CRISPR-Cas Systems in Multicellular Cyanobacteria

Shengwei Hou[1], Manuel Brenes-Álvarez[2], Viktoria Reimann[1], Omer S. Alkhnbashi[3], Rolf Backofen[3,4,5], Alicia M. Muro-Pastor[2] and Wolfgang R. Hess[1,6*]

[1]University of Freiburg, Faculty of Biology, Genetics and Experimental Bioinformatics, Schänzlestr. 1, D-79104 Freiburg, Germany;

[2]Instituto de Bioquímica Vegetal y Fotosíntesis, Consejo Superior de Investigaciones Científicas and Universidad de Sevilla, E-41092 Seville, Spain;

[3]Bioinformatics group, Department of Computer Science, University of Freiburg, Georges-Köhler-Allee 106, 79110 Freiburg, Germany;

[4]Center for Biological Systems Analysis (ZBSA), University of Freiburg, Habsburgerstr. 49, D-79104 Freiburg, Germany;

[5]BIOSS Centre for Biological Signaling Studies, University of Freiburg, Schänzlestr. 18, D-79104 Freiburg, Germany

[6]Freiburg Institute for Advanced Studies, University of Freiburg, Albertstr. 19, D-79104 Freiburg, Germany

***Corresponding author:** Wolfgang R. Hess, Phone: +49-761-203-2796; FAX: +49-761-203-2745; Email: wolfgang.hess@biologie.uni-freiburg.de;

Running head: CRISPR-Cas Systems in Multicellular Cyanobacteria

Keywords: CRISPR, cyanobacteria, heterocyst, nitrogen fixation, programmed DNA recombination

The authors declare the absence of conflicts of interest.

Abstract

Novel CRISPR-Cas systems possess substantial potential for genome editing and manipulation of gene expression. The types and numbers of CRISPR-Cas systems vary substantially between different organisms. Some filamentous cyanobacteria harbor >40 different putative CRISPR repeat-spacer cassettes, while the number of *cas* gene instances is much lower. Here we addressed the types and diversity of CRISPR-Cas systems and of CRISPR-like repeat-spacer arrays in 171 publicly available genomes of multicellular cyanobacteria. The number of 1328 repeat-spacer arrays exceeded the total of 391 encoded Cas1 proteins suggesting a tendency for fragmentation or the involvement of alternative adaptation factors. The model cyanobacterium *Anabaena* sp. PCC 7120 contains only three *cas1* genes but hosts at least three Class 1, one Class 2 and five orphan repeat-spacer arrays, all of which exhibit crRNA-typical expression patterns suggesting active transcription, maturation and incorporation into CRISPR complexes. The CRISPR-Cas system within the element interrupting the *Anabaena* sp. PCC 7120 *fdxN* gene, as well as analogous arrangements in other strains, occupy the genetic elements that become excised during the differentiation-related programmed site-specific recombination. This fact indicates the propensity of these elements for the integration of CRISPR-*cas* systems and points to a previously not recognized connection. The possible Class 2 effector protein gene *all3613* is linked to a short repeat-spacer array and a single tRNA gene, similar to its homologs in other cyanobacteria. The diversity, high number and presence of CRISPR-Cas systems in DNA elements that are programmed for homologous recombination make filamentous cyanobacteria a prolific resource for their study.

Introduction

Genetic tools based on CRISPR-Cas technology are currently the most popular technology for the manipulation of gene expression and genome editing. In most of these approaches, the CRISPR-Cas Type II enzyme Cas9 is used together with a guide RNA to target specific regions in chromosomal DNA [1]. In addition,the large potential exists for alternative CRISPR-Cas systems and novel applications, e.g., in the markerless generation of point mutations using Type I and Type III CRISPR-Cas systems for genome editing [2] or the use of the Type V-A Cas12a (previously known as Cpf1) for the rapid engineering of markerless knock-ins, knock-outs and point mutations [3,4]. Such facts underline that the search for additional types of CRISPR-Cas modules can lead to productive innovation.

Currently, six major types of CRISPR-Cas systems are known, which belong to two major classes and can be further subdivided into multiple subtypes [5]. The functions of the diverse genes and gene products involved in these systems can be classified into three primary functions: adaptation, processing and interference [6]. During adaptation, CRISPR-associated (Cas) proteins excise the protospacer sequence from an invader DNA directly or after reverse transcription of RNA into cDNA [7] and insert it into the first repeat of the CRISPR locus. The CRISPR RNAs (crRNAs) are transcribed from the repeat-spacer array in the form of a long precursor (pre-crRNA) that is processed into the individual crRNAs each consisting of a single spacer sequence and part of the adjoining repeat sequences. During the interference stage, sequences on either invading DNA elements or their transcripts become recognized by crRNAs as guides for the Cas protein complexes that cleave the targeted nucleic acid.

CRISPR-Cas systems have been classified into two classes with regard to the complexity of the effector ribonucleoprotein complexes. Class 1 systems consist of several different subunits, whereas Class 2 systems utilize a single modularized large protein, such as Cas9, Cas12a or Cas13a [5]. Proteins implicated in adaptation are the endonuclease Cas1, Cas2 and, in some systems, Cas4, which facilitates the integration of PAM-compatible spacers [8,9]. The PAM (protospacer adjacent motif) is a short sequence motif in the target DNA that flanks the crRNA-DNA duplex and is crucial for avoiding self-cleavage [10,11]. Despite the impressive general variation in gene content and sequence diversity among different types of CRISPR-Cas systems, all systems have been assumed until very recently to possess a single Cas1 protein, which is less diverse than other Cas proteins and therefore has served as a marker for CRISPR loci. However, this notion has been challenged by recent observations of C2c (Class 2 candidate) systems lacking the *cas1* gene, as they apparently only contain a CRISPR array and single gene encoding a large protein with no sequence similarity to Cas12a, Cas12b, Cas13a, or Cas9 [5]. Additionally, it was speculated that these systems might rely on an adaptation module (*cas1*-*cas2*) encoded elsewhere in the genome [12,13]. Therefore, the detection of putative novel CRISPR systems is not trivial: the numbers and types of CRISPR-Cas systems vary greatly, even between closely related strains, the similarity between Cas proteins can be very remote, and the existence of direct sequence repeats may also relate to different (non-CRISPR) genetic elements.

Cyanobacteria are the only bacteria that perform oxygenic photosynthesis. They occur in widely different environments as long as there is at least some light. Many cyanobacteria are also able to convert atmospheric nitrogen, N_2, into organic biomass, hence sustaining a diazotrophic lifestyle. This process, called N_2 fixation, is catalyzed by nitrogenase, an enzyme that can be irreversibly damaged by oxygen [14]. The need to protect the oxygen-sensitive nitrogenase from photosynthetically produced oxygen has probably driven the evolution of heterocysts, a type of differentiated cells providing a microoxic environment compatible with N_2 fixation in some filamentous cyanobacteria. The evolution of this specialized cell type has driven the division of cellular functions and processes between heterocysts and the other, vegetative, cells along the filaments [15]. Heterocysts transfer fixed nitrogen to the neighboring vegetative cells whereas vegetative cells provide heterocysts with photosynthetically fixed carbon in return. Hence, heterocyst-forming cyanobacteria are true multicellular organisms. Heterocysts are formed in a complex differentiation process that includes the programmed site-specific deletion of large genetic elements that interrupt the reading frames of critical genes by homologous recombination between direct repeats [15–17].

It has been suggested that the Cas9 effector proteins of Class 2 CRISPR-Cas systems evolved from a type of TnpB-like transposase with an HNH nuclease insert that is particularly abundant in cyanobacteria [12]. Type I and Type III (Class 1) CRISPR-Cas systems are frequent in cyanobacteria [18]; however, no proteins with sequence similarity to the hitherto characterized Class 2 effectors such as Cas12a, Cas12b, Cas13a or Cas9 have been identified thus far. Therefore, novel Class 2 CRISPR systems could exist in some cyanobacteria, a view consistent with multiple instances of CRISPR-Cas candidate systems classified as subtype V-U [5,13].

The genomes of multicellular cyanobacteria are complex (up to 12.29 Mb, >10,000 annotated genes) and rich in the number of transposable elements and transposase genes including some encoding TnpB-like transposases. Therefore, we scanned 171 publicly available genomes of multicellular cyanobacteria for the presence of

CRISPR-like repeat-spacer cassettes (**Table S1**). We report a high number of CRISPR-Cas candidate systems, including some with likely Class 2 effector proteins that are associated with a repeat-spacer array that is almost invariably adjacent to a tRNA gene. We then focus on *Anabaena* (*Nostoc)* sp. PCC 7120 (from here: *Anabaena* 7120) in greater detail and demonstrate crRNA-typical expression patterns for three Class 1, one Class 2 and five orphan repeat-spacer arrays in this the well-established model for filamentous cyanobacteria.

