



(51) International Patent Classification:

C07K 14/33 (2006.01)

C12N 15/90 (2006.01)

C12N 15/10 (2006.01)

(21) International Application Number:

PCT/EP2019/057786

(22) International Filing Date:

27 March 2019 (27.03.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

18164238.0

27 March 2018 (27.03.2018)

EP

(71) Applicant: DANMARKS TEKNISKE UNIVERSITET [DK/DK]; Anker Engelunds Vej 101 A, 2800 Kgs. Lyngby (DK).

(72) Inventors: VAZQUEZ-URIBE, Ruben; Arnestedet 5, 2720 Vanløse (DK). VAN DER HELM, Eric; Søndre Fasanvej 66, 2. tv., 2000 Frederiksberg (DK). SOMMER, Morten Otto Alexander; Duntzfelts Allé 22, 2900 Hellerup (DK). MISIAKOU, Maria-Anna; c/o Nuria Pelitero, Fuglevangsvej 7, 3. th., 1962 Frederiksberg C (DK). LEE, Sang Woo; Buddingevej 281, 2860 Soborg (DK).

(74) Agent: GUARDIAN IP CONSULTING I/S; Diplomvej, Building 381, 2800 Kgs. Lyngby (DK).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM,

AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

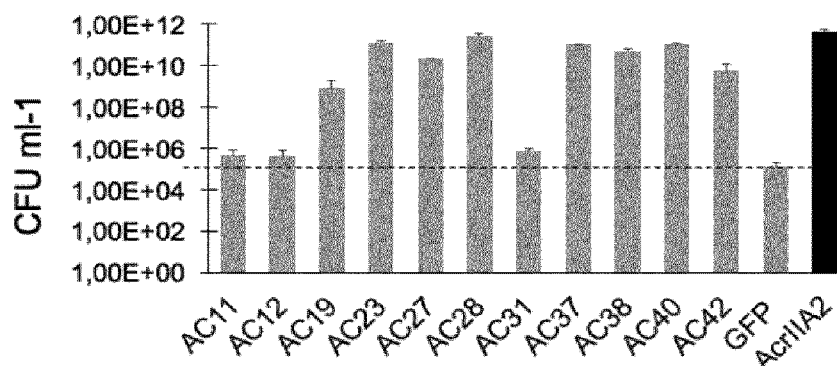
Published:

— with international search report (Art. 21(3))

— with sequence listing part of description (Rule 5.2(a))

(54) Title: INHIBITORS OF CRISPR-CAS ASSOCIATED ACTIVITY

Figure 6



(57) Abstract: The invention relates to Acr proteins capable of regulating CRISPR-Cas associated activity, and their use in regulating CRISPR-Cas associated activity within a biological cell. The invention further provides nucleic acid molecules and self-replications genetic elements adapted for expression of said Acr proteins, or for their co-expression with a cognate CRISPR-Cas system in a biological cell.

TITLE: Inhibitors of CRISPR-Cas associated activity**Technical field of the invention**

The invention relates to Acr proteins capable of regulating CRISPR-Cas associated activity, and their use in regulating CRISPR-Cas associated activity within a biological cell. The invention further provides nucleic acid molecules and self-replications genetic elements adapted for expression of said Acr proteins, or for their co-expression with a cognate CRISPR-Cas system in a biological cell.

Background of the invention

CRISPR-Cas systems originate from bacteria where they function as adaptive immune systems that cleave foreign DNA or RNA in a sequence specific manner. Components of type II CRISPR-cas systems, such as Cas9 from *Streptococcus pyogenes*, have been exploited for precise and programmable gene editing with a substantial impact on life sciences in spite of concerns about off target activity. Given the significant impact of CRISPR-Cas systems, substantial efforts have been made to discover new systems in cultured and recently uncultured bacterial species. CRISPR-Cas technologies provide effective biotechnological tools that broadly impact life sciences. Applications range from treating human genetic diseases to eliminating pathogens using gene-drives. Beyond genome editing, CRISPR-Cas systems have several other applications, including programmable transcriptional regulators, and ultra-sensitive detection devices. However, there is increasing concern about the safety and precision of CRISPR-Cas. Indeed, the 2016 worldwide threat assessment of U.S. intelligence listed genome editing technologies as a potential weapon of mass destruction and proliferation. Accordingly, more tools are needed to control the CRISPR-Cas technology in order to reduce potential risks of genome editing, and to contain their spread and potential harm to other organisms.

A promising method to control CRISPR-Cas activity is via anti-CRISPR (ACR) proteins. These naturally occurring inhibitors can potentially inactivate CRISPR-Cas systems when needed, or decrease off-target effects, without significantly affecting on-target efficiency. ACRs were initially discovered for type I CRISPR systems. Subsequent bioinformatic mining identified ACR activity against type II-C Cas9 from *Neisseria meningitidis* and type II-A Cas9 from *Streptococcus pyogenes* (spCas9). More recently an additional ACR against spCas9 was identified by cloning and testing multiple genes from a phage that was able to escape CRISPR-based immunity from *Streptococcus thermophilus*. Given the abundance of CRISPR-Cas systems in bacteria as well as the abundance of uncharacterized phages, it is likely that we currently have only revealed

a minute proportion of ACRs in the environment. Accordingly, there exists a need for for identifying further ACRs, that can be used to regulate CRISPR-cas in its diverse applications.

5 **Summary of the invention**

A first embodiment of the invention provides a nucleic acid molecule encoding an anti-CRISPR protein selected from the group consisting of an:

- i. AcrIIA8 protein,
- ii. AcrIIA6 protein,
- 10 iii. AcrIIA9 protein and
- iv. AcrIIA7 protein.

In respect to the first embodiment, the invention further provides a self-replicating genetic element comprising the nucleic acid molecule encoding said anti-CRISPR protein, wherein:

- a. the amino acid sequence of said AcrIIA8 protein has at least 80% sequence identify to an amino acid sequence selected from the group consisting of SEQ ID Nos.: 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, amd 219;
- 20 b. the amino acid sequence of said AcrIIA6 protein has at least 80% sequence identify to an amino acid sequence selected from the group consisting of SEQ ID Nos.: 221, 223 and 225;
- c. the amino acid sequence of said AcrIIA9 protein has at least 80% sequence identify to an amino acid sequence selected from the group consisting of SEQ ID Nos.: 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, and 247;
- 25 d. the amino acid sequence of said AcrIIA7 protein has at least 80% sequence identify to an amino acid sequence selected from the group consisting of SEQ ID Nos.: 249, 251, 253, 255 and 257; and

wherein said nucleic acid molecule is operably linked to a promoter.

30

In respect to the first embodiment, the invention further provides a biological cell comprising either said nucleic acid molecule encoding an anti-CRISPR protein, or comprising said self-replicating genetic element comprising the nucleic acid molecule operably linked to a promoter and where the nucleic acid molecule encodes said anti-CRISPR protein.

35

Furthermore, said first embodiment may further include a CRISPR-Cas expression system comprising: a first gene encoding a polypeptide or a first gene cluster encoding a polypeptide complex, said polypeptide or polypeptide complex having RNA-guided

endonuclease activity, wherein said gene or gene cluster is operably linked to a promoter; and a nucleic acid molecule operably linked to a promoter, said nucleic acid molecule encoding one or more RNA molecules,
wherein said CRISPR-Cas expression system is comprised on said self-replicating
5 genetic element or on a further self-replicating genetic element, wherein said one of more RNA molecules are capable of guiding said polypeptide or polypeptide complex having RNA-guided endonuclease activity to a target gene in a cell; and wherein expression of said anti-CRISPR protein is capable of regulating said CRISPR-Cas system when co-expressed in the cell.

10

A second embodiment of the invention provides an isolated anti-CRISPR protein selected from the group consisting of:

- i. AcrIIA8 protein,
- ii. AcrIIA6 protein,
- 15 iii. AcrIIA9 protein and
- iv. AcrIIA7 protein.

20

A third embodiment of the invention provides for the use of the nucleic acid molecule of the first embodiment, or an anti-CRISPR protein of the second embodiment to regulate CRISPR-Cas associated activity. This use may be applied to regulating CRISPR-Cas associated activity expressed in a biological cell selected from the group consisting of a bacterial, yeast, fungus, insect, and plant cell. Alternatively, this use may be applied to a mammalian cell (e.g. somatic cell).

25

A fourth embodiment of the invention provides a method for *in vivo* regulation of a CRISPR-Cas system, comprising

- a. providing an isolated biological cell comprising a target gene;
- b. contacting said cell with a molecule, wherein said molecule is selected from the group consisting of:
30
 - i. an anti-CRISPR protein of the invention,
 - ii. a nucleic acid molecule encoding and capable of expressing an anti-CRISPR protein of the invention, and
 - iii. a self-replicating genetic element of the invention comprising said nucleic acid molecule encoding and capable of expressing said anti-
35 CRISPR protein,

wherein said cell either comprises or is additionally contacted with a CRISPR-Cas expression system comprising:

- iv. a first gene encoding a polypeptide or a first gene cluster encoding a polypeptide complex, said polypeptide or polypeptide complex having RNA-guided endonuclease activity, wherein said gene or gene cluster is operably linked to a promoter; and
- 5 v. a nucleic acid molecule operably linked to a promoter, said nucleic acid molecule encoding one or more RNA molecules capable of guiding said polypeptide or polypeptide complex having RNA-guided endonuclease activity to said target gene;
- 10 wherein expression of said anti-CRISPR protein in said cell is capable of regulating the activity of the CRISPR-Cas system expressed by the Cas expression system. This method may be applied to an isolated biological cell selected from the group consisting of a bacterial, yeast, fungus, insect, and plant cell or alternatively to a mammalian cell (e.g. somatic cell).
- 15 A fifth embodiment of the invention provides a composition comprising an anti-CRISPR protein of the invention, or a nucleic acid molecule encoding and capable of expressing an anti-CRISPR protein of the invention, for use as a medicament. Said composition is also for use in regulating CRISPR-Cas associated activity in a somatic mammalian cell.
- 20 A sixth embodiment of the invention provides a pharmaceutical kit comprising said aforementioned composition in combination with a CRISPR-Cas expression system for use as a medicament, wherein said expression system comprises: a first gene encoding a polypeptide or a first gene cluster encoding a polypeptide complex, said polypeptide or polypeptide complex having RNA-guided endonuclease activity, wherein
- 25 said gene or gene cluster is operably linked to a promoter; and a nucleic acid molecule operably linked to a promoter, said nucleic acid molecule encoding one or more RNA molecules capable of guiding said polypeptide or polypeptide complex having RNA-guided endonuclease activity to said target gene; and wherein said anti-CRISPR protein is for use in regulating CRISPR-Cas associated activity in a somatic mammalian cell.
- 30 cell.

Description of the invention

Figures:

- 35 **Figure 1** is a cartoon illustrating a host cell comprising a genetic circuit and a member of a self-replicating library of non-host DNA, used in the detection of inhibitors of CRISPR-Cas systems. The host cell comprises a gene or gene cluster encoding a protein or protein complex having RNA-guided endonuclease activity, a nucleic acid molecule encoding chimeric gRNA (sgRNA) or a CRISPR array, as well as a reporter gene encoding a selection marker. Cells of the host are transformed with a self-

replicating library of non-host DNA. Expression of polypeptides encoded by the non-host DNA that inhibit inactivation of the reporter gene or its transcript by the RNA-guided endonuclease are detected by means of the selection marker.

Figure 2 shows a cartoon (A) illustrating the pCasens3 and pDual3 plasmids and their interaction with an anti-CRISPR protein encoded by a positive control plasmid. Plasmid pCasens3 (middle) comprises: the *Streptococcus pyogenes* Cas9 gene (spCas9) [SEQ ID No.:150] cloned into the backbone of the plasmid pSEVA47. Additionally, a sigma70 constitutive promoter (P₂₃₁₀₀) [SEQ ID No.:148], adjacent to a DNA molecule encoding a theophylline translational riboswitch [SEQ ID No.:149] were inserted upstream of the spCas9 gene in pSEVA47, in order to control its expression. pSEVA47 comprises a low copy number origin of replication, SC101 [SEQ ID No.:155]; and the antibiotic resistance gene aadA [SEQ ID No.:153] under the control of its native (constitutive) promoter and ribosomal binding site, conferring resistance against Spectinomycin.

Plasmid pDual3 (to the right) comprises a chloramphenicol resistance gene (CmR) [SEQ ID No.:164] operably linked to an upstream constitutive promoter and RBS; and a DNA sequence encoding a sgRNA [SEQ ID No.:161] whose transcription/expression is regulated by the inducible pBAD promoter [SEQ ID No.: 158], located upstream of the respective sequences. The plasmid further comprises an araC L-arabinose sensor gene [SEQ ID No.:159] that regulates the pBAD promoter in response to the metabolite, L-arabinose. pDual3 comprises a low copy number origin of replication, p15A [SEQ ID No.:163]. A positive control plasmid comprising a gene [SEQ ID No.:146] encoding ACRIIA2 protein under the control of a constitutive pTET promoter [SEQ ID No.:139].

Cartoon in figure 2B illustrates the negative control plasmid comprising a gene [SEQ ID No.:140] encoding Green Fluorescent Protein (GFP) under the control of a constitutive pTET promoter [SEQ ID No.:139]. Both negative and positive control plasmids of Figure 2 further comprise a kanamycin resistance gene [SEQ ID No.:144] under the control of its native (constitutive) promoter and ribosomal binding site and a colE1 origin of replication [SEQ ID No.:143].

Figure 3 shows (A) a cartoon illustrating the molecular mechanism whereby in *E. coli* cells comprising the Genetic Circuit (*E. coli*-GC), the genetic circuit is induced to express the cas9 gene and sgRNA, with the formation of a Cas9-sgRNA complex, and the endonuclease-mediated cleavage of the reporter gene (Cat), leading to cell death when the cells are exposed to chloramphenicol. The cartoon further shows that co-expression of an acrIIA2 gene in the *E. coli*-GC cells can inhibit Cas9-sgRNA inactivation of the Cat gene, leading to cell survival under selective conditions. (B) is a histogram showing the number of colony forming units (CFU.ml⁻¹) of *E. coli*-GC cells comprising a negative control plasmid (expressing GFP) or the positive control plasmid

(expressing ACRIIA2) following cultivation on selective media, where the genetic circuit is either induced or non-induced.

Figure 4 is a cartoon showing the use of host cells comprising the genetic circuit to screen a metagenomics DNA expression library for anti-CRISPR proteins.

5 **Figure 5** is a cartoon illustrating a metagenomics DNA expression plasmid comprising a fragment of (a library of) metagenomics DNA fragments cloned down-stream of a constitutive pTET promoter [SEQ ID No.:139]; a kanamycin resistance gene under the control of its native (constitutive) promoter and ribosomal binding site; together with the genetic circuit comprising plasmids pCasens3 and pDual3.

10 **Figure 6** is a histogram showing the number of colony forming units (CFU.ml⁻¹) of *E. coli*-GC cells comprising selected metagenomics inserts, designated AC11 – AC42; as compared to a negative control plasmid (expressing GFP) and the positive control plasmid (expressing ACRIIA2), following cultivation on selective media, where the genetic circuit is induced.

15 **Figure 7** (A) shows a 1% agarose gel comprising the size-separated cleavage products resulting from an *in vitro* cleavage of linear double-stranded DNA template by the spCas9-sgRNA complex in the presence or absence of putative inhibitors of spCas9 activity. Said DNA template comprises the same 20 base pair target sequence recognised spCas9-sgRNA complex as used in the *in vivo* assay of figure 6. (B) is a graphical presentation of the binding strength of each putative inhibitor of spCas9 (ACR) to biotinylated spCas9 showed as K_d determined using biolayer interferometry.

20 **Figure 8** Table showing the sequence identities between the isolated ACRIIA 1-9 polypeptides in matrix format.

25 **Figure 9** Diagram showing the computed hierarchical clustering of members of the ACRIIA1 - 9 protein families.

30 **Figure 10** (A) Diagram showing the distribution of AcrIIA genes across genomic and metagenomic datasets, where hosts of all identified AcrIIAs are mapped to the tree of life (Hug, LA et al., 2016). Tree branches correspond to all distinct bacterial phyla, (NCBI Taxonomy) carrying AcrIIA's, where previously identified AcrIIAs are given in italics. (B) Quantification of AcrIIA genes and their homologues in publically available viromes (MetaVir). The x axis corresponds to the number of raw reads with homology to an AcrIIA and the y axis to the distinct virome datasets stacked by habitat and grouped by AcrIIA. Bars are assigned shades of gray according to the type of habitat the datasets represent; showing localization of AcrIIA9 in respectively human, mouse, wastewater and freshwater samples; and AcrIIA7 in respectively human, insect, food, coral, wastewater, freshwater, seawater, hypersaline, and soil samples.

35 **Figure 11** Genomic context of AcrIIA8 homologues. The AcrIIA8-surrounding genomic neighbourhood across all relevant reference genomes is shown mapped to the AcrIIA8 protein tree. Homologous genomic regions were identified by MultiGeneBlast

(Medema, M. H., et al., 2013) using the genome carrying the closest homologue to AcrIIA8 as a search query, here displayed at the top. Homology to the queried genes is denoted by shapes having a common shading/hatching; shaded/hatched shapes that are not part of the legend correspond to uncharacterized proteins; White shapes correspond to genes below the homology threshold.

Figure 12 Alignment of the amino acid sequence of members of the AcrIIA8 anti-CRISPR protein family. The aligned proteins are all approximately 90-110 amino acids in length; where ArcIIA8 (encoded by ac27-1 gene) shares a sequence identity of 81% to WO_009270720.1 derived from *Erysipelotrichaceae*; 48% to WP_009270720.1; and between 29% and 40% to the other members of this protein family.

Figure 13 Alignment of the amino acid sequence of members of the AcrIIA6 anti-CRISPR protein family. The aligned proteins are all approximately 109 amino acids in length; wherein ArcIIA6 shares a sequence identity of 99% to WP_058324878.1 derived from *Sinorhizobium* sp. GL28 as well as to KGM08958.1 derived from *Cellulomonas carbonis* T26, based on the aligned regions.

Figure 14 Alignment of the amino acid sequence of members of the AcrIIA9 anti-CRISPR protein family. The aligned proteins are all approximately 140-145 amino acids in length; wherein ArcIIA9 shares a sequence identity of at least 78% with the selected members of the family; ranging from a sequence identity of 76% to WP_009036714.1 derived from *Bacteroides* sp. D20, up to 93% to WP_004294380.1 derived from *Bacteroides fragilis*. The aligned amino acid sequences share a PcfK domain spanning their N-terminal portion.

Figure 15 Alignment of the amino acid sequence of members of the AcrIIA7 anti-CRISPR protein family. ArcIIA7 shares a sequence identity of at least 60% with the selected members of the family. The aligned amino acid sequences share a DUF2829 domain.

Figure 16 Cartoon showing design of T7E1 assay for Anti-CRISPR protein activity in a human cell line (left hand) and image of agarose gel showing DNA fragments following T7 endonuclease I (T7E1) cleavage of Cas9 target DNA amplified from said cell line.

The cartoon shows a human cell line (HEK293T) expressing Cas9, that together with guide RNA (crRNA) introduces a double-stranded break in a target locus on the genome, which is then repaired by a non-homologous end joining mechanism (NHEJ), leaving insertion or deletion mutations (indels). Genomic DNA is extracted from the cells; the target locus is then amplified; and the PCR products are then denatured and annealed to generate mismatched DNA hybrids around the target site. T7 endonuclease I cleaves the mismatched DNA hybrids resulting in cleaved products detectable on an agarose gel. The agarose gel (right hand image) shows that the proportion of target locus DNA in a human cell line (HEK293T) that is cleaved by CRISPR-Cas9-mediated double-stranded breaks and detected by the T7E1 assay, is

inhibited by co-expression of the anti-CRISPR proteins AcrIIA6, AcrIIA7, AcrIIA8, AcrIIA9, and AcrIIA2.

Abbreviations and definition of terms:

5 **Amino acid sequence identity:** The term "sequence identity" as used herein, indicates a quantitative measure of the degree of homology between two amino acid sequences of substantially equal length. The two sequences to be compared must be aligned to give a best possible fit, by means of the insertion of gaps or alternatively, truncation at the ends of the protein sequences. The sequence identity can be
10 calculated as $((N_{\text{ref}} - N_{\text{dif}})100)/(N_{\text{ref}})$, wherein N_{dif} is the total number of non-identical residues in the two sequences when aligned and wherein N_{ref} is the number of residues in one of the sequences. Sequence identity between each AcrIIA and its homologues is calculated by the BLAST program e.g. the BLASTP program (Pearson W.R and D.J. Lipman (1988)) (www.ncbi.nlm.nih.gov/cgi-bin/BLAST). The identity
15 matrix between AcrIIA1-9 is determined by the sequence alignment method ClustalW with default parameters as described by Thompson J., et al 1994, available at <http://www2.ebi.ac.uk/clustalw/>.

