., Fineran,P.C., Maxwell,K.L. and Davidson,A.R. (2016) Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nat. Microbiol.*, **1**, 16085.
6. Rauch,B.J., Silvis,M.R., Hultquist,J.F., Waters,C.S., McGregor,M.J., Krogan,N.J. and Bondy-Denomy,J. (2017) Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. *Cell*, **168**, 150–158.
7. Hynes,A.P., Rousseau,G.M., Lemay,M.L., Horvath,P., Romero,D.A., Fremaux,C. and Moineau,S. (2017) An anti-CRISPR from a virulent streptococcal phage inhibits Streptococcus pyogenes Cas9. *Nat. Microbiol.*, **2**, 1374–1380.
8. Marino,N.D., Zhang,J.Y., Borges,A.L., Sousa,A.A., Leon,L.M., Rauch,B.J., Walton,R.T., Berry,J.D., Joung,J.K., Kleinstiver,B.P., *et al.* (2018) Discovery of widespread type I and type V CRISPR-Cas inhibitors. *Science*, **362**, 240–242.
9. Hynes,A.P., Rousseau,G.M., Agudelo,D., Goulet,A., Amigues,B., Loehr,J., Romero,D.A., Fremaux,C., Horvath,P., Doyon,Y., *et al.* (2018) Widespread anti-CRISPR proteins in virulent bacteriophages inhibit a range of Cas9 proteins. *Nat. Commun.*, **9**, 2919.
10. Uribe,R. V, van der Helm,E., Misiakou,M.A., Lee,S.W., Kol,S. and Sommer,M.O.A. (2019) Discovery and Characterization of Cas9 Inhibitors Disseminated across Seven Bacterial Phyla. *Cell Host and Microbe*, **25**, 233–241.
11. Watters,K.E., Shivram,H., Fellmann,C., Lew,R.J., McMahon,B. and Doudna,J.A. (2020) Potent CRISPR-Cas9 inhibitors from Staphylococcus genomes. *Proc. Natl. Acad. Sci. USA*, **117**, 1–9.
12. Lee,J., Mir,A., Edraki,A., Garcia,B., Amrani,N., Lou,H.E., Gainetdinov,I., Pawluk,A., Ibraheim,R., Gao,X.D., *et al.* (2018) Potent Cas9 inhibition in bacterial and human cells by AcrIIC4 and AcrIIC5 anti-CRISPR proteins. *mBio*, **9**, 1–17.
13. Bhoobalan-Chitty,Y., Johansen,T.B., Di Cianni,N. and Peng,X. (2019) Inhibition of Type III CRISPR-Cas Immunity by an Archaeal Virus-Encoded Anti-CRISPR Protein. *Cell*, **179**, 448-458.e11.
14. Kyle E. Watters, Christof Fellmann, Hua B. Bai, Shawn M. Ren and Jennifer A. Doudna (2018) Systematic discovery of natural CRISPR-Cas12a inhibitors. *Science*, **362**, 236–239.