Results

CRISPR-Cas systems are frequent in multicellular cyanobacteria

When searching for arrangements of direct repeats that match the criteria of CRISPR repeat-spacer arrays with relaxed parameters, some cyanobacteria, e.g., *Tolypothrix bouteillei* VB521301, host many, up to 44 widely different CRISPR-Cas systems (**Table 1**, see **Table S1** for the full results). Together with Cas2, the Cas1 DNA-specific endonuclease makes up the core machinery of the CRISPR adaptation process. Therefore Cas1 is, with very few exceptions [5], almost universally conserved among different types of CRISPR systems. The here studied genomes of 171 filamentous cyanobacteria contain altogether 391 *cas1* genes. Among the 391 deduced Cas1 proteins are 40 which are fused to a reverse transcriptase (RT) domain and 31 which are fused to a Cas4 protein domain. Such gene fusions are in line with findings that Cas4 promotes the integration of spacers from invading DNA with the correct PAM [8,9] and that the fused RT domains facilitate the direct CRISPR spacer acquisition from RNA [7]. Phylogenetic analysis of our set of Cas1 proteins yielded seven distinct clusters (**Figure 1** and **Figure S1**). Three of these clusters consist of free-standing Cas1 sequences, whereas two clusters each contain Cas1-Cas4 or Cas1-RT fusions. The three clusters of Cas1 proteins encoded by free-standing genes each contain one of the three Cas1 proteins of *Synechocystis* sp. PCC 6803. We included them in this analysis because the unicellular *Synechocystis* sp. PCC 6803 is the best-studied model for CRISPR-Cas systems in cyanobacteria [8,19–23]. The clustering of the three *Synechocystis* sp. PCC 6803 Cas1 sequences suggests that they match well to the three major groups of Cas1 proteins lacking other fused domains of filamentous cyanobacteria. The set of 391 Cas1 proteins is available in **Supplemental Dataset 1**.

We noticed a striking discrepancy in the number of *cas1* genes and the number of repeat-spacer arrays. There are 5 *cas1* genes and 44 repeat-spacer instances in *Tolypothrix bouteillei* VB521301, 12 *cas1* genes and 29 repeat-spacer arrays in *Scytonema hofmannii* sp. PCC 7110, 5 and 15 in *Aphanizomenon flos-aquae* NIES-81, 3 and 14 in *Rivularia* sp. PCC 7116, 1 and 6 in *Nostoc punctiforme* PCC 73102, and 3 and 11 in *Anabaena* 7120 (**Table 1**), respectively. The genome sequences of aforementioned organisms are complete or in draft state with rigorous quality control [24] , excluding assembly artefacts as a possible source of overestimation. To look further into this obvious discrepancy, we filtered the available sequences (**Table S1**) for completion of sequencing, yielding 36 finished genome sequences. From these, we chose three representative examples for which we performed a detailed re-annotation of the CRISPR-Cas systems (available upon request). *Calothrix* sp. PCC 7507 possesses 10 repeat-spacer arrays, 9 *cas1* genes and four identifiable CRISPR-*cas* loci (**Figure 2**). However, three *cas1* genes are fragmented and constitute pseudogenes. The remaining six include two gene copies encoding a Cas1-Cas4 and a Cas1-RT fusion (**Figure 2**). *Rivularia* sp. PCC 7116 and *Nostoc* sp.

PCC 7107 both have 14 separate instances of repeat-spacer arrays but possess only three *cas1* genes. Hence, in all these cases the numbers of repeat-spacer arrays is larger than the number of *cas1* gene copies. Moreover, the two instances of genes encoding Cas1-RT fusions (in *Calothrix* sp. PCC 7507 and in *Nostoc* sp. PCC 7107) co-locate with an additional free-standing *cas1* gene and are part or very close to subtype III-D systems (**Figure 2**). From these observations we conclude that genes encoding Cas1-RT fusions may become integrated in addition to existing *cas1-cas2* adaptation modules and that the number of repeat-spacer arrays regularly exceeds the number of *cas1* gene copies.

Unicellular cyanobacteria do not share this feature, e.g., the model cyanobacteria *Synechocystis* sp. PCC 6803 and PCC 6714 each harbor three different *cas1* genes, matching the number of three different repeat-spacer arrays [19,20]. Therefore, the *cas1*-lacking systems in multicellular cyanobacteria might rely on adaptation modules encoded elsewhere in the genome [12,13] or depend on other mechanisms for recombination.

In the model *Anabaena* 7120, five DNA recombinase proteins are involved in the recombination in heterocyst differentiation. XisA mediates the excision of the 11 kb element from the *nifD* gene [25–27], the three-subunit enzyme encoded by the *xisF, xisH* and *xisI* genes excises the 55 kb element from the *fdxN* gene [28–30] (see **Figure 3** for their location), while the XisC recombinase deletes the 10.5 kb element from the *hupL* gene [31,32]. The XisI recombinase has recently been identified as a candidate protein for an anti-phage defense system based on its pfam08869 domain [33]. Moreover, there are at least eight additional recombinase genes in the genome of *Anabaena* 7120 (*all3124, alr0083, alr0084, alr2075, alr3224, alr3645, asl0560* and *asl0561*). The involvement of host-encoded factors such as IHF (integration host factor) in Cas1-Cas2 mediated adaptation has been reported for some types of CRISPR systems [34–36]. Therefore, it is tempting to speculate that one or several of the cyanobacterial recombinases are involved in CRISPR adaptation by functionally replacing the Cas1-Cas2 integrase complex that normally is facilitating the site-specific integration of new spacers into the CRISPR array [37,38].

CRISPR-Cas systems are present in genetic elements that are excised during cell differentiation by homologous recombination

The appearance of fusions between the Cas1 protein and an RT domain is typical of certain types of CRISPR-Cas systems in cyanobacteria [39]. We observed that the presence of genes encoding RT_Cas1 fusions is frequently linked to the occurrence of two separate CRISPR repeat-spacer units framing the *cas1*-RT gene on the forward and reverse DNA strands (e.g., in *Anabaena* 7120, **Figure 3**). This suggests that an unknown DNA recombination event is involved in the evolution of some of the cassettes that contain these genes. The *cas1*-RT gene in *Anabaena* 7120 is not fused to a *cas6* gene encoding the maturation endoribonuclease activity as observed in some other bacteria, e.g., *Marinomonas* [7,39].

Heterocyst differentiation includes the deletion of large genetic elements that interrupt the reading frames of critical genes. In different cyanobacteria, there are altogether more than ten different genes known that can be interrupted by such elements. Some of the frequently interrupted genes are the *nifH*, *nifD* and *nifK* (encoding nitrogenase Fe protein and subunits alpha and beta), *hupL* (encoding a subunit of heterocyst-specific uptake hydrogenase), *fdxN* (heterocyst ferredoxin) and *hglE* (encoding heterocyst glycolipid synthase) genes [25,28–32]. We observed that in several cases, CRISPR-*cas* systems are associated with these genetic elements that are precisely excised from the genome during the last steps of the differentiation of heterocyst cells. In *Anabaena* 7120, this is the case for the *fdxN* element (**Figure 3**). Similar CRISPR arrangements can be found in the *nifK* elements of *Calothrix* sp. 7102 (**Figure 3**), *Calothrix* sp. HK-06 and *Calothrix* sp. 7103, *nifD* element of *Tolypothrix* PCC 7601, and *nifH* and *hglE* elements of *Calothrix* sp. PCC 6303. The fact that different types of CRISPR-Cas systems are present in different elements suggests that they evolved independently from each other. Thus, these elements constitute a preferred site for the integration and hosting of mobile CRISPR-Cas cassettes, as was observed previously for certain types of mobile genetic elements [5]. It should be noted that the CRISPR-Cas cassettes together with the elements in which they reside are eliminated during heterocyst differentiation. These facts point further to a previously not recognized connection between CRISPR-Cas cassettes and the genetic mechanisms involved in heterocyst differentiation.