Preferably, the numbers of substitutions, insertions, additions or deletions of one or more amino acid residues in the polypeptide as compared to its comparator
20 polypeptide is limited, i.e. no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 substitutions, no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 insertions, no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 additions, and no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 deletions. Preferably the substitutions are conservative amino acid substitutions: limited to exchanges within members of group 1: Glycine, Alanine, Valine, Leucine, Isoleucine;
25 group 2: Serine, Cysteine, Selenocysteine, Threonine, Methionine; group 3: proline; group 4: Phenylalanine, Tyrosine, Tryptophan; Group 5: Aspartate, Glutamate, Asparagine, Glutamine.

Cis element: is a sequence of DNA that is capable of conferring inducible expression of a gene that is functionally linked to a promoter; where the cis element may induce
30 expression by regulating transcription of the gene or by regulating translation of transcript product of the gene.

Cognate: is used to refer to interacting pairs of functional entities; more specifically that each protein or protein cluster having RNA-guided endonuclease activity has a cognate mat-crRNA, and in the case of Type II CRISPR-Cas systems both a cognate
35 mat-crRNA and cognate tracrRNA, that are required for their activity.

CRISPR system: is a Clustered Regularly Interspaced Short Palindromic Repeats is bacterial immune system that cleave DNA or RNA in a sequence specific manner, mediated by a RNA guided nuclease.

***E. coli* Top10:** is *E. coli* having chromosomal Genotype mcrA, Δ (mrr-hsdRMS-mcrBC), Phi80lacZ(del)M15, Δ lacX74, deoR, recA1, araD139, Δ (ara-leu)7697, galU, galK, rpsL(SmR), endA1, nupG

gi number: (genInfo identifier) is a unique integer which identifies a particular sequence, independent of the database source, which is assigned by NCBI to all sequences processed into Entrez, including nucleotide sequences from DDBJ/EMBL/GenBank, protein sequences from SWISS-PROT, PIR and many others.

Nucleic acid molecule: according to the invention comprises a gene whose nucleic acid sequence encodes a AcrIIA protein of the invention.

Promoter: is a region of DNA that initiates transcription of a particular gene. Promoters are located near the transcription start sites of genes, on the same strand and upstream on the DNA (towards the 5' region of the sense strand). A promoter that is "functionally linked" to a gene is capable of driving expression of said gene. A promoter may be drive expression constitutively or only in response to an inducer.

RBS: Ribosomal Binding Site is a sequence of nucleotides upstream of the start codon of an mRNA transcript that is responsible for the recruitment of a ribosome during the initiation of protein translation.

Heterologous gene and heterologous DNA molecule: have a different genetic origin from the recombinant cell in which they are expressed; and this also applies to the protein or transcript that they encode. The nucleotide sequence of the heterologous gene or heterologous DNA molecule may be optimized (e.g. codon optimization) with respect to the recombinant cell in which they are expressed. Heterologous gene and heterologous DNA molecule may be located on (and therefor be a part of) the chromosome or a episome of the recombinant cell, and may be inserted into this location by recombinant DNA cloning.

Detailed description of the invention

The invention provides isolated nucleic acid molecules encoding anti-CRISPR proteins; their encoded anti-CRISPR proteins; and the use of the isolated nucleic acid molecules and respective anti-CRISPR proteins in methods for regulating CRISPR-Cas associated activity. The regulation conferred by these proteins may provide an enhancement, or more preferably an inhibition of CRISPR-Cas associated activity. The anti-CRISPR proteins of the invention have broad application, since they can be used to regulate more than one CRISPR-Cas system. Examples of systems having CRISPR-Cas associated activity, their component molecules, and their expression are detailed in section III.

The anti-CRISPR proteins and their encoding nucleic acid molecules were identified in soil or human gut fecal metagenomics libraries (Example 3, figures 4 and 5) by means of a high through-put screening assay (as described in Example 1 and 2, Figures 1-3). The identified anti-CRISPR proteins were classified as ArcIIA6, ArcIIA7 ArcIIA8 and ArcIIA9, based on their amino acid sequence encoded by their respective nucleic acid molecules. The nucleic acid molecules encoding anti-CRISPR proteins of the invention may comprise a promoter sequence and be capable of expression of said proteins *in vitro* or *in vivo*.

I Nucleic acid molecules encoding anti-CRISPR proteins

The anti-CRISPR proteins, encoded by the isolated nucleic acids molecules of the invention and their homologues, are members of 4 families of anti-CRISPR proteins, that are both distinct from each other and from all previously identified anti-CRISPR proteins. Sequence identity between the anti-CRISPR proteins, ArcIIA6, ArcIIA7 ArcIIA8 and ArcIIA9 respectively and when compared to the previously identified AcrIIA1 – 5, is less than 20% (Figure 8). Furthermore, all the members of the four protein families, encoded by genes identified in reference genomes, cluster in discrete groups that are non-overlapping with each other or with any of AcrIIA1 – 5, confirming the classification of the identified anti-CRISPR proteins and their family members as providing 4 new families of anti-CRISPR proteins (Figure 9). A surprising characteristic of the identified genes encoding the anti-CRISPR proteins of the invention is their wide distribution in natural microbial populations and habitats. While genes encoding AcrIIA1 – 5 are confined to *Firmicutes* bacteria, whose homologues are confined to a few species, the genes encoding members of the anti-CRISPR ArcIIA6, ArcIIA7 and ArcIIA9 protein families have a wide host range (Figure 10A). Additionally, genes and their homologues encoding ArcIIA7 and ArcIIA9 protein families were found in a wide range of habitats, based on analysis of publically available viromes datasets (Figure 10B).

II Structure of the ArcIIA8 anti-CRISPR proteins of the invention

A nucleic acid molecule [SEQ ID No.: 168] encoding ArcIIA8 [SEQ ID No.: 169] was identified in a gut metagenomic library. The genes encoding the AcrIIA8 protein family comprise at least 52 homologues present in 68 reference genomes of 38 unique species of *Firmicutes*. A common characteristic of the AcrIIA8 protein family is the location of their genes, which are all located within prophage DNA fragments present in the genomes of the *Firmicutes* species. The highly conserved synteny of these prophage DNA regions, and co-localisation of the AcrIIA8 genes within the region is seen in the alignment in Figure 11. The identification of the AcrIIA8 homologues is based on their highly conserved genetic origin (figure 11), clustering of the members

of this protein family (Figure 9), and on the conservation of their primary protein structure as seen from the alignment of their sequences (Figure 12).

Accordingly, the invention provides an isolated nucleic acid molecule encoding an anti-CRISPR protein belonging to the ArcIIA8 family, wherein the amino acid sequence of

5 said protein has at least 80, 81, 82, 83, 84, 85, 86, 87, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% amino acid sequence identity to an amino acid sequence selected from among SEQ ID No's: 169 [ArcIIA8]; 171 [WP_009270720.1], 173 [WP_008690853.1]; 175 [WP_076236714.1], 177 [WP_034586516.1], 179 [WP_035290260.1], 181 [WP_004605840.1], 183 [WP_003465204.1], 185 [WP_003477332.1], 187 [WP_079286774.1], 189 [WP_055225583.1], 191 [WP_077229201.1], 193 [WP_009242330.1], 195 [WP_048571017.1], 197 [WP_073005556.1], 199 [WP_049041135.1], 201 [WP_045518524.1], 203 [WP_061327540.1], 205 [WP_033066342.1], 207 [WP_071647430.1], 209 [WP_089966494.1], 211 [WP_041082932.1], 213 [WP_069680440.1], 215 [WP_061301923.1], 217 [WP_079481466.1], and 219 [WP_099313953.1].

In one preferred embodiment, the amino acid sequence of the encoded anti-CRISPR protein belonging to the ArcIIA8 family has at at least 80, 81, 82, 83, 84, 85, 86, 87, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% amino acid sequence identity to an amino acid sequence selected from among SEQ ID No.: 169 [ArcIIA8]; 171 [WP_009270720.1], 173 [WP_008690853.1].

III Structure of the ArcIIA6 anti-CRISPR proteins of the invention

A nucleic acid molecule [SEQ ID No.: 220] encoding ArcIIA6 [SEQ ID No.: 221] was 25 identified in a soil metagenomic library. The genes encoding the ArcIIA6 protein family includes at least 2 homologues present in *Cellulomonas carbonis* T26 (*Actinobacteria*) and *Sinorhizobium* sp. GL28 (*Proteobacteria*). An alignment revealing their highly conserved structure is shown in Figure 13.

Accordingly, the invention provides an isolated nucleic acid molecule encoding an anti-CRISPR protein belonging to the ArcIIA6 family, wherein the amino acid sequence of 30 said protein has at at least 80, 81, 82, 83, 84, 85, 86, 87, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% amino acid sequence identity to an amino acid sequence selected from among SEQ ID No.: SEQ ID No.: 221 [ArcIIA6], 223 [WP_058324878.1] and SEQ ID No.: 225 [KGM08958.1].

IIII Structure of the ArcIIA9 anti-CRISPR proteins of the invention

A nucleic acid molecule [SEQ ID No.: 226] encoding ArcIIA9 [SEQ ID No.: 227] was identified in a human gut fecal metagenomic library. The ArcIIA9 protein family is very large, encoded by genes distributed across multiple families of *Bacteroidetes*. Its

members are encoded by genes in several *Bacteroides* species and the protein members from three *Parabacteroides* species share 100% amino acid sequence identity. Gene homologues are also found in genomes of the recently defined phylum *Balneolaeota*, as well as the phyla *Firmicutes*, *Proteobacteria* and *Actinobacteria*. An amino acid sequence alignment with ten members of the AcrIIA9 protein family is shown in Figure 14; where all members share in common a highly conserved PcfK domain that is characterised by completely conserved residues (D and L).

Accordingly, the invention provides an isolated nucleic acid molecule encoding an anti-CRISPR protein belonging to the ArcIIA9 family, wherein the amino acid sequence of said protein has at at least 80, 81, 82, 83, 84, 85, 86, 87, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% sequence identity to an amino acid sequence selected from among SEQ ID No.: SEQ ID No.: 227 [ArcIIA9], 229 [WP_009036714.1], 231 [WP_072067265.1], 233 [WP_074707943.1], 235 [WP_026367493.1], 237 [WP_008152072.1], 239 [WP_009040414.1], 241 [WP_008654998.1], 243 [EKC79239.1], 245 [WP_055173355.1], 247 [WP_004294380.1].

Iiv Structure of the ArcIIA7 anti-CRISPR proteins of the invention

A nucleic acid molecule [SEQ ID No.: 248] encoding ArcIIA7 [SEQ ID No.: 249] was identified in a human gut fecal metagenomic library. The AcrIIA7 family is very large, being distributed across 6 distinct phyla (*Firmicutes*, *Proteobacteria*, *Bacteroidetes*, *Actinobacteria*, *Cyanobacteria* and *Spirochaetes*). The AcrIIA7 gene family has homologues in viral reference genomes, including all three tailed bacteriophage families (i.e. *Siphoviridae*, *Myoviridae* and *Podoviridae*), confirming its ubiquitousness. An amino acid sequence alignment with five selected AcrIIA7 protein family members is shown in Figure 15.

Accordingly, the invention provides an isolated nucleic acid molecule encoding an anti-CRISPR protein belonging to the ArcIIA7 family, wherein the amino acid sequence of said protein has at at least 80, 81, 82, 83, 84, 85, 86, 87, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% amino acid sequence identity to an amino acid sequence selected from among SEQ ID No.: 249 [ArcIIA7], 251 [WP_081033090.1], 253 [WP_006800621.1], 255 [WP_005795395.1], and 257 [EKC44436.1].

Iv Functional properties of the anti-CRISPR proteins of the invention

The anti-CRISPR proteins ArcIIA6, ArcIIA7 ArcIIA8 and ArcIIA9 and their identified family members (Ii – Iiv), are capable of regulating the activity of a CRISPR-Cas system. The ability of the anti-CRISPR proteins ArcIIA6, ArcIIA7 ArcIIA8 and ArcIIA9 to regulate a CRISPR-Cas system is illustrated in Example 3 and 5, where an *in vivo* assay demonstrated regulation of CRISPR-Cas9 editing of a target DNA sequence by each of these anti-CRISPR proteinin a microbial cell (Figure 6), and in a human cell

(Figure 16). A regulatory effect of said anti-CRISPR proteins on CRISPR-Cas9 editing was further demonstrated in an *in vitro* assay where inhibition of CRISPR-Cas9 nuclease cleavage of a target DNA sequence was detected (Figure 7A). Regulation of CRISPR-Cas systems may employ a range of molecular mechanisms, one involving the direct binding of the anti-CRISPR protein with components of the CRISPR-Cas system. The binding affinity of the anti-CRISPR proteins of the invention with spCas9 was determined; and surprisingly showed that AcrIIA8, in particular, had a higher binding affinity for spCas9 than AcrIIA2 (Figure 7B).

II A self-replicating genetic element comprising the nucleic acid molecule encoding an anti-CRISPR protein, and cells comprising the genetic element.

III Anti-CRISPR protein expression vectors

The invention further provides a self-replicating genetic element comprising a nucleic acid molecule encoding an anti-CRISPR protein of the invention (a member of one of said ArcIIA6, ArcIIA7, ArcIIA8 and ArcIIA9 protein families); where said nucleic acid molecule is operably linked to a promoter. Preferably the promoter is a heterologous promoter capable of directing gene expression in a selected host cell; where the promoter may drive constitutive or inducible expression, (e.g. a promoter in combination with a regulating transactivating element (e.g. riboswitch or inducible transcription factor)). The self-replicating genetic element, comprising said nucleic acid molecule linked to a promoter, facilitates expression of the anti-CRISPR protein either *in vitro* or *in vivo*, where the choice and design of the genetic element may be adapted for expression in a given cell-type or tissue-type. A suitable self-replicating genetic element for expression in bacteria includes a plasmid with an origin of replication (e.g. p15A as described in Example 1) and either a constitutive or inducible promoter to control expression of the nucleic acid molecule encoding the operably linked anti-CRISPR protein. A suitable self-replicating genetic element for expression in a mammalian cell typically contains both prokaryotic sequences that facilitate the propagation of the vector in bacteria and one or more eukaryotic transcription units that are expressed only in eukaryotic cells, as known to one skilled in the art. Delivery and expression of a nucleic acid molecule encoding the operably linked anti-CRISPR protein in mammalian cells may also be facilitated using viral vectors adapted for the purpose, for example a Lentivirus or Adenovirus, as known to one skilled in the art.

III CRISPR-Cas and anti-CRISPR protein co-expression vectors

The invention further provides a self-replicating genetic element (as described in III) comprising both a nucleic acid molecule encoding an anti-CRISPR protein of the invention and a CRISPR-Cas expression system; capable of directing the expression of both the anti-CRISPR protein and the components of the CRISPR-Cas system, as

detailed in section III. Alternatively, the anti-CRISPR protein of the invention and the CRISPR-Cas expression system may each be encoded by their respective nucleic acid molecules, said respective molecules being comprised on separate self-replicating genetic elements.

5

IIiii Biological cells expressing an anti-CRISPR protein

The invention further provides a biological cell comprising a nucleic acid molecule encoding an anti-CRISPR protein of the invention, or a self-replicating genetic element comprising said nucleic acid molecule. The cell may be a prokaryotic or eukaryotic cell, and may be present in, or derived from, a bacterium, yeast, fungus, insect, plant or mammal. When the biological cell is a eukaryotic cell, the cell may be an isolated cell or it may be present in a biological tissue; where the tissue may be healthy, diseased, and/or have genetic mutations. The biological tissue may include any single tissue (e.g., a collection of cells that may be interconnected) or a group of tissues making up an organ or part or region of the body of an organism and can include lung tissue, skeletal tissue, and/or muscle tissue. Exemplary tissues include, but are not limited to those derived from liver, lung, thyroid, skin, pancreas, blood vessels, bladder, kidneys, brain, biliary tree, duodenum, abdominal aorta, iliac vein, heart and intestines, including any combination thereof.

10

15

20

IIiv Isolated anti-CRISPR protein of the invention

The invention further provides an isolated anti-CRISPR protein expressed by the nucleic acid molecule of the invention, wherein the protein is a member of a protein family selected from the group ArcIIA6 protein, ArcIIA7 protein, ArcIIA8 protein and ArcIIA9 protein families. Methods for isolating an expressed anti-CRISPR protein of the invention are known to the skilled man and are illustrated in Example 3.5.

25

III Use of the nucleic acid molecule, or anti-CRISPR protein of the invention to regulate CRISPR-Cas associated activity

The anti-CRISPR proteins, encoded by the nucleic acid molecules of the invention, find use in a method of regulating CRISPR-Cas associated activity. For example they may be used in a method to regulate the activity of a CRISPR-Cas system, wherein said system is selected from the group consisting of Type 1, Type II, Type III; Type IV, Type V, and Type VI.

30

35

The CRISPR-Cas associated activity that can be regulated by the anti-CRISPR proteins of the invention share the common functional property of modifying or cleaving a target DNA molecule or its transcript; and thereby allowing genetic material to be added, removed, or altered at particular locations in the genome of a cell or in a DNA

molecule. The genome of a cell or the DNA molecule comprising the target DNA to be modified or cleaved by CRISPR-Cas associated activity, be means of a system that can be present in a biological cell as defined in section IIIiii. In general terms, such a system is located in a CRISPR locus consists of a CRISPR array, comprising short variable DNA sequences (called spacer sequence(s)) interspersed by short palindromic repeat sequences (called repeat sequence(s)), this being flanked by diverse cognate cas genes. The CRISPR-Cas systems may for example be those that have been classified into 2 classes; whereby members of Class 1 comprise a multi-subunit protein complex, each subunit encoded by a gene in the CRISPR locus; while members of Class 2 comprise a single multidomain protein encoded by a single gene. Class 1, includes type I, III and IV CRISPR-Cas systems; while Class 2, includes type II, V and VI CRISPR-Cas systems. Thus, in total there are six types, which have been further classified into over 28 subtypes; based on their protein structure, signature Cas genes and operon arrangements (Koonin, E. V. et al., 2017).

The genetic loci encoding Class 1, type I CRISPR-Cas systems share in common a signature gene *cas3* (or its variant *cas3'*) encoding a nuclease that both unwinds and cleaves target double-stranded DNA (dsDNA) and RNA-DNA duplexes. Type I systems comprise seven subtypes, I-A to I-F, and I-U. The genetic loci comprising Class 1, type III CRISPR-Cas systems all contain the signature gene *cas10*, as well genes encoding *Cas5* gene and several paralogues of *Cas7* that co-transcriptionally target RNA and DNA. The Type III systems comprise four subtypes, III-A to III-D. The loci comprising Class 1, type IV CRISPR-Cas system comprise two subtypes, IV-A and IV-B; where the *Csf1* can serve as as signature gene.

The genetic loci encoding the Class 2, type II CRISPR-Cas systems, comprise the signature *Cas9* gene encoding a multidomain protein that combines the functions of a crRNA-effector complex and cleaves target DNA. All Type II loci are characterised by a nucleic acid sequence encoding a *tracrRNA*, which is partially complementary to the repeats within the respective CRISPR array. Type II systems comprise three subtypes, II-A to II-C. Class 2, Type V CRISPR-Cas systems comprise the signature gene *Cpf1*, also encoding a single multidomain protein. Unlike *Cas9*, the *Cpf1* gene is commonly located outside a CRISPR-Cas locus. Type V systems are currently divided into five subtypes, V-A to V-E.

For the purpose of expression of members of the Class 2 typeII CRISPR-Cas systems in a cell, it is sufficient to express the signature Cas gene encoding the multidomain protein having RNA guided-endonuclease activity. Expression of all other Class 1 and 2 CRISPR-Cas systems can be achieved by co-expressing the several genes encoding the

Cas proteins that constitute the multi-subunit protein complex having RNA guided-endonuclease activity.

When expression of the CRISPR locus is activated, transcription of the CRISPR array
5 yields a pre-crRNA, that is processed into individual mature crRNAs (mat-crRNA). Each mat-crRNA sequence, aided by its cognate single or multi-subunit Cas protein(s) (effector complex), functions as a guide to specifically target the effector complex to modify or cleave a target nucleic acid molecule (DNA or RNA). Processing of the pre-crRNA is mediated by either an endonuclease subunit of the effector complex; or in the
10 case of type II CRISPR-cas system by a combination of a RNase III protein together with a trans-activating CRISPR RNA (tracrRNA); where tracrRNA serves to recruit the components of the type II effector complex.

For the purpose of expressing of any member of the Class 1 or 2 CRISPR-Cas systems
15 in a cell, an individual mat-crRNA comprising at least one spacer sequence, flanked at least at one end (preferably at both ends) by a palindromic repeat sequence, is sufficient to specifically guide the cognate effector complex to cleave a reporter nucleic acid molecule (DNA or RNA). Alternatively, a complete CRISPR array (pre-crRNA) can be expressed. The target nucleic acid molecule comprises a sequence that is
20 complementary to the spacer sequence. A tracrRNA is additionally necessary for the expression of a Type II CRISPR-cas system. Synthetic DNA molecules can conveniently be engineered to express a RNA transcript comprising a spacer sequence flanked, at least at one end, by a palindromic repeat sequence (i.e. a mat-RNA), for use in the expression of the Class 1 or 2 CRISPR-Cas systems. However, in the case a Type II
25 CRISPR-cas system, the RNA transcript of the synthetic DNA molecule can comprise a a mat-RNA fused with a tracrRNA sequence, also called single-guide RNA (sgRNA).

Accordingly, expression of a CRISPR-Cas system in a biological cell can be achieved by providing:

- 30 a. a first gene encoding a polypeptide, or a first gene cluster encoding a polypeptide complex, said polypeptide or polypeptide complex having RNA-guided endonuclease activity, wherein said gene or gene cluster is operably linked to a promoter; and
b. a nucleic acid molecule operably linked to a promoter, wherein said nucleic acid molecule encodes one or more RNA molecules capable of guiding said polypeptide or
35 polypeptide complex having RNA-guided endonuclease activity to a target gene or its transcript,
wherein the polypeptide or polypeptide complex having RNA-guided endonuclease activity is selected from the group of CRISPR-Cas systems consisting of Type 1, Type

II, Type III; Type IV, Type V, and Type VI and is capable of modifying or inactivating said target gene or its transcript.

More specifically, said nucleic acid molecule may encode any one of:

- 5 - a pre-crRNA molecule (encoded by a DNA molecule comprising a CRISPR array) wherein said pre-crRNA comprises a spacer sequence complementary to a nucleic acid sequence of at least 15 nucleotides of at least one strand of said target gene or its transcript;
- a mat-crRNA molecule comprising at least one spacer sequence flanked by at least one palindromic repeat sequence, wherein the spacer sequence is complementary to a nucleic acid sequence of at least 15 nucleotides of at least one strand of said target gene or its transcript;
- said pre-crRNA molecule and a trans-activating crRNA; or
- said mat-crRNA molecule and a trans-activating crRNA (or a fusion thereof),
- 15 wherein the palindromic repeat sequence and the trans-activating crRNA are cognate with respect to the polypeptide or polypeptide complex having RNA-guided endonuclease activity.

- Examples of the components of polypeptide or polypeptide complex that make up the CRISPR-Cas systems that may be expressed in a cell, and be regulated by the anti-CRISPR proteins of the invention are listed in Table 1 below, together with their DNA coding sequences. In a preferred embodiment, the amino acid sequence of the polypeptide or polypeptides of a protein complex having RNA-guided endonuclease activity has at least 70, 75, 80, 81, 82, 83, 84, 85, 86, 87, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% amino acid sequence identity to an amino acid sequence selected from Table 1. These polypeptide or polypeptide complexes each selectively associate with a cognate palindromic repeat sequence and, where applicable also a cognate tracrRNA. Thus, when the polypeptide or polypeptide complex is selected from Table 1, then the cognate palindromic repeat sequence and cognate tracrRNA are encoded by the nucleotide sequences listed in the adjacent boxes in Table 1.

Type - Sub-type	Strain designation & Genome ID	Locus id	Cas protein or protein complex sequences		Cognate palindromic repeat SEQ ID No.:	tracrRNA SEQ ID No.:
			Protein id	SEQ ID No.:		
I A	<i>Archaeoglobus fulgidus</i> DSM 4304	AF1870	WP_010879363.1	SEQ ID No: 1	SEQ ID: 109	N/A
		AF1871	WP_010879364.1	SEQ ID No: 2		
		AF1872	WP_010879365.1	SEQ ID No: 3		

	NC_000917.1	AF1873 AF1874 AF1875 AF1876 AF1877 AF1878 AF1879	WP_010879366.1 SEQ ID No: 4 WP_010879367.1 SEQ ID No: 5 WP_010879368.1 SEQ ID No: 6 WP_010879369.1 SEQ ID No: 7 WP_010879370.1 SEQ ID No: 8 WP_010879371.1 SEQ ID No: 9 WP_010879372.1 SEQ ID No: 10		
I B	<i>Clostridium difficile</i> 630 NC_009089.1	CD630_29770* CD630_29820 CD630_29810 CD630_29800 CD630_29780 CD630_29790	YP_001089495.1 SEQ ID No: 11 YP_001089500.1 SEQ ID No: 12 YP_001089499.1 SEQ ID No: 13 YP_001089498.1 SEQ ID No: 14 YP_001089496.1 SEQ ID No: 15 YP_001089497.1 SEQ ID No: 16	SEQ ID: 110	N/A
I C	C-125 DNA <i>Bacillus halodurans</i> NC_002570.2	BH0336 BH0337 BH0338 BH0339 BH0340 BH0341 BH0342	WP_010896517 SEQ ID No: 17 WP_010896518.1 SEQ ID No: 18 WP_010896519.1 SEQ ID No: 19 WP_010896520.1 SEQ ID No: 20 BAB04059.1 SEQ ID No: 21 WP_010896522.1 SEQ ID No: 22 WP_010896523.1 SEQ ID No: 23	SEQ ID: 111	N/A
I D	<i>Cyanothece</i> sp. 8802 NC_013161.1	Cyan8802_0527 Cyan8802_0526 Cyan8802_0525 Cyan8802_0524 Cyan8802_0523 Cyan8802_0522 Cyan8802_0521 Cyan8802_0520	WP_015783167.1 SEQ ID No: 24 WP_015783166.1 SEQ ID No: 31 WP_015783165.1 SEQ ID No: 30 WP_015783164.1 SEQ ID No: 29 WP_015783163.1 SEQ ID No: 28 WP_015783162.1 SEQ ID No: 27 WP_015783161.1 SEQ ID No: 26 WP_015783160.1 SEQ ID No: 25	SEQ ID: 112	N/A
I E	<i>Escherichia coli</i> K-12 NC_000913.3	ygcB ygbF	NP_417241.1 SEQ ID No: 32 NP_417234.2 SEQ ID No: 33	SEQ ID: 113	N/A
I F	<i>Yersinia</i> <i>pseudotuberculosis</i> YPIII NC_010465.1	YPK_1644 YPK_1645 YPK_1646 YPK_1647 YPK_1648 YPK_1649	WP_072080408.1 SEQ ID No: 34 WP_012303962.1 SEQ ID No: 35 WP_012303963.1 SEQ ID No: 36 WP_012303964.1 SEQ ID No: 37 WP_011192624.1 SEQ ID No: 38 WP_002211866.1 SEQ ID No: 39	SEQ ID: 114	N/A

I F* (variant)	<i>Shewanella putrefaciens</i> CN-32 NC_009438.1	Sputcn32_1819 Sputcn32_1820 Sputcn32_1821 Sputcn32_1822 Sputcn32_1823	WP_011789486.1 SEQ ID No: 40 WP_000989810.1 SEQ ID No: 41 WP_011919225.1 SEQ ID No: 42 WP_011919226.1 SEQ ID No: 43 WP_011919227.1 SEQ ID No: 44	SEQ ID: 115	N/A
II A	<i>Streptococcus pyogenes</i> MGAS10270 NC_008022.1	RS04390	WP_020905136.1 SEQ ID No: 45	SEQ ID: 116	SEQ ID: 134
II B	<i>Legionella pneumophila</i> str. Paris NC_006368.1	Lpp0160	WP_011212792.1 SEQ ID No: 46	SEQ ID: 117	SEQ ID: 135
II C	<i>Neisseria meningitidis</i> 8013 NC_017501.1	NMV_RS09835	WP_014574210.1 SEQ ID No: 47	SEQ ID: 118	SEQ ID: 136
III A	<i>Staphylococcus epidermidis</i> RP62A NC_002976.3	SERP2455 SERP2456 SERP2457 SERP2458 SERP2459 SERP2460 SERP2461 SERP2462 SERP2463	WP_002486027.1 SEQ ID No: 48 WP_002486034.1 SEQ ID No: 49 WP_002486031.1 SEQ ID No: 50 WP_002486041.1 SEQ ID No: 51 WP_002486018.1 SEQ ID No: 52 WP_002486044.1 SEQ ID No: 53 WP_002486045.1 SEQ ID No: 54 WP_002469399.1 SEQ ID No: 55 WP_002486040.1 SEQ ID No: 56	SEQ ID: 119	N/A
III B	<i>Pyrococcus furiosus</i> DSM 3638 NC_003413.1	PF1124 PF1125 PF1126 PF1127 PF1128 PF1129 PF1130 PF1131	WP_011012264.1 SEQ ID No: 57 WP_011012265.1 SEQ ID No: 58 WP_011012266.1 SEQ ID No: 59 WP_011012267.1 SEQ ID No: 60 WP_011012268.1 SEQ ID No: 61 WP_011012269.1 SEQ ID No: 62 WP_011012270.1 SEQ ID No: 63 WP_011012271.1 SEQ ID No: 64	SEQ ID: 120	N/A
III B* (variant)	<i>Marinomonas mediterranea</i> MMB_1 NC_015276.1	Marme_0668 Marme_0669 Marme_0670 Marme_0671 Marme_0673 Marme_0674	WP_013659857.1 SEQ ID No: 65 WP_013659858.1 SEQ ID No: 66 WP_013659859.1 SEQ ID No: 67 WP_013659861.1 SEQ ID No: 68 WP_083799838.1 SEQ ID No: 69 WP_013659863.1 SEQ ID No: 70	SEQ ID: 121	N/A

		Marme_0675 Marme_0676 Marme_0677	WP_013659864.1 SEQ ID No: 71 WP_013659865.1 SEQ ID No: 72 WP_013659866.1 SEQ ID No: 73		
III C	<i>Methanothermobacter thermautotrophicus str. Delta H</i> NC_000916.1	MTH323 MTH324 MTH325 MTH326 MTH327 MTH328	WP_010875962.1 SEQ ID No: 74 WP_010875963.1 SEQ ID No: 75 WP_010875964.1 SEQ ID No: 76 WP_010875965.1 SEQ ID No: 77 WP_010875966.1 SEQ ID No: 78 WP_010875967.1 SEQ ID No: 79	SEQ ID: 122	N/A
III D	<i>Synechocystis sp. 6803</i> NC_005230.1	sII7063 sII7064 sII7065 sII7066 sII7067 slr7068	WP_011153732.1 SEQ ID No: 80 WP_011153733.1 SEQ ID No: 81 WP_011153734.1 SEQ ID No: 82 WP_011153735.1 SEQ ID No: 83 WP_011153736.1 SEQ ID No: 84 WP_011153737.1 SEQ ID No: 85	SEQ ID: 123	N/A
IV	<i>Thioalkalivibrio sp. K90mix</i> NC_013930.1	TK90_2699 TK90_2700 TK90_2701 TK90_2702 TK90_2703	WP_013006550.1 SEQ ID No: 86 WP_013006551.1 SEQ ID No: 87 WP_018940624.1 SEQ ID No: 88 WP_013006553.1 SEQ ID No: 89 WP_013006554.1 SEQ ID No: 90	SEQ ID: 124	N/A
V A	<i>Francisella cf. novicida Fx1</i> NC_017450.1	FNFX1_1431	WP_014550095.1 SEQ ID No: 91	SEQ ID: 125	N/A
V B	<i>Alicyclobacillus acidoterrestris ATCC 49025</i> NZ_AURB01000127.1	N007_06525 N007_06530 N007_06535	WP_021296342.1 SEQ ID No: 92 WP_021296343.1 SEQ ID No: 93 WP_021296344.1 SEQ ID No: 94	SEQ ID: 126	SEQ ID: 137
V C	<i>Oleiphilus sp. HI0009</i> <i>HI0009_c113</i> LWEK01000147.1	A3715_16885 A3715_16890	KZX85786.1 SEQ ID No: 95 KZX85787.1 SEQ ID No: 96	SEQ ID: 127	N/A
V D	<i>Bacterium CG09_39_24</i> <i>isolate CG09_39_24</i> MOEH01000029.1	BK003_02070 BK003_02075	OJI08768.1 SEQ ID No: 97 OJI08769.1 SEQ ID No: 98	SEQ ID: 128	N/A
V E	<i>Deltaproteobacteria bacterium</i> <i>GWA2_43_19</i> <i>gwa2_scaffold_7984</i>	A2Z89_08250 A2Z89_08255 A2Z89_08260 A2Z89_08265	OGP07438.1 SEQ ID No: 99 OGP07439.1 SEQ ID No: 100 OGP07440.1 SEQ ID No: 101 OGP07441.1 SEQ ID No: 102	SEQ ID: 129	SEQ ID: 138

	MGPG01000094.1				
VI A	<i>Leptotrichia shahii</i> DSM 19757 B031DRAFT_scaffold_9.10 NZ_KB890278.1	B031_RS011044 5	WP_018451595.1 SEQ ID No: 103	SEQ ID: 130	N/A
VI B1	<i>Prevotella buccae</i> ATCC 33574 SCAFFOLD1 NZ_GL586311.1	HMPREF6485_R S00335 HMPREF6485_R S00340	WP_004343973.1 SEQ ID No: 104 WP_004343974.1 SEQ ID No: 105	SEQ ID: 131	N/A
VI B2	<i>Bergeyella zoohelcum</i> ATCC 43767 JH932293.1	HMPREF9699_0 2005 HMPREF9699_0 2006	EKB54193.1 SEQ ID No: 106 EKB54194.1 SEQ ID No: 107	SEQ ID: 132	N/A
VI C	<i>Fusobacterium perfoetens</i> ATCC 29250 NZ_JHXL01000011.1	T364_RS010511 0	WP_027128616.1 SEQ ID No: 108	SEQ ID: 133	N/A

N/A = not applicable or required; * non-essential Cas component of Type 1B

5 The target gene may comprise a protospacer adjacent motif (PAM) located adjacent to the sequence complementary to the mat-crRNA spacer sequence, in particular where the RNA-guided endonuclease polypeptide is a Type II A Cas (e.g. Cas9). A canonical PAM is the sequence 5'-NGG-3' where "N" is any nucleobase.

In one embodiment, the CRISPR-Cas system regulated by an anti-CRISPR protein of the invention is a Class 2 type II A Cas9 system.

10 In one embodiment, the CRISPR-Cas system regulated by an anti-CRISPR protein of the invention is a Class 2 type V A system comprising the signature gene Cpf1 encoding a Cpf1 protein (e.g. WP_014550095.1).

In one embodiment, the CRISPR-Cas system regulated by an anti-CRISPR protein of the invention is a Class 1 type I system comprising the signature gene Cas3 encoding
15 a Cas3 protein.

Each of said first gene, first gene cluster, and nucleic acid molecule may independently be a component of a self-replicating genetic element (episome) that is stably replicated in a biological cell or is an integrated component of its chromosome.

5 IV Applications of the nucleic acid molecule of the invention and its encoded anti-CRISPR protein

The anti-CRISPR proteins and their respective encoding nucleic acid molecule, capable of regulating CRISPR-Cas gene editing activity, provide an advantageous adjunct (e.g. by co-administration) in permitting safe and practical biological therapeutics through
10 spatial or temporal control of CRISPR-Cas activity; controlling CRISPR-Cas-based gene drives in wild populations to reduce the ecological consequences of such forced inheritance schemes; and contributing to general research into various biotechnological, agricultural, and medical applications of gene editing technologies.

15 IVi. Temporal control to minimize Off-Target Effects

An anti-CRISPR protein of the invention, and/or its respective encoding nucleic acid molecule, find a use in a method of controlling the duration of CRISPR-Cas activity. For example, a CRISPR-Cas expression system delivered into a cell, can be expressed, and allowed to function for a predetermined duration shown empirically to yield
20 efficient on-target editing, and then inactivated through the co- or post-delivery of the anti-CRISPR protein or its encoding nucleic acid molecule. In some cases it could be important to stop Cas9 activity after the initial rounds of mitosis, otherwise it could give rise to mosaic genotypes.

25 IVii Control Of Spatial Activity

An anti-CRISPR protein of the invention, and/or its respective encoding nucleic acid molecule, find a use in a method of controlling the spacial activity of a CRISPR-Cas having two or more tissue or cell type specific activities where the anti-CRISPR protein selectively regulates at least one of said plurality of tissue- or cell type-specific
30 activities. The co- or post-delivery of the anti-CRISPR protein or its encoding nucleic acid molecule to a cell with a CRISPR-Cas expression system will result in CRISPR-Cas activity being confined to a selected subset of possible cell types or tissues.

IViii Gene Drives

35 An anti-CRISPR protein of the invention, and/or its respective encoding nucleic acid molecule, find a use in a method of controlling CRISPR-Cas mediated gene drives used to modify the genomes of entire populations of organisms for potentially beneficial purposes. For examples use in i) preventing the spread of genes in insect vectors of infectious disease (malaria, dengue, etc.); ii) preventing the ability of a general

population to act as a host for a pathogen; or iii) rendering a specific population sub-fertile or even sterile. By introducing the nucleic acid sequence encoding the anti-CRISPR of the invention into a "reserve population" it would be possible to counter the potential for negative effects of an uncontrolled gene drive.

5

IViv Pharmaceutical compositions

A composition comprising an anti-CRISPR protein of the invention, and/or its respective encoding nucleic acid molecule is provided for use as a medicament, where the composition may optionally comprise ingredients required for administration of said composition to a subject in need thereof. Said composition finds application when provided as an adjunct (e.g. co-administration) to the administration of a CRISPR-Cas expression system to the subject, whereby said anti-CRISPR protein is one that is capable of regulating the CRISPR-Cas gene editing activity of said CRISPR-Cas expression system such that target DNA in the subject is edited, while limiting or preventing off-target DNA editing. The DNA in the subject that is the target for DNA editing by the CRISPR-Cas expression system is DNA in a somatic cell (and not a germ-line cell).

EXAMPLES

20 Example 1 Construction and cloning of a genetic circuit for screening for polypeptide inhibitors of CRISPR-Cas associated activity

1.1 Construction of pCasens3 and pDual3 plasmids

The pCasens3 and pDual3 plasmids, illustrated in Figure 2A (central and right-hand plasmids respectively), were constructed as follows:

Plasmid pCasens3 containing the *Streptococcus pyogenes* Cas9 gene (spCas9) [SEQ ID No.:] was constructed in a single step by using USER cloning (Genee et al. 2014). A DNA fragment containing spCas9 gene [SEQ ID No.:150] and its endogenous terminator [SEQ ID No.:152] was PCR amplified from a plasmid DS-SPcas (supplied by addgene: Plasmid #48645) (Esvelt et al. 2013); and the amplified DNA fragment was cloned into the backbone of the plasmid pSEVA47 (as described by Martinez-Garcia et al. 2014). The plasmid pSEVA47 contains: the low copy number origin of replication pSC101 [SEQ ID No.:155], and the antibiotic resistance gene aadA [SEQ ID No.:153] that confers resistance against Spectinomycin and a restriction site for insertion of a gene. A DNA molecule [SEQ ID No.:149] encoding a theophylline translational riboswitch was placed in front of Cas9 using a long forward primer:

5'AAGTCTAGCGAACCGCACTTAATACGACTCACTATAGGTACCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCCTGCTAAGGTAACAACAAGATGATGGATAAGAAATACTCAATAGGCTTAGATATCGGCAC-3' [SEQ ID No.:156]). Additionally, a sigma70 constitutive

promoter (J23100; as defined by <http://parts.igem.org/Promoters/Catalog/Anderson>) was also introduced using a reverse primer in substitution for the endogenous of spCas9 promoter:

5'-ctctagTagctagcactgtacctaggactgagctagccgtcaaGTTAGCTGTGCTCTAGAAGCTAGCAG-3'
[SEQ ID No.:157].

Plasmid pDual3 for arabinose inducible sgRNA expression was constructed by using USER cloning (Genee et al. 2014) to insert the respective genes into the backbone of the plasmid was from pSEVA3610, described by Martínez-García et al., 2015. The chimeric gRNA encoding sequence [SEQ ID No.:161] with a TrnB terminator [SEQ ID No.:162] and a chloramphenicol resistance gene (CmR) [SEQ ID No.:164] were cloned downstream of the inducible pBAD promoter [SEQ ID No.:158] and araC gene [SEQ ID No.:159] encoding an L-arabinose-inducible transcription factor. In the absence of L-arabinose, the AraC protein binds to operator sites within pBAD effectively repressing transcription. On addition of L-arabinose, AraC protein binds to L-arabinose to form a complex having a DNA-binding conformation that activates pBAD leading to induction of transcription of its cognate genes, namely sgRNA and CmR. The plasmid comprises a low copy number origin of replication, p15A [SEQ ID No.:163]. The first twenty nucleotide sequence of the encoded gRNA transcript is complementary to nucleotides 96:115 of the + strand in the CmR gene having [SEQ ID No.:164].