Characteristics of CRISPR-Cas systems in the model *Anabaena* 7120

Based on the number of *cas1* genes, there are three different Class 1 CRISPR-Cas systems in *Anabaena* 7120 (**Table 1**). However, a search for interspaced direct repeats showed that there are at least 11 CRISPR and CRISPR-like repeat-spacer cassettes (designated here CR_1 to CR_11) in *Anabaena* 7120, with more than 100 spacers (**Table 2, Figure 4**), all located in the 6,413,771-bp chromosome, whereas the seven plasmids are free of such cassettes. The presence of multiple small dispersed repeat (SDR) sequences was previously reported for *Nostoc punctiforme* and related cyanobacteria [40]; however, these SDR sequences are different from the CRISPR repeats presented here. The repeat-spacer cassettes CR_1 to CR_11 could be fragmented versions of a lower number of functional CRISPR-Cas systems, pseudogenized versions or belong to novel types of such systems. To test their transcription as an indicator of functionality, we isolated RNA from four different cultures and hybridized specific single-stranded RNA probes after gel electrophoretic separation and blotting. The observed lengths of mature crRNAs correlated well with the theoretically expected lengths of ~44 nt (±5 nt) for double processing (e.g. CR_4, 7, 8, 11 in **Figure 5**) or ~73 nt (±5 nt) in case of single processing (e.g. CR_1, 2, 3 in **Figure 5**).Hence, the results showed that all of the elements are transcribed and exhibit the typical pattern of precursor accumulation, processing intermediates and accumulated crRNAs (**Figure 5**). Thus, they are likely part of functional CRISPR-Cas systems. We included RNA from cultures grown for a nitrogen starvation time course that would be long enough to trigger heterocyst differentiation because of the affiliation of CRISPR-Cas systems, such as CR_2 and CR_3, with elements directly affected by this process. However, remarkable nitrogen-dependent differences in crRNA accumulation were not detected over the here applied time course of 32 h (**Figure 5**), indicating these CRISPR cassettes were actively transcribed in the vegetative cells independent of nitrogen availability.

Organisms possessing CRISPR-Cas systems become immune to phage or other invading DNA by the insertion of DNA sequences (spacers) into the leader-repeat junction (i.e., at the 5’ end of the repeat-spacer array) in a site-specific process called adaptation. The leader region, especially its 3' end, is indispensable for this adaptation [41–45]. Therefore, it must contain sequence determinants important for adaptation. However, the lengths of CRISPR leaders vary greatly in size, from 47 nt in some bacteria to several hundred nt in some hyperthermophilic archaea. Moreover, they possess long regions of low complexity sequence, show only limited sequence conservation [46] and therefore are difficult to predict [47]. Using differential RNA-seq, we previously experimentally defined a genome-wide map of more than 10,000 transcriptional start sites (TSS) of *Anabaena* 7120 at single-nucleotide resolution [48].Therefore, it is possible to precisely map the first transcribed nucleotide and to infer the length of the transcribed part of the leader when the element was expressed. This was possible for all 11 repeat-spacer cassettes. The length of the transcribed leaders varied from 57 to 3,616 nt (**Table 2**).

When judged by the association with known *cas1* genes, the arrays CR_1 to CR_4 represent classical Class 1 CRISPR elements (the two repeat-spacer arrays CR_2 and CR_3 frame the RT_*cas1* gene and belong to the same element, as depicted in **Figure 3** and **4**). Thus, there are at least three distinct Class 1 systems and one Class 2 (CR_9) system. The repeats CR_6 and CR_7 can be joined because they are only separated by the insertion of a 134 nt long miniature inverted repeat element (MITE) in repeat 9 of an originally contiguous array, yielding a total of five orphan repeat-spacer arrays.

Some repeats might be unified according to the similarities among their sequences, lumping CR_5 together with CR_2 and CR_3 in one group and CR_4, CR_6, CR_7, CR_8 and CR_11 in another (**Figure 4**), leaving CR_1, CR_9 and CR_10 separate. This is consistent with their assignment to distinct structural motif, sequence and super families as classified by the CRISPRmap algorithm [49] (**Table 2**). However, this unification might be an oversimplification. Hence, even when judged in a very restrictive way, there are at least five different types of arrays in total.

Novel CRISPR-Cas systems have substantial potential for genome editing and manipulation of gene expression. Therefore, it is interesting that one of the remaining systems, CR_9, is associated with a gene encoding All3613, a relatively large protein of unknown function. This protein is significantly similar in its C-terminal region to a subset of TnpB proteins encoded by transposons of the IS605 family, a feature typically associated with the Class 2 effector proteins Cas12b (C2c1) and C2c3 [12]. Among the studied 171 genomes of filamentous cyanobacteria, we found 86 All3613 homologs with a bit score ≥100, of which 29 were associated with a CRISPR array. This percentage is higher than expected by chance, supporting the idea that All3613 represents a novel type of CRISPR effector. This view was further supported when All3613 was analyzed by the HHpred algorithm [50], identifying a highly supported ~200 residues long similarity of the C terminus to the Cas12a (Cpf1) proteins of *Lachnospiraceae* bacterium ND2006 (probability 98.82, E-value $3.3e^{-10}$), *Acidaminococcus* sp. BV3L6 (probability 98.58, E_value $2.2e^{-9}$) and of *Francisella tularensis* subsp. novicida (probability 98.57, E-value $7.7e^{-9}$). All three proteins have been well characterized as single RNA-guided Type V effector proteins [4,51,52]. Nevertheless, proteins such as All3613 are with 648 amino acids substantially shorter than these effectors (e.g., Cas12a (Cpf1) of *Lachnospiraceae* bacterium ND2006 is 1231 residues long). Therefore, it is important that All3613 as well as many of its homologs are directly adjacent to a repeat-spacer cassette and that this cassette is expressed (**Figure 5**). A likely paralogous gene with *all3613* is *alr2691,* encoding a protein that in a pairwise alignment exhibits 40% identical and 60% similar amino acid residues with All3613 (bit score of 447). However, *alr2691* is not connected to a repeat-spacer array anywhere close in the genome.

The number of direct repeats in the CRISPR arrays associated with *all3613* homologs in cyanobacteria was relatively low, with a maximum count of 13, mean count of 6, and median count of 5, pointing to a possibly inefficient insertion process of new spacers. We observed that the majority of these CRISPR arrays (24/29) were adjacent to a tRNA gene, for example, *trnT*(CGT) in *Anabaena* 7120, *trnV*(GAC) in *Nostoc punctiforme* PCC 73102, *trnA*(CGC) in *Aphanizomenon flos-aquae* NIES 81 and *trnA*(GGC) and *trnR*(CCG) in two instances in *Tolypothrix* sp. PCC 7601 (**Figure 6**). Hence, we speculate that All3613 or its homologues in other cyanobacteria constitute effector protein candidates for a novel type of CRISPR system. It cannot be excluded from consideration that the immediately adjacent tRNA genes served as integration sites of the respective cassettes. But the association with multiple different tRNA genes is puzzling in this regard (**Figure 6**). Although none of these putative Class 2 systems is directly associated with a *cas1* gene, the assignment of these regions as uncharacterized CRISPR system is consistent with recent biocomputational analyses, which suggested that homologs of All3613 (Ava_2196 in *Trichormus variabilis* ATCC 29413 (previously called *Anabaena variabilis* sp. PCC 8801) and protein WP_027402996.1 in *Aphanizomenon flos-aquae* NIES 81) are the core subunits of a Class 2 system called C2c5 [5,13]. Because All3613 and its homologs are shorter than the characterized Class 2 single effectors we looked for the possible synteny with other genes. Except the tRNA genes following the arrays, we only identified a gene encoding a MerR-type transcriptional regulator that is frequently located directly adjacent to the *all3613/c2c5* gene (**Figure 6**).

Summary and perspective

Certain filamentous cyanobacteria appear to be a rich source of CRISPR-Cas systems. In the case of *Anabaena* 7120, we show expression of the separate instances of repeat-spacer arrays by Northern hybridization with a pattern typical of the accumulation of crRNAs. We report that different types of CRISPR-Cas systems are encoded in different types of the genetic elements that are recombined during the differentiation of heterocysts, suggesting their independent evolution. All3613 is predicted as a possible effector protein of the C2c5 type of Class 2 CRISPR-Cas system.

Material and Methods

The wild type strain of *Anabaena* 7120 was bubbled with an air/CO_2 mixture (1% v/v) and grown photoautotrophically at 30 °C in $BG11_0C$ medium [53] lacking $NaNO_3$ but containing 6 mM NH_4Cl, 10 mM $NaHCO_3$ and 12 mM tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid-NaOH buffer (pH 7.5) until exponential phase. In order to induce nitrogen deficiency, cells grown in the presence of ammonium were collected by filtration, washed with and resuspended in nitrogen-free $BG11_0C$. Four RNA samples were isolated from cells taken at 0h, 8h, 24h and 32h after removing combined nitrogen from the media.

Total RNA was isolated using hot phenol as described [54] with some modifications. Hot phenol was added to the cells immediately after addition of lysis buffer and incubation was carried out at 65 °C for 5 min. Further extractions were carried out with hot phenol, phenol:chloroform (1:1) and chloroform, followed by RNA precipitation by addition of one volume of isopropanol. CRISPR-related transcript accumulation was analyzed by Northern hybridization using single-stranded radioactively labelled RNA probes transcribed *in vitro* from PCR-generated templates (see **Table S2** for primers) as described [55]. The oligonucleotide for the detection of 5S rRNA was ^{32}P labelled using polynucleotide kinase (Thermo Fisher) and γ-ATP.