Positive strand nucleotides 96:115: CTATAACCAGACCGTTCAGC [SEQ ID No.:166]

1.2 Construction of positive and negative control plasmids for testing the genetic circuit

Control plasmids for testing the genetic circuit include a positive control plasmid (left-hand plasmid in Figure 2A) comprising an *acrIIa2* gene [SEQ ID No.:159] capable of expression of the anti-CRISPR protein, AcrIIA2, and a negative control plasmid (Figure 2B) comprising a *gfp* gene [SEQ ID No.:140] with *rrnB* T1 terminator [SEQ ID No.:142] capable of expression of GFP; each under the control of a constitutive pTET promoter (lacking the TET repressor) and RBS [SEQ ID No.:139]. Both plasmids have a ColE1 ori [SEQ ID No.:143] and kanamycin resistance gene [SEQ ID No.:144]. When the control plasmids are transformed into host cells, comprising the genetic circuit (i.e. pCasens3 and pDual3 plasmids as described in Example 1.1); the expression of the anti-CRISPR protein, AcrIIA2 is capable of inhibiting Cas9 mediated inactivation of the CmR gene.

1.3 Host cells transformed with pCasens3 and pDual3 plasmids

E. coli TOP10 were made electro-competent and transformed with the genetic circuit comprising the plasmids: pCasens3 and pDual3 (obtained in example 1.1) as follows:

- 5 i. A 40 µl sample of electrocompetant *E. coli* TOP10 cells were transferred to microcentrifuge tubes; to which a 1 µl sample of pCasens3 and pDual3 DNA was added; and this DNA-cell mixture was then transferred to a cold cuvette and electroporated with a pulse having 1.8 kv, 200 ohms and 25 µF.
- 10 ii. Immediately thereafter, a 975 µl volume of 37°C SOC [Ausubel F.M. et al. (1987)] was added to the electroporated cells; and then incubated on rollers at 37°C for 1 h;
- 15 iii. The recovered transformed cells were diluted $10^0 - 10^8$; plated on selective solid media comprising Luria Broth agar supplemented with: 50 µg/mL spectinomycin; 30 µg/mL chloramphenicol and 50 µg/mL kanamycin; and incubate overnight at 37°C.
- iv. Transformed cell colonies comprising the pCasens3 and pDual3 plasmids (*E. coli* genetic circuit [*E. coli*-GC]) were selected; amplified and stored.

1.4 *E. coli*-GC cells transformed with control plasmids

- 20 *E. coli* TOP10 containing the genetic circuit comprising the plasmids: pCasens3 and pDual3 (*E. coli*-GC) were made electro-competent and transformed with either the negative control plasmid expressing GFP or positive control comprised a plasmid expressing AcrIIA2, as described in Example 1.2, employing the transformation protocol described in Example 1.3, but where the host cells were *E. coli*-GC.

25 **Example 2 *E. coli*-GC host cells provide an effective screening tool for polypeptide inhibitors of CRISPR-Cas associated activity**

2.1 *E. coli*-GC host cells detect proteins having anti-CRISPR activity

- 30 *E. coli*-GC host cells transformed with the positive control plasmid expressing AcrIIA2 (Figure 2A), or the negative control expressing GFP (Figure 2B); were screened for anti-CRISPR activity as follows:

- 35 i. Overnight cultures of said *E. coli* host cells were prepared by culture in 3ml 2xYT media supplemented with the following antibiotics (50 µg/mL spectinomycin; 30 µg/mL chloramphenicol and 50 µg/mL kanamycin) at 37°C and 250RPM;
- ii. Said cultures were used to prepared dilutions from 10^0 - 10^7 and 5µl volumes of each dilution was spotted onto a 1st and 2nd set of agar plates, in order to determine both cell death and cell viability, having the following agar composition:

- 1st set of plates used to determine cell death comprised LB agar supplemented with: 2mM theophylline; 1% arabinose and antibiotics (50 µg/mL spectinomycin; 30 µg/mL chloramphenicol and 50 µg/mL kanamycin);
- 2nd set of plates used to determine cell viability comprised LB agar supplemented with: 2mM theophylline; 1% arabinose; and antibiotics (50 µg/mL spectinomycin; and 50 µg/mL kanamycin).

The survival of *E. coli*-GC host cells comprising the negative control plasmid, and grown under conditions that induce expression of the genetic circuit, is significantly reduced (see Figure 3B). However, expression of the, AcrIIA2 is seen to enable cell growth and survival, consistent with the known anti-CRISPR properties of this protein (Figure 3A and B).

Example 3 Construction and screening of Metagenomic Libraries for anti-CRISPR proteins

3.1 Construction of expression metagenome libraries from diverse sources

Functional metagenomic libraries from human, cow and pig fecal samples and soil samples were constructed as described previously [Sommer et al. 2009; Genée et al., 2016]. The steps of the procedure were as follows:

- i) total DNA was isolated from each of 5 g of fecal samples and a soil sample using the PowerMax Soil DNA Isolation Kit (Mebio Laboratories Inc.); where after total DNA isolated from each sample was treated as follows:
- ii) the extracted DNA, in a total sample volume of 200µl in minitubes, was fragmented by sonication into fragments of an average size of 4-5 kb using a Covaris E210 (Massachusetts, USA);
- iii) the sheared DNA (200µl) was size-selected by separation on a 1 % agarose gel with DNA dye and DNA migrating in the 3-6 kb region was extracted using a Qiagen gel extraction kit;
- iv) the gel-extracted DNA fragments were end-repaired using an End-It end repair kit (Epicentre) and selected using PCR clean up column; and
- v) the end-repaired fragments were blunt-end cloned into the HincII site, downstream of a constitutive pTET promoter (lacking the tet repressor) and RBS [SEQ ID No.:139] in the expression 100-200 ng pZE21 plasmid in a 1:5 ratio [Lutz and Bujard, 1997] using the Fast Link ligation kit (Epicentre) over-night and the reaction was cleaned with a spin column and eluted in a 15 µl twice;
- vi) 2 µl cleaned ligation mix was transformation into 50 µl *E. coli* top10 cells by electroporation; and cells in the transformation mix were recovered in 1 ml SOC medium for 2 h; using multiple electroporations to increase library size;

- vii) the transformation mix were plated in 1:100 and 1:1000 dilutions on kanamycin 50 µg/ml plates to create a library;
- viii) the library was amplified by inoculation into 10 ml LB+Kan 50 and grown over night; 5 ml 50 % glycerol is added and the library is aliquoted into 1 ml portions and frozen at -80 C.

In brief, plasmid DNA, comprising the amplified metagenomic library was extracted and recovered using Machery-Negel plasmid DNA purification kit as follows:

1. An overnight culture of bacteria comprising the metagenomic library was grow in 3 ml LB medium supplemented with 50 µg/mL kanamycin;
2. the culture was centrifuged, and the pelleted bacteria were resuspend the bacteria in buffer, to which a denaturing solution was added;
3. the proteins and genomic DNA were pelleted by centrifugation, and the plasmid-containing supernatant was recovered and purified through the Nucleospin ® plasmid column on the kit; and bound plasmid DNA was eluted from the column using water or a neutral buffer such as Tris:EDTA, to provide a plasmid DNA comprising the metagenomic DNA expression library.

3.2 Transformation and screening of Metagenomic DNA expression library into host cells, comprising genetic circuit

E.coli TOP10 containing the genetic circuit comprising the plasmids: pCasens3 and pDual3 (E. coli-GC) were made electro-competent and transformed with plasmids for expression of each metagenomic library (obtained in example 3.1) as follows:

- i. E.coli TOP10 cells (supplied by ThermoFischer) containing the plasmids: pCasens3 and pDual3 were cultured in 3 ml of 2xYT media with 50 µg/mL spectinomycin and 30 µg/mL chloramphenicol for 16-18 hours at 37°C and 250 rpm;
- ii. the culture from (i) was inoculated 1:100 in 3 ml of pre-warmed 2xYT media with 50 µg/mL spectinomycin and 30 µg/mL chloramphenicol; and cultured at 37°C at 250 rpm until the cell density reached 0.5-0.7 OD600;
- iii. the cultures from (ii) were cooled on ice; and the cells harvested by centrifugation at 10 000 xg for 1 min;
- iv. the cells from (iii) were washed by re-suspension in cold water and then harvested again by centrifugation; then the cells were re-suspended in 1 ml glycerol and harvested by centrifugation; and finally the cells were re-suspended in the residual water and then placed on ice, ready for electroporation.

- v. A 40 µl sample of the electrocompetant cells from (iv) were transferred to microcentrifuge tubes; to which a 1 µl sample of the metagenomic library was added; and this DNA-cell mixture was then transferred to a cold cuvette and electroporated with a pulse having 1.8 kv, 200 ohms and 25 µF.
- 5 vi. Immediately thereafter, a 975 µl volume of 37°C SOC [Ausubel F.M. et al. (1987)] was added to the electroporated cells from (v), which were mixed and transferred to a 15 ml- tube; and incubated on rollers at 37°C for 1 h;
- vii. serial dilutions from 10⁰- 10⁸ from the recovered cells from (vi) were made; and 100ul of each dilution were plated on selective solid media comprising
10 Luria Broth agar supplemented with: 2Mm theophylline; 1% arabinose; 50 µg/mL spectinomycin; 30 µg/mL chloramphenicol and 50 µg/mL kanamycin; and incubated overnight at 37°C.
- viii. Colonies growing on the plates were collected by adding 1 ml H₂O, after which the colonies were scraped off the plate with a sterile loop. The bacterial cells
15 were then pelleted and washed in water by centrifugation and the plasmid DNA in collected bacterial cells was extracted using a Machery-Negel plasmid DNA purification kit.

20 3.3 Detection of metagenomic DNA encoding proteins having anti-CRISPR activity

E.coli host cells containing the genetic circuit, transformed with the library of metagenomic DNA expression plasmids (Figure 5) prepared as described in Example 3.1, were screened for expression of proteins having anti-CRISPR activity (as
25 described in Example 3.2 and figure 4). Colonies that appeared on the selective growth medium were selected and pooled; their metagenomic inserts were extracted, barcoded per library and sequenced using nanopore technology to obtain full inserts (22). The resulting contigs were annotated with blastx and manually curated. The selected metagenomic DNA insert (39 inserts) were each re-cloned in the
30 metagenomics DNA expression plasmid (Figure 5, plasmid to the left); and then transformed into cells of the *E. coli* strain containing the genetic circuit comprising the plasmids: pCasens3 and pDual3 (*E. coli*-GC). Eleven of the individually tested inserts showed anti-CRISPR activity that was greater than the negative control (GFP) and several were similar to that of AcrIIA2 in the selection assay (Figure 6).

35

3.4 Identification of ORFs for polypeptides having ACR activity

Identification of putative ORFs from inserts with ACR activity were detected using MetaGeneMark v3.25 (Zhu, W., et al., 2010) with the flags `gmhmm -m MetaGeneMark\_v1.mod -f G`. In parallel the inserts were re-annotated using blastx

using the NT database accessed at the 17th of May 2017. Based on the combined annotation, potentially 16 biologically active ORFs were manually identified (ie. >80% of the subject gene present, no missing N-terminus) and USER primers were designed to clone the 16 identified ORFs into the pNIC28-Bsa4 plasmid (Savitsky, P, et al., 2010).

3.5 Expression of novel ACR polypeptides

Nine of the cloned ORFs in the pNIC28-Bsa4 plasmid (from 3.4), were expressed in *E. coli* strain BL21 (AI) grown in the 2xYT medium at 18°C for 16 hours following induction with 1% arabinose. Proteins were purified by a combination of affinity, ion exchange, and size exclusion chromatography steps.

Briefly, cells were lysed by three passes through an EmulsiFlex-C5 homogenizer (Avestin, Mannheim, Germany) at 10 000–15 000 psi, any debris and unbroken cells were removed by centrifuging at 18 000 g at 4°C for 30 minutes. The supernatant was loaded onto nickel-nitrilotriacetic acid (Ni²⁺-NTA) resin columns (HiTrap, GE Healthcare, Chicago, IL, USA) on an Äkta Pure system connected to an F9-C fraction collector (GE Healthcare). The expressed proteins were eluted by increasing the imidazole concentration in a stepwise manner to 25 mM, 50 mM, 75 mM and finally 500 mM. After pooling and concentration, protein samples were buffer exchanged into IEX start buffer. Anion exchange was performed on a HiTrap Q FF column (GE Healthcare) in 20 mM phosphate buffer pH 7.0 (AC12-1, AC19-2, AC28-1, AC42-1, AcrIIA2 and GFP) or 20 mM TRIS-HCl pH 8.0 (AC23-2 and AC27-1). Cation exchange was performed on a HiTrap SP FF column (GE Healthcare) in 20 mM phosphate buffer pH 7.0 (AC23-1 and AC27-2). The fractions containing the protein of interest were pooled, concentrated, flash frozen, and stored at -80°C. Purity analysis was performed using a Coomassie-stained SDS-PAGE gel analyzed by ImageQuant TL software (GE Healthcare).

MBP-Cas9 was expressed from plasmid pMJ806 and was essentially purified as described (41) with some modifications. After performing expression and His-tag affinity purification as described above, MBP-Cas9 was further purified using an MBPTrap HP column (GE Healthcare). After cleavage with AcTEV protease (ThermoFisher Scientific, Waltham, MA, USA) during overnight dialysis and negative His-tag affinity purification, the sample was loaded onto a Superdex 200 Increase 10/300 GL column (GE Healthcare) equilibrated with 50 mM Tris-HCl pH 7.5, 150 mM NaCl. The fractions containing Cas9 were pooled, concentrated and biotinylated using EZ-Link™ Sulfo-NHS-LC-Biotinylation Kit (ThermoFisher Scientific). After buffer exchange, samples were flash frozen and stored at -80°C.

3.6 Functional characterisation of novel ACR polypeptides

The nine expressed proteins were functionally characterized as follows:

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

Four expressed proteins (designated AC19-2, AC23-2, AC27-1 and AC42-1) were identified as ACR polypeptides, that inhibited cleavage of a target DNA molecule by the spCas9:gRNA complex in the *in vitro* assay (Figure 7A).

The *in vitro* direct binding assay to spCas9 was carried out using biolayer interferometry. Equimolar gRNA was mixed together with biotinylated Cas9 and incubated at 25°C for 15min to form a biotinylated Cas9:gRNA complex. Streptavidin biosensors (Pall FortéBio) were pre-equilibrated in PBS buffer for 600s, loaded with a
20 biotinylated Cas9:gRNA complex at optimal concentrations and times, and brought to baseline in kinetics buffer (1X PBS, 0.02% Tween-20, 0.1% BSA, 75mM KCl, 10mM MgCl₂) for 300s. Association with anti-CRISPR proteins was measured in the same kinetics buffer for 600s, and then dissociation was measured in the kinetics buffer without anti-CRISPR proteins for 1000s. All biolayer interferometry experiments were
25 performed on Octet RED96 system (Pall FortéBio) in 96-well microplates at 30°C with 200 µl volume. Binding kinetics were calculated using the FortéBio Data Analysis v7.1 software by fitting the association and dissociation data to a 1:1 model.

The proteins AC19-2, AC27-1 and AC42-1 showed binding affinity to Cas9:gRNA complex when Mg²⁺ was added to the running buffer (Figure 7B). The observed K_d for AC19-2, AC27-1 and AC42-1 was 47, 12 and 810 nM, respectively. Surprisingly, AC27-1 showed a stronger binding affinity to Cas9:gRNA complex than AcrIIA2. An absence of detectable binding between AC27-1 and spCas9 and the lower binding affinity detected between AC42-1 and Cas9:gRNA complex using biolayer
30 interferometry, suggests that these two ACR proteins as CRISPR-Cas inhibitors share a different mode of action. The 4 proteins, that clearly display anti-spCas9 activity, based on the *in vivo* and *in vitro* assays were renamed AcrIIA6 (AC19-2), AcrIIA7 (AC23-2), AcrIIA8 (AC27-1) and AcrIIA9 (AC42-1) (Figure 7B). AcrIIA6 originates
35

from a soil metagenomic library whereas AcrIIA7, AcrIIA8 and AcrIIA9 are derived from human gut fecal metagenomic libraries.

Example 4 The ACRs of the invention are distinct families of ACR IIA proteins

5

The four ACR proteins of the invention (ACRIIA6, ACRIIA7, ACRIIA8 and ACRIIA9) were classified as AcrIIA proteins based on their demonstrated ability to inhibit CRISPR-Cas systems belong to Class 2 subtype IIA. Each of the four ACR proteins are representatives of a new family of ACRIIA proteins, each representative having less than 25% sequence identity with each other, as well as any of the ACRIIA2-5 proteins (Figure 8). Genes encoding the four new ACRIIAs 6-9 families, were identified by interrogating metagenomics datasets. The ACRIIAs 6-9 protein families were found to be structurally distinct families of AcrIIA, where the members of each family share a high degree of structural homology, as seen from the cluster analysis shown in Figure 9.

10

Computational analysis of the distribution of the distinct AcrIIA gene families in reference genomes revealed that their host range was surprisingly highly diverse, in strong contrast to known AcrIIA 2-5, which are all confined to Firmicutes genomes (Figure 10A). Furthermore, the AcrIIAs 6 - 9 genes and their homologues were both more abundant and diversely distributed across multiple environments compared to previously characterized AcrIIAs (Figure 10B).

15

4.1 Structural characteristics of ACRIIA 8

20

Fifty-two homologous AcrIIA8 genes were identified in 68 reference genomes of 38 unique species of *Firmicutes* comprising about 1% of the 3257 *Firmicutes* species. An AcrIIA8 homologue was also found in a *Listeria monocytogenes* genome, co-existing with AcrIIA1, AcrIIA2, and AcrIIA4. All of the identified AcrIIA8 homologues share a viral (prophage) origin since they were all localized within highly conserved prophage regions in the genome of diverse *Firmicutes* family members (Figure 11). Based on both the highly conserved synteny of the viral fragments comprising the AcrIIA8 genes; and the conserved amino acid sequence of the encoded proteins, it is concluded that the proteins encoded by the AcrIIA8 genes shown in figure 11 are ACRII8 proteins having anti-CRISPR-Cas activity.

25

30

4.2 Structural characteristics of ACRIIA 6

Homologues of the soil-derived AcrIIA6 gene were found in the genomes of soil isolates *Cellulomonas carbonis* T26 (*Actinobacteria*) and *Sinorhizobium* sp. GL28

35

(Proteobacteria), as seen in Figure 10 and 13. Both genes had a very high degree of homology to AcrIIA6 gene (94% identity at nucleotide level) and its encoded ACRIIA6 protein (99% sequence identity).

5 **4.3 Structural characteristics of ACRIIA 7**

Homologues of the AcrIIA7 gene are widely distributed across 6 distinct phyla (*Firmicutes*, *Proteobacteria*, *Bacteroidetes*, *Actinobacteria*, *Cyanobacteria* and *Spirochaetes*); but the majority are carried by *Firmicutes* representatives, predominantly by *S. pneumoniae* strains (56% of all strains). The second most represented phylum is *Proteobacteria*, including a variety of classes (*Alpha*-, *Beta*-, *Gamma*-, *Delta*-, *Epsilon-proteobacteria*), as seen in figure 10. Notably, AcrIIA7 was present on several viral reference genomes, including all three tailed bacteriophage families (i.e. *Siphoviridae*, *Myoviridae* and *Podoviridae*), reflecting its ubiquitousness. The ACRIIA7 proteins cluster together with *Bacteroidetes* representatives, most likely *B. dorei* being its original host (NCBI EL88_22925), while its closest family member is encoded by a gene derived from a human gut fecal virome sample. The ACRIIA7 proteins belong to the uncharacterized DUF2829 domain superfamily (Helix-turn-helix XRE-family like proteins) as seen in Figure 15.

20 **4.4 Structural characteristics of ACRIIA 9**

AcrIIA9 gene and its homologues are widely distributed across two different families of Bacteroidetes, spanning several *Bacteroides* species and their respective genes in three *Parabacteroides* species encode proteins with 100% sequence identity. Homologues were found in genomes from the recently defined phylum *Balneolaeota*, as well as the phyla *Firmicutes*, *Proteobacteria* and *Actinobacteria* (figure 10). Representatives from the two latter phyla (*Sinorhizobium* sp. GL28 and *C. carbonis* T26, respectively) carry both identical AcrIIA9 homologues, together with identical AcrIIA6 homologues. Indicating that they provide a functional genetic module to counteract type II CRISPR-Cas immune system. The ACRIIA9 proteins are members of the uncharacterized PcfK superfamily; and their members show a high degree of sequence conservation as seen in figure 14.

Example 5 Detetion of ACR activity in human cells

Cells of the human cell line, HEK293T, stably expressing a Cas9 gene from an AAS1 site (supplied by GeneCopoeia, MD, USA), with or without transfection with Alt-R crRNA/tracrRNA (supplied by Integrated DNA technologies, IA, USA) using PolyFect Transfection Reagent (Qiagen, Venlo, Netherlands), were used to observe the effect of Acr proteins on guide Cas9-mediated gene editing of a target locus in the host cell genome. The candidate Acr proteins were transfected into cells of the human cell line,

before crRNA transfection, by using an Xfect protein transfection reagent (Takara, Japan). The cells were then cultured for 24 hours, allowing for Cas9-induced double-strand breaks introduced in the targeted locus on the chromosome of the human cell to be repaired by non-homologous end joining mechanism (NHEJ) leaving insertion or deletion mutation (indels). A T7E1 assay was then used, according to the manufacturer's protocol (New England Biolabs, MA, USA), to detect indels in the targeted locus in the genomic DNA extracted from the human cells. The assay was performed by amplifying the target locus by PCR reaction; denaturing and re-annealing the PCR products to create mismatched DNA region around the target site. T7 endonuclease I was then used to detect and cleave the mismatched DNA region resulting in cleaved product appeared on the agarose gel.

A reduction in CRISPR-Cas9 mediated double-stranded breaks in the host cells was detected using the T7E1 assay in cells transfected with Acr proteins AcrIIA6, AcrIIA7, AcrIIA8, AcrIIA9 and AcrIIA2 (Figure 16).

References

Ausubel F.M. et al. (1987) *Current Protocols in Molecular Biology* ;

Esvelt, K. M., Mali, P., Braff, J. L., Moosburner, M., Yang, S. J., & Church, G. M.

(2013). Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature Methods*, 10(11), 1116–21. <https://doi.org/10.1038/nmeth.2681>

Genee, H. J., Bonde, M. T., Bagger, F. O., Jespersen, J. B., Sommer, M. O. a, Wernersson, R., & Olsen, L. R. (2014). Software-supported USER cloning strategies for site-directed mutagenesis and DNA assembly. *ACS Synthetic Biology*, 4(3), 342–349. <https://doi.org/10.1021/sb500194z>

Genee H J, Bali A P, Petersen SD, Siedler, S, Bonde MT, Gronenberg LS, Kristensen, M and Harrison SJ and MOA Sommer (2016) Functional mining of transporters using synthetic selections. *Nature Chemical Biology* 12, 1015–1022;

Hug, L. A., B.J. Baker, K. Anantharaman, C.T. Brown, A.J. Probst, C.J. Castelle, C.N. Butterfield, A.W. Hermsdorf, Y. Amano, K. Ise, Y. Suzuki, N. Dudek, D.A. Relman, K.M. Finstad, R. Amundson, B.C. Thomas, J.F. Banfield (2016) A new view of the tree of life. *Nat. Microbiol.* 1, 1–6.

Koonin, E. V., Makarova, K. S., & Zhang, F. (2017) Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology*, 37, 67–78.

<https://doi.org/10.1016/j.mib.2017.05.008>

Lutz, R., Bujard, H., 1997. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR / O , the TetR / O and AraC / I 1 -I 2 regulatory elements. *Pharmacia* 25: 1203–1210.

Martínez-García, E., Aparicio, T., Goñi-Moreno, A., Fraile, S., & De Lorenzo, V. (2015). SEVA 2.0: An update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities. *Nucleic Acids Research*, 43(D1), D1183–D1189. <https://doi.org/10.1093/nar/gku1114>

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

Pawluk, A., N. Amrani, Y. Zhang, B. Garcia, Y. Hidalgo-Reyes, J. Lee, A. Edaki, M. Shah, E.J. Sontheimer, K.L. Maxwell, A.R. Davidson (2016) Naturally Occurring Off-Switches for CRISPR-Cas9 Article Naturally Occurring Off-Switches for CRISPR-Cas9.

- 10 Cell. 167, 1829–1834.e9.

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

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

Zhu, W., A. Lomsadze, M. Borodovsky, (2010) Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* 38, 1–15

Claims

1. A nucleic acid molecule encoding an anti-CRISPR protein selected from the group consisting of an:
- 5 a. AcrIIA8 protein,
 b. AcrIIA6 protein,
 c. AcrIIA9 protein and
 d. AcrIIA7 protein.
- 10 2. A self-replicating genetic element comprising the nucleic acid molecule according to claim 1, wherein
- a. the amino acid sequence of said AcrIIA8 protein has at least 80% sequence identify to an amino acid sequence selected from the group consisting of SEQ ID Nos.: 169, 171, 173, 175, 177, 179, 181, 183,
15 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, amd 219;
- b. the amino acid sequence of said AcrIIA6 protein has at least 80% sequence identify to an amino acid sequence selected from the group consisting of SEQ ID Nos.: 221, 223 and 225;
- 20 c. the amino acid sequence of said AcrIIA9 protein has at least 80% sequence identify to an amino acid sequence selected from the group consisting of SEQ ID Nos.: 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, and 247;
- d. the amino acid sequence of said AcrIIA7 protein has at least 80% sequence identify to an amino acid sequence selected from the group consisting of SEQ ID Nos.: 249, 251, 253, 255 and 257,
25 wherein said nucleic acid molecule is operably linked to a promoter.
- 30 3. The self-replicating genetic element according to claim 2; and a CRISPR-Cas expression system comprising:
- a. a first gene encoding a polypeptide or a first gene cluster encoding a polypeptide complex, said polypeptide or polypeptide complex having
35 RNA-guided endonuclease activity, wherein said gene or gene cluster is operably linked to a promoter; and
- b. a nucleic acid molecule operably linked to a promoter, said nucleic acid molecule encoding one or more RNA molecules,

- wherein said CRISPR-Cas expression system is comprised on said self-replicating genetic element or on a further self-replicating genetic element, wherein said one of more RNA molecules are capable of guiding said polypeptide or polypeptide complex having RNA-guided endonuclease activity to a target gene in a cell; and wherein expression of said anti-CRISPR protein is capable of regulating said CRISPR-Cas system when co-expressed in the cell.
- 5
4. The self-replicating genetic element according to claim 3, wherein
- 10
- a. said first gene encodes a Cas9 polypeptide having RNA-guided endonuclease activity, and wherein
- b. one of said RNA molecules is a mat-crRNA and one of said RNA molecules is a cognate trans-activating crRNA, or said one RNA molecule is a fusion of a mat-crRNA and a cognate trans-activating
- 15
- crRNA.
5. A biological cell comprising:
- a. a nucleic acid molecule according to claim 1;
- or
- 20
- b. a self-replicating genetic element according to any one of claims 2 to 4.
6. An isolated anti-CRISPR protein encoded by the nucleic acid molecule according to claim 1 or encoded by the self-replicating genetic element according to claim 2, wherein said protein is selected from the group consisting of:
- 25
- a. AcrIIA8 protein,
- b. AcrIIA6 protein,
- c. AcrIIA9 protein and
- 30
- d. AcrIIA7 protein.
7. Use of a nucleic acid molecule according to claim 1, or an anti-CRISPR protein according to claim 6 to regulate CRISPR-Cas associated activity.
8. The use of a nucleic acid molecule or an anti-CRISPR protein to regulate the CRISPR-Cas associated activity according to claim 7, wherein said activity is provided by a CRISPR-Cas system selected from the group consisting of Type II and Type V and wherein said system comprises:
- 35

- a. a polypeptide or polypeptide complex having RNA-guided endonuclease activity, and
- b. one or more RNA molecules capable of guiding said polypeptide or polypeptide complex having RNA-guided endonuclease activity to a target gene or its transcript,

wherein the amino acid sequence(s) of said polypeptide or each member of said polypeptide complex having RNA-guided endonuclease activity has at least 80% amino acid sequence identity to an amino acid sequence or a corresponding member of a set of amino acid sequences selected from among Group A; and

wherein one of said RNA molecules comprises at least one cognate palindromic repeat sequence and one of said RNA molecules comprises a cognate trans-activating crRNA or one of said RNA molecules comprises a fusion at least one cognate palindromic repeat sequence and one cognate trans-activating crRNA, wherein said repeat sequence is encoded by a nucleotide sequence at a corresponding position in the list of Group B; and wherein said cognate trans-activating crRNA is encoded by a nucleotide sequence at a corresponding position in the list of Group C

	Group A	Group B	Group C
Type II A	SEQ ID No.: 45	SEQ ID No.: 116	SEQ ID No.:134
Type II B	SEQ ID No.: 46	SEQ ID No.: 117	SEQ ID No.:135
Type II C	SEQ ID No.: 47	SEQ ID No.: 118	SEQ ID No.:136
Type V B	SEQ ID No.: 92-94	SEQ ID No.: 126	SEQ ID No.:137
Type V E	SEQ ID No.: 99-102	SEQ ID No.: 129	SEQ ID No.:138.

9. The use of a nucleic acid molecule or an anti-CRISPR protein to regulate the CRISPR-Cas associated activity according to claim 7, wherein said activity is provided by a CRISPR-Cas system selected from the group consisting of Type 1, Type III, Type IV, Type V, and Type VI and wherein said CRISPR-Cas system comprises:

- a. a polypeptide or polypeptide complex having RNA-guided endonuclease activity, and
- b. one or more RNA molecules capable of guiding said polypeptide or polypeptide complex having RNA-guided endonuclease activity to a target gene or its transcript,

wherein the amino acid sequence of each member of said polypeptide complex having RNA-guided endonuclease activity has at least 80% amino acid sequence identity to a corresponding member of a set of amino acid sequences

selected from among Group A;

wherein said one or more RNA molecules comprises at least one cognate palindromic repeat sequence, wherein said repeat sequence is encoded by a nucleotide sequence at a corresponding position in the list of Group B

	Group A	Group B
Type I A	SEQ ID No.: 1-10	SEQ ID No.: 109
Type I B	SEQ ID No.: 12-16	SEQ ID No.: 110
Type I C	SEQ ID No.: 17-23	SEQ ID No.: 111
Type I D	SEQ ID No.: 24-31	SEQ ID No.: 112
Type I E	SEQ ID No.: 32-33	SEQ ID No.: 113
Type I F	SEQ ID No.: 34-39	SEQ ID No.: 114
Type I F*	SEQ ID No.: 40-44	SEQ ID No.: 115
Type III A	SEQ ID No.: 48-56	SEQ ID No.: 119
Type III B	SEQ ID No.: 57-64	SEQ ID No.: 120
Type III B*	SEQ ID No.: 65-73	SEQ ID No.: 121
Type III C	SEQ ID No.: 74-79	SEQ ID No.: 122
Type III D	SEQ ID No.: 80-85	SEQ ID No.: 123
Type IV	SEQ ID No.: 86-90	SEQ ID No.: 124
Type V A	SEQ ID No.: 91	SEQ ID No.: 125
Type V C	SEQ ID No.: 95-96	SEQ ID No.: 127
Type V D	SEQ ID No.: 97-98	SEQ ID No.: 128
Type VI A	SEQ ID No.: 103	SEQ ID No.: 130
Type VI B1	SEQ ID No.: 104-105	SEQ ID No.: 131
Type VI B2	SEQ ID No.: 106-107	SEQ ID No.: 132
Type VI C	SEQ ID No.: 108	SEQ ID No.: 133.

5

- 10.** The use of a nucleic acid molecule or an anti-CRISPR protein to regulate CRISPR-Cas associated activity according to any one of claims 7 to 9, wherein said activity is expressed in a biological cell selected from the group consisting of a bacterial, yeast, fungus, insect, and plant cell.

10

- 11.** A method for *in vivo* regulation of a CRISPR-Cas system, comprising
- a. providing a biological cell comprising a target gene;
 - b. contacting said cell with a molecule, wherein said molecule is selected from the group consisting of:
 - i. an anti-CRISPR protein according to claim 6,

15

- ii. a nucleic acid molecule encoding and capable of expressing an anti-CRISPR protein according to claim 1, and
- iii. a self-replicating genetic element comprising the nucleic acid molecule encoding and capable of expressing an anti-CRISPR protein according to any one of claims 2 to 4,

wherein said cell either comprises or is additionally contacted with a CRISPR-Cas expression system comprising:

- iv. a first gene encoding a polypeptide or a first gene cluster encoding a polypeptide complex, said polypeptide or polypeptide complex having RNA-guided endonuclease activity, wherein said gene or gene cluster is operably linked to a promoter; and
- v. a nucleic acid molecule operably linked to a promoter, said nucleic acid molecule encoding one or more RNA molecules capable of guiding said polypeptide or polypeptide complex having RNA-guided endonuclease activity to said target gene;

wherein expression of said anti-CRISPR protein in said cell is capable of regulating said CRISPR-Cas system.

12. The method for *in vivo* regulation of a CRISPR-Cas system according to claim 11, wherein said polypeptide or polypeptide complex having RNA-guided endonuclease activity is selected from the group of CRISPR-Cas systems consisting of Type 1, Type II, Type III; Type IV, Type V, and Type VI.

13. The method for *in vivo* regulation of a CRISPR-Cas system according to claim 11 or 12, wherein

- i. said first gene encodes a Cas9 polypeptide having RNA-guided endonuclease activity, and wherein
- ii. one of said RNA molecules is a mat-crRNA and one of said RNA molecules is a cognate trans-activating crRNA, or said one RNA molecule is a fusion of a mat-crRNA and a cognate trans-activating crRNA.

14. The method for *in vivo* regulation of a CRISPR-Cas system according to any one of claims 11 to 13, wherein the amino acid sequence(s) of said polypeptide or each member of said polypeptide complex having RNA-guided endonuclease activity has at least 80% amino acid sequence identity to an amino acid sequence or a corresponding member of a set of amino acid sequences

- 5 selected from among Group A; and
 wherein one of said RNA molecules comprises at least one cognate palindromic repeat sequence and one of said RNA molecules comprises a cognate trans-activating crRNA or one of said RNA molecules comprises a fusion at least one cognate palindromic repeat sequence and one cognate trans-activating crRNA, wherein said repeat sequence is encoded by a nucleotide sequence at a corresponding position in the list of Group B; and wherein said cognate trans-activating crRNA is encoded by a nucleotide sequence at a corresponding position in the list of Group C

	Group A	Group B	Group C
Type II A	SEQ ID No.: 45	SEQ ID No.: 116	SEQ ID No.:134
Type II B	SEQ ID No.: 46	SEQ ID No.: 117	SEQ ID No.:135
Type II C	SEQ ID No.: 47	SEQ ID No.: 118	SEQ ID No.:136
Type V B	SEQ ID No.: 92-94	SEQ ID No.: 126	SEQ ID No.:137
Type V E	SEQ ID No.: 99-102	SEQ ID No.: 129	SEQ ID No.:138.

10

- 15 **15.** The method for *in vivo* regulation of a CRISPR-Cas system according to any one of claims 11 to 13, wherein the amino acid sequence of each member of said polypeptide complex having RNA-guided endonuclease activity has at least 80% amino acid sequence identity to a corresponding member of a set of amino acid sequences selected from among Group A;
 wherein said one or more RNA molecules comprises at least one cognate palindromic repeat sequence, wherein said repeat sequence is encoded by a nucleotide sequence at a corresponding position in the list of Group B

	Group A	Group B
Type I A	SEQ ID No.: 1-10	SEQ ID No.: 109
Type I B	SEQ ID No.: 12-16	SEQ ID No.: 110
Type I C	SEQ ID No.: 17-23	SEQ ID No.: 111
Type I D	SEQ ID No.: 24-31	SEQ ID No.: 112
Type I E	SEQ ID No.: 32-33	SEQ ID No.: 113
Type I F	SEQ ID No.: 34-39	SEQ ID No.: 114
Type I F*	SEQ ID No.: 40-44	SEQ ID No.: 115
Type III A	SEQ ID No.: 48-56	SEQ ID No.: 119
Type III B	SEQ ID No.: 57-64	SEQ ID No.: 120
Type III B*	SEQ ID No.: 65-73	SEQ ID No.: 121
Type III C	SEQ ID No.: 74-79	SEQ ID No.: 122
Type III D	SEQ ID No.: 80-85	SEQ ID No.: 123

Type IV	SEQ ID No.: 86-90	SEQ ID No.: 124
Type V A	SEQ ID No.: 91	SEQ ID No.: 125
Type V C	SEQ ID No.: 95-96	SEQ ID No.: 127
Type V D	SEQ ID No.: 97-98	SEQ ID No.: 128
Type VI A	SEQ ID No.: 103	SEQ ID No.: 130
Type VI B1	SEQ ID No.: 104-105	SEQ ID No.: 131
Type VI B2	SEQ ID No.: 106-107	SEQ ID No.: 132
Type VI C	SEQ ID No.: 108	SEQ ID No.: 133.

- 5 **16.** The method for *in vivo* regulation of a CRISPR-Cas system according to any one of claims 11 to 15, wherein biological cell is selected from the group consisting of a bacterial, yeast, fungus, insect, and plant cell.
- 10 **17.** A composition comprising an anti-CRISPR protein according to claim 6, or a nucleic acid molecule encoding and capable of expressing an anti-CRISPR protein according to claim 1, for use as a medicament.
- 15 **18.** The composition according to claim 17 for use as a medicament, wherein said composition is for use in regulating CRISPR-Cas associated activity in a somatic mammalian cell.
- 20 **19.** A pharmaceutical kit comprising the composition according to claim 17 in combination with a CRISPR-Cas expression system for use as a medicament, wherein said expression system comprises:
- 25 a. a first gene encoding a polypeptide or a first gene cluster encoding a polypeptide complex, said polypeptide or polypeptide complex having RNA-guided endonuclease activity, wherein said gene or gene cluster is operably linked to a promoter; and
- b. a nucleic acid molecule operably linked to a promoter, said nucleic acid molecule encoding one or more RNA molecules capable of guiding said polypeptide or polypeptide complex having RNA-guided endonuclease activity to said target gene; and
- wherein said anti-CRISPR protein is for use in regulating CRISPR-Cas associated activity in a somatic mammalian cell.