Phylogenetic analysis. The maximum likelihood tree was constructed based on 391 Cas1 proteins from 171 sequenced filamentous cyanobacteria and 3 Cas1 proteins from *Synechocystis* sp. PCC 6803. These Cas1 protein sequences were separately aligned with Clustal Omega v1.2.4 [56] and MAFFT E-INS-i v7.313 [57] with default parameters. The resulting alignments were merged using TrimAl v1.4.rev15 [58] with a minimum consistency score 0.5 (ct=0.5) and only columns with a gap percentage <50% were kept (gt=0.5) for further phylogenetic analysis. ProtTest v3.4 [59] was used to find the best amino acid replacement model with best tree search operation of NNI and SPR (-s BEST) and empirical frequency estimation (-F). Based on Bayesian information criterion (BIC), the estimated best model LG+G (-m PROTGAMMALG) was chosen to infer the maximum likelihood tree using RAxML v8.1.20 [60] with 20 best-scoring maximum likelihood searches and 1000 fast bootstrap searches (-f a -# 1000) from a random seed 12345 (-p 12345). The final phylogenetic tree was visualized using FigTree v1.4.2 (available at: http://tree.bio.ed.ac.uk/software/figtree/) and iTOL v4 online server (available at: http://itol.embl.de/ [61]). The 16S rRNA gene tree was constructed based on a MAFFT E-INS-i v7.313 [57] alignment using RAxML v8.1.20 [60] with GTR nucleotide substitution model and GAMMA model of rate heterogeneity (-m GTRGAMMA). The 16S rDNA sequence of *Gloeobacter violaceus* PCC 7421 was used as an outgroup to root the 16S rDNA tree. 1000 fast bootstrap searches were done with the same setting as used in the Cas1 tree (-f a -# 1000 -p 12345).

Genome annotation and identification of CRISPR-Cas containing interruption elements. The genome sequences of filamentous cyanobacteria used in this study were downloaded from NCBI on March 25th, 2017 using phyloutils v1.0 (available at: https://github.com/housw/phyloutils). To keep the genome annotations consistent, all the genomes used in this study were re-annotated using Prokka v1.12-beta [62] with phyloutils wrapper. CRISPR cassettes were predicted using MinCED v0.2.0 with default parameters (MinCED is available at https://github.com/ctSkennerton/minced). Protein domains were predicted using pfam_scan.pl v1.6 (available at ftp://ftp.ebi.ac.uk/pub/databases/Pfam/Tools/) against Pfam release 30.0 [63] with an E-value cutoff of 1e^{-5}. Interruption elements were scanned against all genomes using a *xis* gene-anchored algorithm with modifications [64], which required to determine the *xis* genes as a first step. In brief, for each genome, the *xis* genes were identified by searching all protein sequences against the previously identified Xis proteins [64] using blastP with an E-value cutoff of 1e^{-20}. Then, the DNA sequences 3 kb upstream and 3 kb downstream of identified *xis* genes were extracted to check partial coding regions against all full-length protein sequences of *Anabaena* sp. PCC 7120 using blastX with default parameters. The extracted *xis*-containing regions were extended accordingly to cover the full length of surrounding overlapping genes. When a partial hit was identified, the target reference protein sequence was used as query to search against the whole genome sequence to find the other parts using tblastN. After that, all the partial coding regions were translated and aligned against the reference proteins to compose the full-length proteins. If successful, the excised regions were further checked for CRISPR cassettes. Motifs, families and super families of CRISPR direct repeats were identified using the CRISPRmap [49] online server (http://rna.informatik.uni-freiburg.de/CRISPRmap/Input.jsp).

Abbreviations

Cas, CRISPR associated sequences; CRISPR, Clustered Regularly Interspaced Short Palindromic Repeats; C2c, Class 2 candidate; SDR, small dispersed repeat; TSS, transcriptional start site; UTR, untranslated region.

Funding information

Financial support for this work was provided by the German Research Foundation (DFG) program FOR1680 “Unravelling the Prokaryotic Immune System” (grants HE 2544/8-2 and BA 2168/5-2) to WRH and RB, by grant HE 2544/13-1, by the Ministerio de Economía y Competitividad (grant BFU2013-48282-C2-1) and the Agencia Estatal de Investigación (grant BFU2016-74943-C2-1-P) to AM, both cofinanced by FEDER; a predoctoral contract (FPU014/05123) and a short term research stay grant (EST16/00088) by the Ministerio de Educación, Cultura y Deportes to MB and by a China Scholarship Council grant to S.H., all of which are greatly acknowledged.

Author contributions

SH and OSA performed bioinformatics analyses, MBA and VR provided the experimental data, SH, OSA, MBA, RB, AMP and WRH analyzed data and WRH drafted the manuscript with contributions from all authors.

References

1. Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science. 2014. 346:1258096.

2. Li Y, Pan S, Zhang Y, Ren M, Feng M, Peng N, Chen L, Liang YX, She Q. Harnessing Type I and Type III CRISPR-Cas systems for genome editing. Nucleic Acids Res. 2016. 44:e34.

3. Ungerer J, Pakrasi HB. Cpf1 is a versatile tool for CRISPR genome editing across diverse species of Cyanobacteria. Sci Rep. 2016. 6:39681.

4. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell. 2015. 163:759–71.

5. Koonin EV, Makarova KS, Wolf YI. Evolutionary Genomics of Defense Systems in Archaea and Bacteria. Annu Rev Microbiol. 2017. 71:233–61.

6. Garrett RA, Vestergaard G, Shah SA. Archaeal CRISPR-based immune systems: exchangeable functional modules. Trends Microbiol. 2011. 19:549–56.

7. Silas S, Mohr G, Sidote DJ, Markham LM, Sanchez-Amat A, Bhaya D, Lambowitz AM, Fire AZ. Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. Science. 2016. 351:aad4234.

8. Kieper SN, Almendros C, Behler J, McKenzie RE, Nóbrega FL, Haagsma AC, Vink JNA, Hess WR, Brouns SJJ. Cas4 facilitates PAM-compatible spacer selection during CRISPR adaptation. Cell Rep. 2018. 22:3377–338.

9. Lee H, Zhou Y, Taylor DW, Sashital DG. Cas4-dependent prespacer processing ensures high-fidelity programming of CRISPR arrays. Mol Cell. 2018. 70:48-59.e5.

10. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiol. 2009. 155:733–40.

11. Shah SA, Erdmann S, Mojica FJM, Garrett RA. Protospacer recognition motifs: mixed identities and functional diversity. RNA Biol. 2013. 10:891–9.

12. Shmakov S, Abudayyeh OO, Makarova KS, Wolf YI, Gootenberg JS, Semenova E, Minakhin L, Joung J, Konermann S, Severinov K, et al. Discovery and functional characterization of diverse Class 2 CRISPR-Cas systems. Mol Cell. 2015. 60:385–97.

13. Shmakov S, Smargon A, Scott D, Cox D, Pyzocha N, Yan W, Abudayyeh OO, Gootenberg JS, Makarova KS, Wolf YI, et al. Diversity and evolution of class 2 CRISPR-Cas systems. Nat Rev Microbiol. 2017. 15:169–82.

14. Fay P. Oxygen relations of nitrogen fixation in cyanobacteria. Microbiol Rev. 1992. 56:340–73.

15. Herrero A, Stavans J, Flores E. The multicellular nature of filamentous heterocyst-forming cyanobacteria. FEMS Microbiol Rev. 2016. 40:831–54.

16. Wolk CP. Heterocyst formation. Annu Rev Genet. 1996. 30:59–78.

17. Kumar K, Mella-Herrera RA, Golden JW. Cyanobacterial heterocysts. Cold Spring Harb Perspect Biol. 2010. 2:a000315.

18. Cai F, Axen SD, Kerfeld CA. Evidence for the widespread distribution of CRISPR-Cas system in the phylum Cyanobacteria. RNA Biol. 2013. 10:687–93.

19. Hein S, Scholz I, Voß B, Hess WR. Adaptation and modification of three CRISPR loci in two closely related cyanobacteria. RNA Biol. 2013. 10:852–64.

20. Scholz I, Lange SJ, Hein S, Hess WR, Backofen R. CRISPR-Cas systems in the cyanobacterium *Synechocystis* sp. PCC6803 exhibit distinct processing pathways involving at least two Cas6 and a Cmr2 protein. PloS One. 2013. 8:e56470.

21. Behler J, Sharma K, Reimann V, Wilde A, Urlaub H, Hess WR. The host-encoded RNase E endonuclease as the crRNA maturation enzyme in a CRISPR–Cas subtype III-Bv system. Nat Microbiol. 2018. 3:367–77.

22. Reimann V, Alkhnbashi OS, Saunders SJ, Scholz I, Hein S, Backofen R, Hess WR. Structural constraints and enzymatic promiscuity in the Cas6-dependent generation of crRNAs. Nucleic Acids Res. 2017. 45:915–25.

23. Jesser R, Behler J, Benda C, Reimann V, Hess WR. Biochemical analysis of the Cas6-1 RNA endonuclease associated with the subtype I-D CRISPR-Cas system in *Synechocystis* sp. PCC 6803. RNA Biol. 2018.

24. Zhu T, Hou S, Lu X, Hess WR. Draft genome sequences of nine cyanobacterial strains from diverse habitats. Genome Announc. 2017. 5.

25. Golden JW, Robinson SJ, Haselkorn R. Rearrangement of nitrogen fixation genes during heterocyst differentiation in the cyanobacterium Anabaena. Nature. 1985. 314:419–23.

26. Brusca JS, Chastain CJ, Golden JW. Expression of the *Anabaena* sp. strain PCC 7120 *xisA* gene from a heterologous promoter results in excision of the *nifD* element. J Bacteriol. 1990. 172:3925–31.

27. Henson BJ, Pennington LE, Watson LE, Barnum SR. Excision of the *nifD* element in the heterocystous cyanobacteria. Arch Microbiol. 2008. 189:357–66.

28. Golden JW, Carrasco CD, Mulligan ME, Schneider GJ, Haselkorn R. Deletion of a 55-kilobase-pair DNA element from the chromosome during heterocyst differentiation of *Anabaena* sp. strain PCC 7120. J Bacteriol. 1988. 170:5034–41.

29. Carrasco CD, Ramaswamy KS, Ramasubramanian TS, Golden JW. *Anabaena xisF* gene encodes a developmentally regulated site-specific recombinase. Genes Dev. 1994. 8:74–83.

30. Ramaswamy KS, Carrasco CD, Fatma T, Golden JW. Cell-type specificity of the *Anabaena fdxN*-element rearrangement requires *xisH* and *xisI*. Mol Microbiol. 1997. 23:1241–9.

31. Carrasco CD, Buettner JA, Golden JW. Programmed DNA rearrangement of a cyanobacterial *hupL* gene in heterocysts. Proc Natl Acad Sci USA. 1995. 92:791–5.

32. Carrasco CD, Holliday SD, Hansel A, Lindblad P, Golden JW. Heterocyst-specific excision of the *Anabaena* sp. strain PCC 7120 *hupL* element requires xisC. J Bacteriol. 2005. 187:6031–8.

33. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, Amitai G, Sorek R. Systematic discovery of antiphage defense systems in the microbial pangenome. Science. 2018.

34. Nuñez JK, Bai L, Harrington LB, Hinder TL, Doudna JA. CRISPR immunological memory requires a host factor for specificity. Mol Cell. 2016. 62:824–33.

35. Wright AV, Liu J-J, Knott GJ, Doxzen KW, Nogales E, Doudna JA. Structures of the CRISPR genome integration complex. Science. 2017. 357:1113–8.

36. Yoganand KNR, Sivathanu R, Nimkar S, Anand B. Asymmetric positioning of Cas1-2 complex and Integration Host Factor induced DNA bending guide the unidirectional homing of protospacer in CRISPR-Cas type I-E system. Nucleic Acids Res. 2017. 45:367–81.

37. Xiao Y, Ng S, Nam KH, Ke A. How type II CRISPR–Cas establish immunity through Cas1–Cas2-mediated spacer integration. Nature. 2017. 550:137–41.

38. Nuñez JK, Lee ASY, Engelman A, Doudna JA. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. Nature. 2015. 519:193–8.

39. Silas S, Makarova KS, Shmakov S, Páez-Espino D, Mohr G, Liu Y, Davison M, Roux S, Krishnamurthy SR, Fu BXH, et al. On the origin of reverse transcriptase-using CRISPR-Cas systems and their hyperdiverse, enigmatic spacer repertoires. mBio. 2017. 8:e00897–17.

40. Elhai J, Kato M, Cousins S, Lindblad P, Costa JL. Very small mobile repeated elements in cyanobacterial genomes. Genome Res. 2008. 18:1484–99.

41. Wei Y, Chesne MT, Terns RM, Terns MP. Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in *Streptococcus thermophilus*. Nucleic Acids Res. 2015. 43:1749–58.

42. Erdmann S, Garrett RA. Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. Mol Microbiol. 2012. 85:1044–56.

43. Yosef I, Shitrit D, Goren MG, Burstein D, Pupko T, Qimron U. DNA motifs determining the efficiency of adaptation into the *Escherichia coli* CRISPR array. Proc Natl Acad Sci USA. 2013. 110:14396–401.

44. Erdmann S, Le Moine Bauer S, Garrett RA. Inter-viral conflicts that exploit host CRISPR immune systems of *Sulfolobus*. Mol Microbiol. 2014. 91:900–17.

45. Van Orden MJ, Klein P, Babu K, Najar FZ, Rajan R. Conserved DNA motifs in the type II-A CRISPR leader region. PeerJ. 2017. 5:e3161.

46. Shah SA, Garrett RA. CRISPR/Cas and Cmr modules, mobility and evolution of adaptive immune systems. Res Microbiol. 2011. 162:27–38.

47. Alkhnbashi OS, Shah SA, Garrett RA, Saunders SJ, Costa F, Backofen R. Characterizing leader sequences of CRISPR loci. Bioinforma. 2016. 32:576–85.

48. Mitschke J, Vioque A, Haas F, Hess WR, Muro-Pastor AM. Dynamics of transcriptional start site selection during nitrogen stress-induced cell differentiation in *Anabaena* sp. PCC7120. Proc Natl Acad Sci USA. 2011. 108:20130–20135.

49. Lange SJ, Alkhnbashi OS, Rose D, Will S, Backofen R. CRISPRmap: an automated classification of repeat conservation in prokaryotic adaptive immune systems. Nucleic Acids Res. 2013. 41:8034–44.

50. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J Mol Biol. 2017.

51. Dong D, Ren K, Qiu X, Zheng J, Guo M, Guan X, Liu H, Li N, Zhang B, Yang D, et al. The crystal structure of Cpf1 in complex with CRISPR RNA. Nature. 2016. 532:522–6.

52. Gao P, Yang H, Rajashankar KR, Huang Z, Patel DJ. Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. Cell Res. 2016. 26:901–13.

53. Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. Microbiology. 1979. 111:1–61.

54. Mohamed A, Jansson C. Influence of light on accumulation of photosynthesis-specific transcripts in the cyanobacterium *Synechocystis* 6803. Plant Mol Biol. 1989. 13:693–700.

55. Steglich C, Futschik ME, Lindell D, Voss B, Chisholm SW, Hess WR. The Challenge of Regulation in a Minimal Photoautotroph: Non-Coding RNAs in *Prochlorococcus*. PLOS Genet. 2008. 4:e1000173.

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

57. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in performance and usability. Mol Biol Evol. 2013. 30:772–80.

58. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009. 25:1972–3.

59. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinforma. 2011. 27:1164–5.

60. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014. 30:1312–3.

61. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016. 44:W242-245.

62. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinforma. 2014. 30:2068–9.

63. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016. 44:D279-285.

64. Hilton JA, Meeks JC, Zehr JP. Surveying DNA elements within functional genes of heterocyst-forming cyanobacteria. PLOS ONE. 2016. 11:e0156034.

Figure legends

Figure 1. Phylogenetic analysis of 391 Cas1 sequences from 171 filamentous cyanobacteria. The three Cas1 protein sequences from *Synechocystis* sp. PCC 6803 were included in this unrooted phylogenetic tree, each representing a Cas1 cluster (grey shaded clusters) of 7 major clusters. Dark blue branches within light blue shaded clusters represent Cas1 proteins fused with a Cas4 domain, while red branches with light pink shaded clusters represent Cas1 proteins fused with a RVT_1 domain. The positions of Cas1 proteins from *Anabaena* 7120 are given in brackets. The detailed Cas1 tree with sequence names and bootstrap values is presented in **Figure S1**, the protein sequences can be found in **Supplemental Dataset 1**.