Figure 1

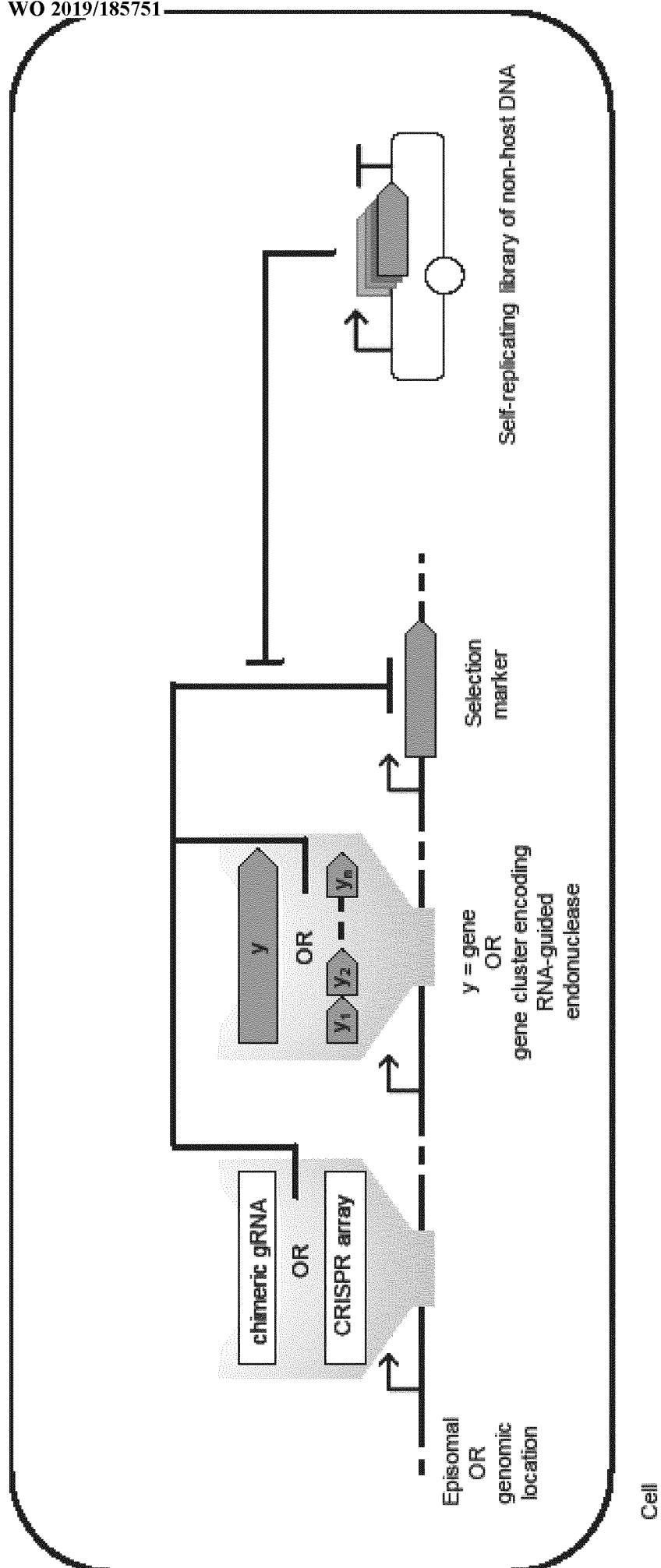


Figure 2 A

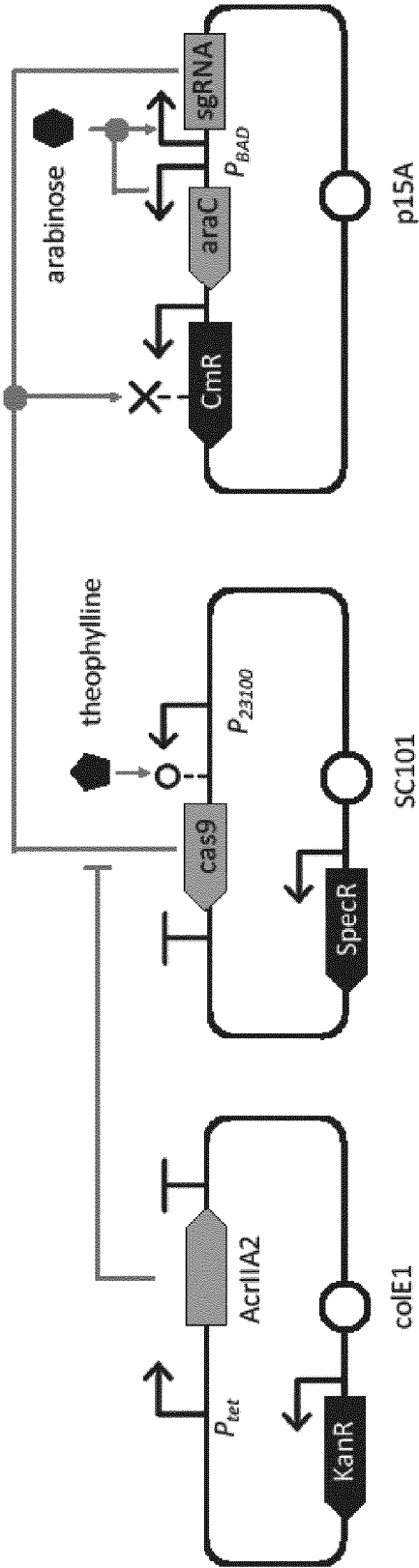


Figure 2B

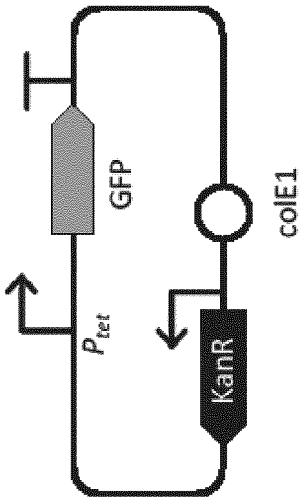
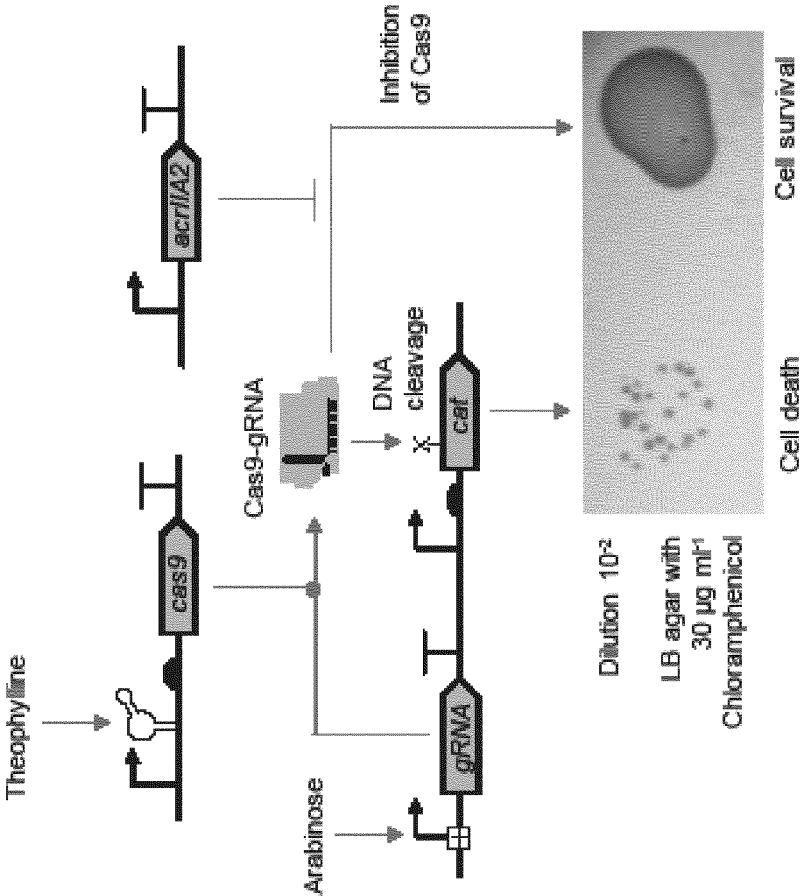


Figure 3

A



B

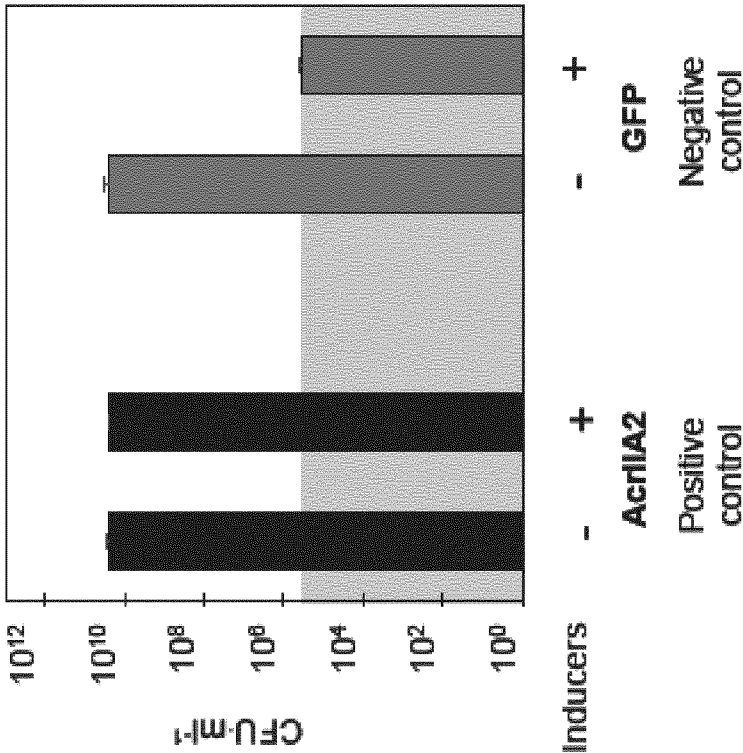


Figure 4

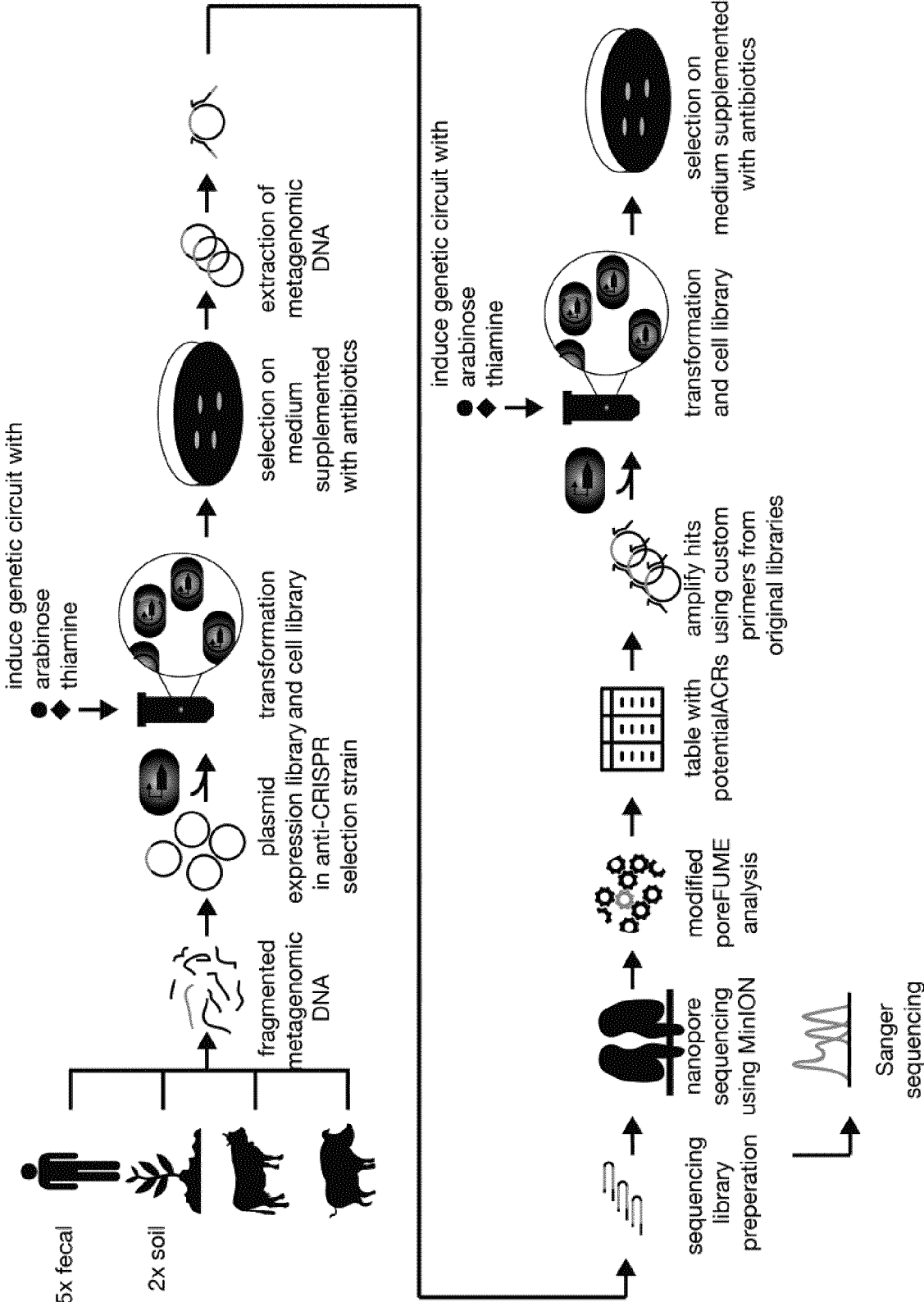


Figure 5

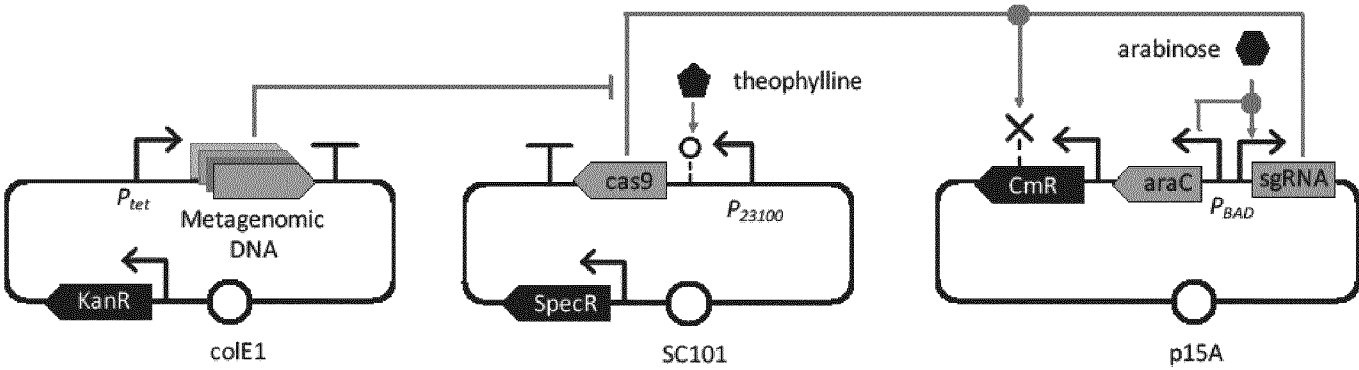


Figure 6

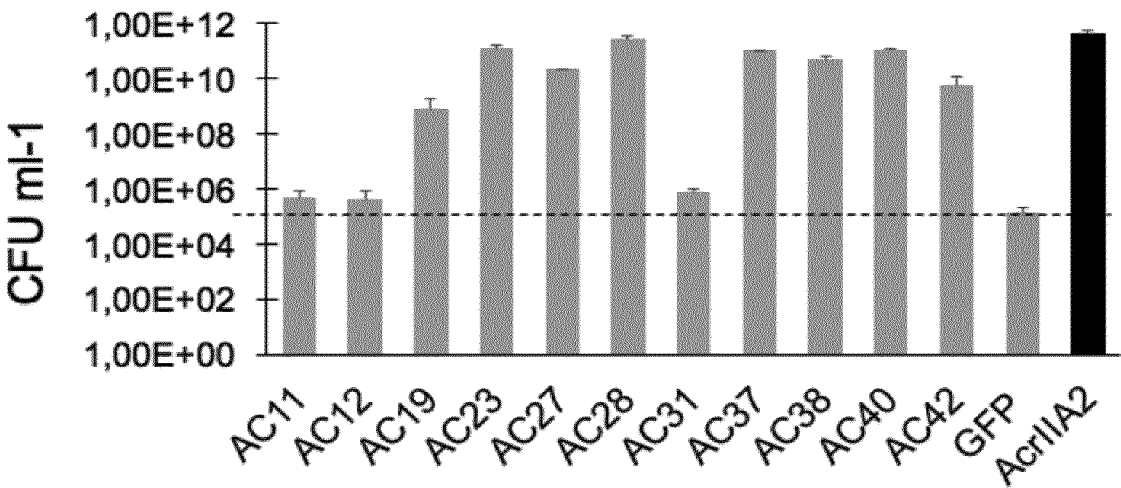


Figure 7

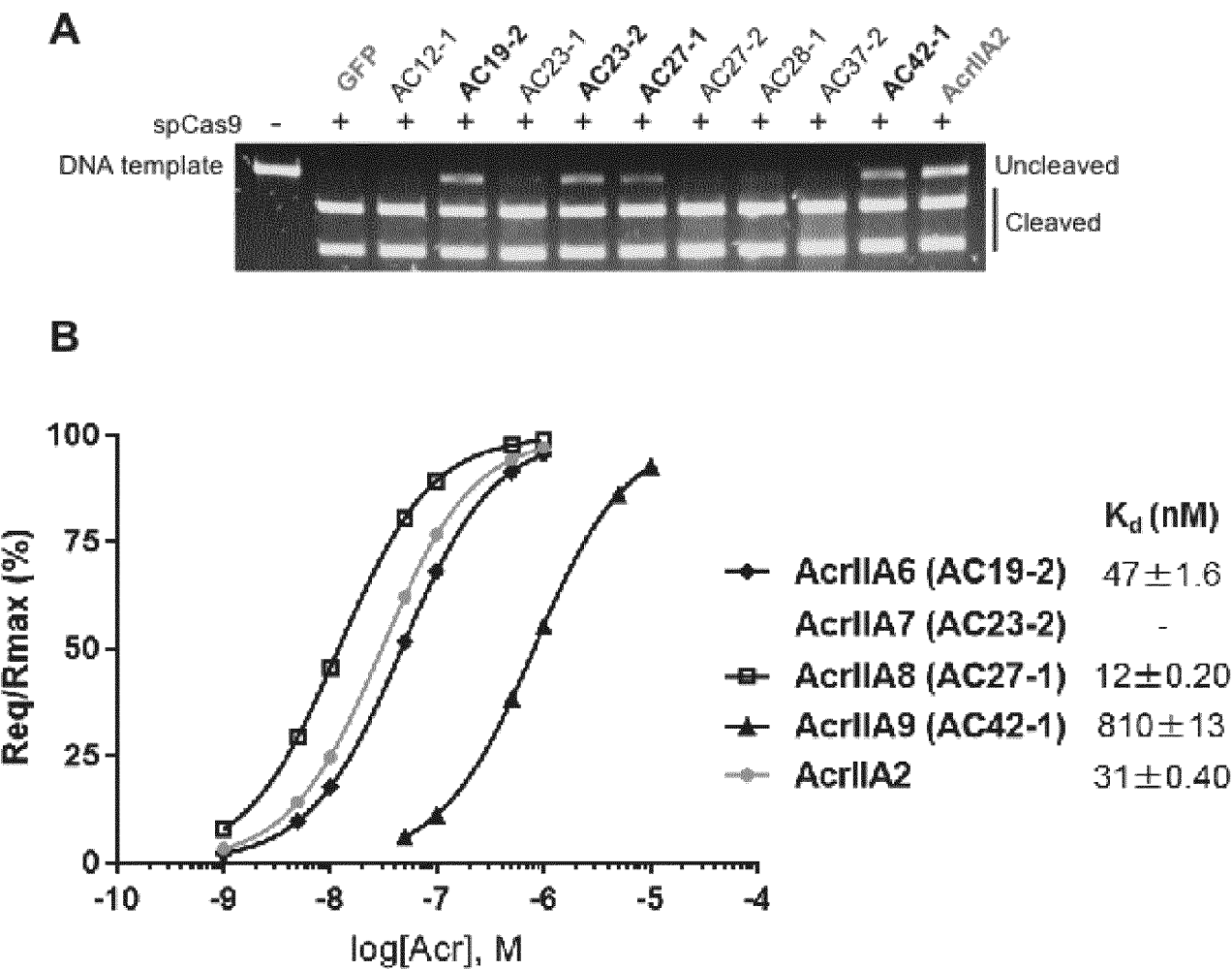


Figure 8

Percent Identity Matrix - created by Clustal2.1

#

	AcrIIA1	AcrIIA5	AcrIIA3	AcrIIA6	AcrIIA7	AcrIIA4	AcrIIA8	AcrIIA2	AcrIIA9
AcrIIA1	100.00	10.00	14.04	18.52	7.84	17.44	9.00	11.76	13.82
AcrIIA5	10.00	100.00	11.30	11.32	11.00	5.95	16.16	8.55	11.45
AcrIIA3	14.04	11.30	100.00	12.50	10.00	14.94	8.08	11.82	16.10
AcrIIA6	18.52	11.32	12.50	100.00	10.99	17.24	8.08	11.54	14.15
AcrIIA7	7.84	11.00	10.00	10.99	100.00	12.00	9.30	8.08	11.88
AcrIIA4	17.44	5.95	14.94	17.24	12.00	100.00	8.05	17.44	12.64
AcrIIA8	9.00	16.16	8.08	8.08	9.30	8.05	100.00	9.28	12.37
AcrIIA2	11.76	8.55	11.82	11.54	8.08	17.44	9.28	100.00	15.00
AcrIIA9	13.82	11.45	16.10	14.15	11.88	12.64	12.37	15.00	100.00