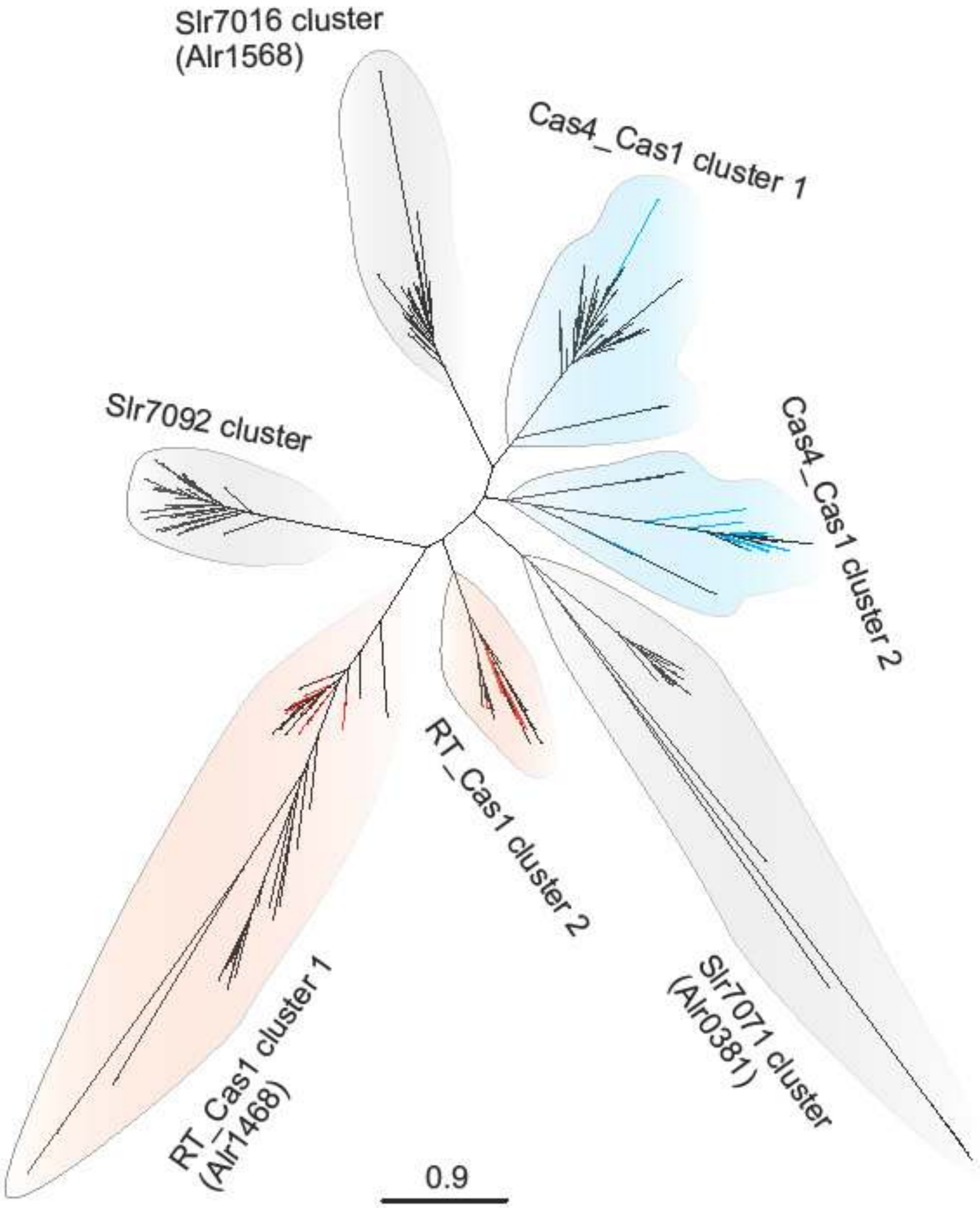

Slr7016 cluster
(Alr1568)
Cas4_Cas1 cluster 1
Slr7092 cluster
Cas4_Cas1 cluster 2
RT_Cas1 cluster 2
RT_Cas1 cluster 1
(Alr1468)
Slr7071 cluster
(Alr0381)
0.9

Figure 2. CRISPR-Cas systems and repeat-spacer arrays in three representative species, (A) *Nostoc* sp. PCC 7107, (B) *Calothrix* sp. PCC 7507 and (C) Rivularia sp. PCC 7116. The location of *cas1* genes is indicated by asterisks, an added "RT_1" or "1_4" indicates RT or Cas4 fusions. Pseudogenes are labelled by a Φ symbol.

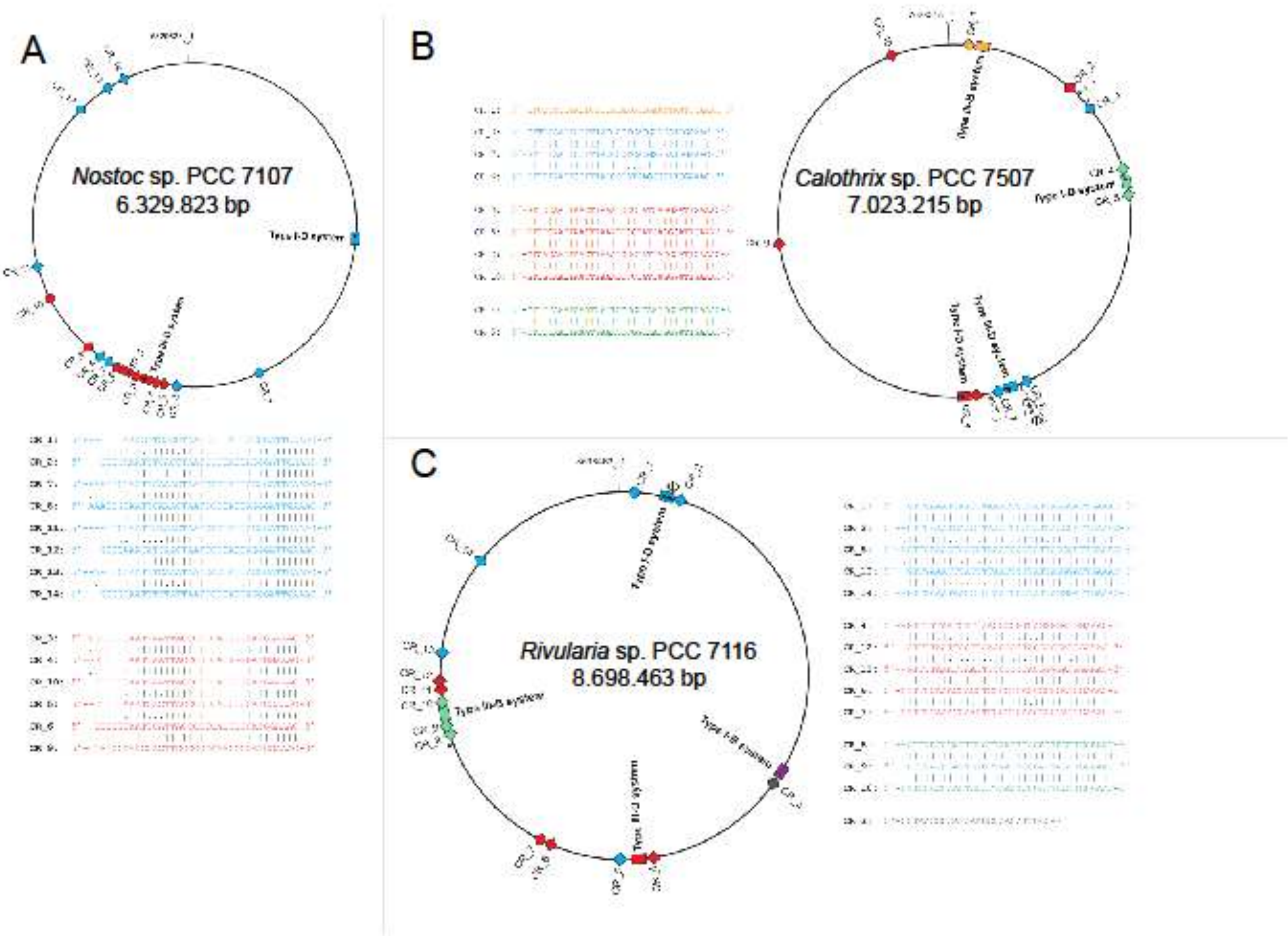


Figure 3. Examples of CRISPR-Cas systems that are encoded on genetic elements that are excised during cell differentiation into nitrogen-fixing heterocysts. The upper example is in the model organism *Anabaena* 7120, in which a CRISPR-Cas system

is present within the *fdxN* element. The gene *alr1468* encoding a reverse transcriptase-Cas1 fusion protein is annotated; details for the repeat spacer arrays CR_2 and CR_3 can be found in **Table 2**. The lower example presents the *nifK* element in *Calothrix desertica* sp. PCC 7102. CRISPR repeat-spacer arrays are labelled “CRISPR”. Note that the *Anabaena* 7120 element contains a *cas1*-RT fusion gene that is framed by split instances of the repeat-spacer arrays CR_2 and CR_3. Recombinase genes are labeled *xis* and *xisAFHI* and colored purple, *cas* genes are colored red, genes related to nitrogen fixation are colored blue, all other genes are in grey.