Figure 9

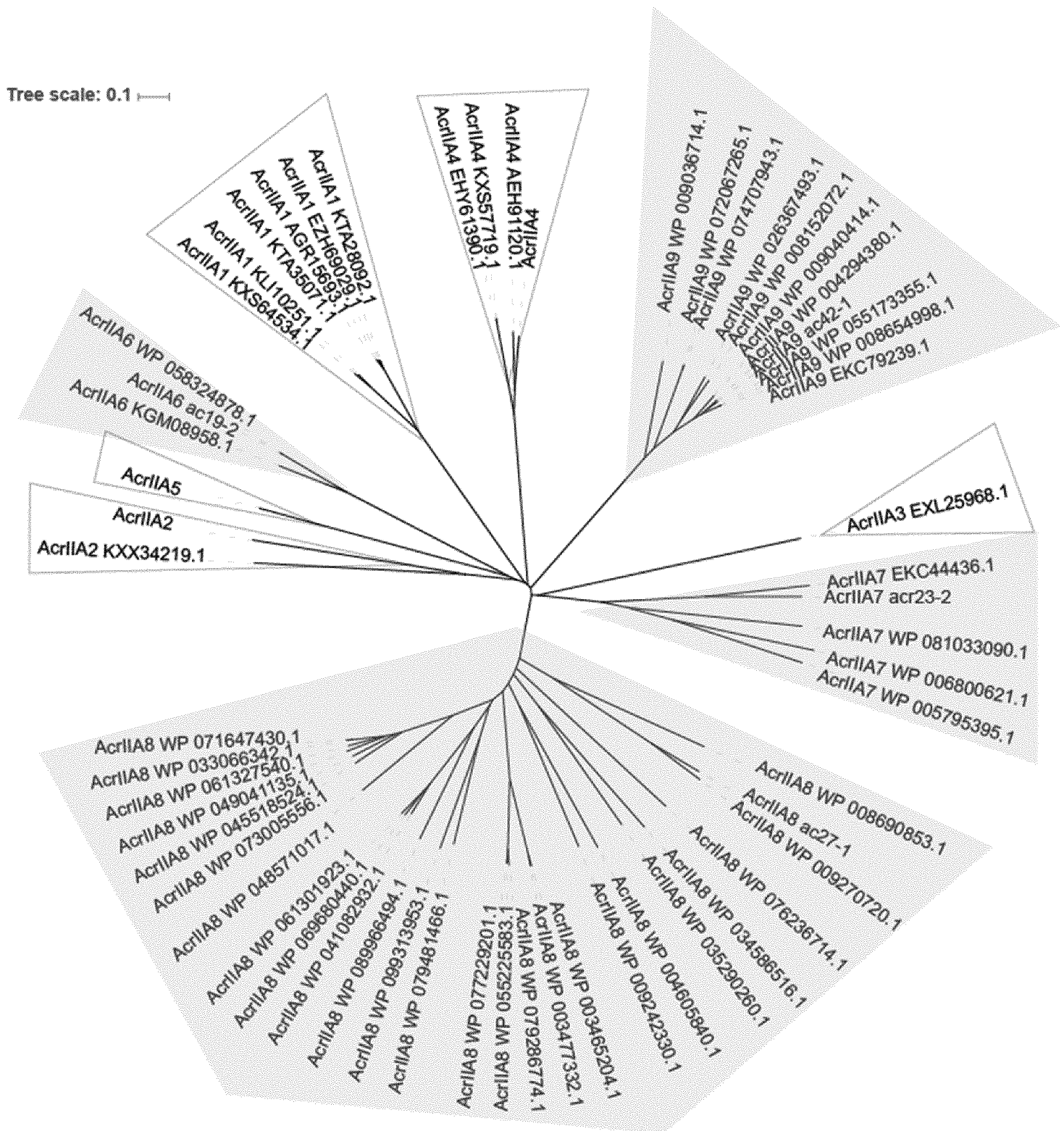


Figure 10 A

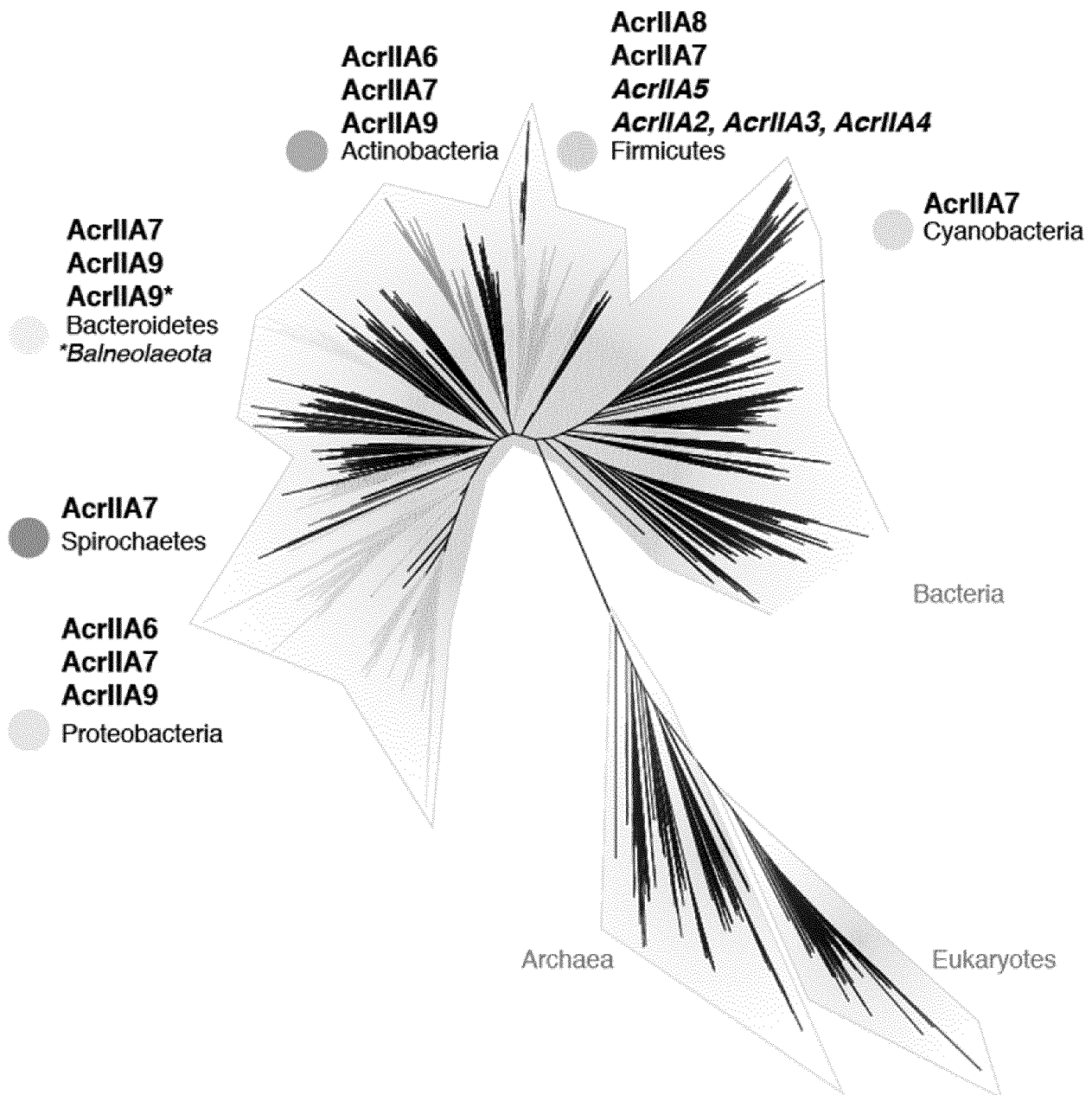
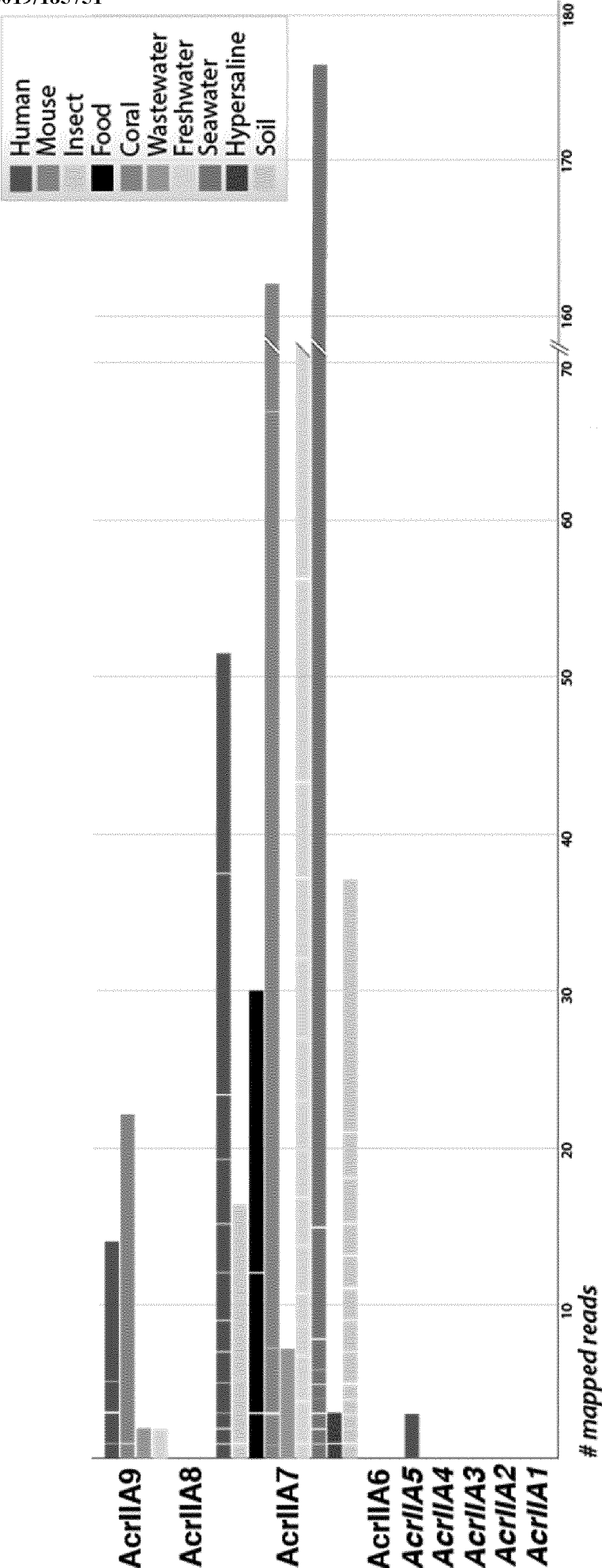
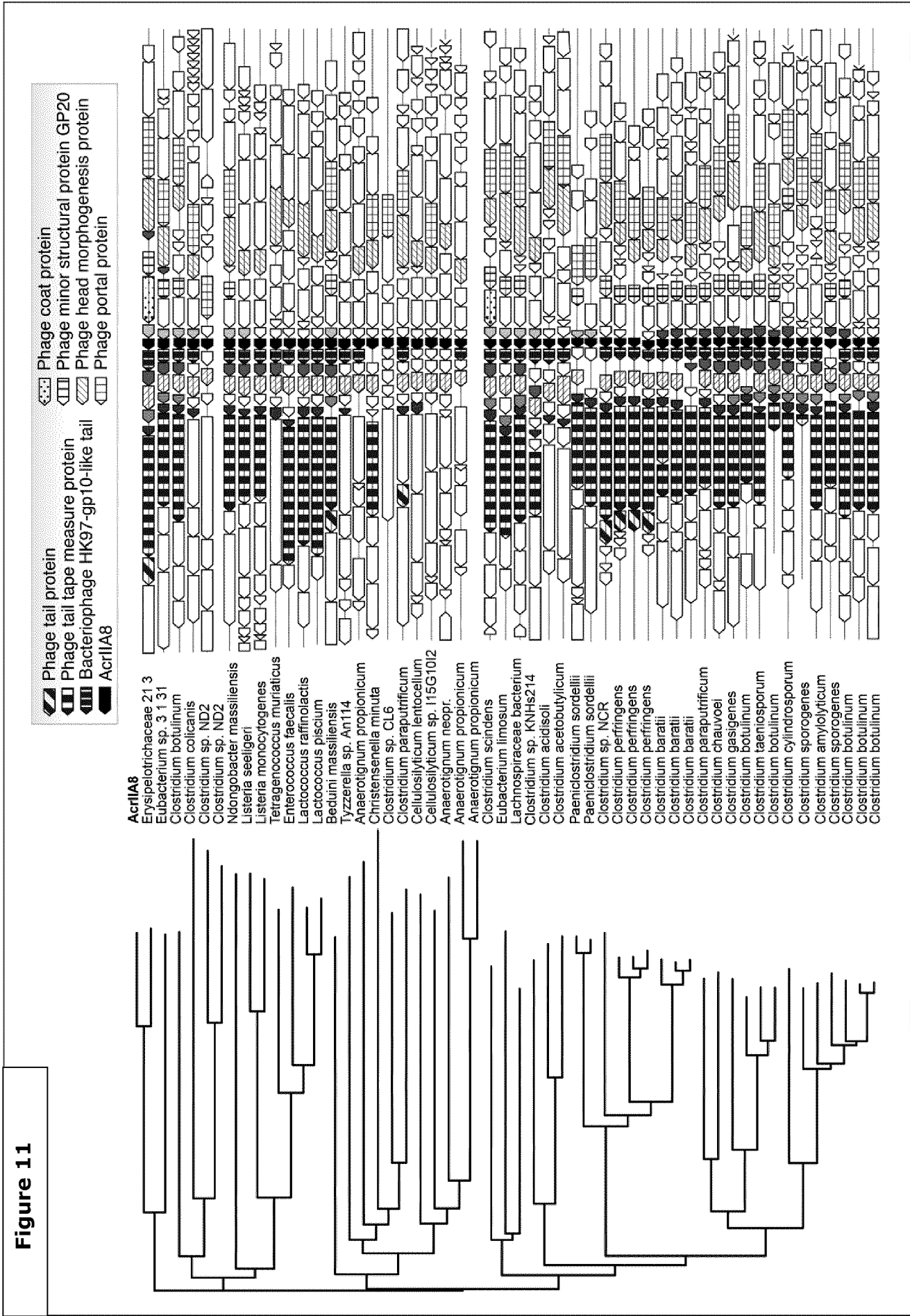


Figure 10 B





AcrIIA8 family alignment

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

gi 495966274 ref WP_008690853.1 AcrlIA8_ac27-1	gi 495966274 ref WP_008690853.1	-MSVFTDFVPAILQINIQEQLSGAPKKTWIDVS--EIEVSIYDNDAFK-S--SQSVKYBE
gi 496569602 ref WP_009270720.1	gi 496569602 ref WP_009270720.1	-MSIFTDMIPAEILLINEYKKGQSGAKHDNYVSVG--RIMVAIYKNNSEFK-N--TGTVKYQ
gi 1133861427 ref WP_076236714.1	gi 1133861427 ref WP_076236714.1	-MSIFTDMLPATLLVNKTKKGKSGAKTDDYVPIG--GIMVAIYKNSSYK-N--INTVKYQ
gi 736575028 ref WP_034586516.1	gi 736575028 ref WP_034586516.1	-MIINREMKQAYLOEKKKIRTQSGATKEDWDKDNQKNIKVAIFSSHEA-NFISDGVKFA
gi 737307379 ref WP_035290260.1	gi 737307379 ref WP_035290260.1	MSCIRRNMKPVTLEKKALVQSSSGQKKETWIDK--TIELAIYVPGSVMN--PNNVLIN
gi 490743532 ref WP_004605840.1	gi 490743532 ref WP_004605840.1	-MIINRNMEKYTLQEKKKVQSKSGQWKETWVDVK--EIDVAIYPANYTILT--SNNIKYS
gi 489560670 ref WP_003465204.1	gi 489560670 ref WP_003465204.1	-MSINRDMRYTIOEKKRVRSPSGGFKPEWVVS--EVDVAIYKIDETV-M--HASELYK
gi 489572886 ref WP_003477332.1	gi 489572886 ref WP_003477332.1	-MSINSNMKSLILOEYRDIITPSMARKKEWVDLEN-PILIAIYETDSRV-N--TNNVRYN
gi 1159736059 ref WP_079286774.1	gi 1159736059 ref WP_079286774.1	-MSINSNMKSLILOENTDIIITPSLARKKEWVDLEN-PILIAIYETDSRV-N--TNNARYN
gi 941889302 ref WP_055225583.1	gi 941889302 ref WP_055225583.1	-MSINGDMKKILLQENREVIKPSGARKKEWIDIEE-PILIAIYETDNRV-N--TSNAIYN
gi 1147451361 ref WP_077229201.1	gi 1147451361 ref WP_077229201.1	-MSINGDMKRILLQENKEVIKPSGARKKEWIDIEE-PILIAIYETDNRV-N--TSNAIYN
gi 496535639 ref WP_009242330.1	gi 496535639 ref WP_009242330.1	-MSINRDMKSYRLQKEETVRTPSGAEKQKIDMG--GVKAAVYKKNDMK-V--AASATYL
gi 873244744 ref WP_048571017.1	gi 873244744 ref WP_048571017.1	-MAIIRDMQPYILEEKVYTRTPSKFLKEVWIEIK--SIDVAIYENDDRI-N--TQSVRFN
gi 1120135380 ref WP_073005556.1	gi 1120135380 ref WP_073005556.1	MPSINRDMYPIILETKVKIRTPSGATKETWDKAI--TIAVAIYDIDDDRI-N--TQSAKFN
gi 895958055 ref WP_049041135.1	gi 895958055 ref WP_049041135.1	-MSINANMEPVLLKETVKTRTSLGATKEKWDTDNPITIQVAIYDIDDDRI-S--TQSVKFN
gi 780209314 ref WP_045518524.1	gi 780209314 ref WP_045518524.1	MPSINRDMKPILLETKIKTRTPSGATKEEKWESN--IIQVAIYDIDDDRI-N--TQSVRFN
gi 1001789565 ref WP_061327540.1	gi 1001789565 ref WP_061327540.1	-MSINSEMKPILLETKVKTRTPSGATKEVWTDNPITIEVAIYDIEDRI-N--TQSVKFN
gi 696532863 ref WP_033066342.1	gi 696532863 ref WP_033066342.1	MPSINRDMKPILLETKVKKRTPSGATKEVWTDNPITIQVAIYDIDDDRI-N--TQSVRFN
gi 1103286903 ref WP_071647430.1	gi 1103286903 ref WP_071647430.1	MASINRDMKPILLETKVKKRTPSGATKEVWGTDPNPITIQVAIYDIDDDRI-N--TQSVRFN
gi 1221872317 ref WP_089966494.1	gi 1221872317 ref WP_089966494.1	-----MKKVKLQVKYIEESPSPAERNEWKDVK--EILVAIYKTNDDL-N--TQSIIRYN
gi 751615000 ref WP_041082932.1	gi 751615000 ref WP_041082932.1	-MSINSNMKPIMLQKKEVKVKSPTGAKKEQWVDVK--VINVAIYKTNDDL-N--TNSTKYN
gi 1069453707 ref WP_069680440.1	gi 1069453707 ref WP_069680440.1	-MSINADMKPIMLQKKKMWKSPSGAKKEQWVDVK--IINVAIYKTNDDL-N--TNSIKYN
gi 1001755700 ref WP_061301923.1	gi 1001755700 ref WP_061301923.1	-MSINSNMKPIMLQKKKGKVKSPSGAPKEQWVDVK--IINVAIFKTNDDL-N--TNSTKYN
gi 1160619369 ref WP_079481466.1	gi 1160619369 ref WP_079481466.1	-MSIISDMNSIMLQVKEKVKKPSGASGEWKDVK--LIDISYHTNDMK-N--TQSVRYN
gi 1271532665 ref WP_099313953.1	gi 1271532665 ref WP_099313953.1	-MSIVSDMKPIMLQRRKKVPTNSGAPKYEWECK--TVDVAIYLVNDMK-N--TQSVRYN

SUBSTITUTE SHEET (RULE 26)

Figure 13

AcrIIA6 family alignment

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

```

AcrIIA6_ac19-2      MDNKFKLKKAINGIEELNFAFDKLTATDYKTTICRIERKMNGLSVDALADSI IASAGTRKT
AcrIIA6_WP_058324878.1 MDNKFKLKKAINGIEELNFAFDKLTATDYKTTICRIERKMNGLSVDALADSI IASAGTRKT
AcrIIA6_KGM08958.1  -----MNGLSVDALADSI IASAGTRKT
                      *****
AcrIIA6_ac19-2      SSEFRIACAWVAAVKGTGDLTVDDYDQLSLDDLELETFGLLFFVGSLE
AcrIIA6_WP_058324878.1 SSEFRIACAWVAAVKGTGDLTVDDYDQLSLDDLELETFGLLFFVGSLE
AcrIIA6_KGM08958.1  SSEFRIACAWVAAVKGTGDLTVDDYDQLSLDDLELETFGLLFFVGSLE
                      *****;*****

```

Figure 14**AcrIIA9 family alignment**

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

***PcfK* superfamily domain**

AcrIIA9_WP_009036714.1	MKGTEQFKQTIKAYLDERAKTDELFAVSYAKENKNLDDCITFVLNQVMAICKEGGCGMTD
AcrIIA9_WP_072067265.1	MKGTEHFKKTIKAYLDNRASDLSLFAVSYAKENKNLDDCITFILNQVKA---SGCCGMTD
AcrIIA9_WP_074707943.1	MKGTEHFKKTIKAYLDNRASDLSLFAVSYAKENKNLDDCITFILNQVKA---NGCCGMTD
AcrIIA9_WP_026367493.1	MKGTEHFKEKTIKRYLDGRAQTDDELFAVSYAKENKNLDDCITFILNQVKA---SGCCGMTD
AcrIIA9_WP_008152072.1	MKGTERFKQTVKEYLDGRAQTDDELFAVSYAKENKNLDDCITFILNQVKA---SGCCGMTD
AcrIIA9_WP_009040414.1	MKGTEHFKEKTIKRYLDKARTDELFAVSYVKENKNLDDCITFILNQVKA---SGCCGMTD
AcrIIA9_WP_008654998.1	MKGTEHFKRRTIKEYLDGKARTDELFAVSYAKENKNLDDCITFILNQVKA---SGCCGMTD
AcrIIA9_EKC79239.1	MKGTEHFKRRTIKEYLDGKARTDELFAVSYAKENKNLDDCITFILNQVKA---SGCCGMTD
AcrIIA9_WP_055173355.1	MKGTEHFKRRTIKEYLDGKARTDELFAVSYAKENKNLDDCITFILNQVKA---SGCCGMTD
AcrIIA9_WP_004294380.1	MKGTEHFQKQTIKEYLDEKARTDKLFAASYAKESKNLDDCITFILNQVKA---SGCCGMTD
AcrIIA9_ac42-1	MKGTEHFQKQTIKEYLDGRAQTDDELFAVSYAKENKNLDDCITFILNQVKA---SGCCGMTD