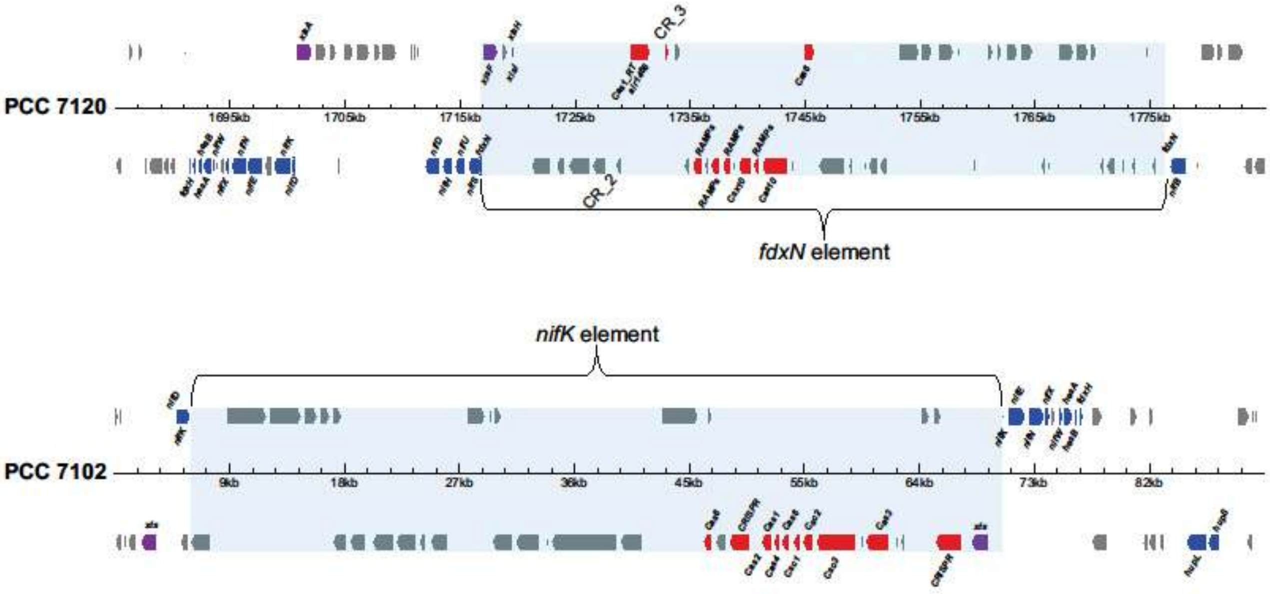


Figure 4. CRISPR-Cas systems in *Anabaena* 7120***.*** Left: Alignments of CRISPR direct repeats. Right: Schematic distribution of CRISPR-*cas* systems, their subtype annotation and the location of repeat-spacer arrays in the chromosome of *Anabaena* 7120. The location of *cas1* genes is indicated by asterisks and the respective gene IDs. For further details, see also **Table 2**.

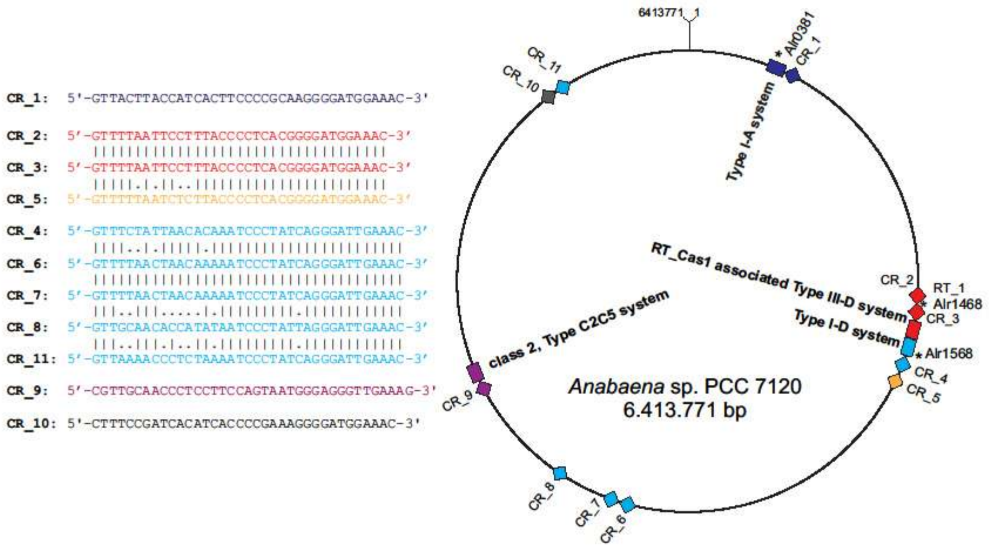


Figure 5. Expression of crRNAs from repeat-spacer arrays in *Anabaena* 7120. Total RNA was isolated from cultures grown under standard conditions for 8, 24 and 32 h after the removal of nitrogen; separated by electrophoresis on denaturing 15% PAA gels; and transferred to nylon membranes. Single-stranded specific RNA probes were used for Northern hybridization. A control hybridization against 5S rRNA was performed to control for equal loading (the same membrane used for CR_3 was re-hybridized with the CR_11 probe later). The size of marker fragments is given on the left.

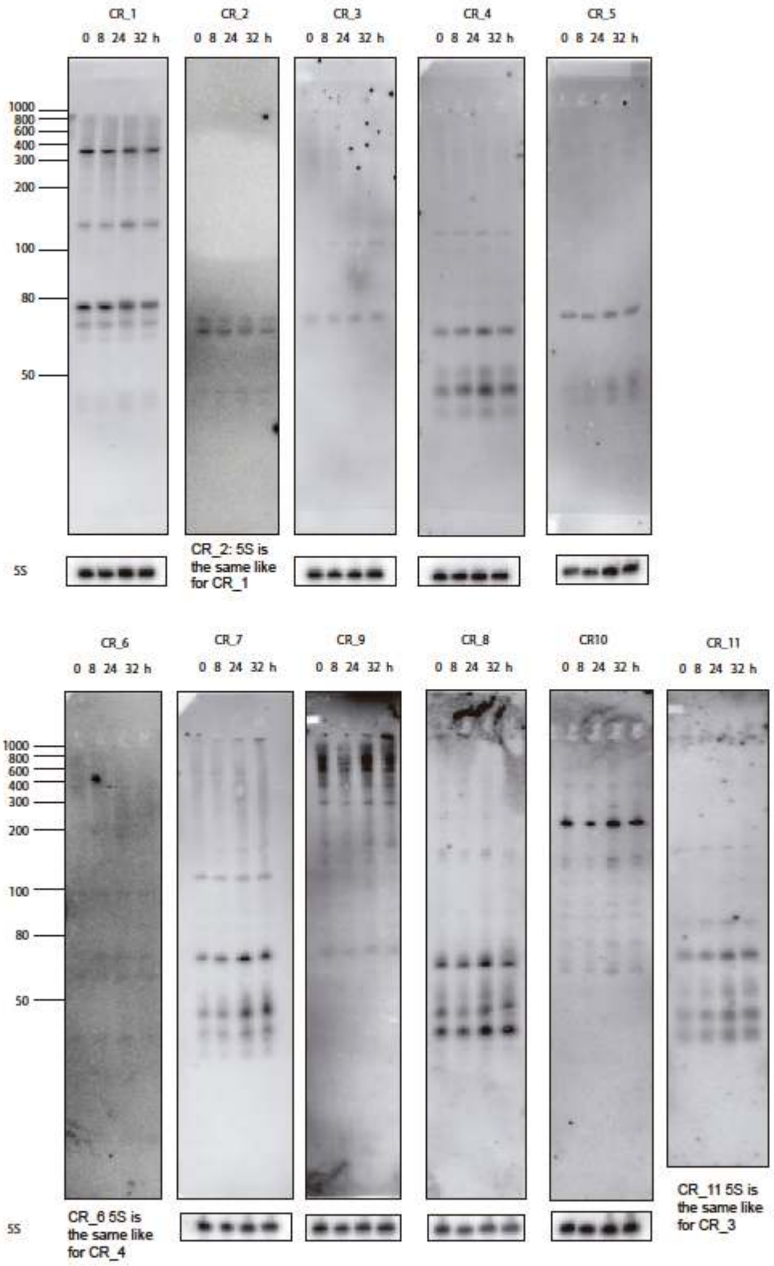
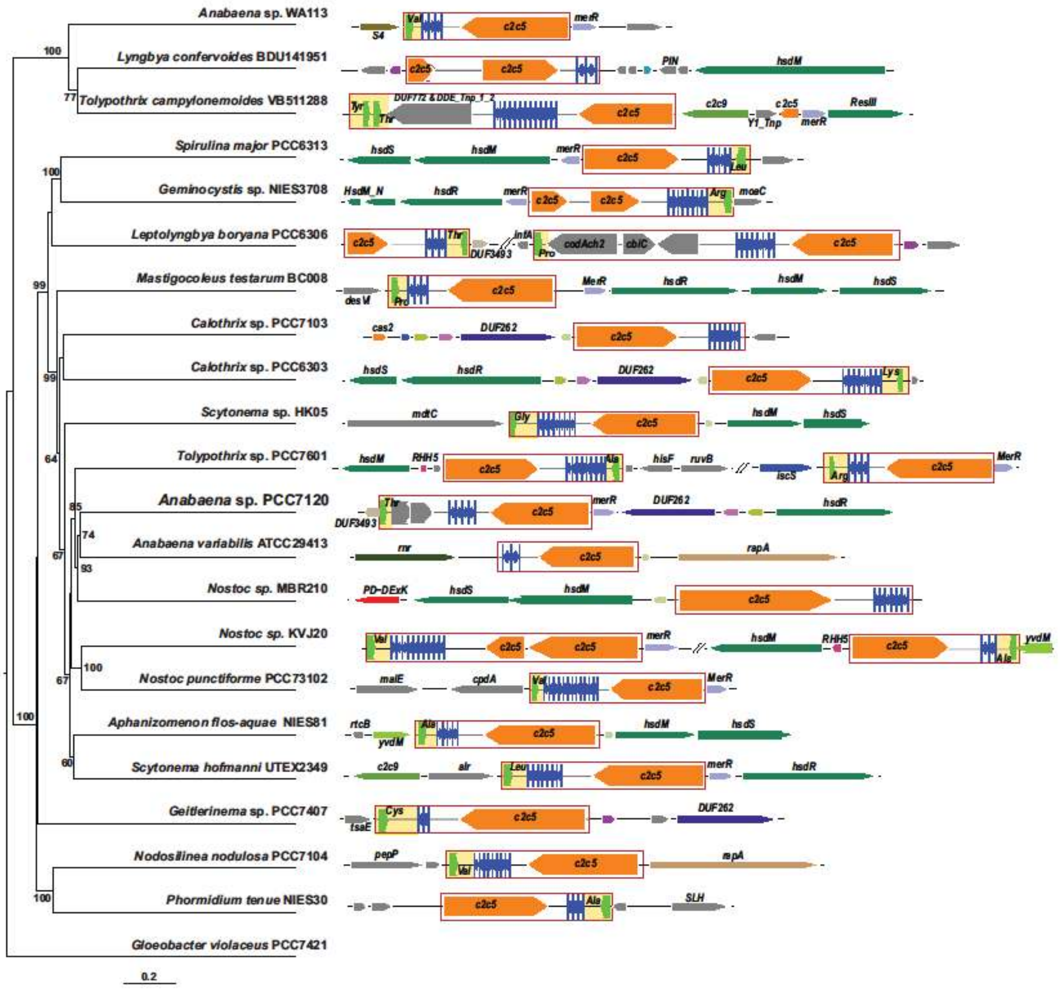