*****. ** *: * ** * . * ** . ****. ** . ** . *****: ***** * . * *****

AcrIIA9_WP_009036714.1	DEVYSLCTHYDYDEDNIEVGKAVNCGVIVNHRVELTPEEQAEAREKALKAYQDEELRKIQQ
AcrIIA9_WP_072067265.1	DEVYSLATHYDYDEDNIDVGSPISCGVVVNHQIELTAEKEQEARQEAIKAYQAEQMRIMQN
AcrIIA9_WP_074707943.1	DEVYSLATHYDYDEDNIDVGSPISCGVVVNHQIELTAEKEQEARQEAIKAYQAEQMRIMQN
AcrIIA9_WP_026367493.1	DEVWSLAIHYDYDEDNIDVGKSISCGIVVNHKVELTEEEKAQARKDAIKAYQEEEMRKMQQ
AcrIIA9_WP_008152072.1	DEVWSLAIHYDYDEDNIDVGKPIISCGIVVNHKVELTDEEKAQARKDAIKAYQEEEIIRKMQQ
AcrIIA9_WP_009040414.1	EEVWSLAIHYDYDEDNIDVGKPIISCGIVVNHKVELTEEEKAQARKDALKAYQEEEMRKIQQ
AcrIIA9_WP_008654998.1	DEVWSLAIHYDYDEDNIDVGKPIISCGVVVNHKVELTEEEKAQARKDALKAYQEEEMRKIQQ
AcrIIA9_EKC79239.1	DEVWSLAIHYDYDEDNIDVGKPIISCGVVVNHKVELTEEEKAQARKDALKAYQEEEMRKIQQ
AcrIIA9_WP_055173355.1	DEVWSLAIHYDYDEDNIDVGKPIISCGVVVNHKVELTEEEKAQARKDALKAYQEEEMRKIQQ
AcrIIA9_WP_004294380.1	DEVWSLAIHYDYDEDNIDVGNPISCGVVVNHKVELTEEEKAQARKDALKAYQEEEMRKIQQ
AcrIIA9_ac42-1	DEVWSLAIHYDYDEDNIDVGNPISCGVVVNHKVELTEEEKAQARKEALKAYQEEEMRKIQQ

: **: **. *****: **. . . . : **: **. : **: **. : **: **. : **: **: *

AcrIIA9_WP_009036714.1	RNSRPKPTPKAVRQEPEQPSLFD
AcrIIA9_WP_072067265.1	RNNKPKVTKP--QPSTPQPSLFD
AcrIIA9_WP_074707943.1	RNNKPKVTKP--QTSTPQPSLFD
AcrIIA9_WP_026367493.1	RHS--KPAKT-QTNQPEPSLFD
AcrIIA9_WP_008152072.1	RNSRPKPAKT-QTSQSELSLFD
AcrIIA9_WP_009040414.1	RHSRSRPTAQAAQCNTQLSLFD
AcrIIA9_WP_008654998.1	RHSRPKPTAKAVQNTQTELFLD
AcrIIA9_EKC79239.1	RHSRPKPTAKAVQNTQTELFLD
AcrIIA9_WP_055173355.1	RHSRPKPTAKAVQNTQTELFLD
AcrIIA9_WP_004294380.1	RHSPKPTAKAAQSNQTELFLD
AcrIIA9_ac42-1	RHSPKPTAKAAQSNQTELFLD

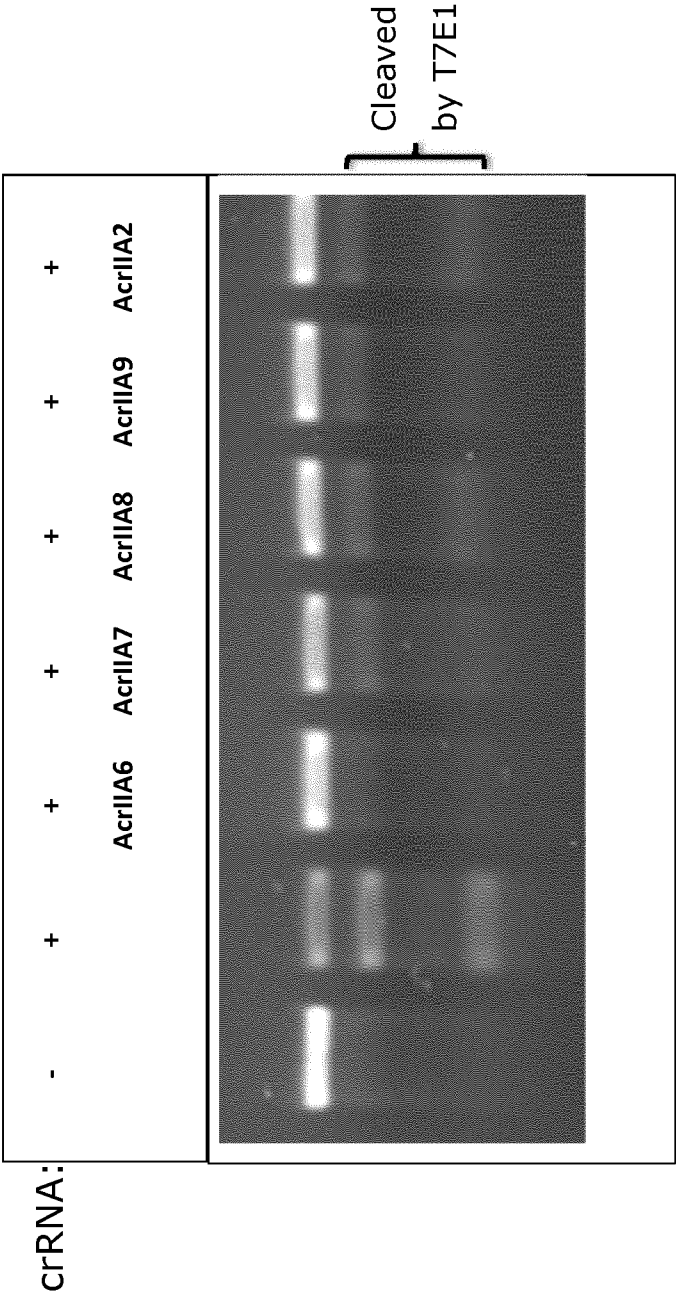
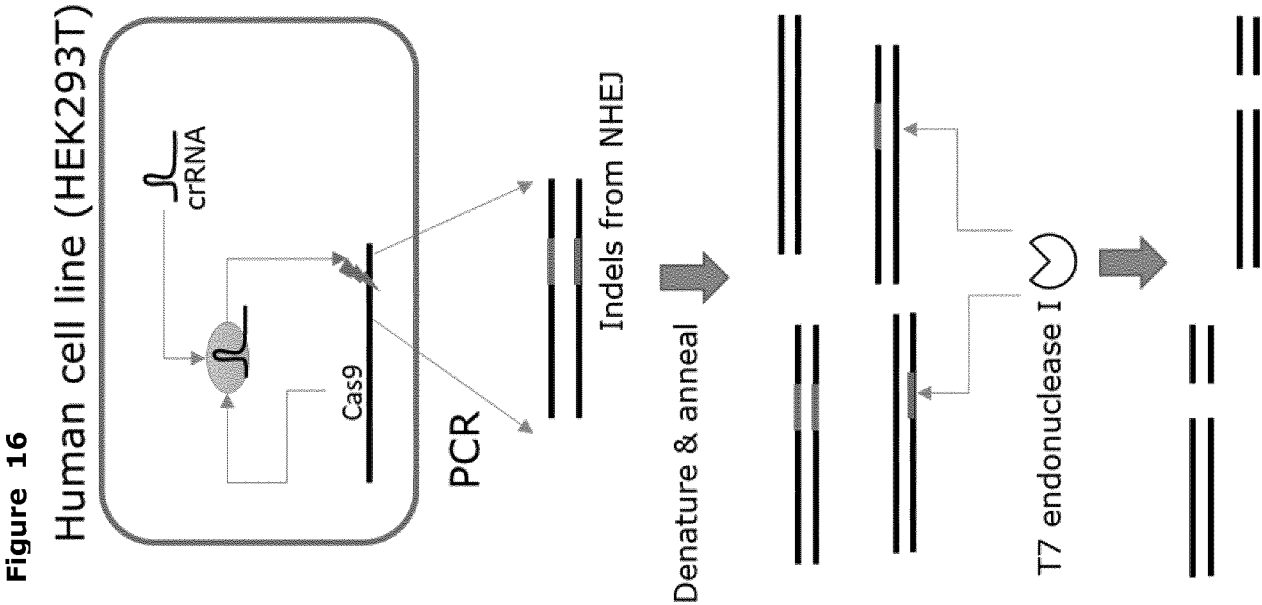
*: . . : . : *****

Figure 15

AcrIIA7 family alignment

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

AcrIIA7_WP_081033090.1	-----MEWICEYEFKKRYRGIMGQ-----MTFGDAIEYLLKK
AcrIIA7_WP_006800621.1	MKTLDKAAEYSAKL--CNQSGNYSKGELETAYVLGANESKYLQYGECGTFGQALEELKQ
AcrIIA7_WP_005795395.1	MKTLDKAAEYAAASVSCNKEACEGLIQTAYILGATTEGELLREETGTFGQALESLKR
AcrIIA7_EKC44436.1	-----MKFGEALEAVKE-----
AcrIIA7_acr23-2	-----MTFGQALESLKR-----
	. ** : * : *
<i>DUF2829 superfamily domain</i>	
AcrIIA7_WP_081033090.1	GCLVTRKGNWGKGMFLFIRPEDTLPLETIVK-AKSLPDAFKDKVLEN-----PN
AcrIIA7_WP_006800621.1	GRLVTRKGNWGKGMFIFMRPADELHVGFVAKDIKSLPQGVKDYHHQDCVDENGKLLDLAK
AcrIIA7_WP_005795395.1	GHLVTRKGNWGKGMFIFMRPADELHIGFVAKDIKSLPQKVKDYYYQDCVDENGNPIELEK
AcrIIA7_EKC44436.1	GKLIARSGWNGKGMFVFQRPEDWLSTDMIVNKVKSLPDSFKKYVND-----
AcrIIA7_acr23-2	GHLVARKGNWGKGMFIFMRPEDSLPTNMIVNQVKSLPESFKRWVANNHGD-----SE
	* * : : . * * * * * : : : * * * * : . * :
AcrIIA7_WP_081033090.1	TESVKFGAYICMKCADGSICNGWLASQTDMLSNNDWMLVLTKDHE
AcrIIA7_WP_006800621.1	EDVVKFTAYLCMKKAADGTIVNGWLASQTDMLSEDWMIFFK-----
AcrIIA7_WP_005795395.1	DDTVKFTAYICMKKAADGSIVNGWLASQTDMLANDWMIFFEF----
AcrIIA7_EKC44436.1	-----
AcrIIA7_acr23-2	TDRIKFTAYLCMKKAADGTIVNGWLASQTDMLANDWVIVE-----



INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2019/057786

A. CLASSIFICATION OF SUBJECT MATTER
INV. C07K14/33 C12N15/10 C12N15/90
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, Sequence Search, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MAXWELL KAREN L: "The Anti-CRISPR Story: A Battle for Survival", MOLECULAR CELL, vol. 68, no. 1, 5 October 2017 (2017-10-05), pages 8-14, XP085207631, ISSN: 1097-2765, DOI: 10.1016/J.MOLCEL.2017.09.002 ----- -/--	1-19



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

30 April 2019

Date of mailing of the international search report

08/07/2019

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Bradbrook, Derek

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2019/057786

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-19(partially)

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2019/057786

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ADAIR L. BORGES ET AL: "The Discovery, Mechanisms, and Evolutionary Impact of Anti-CRISPRs", ANNUAL REVIEW OF VIROLOGY, vol. 4, no. 1, 29 September 2017 (2017-09-29), pages 37-59, XP055483877, USA ISSN: 2327-056X, DOI: 10.1146/annurev-virology-101416-041616 -----	1-19
A	RAUCH BENJAMIN J ET AL: "Inhibition of CRISPR-Cas9 with Bacteriophage Proteins", CELL, CELL PRESS, AMSTERDAM, NL, vol. 168, no. 1, 29 December 2016 (2016-12-29), page 150, XP029882136, ISSN: 0092-8674, DOI: 10.1016/J.CELL.2016.12.009 -----	1-19
A	PAWLUK APRIL ET AL: "Naturally Occurring Off-Switches for CRISPR-Cas9", CELL, CELL PRESS, AMSTERDAM, NL, vol. 167, no. 7, 8 December 2016 (2016-12-08), page 1829, XP029850707, ISSN: 0092-8674, DOI: 10.1016/J.CELL.2016.11.017 -----	1-19
X	DATABASE UniProt [Online] 24 July 2013 (2013-07-24), "SubName: Full=Uncharacterized protein {ECO:0000313 EMBL:CD86037.1}";", XP002782243, retrieved from EBI accession no. UNIPROT:R6VUW4 Database accession no. R6VUW4 sequence -----	1,2,5,6
X	DATABASE UniProt [Online] 21 March 2012 (2012-03-21), "SubName: Full=Uncharacterized protein {ECO:0000313 EMBL:EH080923.1}";", XP002782244, retrieved from EBI accession no. UNIPROT:H1BQC5 Database accession no. H1BQC5 sequence ----- -/--	1,2,5,6

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2019/057786

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE UniProt [Online] 7 September 2016 (2016-09-07), "SubName: Full=Uncharacterized protein {ECO:0000313 EMBL:KEI89326.1}";", XP002782245, retrieved from EBI accession no. UNIPROT:A0A175M4B1 Database accession no. A0A175M4B1 sequence -----	1,2,5,6
X	DATABASE EMBL-WGS [Online] EBI; 28 November 2014 (2014-11-28), "Clostridium acetobutylicum hypothetical protein", XP002782246, Database accession no. KHD34419 the whole document -----	1,2,5,6
X	DATABASE PROTEIN [Online] 23 December 2014 (2014-12-23), "hypothetical protein [Clostridium sp. KNHs214]", XP002782247, retrieved from NCBI Database accession no. WP_035290260 the whole document -----	1,2,5,6
X	DATABASE UniProt [Online] 8 April 2008 (2008-04-08), "SubName: Full=Uncharacterized protein {ECO:0000313 EMBL:EDS08511.1}";", XP002782388, retrieved from EBI accession no. UNIPROT:B0NA18 Database accession no. B0NA18 sequence -----	1,2,5,6
X	DATABASE UniProt [Online] 29 April 2008 (2008-04-29), "SubName: Full=Uncharacterized protein {ECO:0000313 EMBL:EDT14069.1}";", XP002782389, retrieved from EBI accession no. UNIPROT:B1BW36 Database accession no. B1BW36 sequence ----- -/--	1,2,5,6

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2019/057786

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL [Online] 28 October 2016 (2016-10-28), "Clostridium perfringens hypothetical protein", XP002782390, retrieved from EBI Database accession no. A0Y53692 sequence -----	1,2,5,6
X	DATABASE PROTEIN [Online] 16 March 2017 (2017-03-16), "hypothetical protein [Clostridium baratii]", XP002782391, retrieved from NCBI Database accession no. WP_079286774 the whole document -----	1,2,5,6
X	DATABASE PROTEIN [Online] 26 October 2015 (2015-10-26), "hypothetical protein [Clostridium baratii]", XP002782392, retrieved from NCBI Database accession no. WP_055225583 the whole document -----	1,2,5,6
X	DATABASE PROTEIN [Online] 15 February 2017 (2017-02-15), "hypothetical protein [Clostridium baratii]", XP002782393, retrieved from NCBI Database accession no. WP_077229201 the whole document -----	1,2,5,6
X	DATABASE UniProt [Online] 8 March 2011 (2011-03-08), "SubName: Full=Uncharacterized protein {ECO:0000313 EMBL:EFV20804.1}";", XP002782394, retrieved from EBI accession no. UNIPROT:E5XD46 Database accession no. E5XD46 sequence ----- -/--	1,2,5,6

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2019/057786

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE UniProt [Online] 14 October 2015 (2015-10-14), "SubName: Full=Uncharacterized protein {ECO:0000313 EMBL:KMT21593.1}";", XP002782395, retrieved from EBI accession no. UNIPROT:A0A0J8DBC9 Database accession no. A0A0J8DBC9 sequence -----	1,2,5,6
X	DATABASE UniProt [Online] 15 March 2017 (2017-03-15), "SubName: Full=Uncharacterized protein {ECO:0000313 EMBL:SHI90130.1}";", XP002782396, retrieved from EBI accession no. UNIPROT:A0A1M6EXJ7 Database accession no. A0A1M6EXJ7 sequence -----	1,2,5,6
X	DATABASE PROTEIN [Online] 21 July 2016 (2016-07-21), "hypothetical protein [Clostridium sporogenes]", XP002782397, retrieved from NCBI Database accession no. WP_049041135 the whole document -----	1,2,5,6
X	DATABASE PROTEIN [Online] 24 March 2015 (2015-03-24), "hypothetical protein [Clostridium sporogenes]", XP002782398, retrieved from NCBI Database accession no. WP_045518524 the whole document -----	1,2,5,6
X	DATABASE PROTEIN [Online] 27 February 2016 (2016-02-27), "hypothetical protein [Clostridium botulinum]", XP002782399, retrieved from NCBI Database accession no. WP_061327540 the whole document ----- -/--	1,2,5,6

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2019/057786

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE PROTEIN [Online] 13 October 2014 (2014-10-13), "hypothetical protein [Clostridium botulinum]", XP002782400, retrieved from NCBI Database accession no. WP_033066342 the whole document -----	1,2,5,6
X	DATABASE PROTEIN [Online] 11 January 2018 (2018-01-11), "hypothetical protein [Clostridium botulinum]", XP002782401, retrieved from NCBI Database accession no. WP_071647430 the whole document -----	1,2,5,6
X	DATABASE PROTEIN 28 July 2017 (2017-07-28), "hypothetical protein [Clostridium gasigenes]", XP002782402, retrieved from NCBI Database accession no. WP_089966494 the whole document -----	1,2,5,6
X	DATABASE PROTEIN [Online] 30 January 2015 (2015-01-30), "phage head-tail adapter protein [Clostridium botulinum]", XP002782403, retrieved from NCBI Database accession no. WP_041082932 the whole document -----	1,2,5,6
X	DATABASE UniProt [Online] 18 January 2017 (2017-01-18), "SubName: Full=Phage head-tail adapter protein {ECO:0000313 EMBL:A0R24305.1}";", XP002782404, retrieved from EBI accession no. UNIPROT:A0A1D7XLS5 Database accession no. A0A1D7XLS5 sequence ----- -/--	1,2,5,6

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2019/057786

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE PROTEIN [Online] 27 January 2016 (2016-01-27), "hypothetical protein [Clostridium botulinum]", XP002782405, retrieved from NCBI Database accession no. WP_061301923 the whole document -----	1,2,5,6
X	DATABASE UniProt [Online] 10 May 2017 (2017-05-10), "SubName: Full=Uncharacterized protein {ECO:0000313 EMBL:SLK19661.1}";", XP002782406, retrieved from EBI accession no. UNIPROT:A0A1U6JH80 Database accession no. A0A1U6JH80 sequence -----	1,2,5,6
X	DATABASE PROTEIN [Online] 29 October 2017 (2017-10-29), "phage head-tail adapter protein [Clostridium paraputrificum]", XP002782407, retrieved from NCBI Database accession no. WP_099313953 the whole document -----	1,2,5,6
A	WO 2017/160689 A1 (UNIV MASSACHUSETTS [US]) 21 September 2017 (2017-09-21) the whole document -----	1-19
A,P	WO 2018/093990 A1 (UNIV CALIFORNIA [US]) 24 May 2018 (2018-05-24) the whole document -----	1-19

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2019/057786

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2017160689 A1	21-09-2017	CA 3029735 A1	21-09-2017
		CN 109152848 A	04-01-2019
		EP 3429635 A1	23-01-2019
		WO 2017160689 A1	21-09-2017

WO 2018093990 A1	24-05-2018	CA 3042880 A1	24-05-2018
		WO 2018093990 A1	24-05-2018

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-19(partially)

a nucleic acid molecule encoding an anti-CRISPR protein having the sequence of SEQ ID NO: 169, 171, ..., 219 (AcrIIA8), self-replicating element comprising the nucleic acid molecule, biological cell, isolated anti-CRISPR protein, use to regulate CRISPR-Cas associated activity, method for in vivo regulation of a CRISPR-Cas system, medical uses

2. claims: 1-19(partially)

a nucleic acid molecule encoding an anti-CRISPR protein having the sequence of SEQ ID NO: 221, 223, 225 (AcrIIA6), self-replicating element comprising the nucleic acid molecule, biological cell, isolated anti-CRISPR protein, use to regulate CRISPR-Cas associated activity, method for in vivo regulation of a CRISPR-Cas system, medical uses

3. claims: 1-19(partially)

a nucleic acid molecule encoding an anti-CRISPR protein having the sequence of SEQ ID NO: 227, 229,...,247 (AcrIIA9), self-replicating element comprising the nucleic acid molecule, biological cell, isolated anti-CRISPR protein, use to regulate CRISPR-Cas associated activity, method for in vivo regulation of a CRISPR-Cas system, medical uses

4. claims: 1-19(partially)

a nucleic acid molecule encoding an anti-CRISPR protein having the sequence of SEQ ID NO: 249, 251, 253, 255, 257 (AcrIIA7), self-replicating element comprising the nucleic acid molecule, biological cell, isolated anti-CRISPR protein, use to regulate CRISPR-Cas associated activity, method for in vivo regulation of a CRISPR-Cas system, medical uses