Figure 6. Synteny of Class 2 candidate systems. On the left the phylogenetic relationships are drawn based on 16S rRNA sequences, on the right the arrangements of putative CRISPR-*cas* systems are depicted that have an *all3613* homolog next to the array and lack any known genes for adaptation or other known *cas* gene. Note the frequent presence of different tRNA genes adjacent to the repeat-spacer arrays. The units consisting of the *all3613* homolog, the repeat-spacer array

and tRNA gene (if present) are boxed. Known and putative *cas* genes are colored orange. The different tRNA genes are colored in light green and shaded in yellow for visualization, and their cognate amino acid is indicated. Numbers on the phylogenetic tree are bootstrap values and given if ≥60. The position of the model *Anabaena* (*Nostoc)* 7120 is highlighted by larger fonts. 16S rDNA sequence from *Gloeobacter violaceus* PCC7421 was used as an outgroup to root this tree.



Tables

Table 1. GenBank assembly accession numbers and Cas1 protein features of selected cyanobacterial strains. Morphological subsections were assigned according to Rippka et al. [53]. Cas1 gene types were defined as free-standing *cas1, cas4_cas1* or RT_*cas1* fusions. Accession numbers labelled by an asterisk refer to the JGI genome portal, all other refer to Genbank.

Cyanobacterial strains	Subsection	Habitat	Accession	Number of Cas1 genes	Domain types of Cas1	Number of CRISPR cassettes	Identified C2c5 homologs
Anabaena 7120	IV	Freshwater	GCA_000009705.1	3	RT_Cas1, 2x Cas1	11	All3613, Alr2691
Anabaena cylindrica sp. PCC 7122	IV	Freshwater	GCA_000317695.1	5	RT_Cas1, 4x Cas1	13	Anacy_2856, Anacy_0603
Calothrix sp. HK-06	IV	Terrestrial	GCA_001904745.1	1	Cas1	10	0
Calothrix sp. PCC 7507	IV	Terrestrial	GCA_000316575.1	9	Cas4_Cas1, RT_Cas1, 9x Cas1	10	0
Calothrix desertica sp. 7102	IV	Terrestrial	2509887024*	6	Cas4_Cas1, RT_Cas1, 4x Cas1	14	0
Fischerella major NIES-592	V	Hot spring	GCA_001904645.1	2	RT_Cas1, Cas1	4	0
Nostoc sp. PCC 7107	IV	Freshwater	GCA_000316625.1	3	RT_Cas1, 2x Cas1	14	Nos7107_4709
Nostoc calcicola FACHB-389	IV	Terrestrial	GCA_001904715.1	2	2x Cas1	19	0
Nostoc punctiforme PCC 73102	IV	Terrestrial/symbiontic	GCA_000020025.1	1	Cas1	6	Npun_R5656
Limnothrix rosea IAM M-220	III	Marine	GCA_001904615.1	3	RT_Cas1, 2x Cas1	9	0
Phormidium ambiguum IAM M-71	III	Freshwater	GCA_001904725.1	8	RT_Cas1, 7x Cas1	14	0
Rivularia sp. PCC	IV	Marine	GCA_000316665.1	3	Cas1, 2x Cas1	14	0

7116							
Scytonema hofmannii sp. PCC 7110	IV	Terrestrial (limestone)	GCA_000346485.2	12	2x Cas4_Cas1, 2x RT_Cas1, 8x Cas1	29	WA1_24145
Tolypothrix sp. PCC 7601	IV	Terrestrial	GCA_000300115.1	1	Cas1	19	FDUTEX481_03012, FDUTEX481_08898
Tolypothrix bouteillei VB521301	IV	Stone surface	GCA_000760695.2	5	Cas4_Cas1, RT_Cas1, 3x Cas1	44	0

* JGI Taxon ID

Table 2. Predicted repeat-spacer arrays in *Anabaena* 7120. Each array has been numbered (ID), followed by the nucleotide positions of the TSS and the repeat-spacer arrays in the chromosome, the orientation (O) on the forward (+) or reverse (-) strand, the repeat sequence (DR), the number of repeats, with the total number including imperfect repeats in brackets (#), the length of spacers (L) the structural motif family (M), sequence family (F) and super family (S) as classified by the CRISPRmap algorithm [49], followed by the subtype and comments (?, unknown). For the location within the chromosome of *Anabaena* 7120 see also **Figure 4**.

ID	TSS	Start	End	O	transcribed leader	DR	#	L	M	F	S	Subtype, remarks
CR_1	445445	445573	447786	+	128	GTTACTTACCATCACTTCCCCGCAAGGGGATGGAAAC	28 (29)	33-48	4	8	E	subtype I-A, no cas6
CR_2	1728349	1728071	1727817	-	278	GTTTTAATTCCTTTACCCCTCACGGGGATGGAAAC	3 (4)	37-40	4	9	E	subtype III-D; CR_2 and CR_3 belong to one element; RT_cas1 fusion
CR_3	1731975	1732269	1733321	+	294	GTTTTAATTCCTTTACCCCTCACGGGGATGGAAAC	9 (15)	34-40	4	9	E	
CR_4	1836427	1836813	1837723	+	386	GTTTCTATTAACACAAATCCCTATCAGGGATTGAAAC	13	33-41	8	9	E	subtype I-D
CR_5	2179566	2179167	2178606	-	399	GTTTTTAATCTCTTACCCCTCACGGGGATGGAAAC	8 (11)	39-42	4	2	E	?
CR_67	3518141	3518084	3516820	-	57	GTTTTAACTAACAAAAATCCCTATCAGGGATTGAAAC	13 (16)	31-44	8	9	E	134 nt MITE insertion in repeat 9
CR_8	3836504	3840120	3840737	+	3616	GTTGCAACACCATATAATCCCTATTAGGGATTGAAAC	9	33-44	8	9	E	?
CR_9	4362990	4362577	4362255	-	413	CGTTGCAACCCTCCTTCCAGTAATGGGAGGGTTGAAAG	3 (5)	32-35	?	?	?	C2c5, All3613 effector
CR_10	5647342	5647145	5646379	-	197	CTTTCCGATCACATCACCCCGAAAGGGGATGGAAAC	10 (11)	32-45	-	18	C	?
CR_11	5654075	5654133	5654384	+	58	GTTAAAACCCTCTAAAATCCCTATCAGGGATTGAAAC	3	34-36	8	9	E	?

Supplemental Dataset 1: 391 Cas1 sequences of filamentous cyanobacteria

```
>PCC10914_05187  Cas_Cas4+Cas_Cas1
MQTLKYLEFSTKPETIRVSALHALAYCPRLFYLEEVEELYTQDAAVFAGRRLHTELEKQEDEEWEELFLESEE
LGLRGRLDALRTRDGQTIPYEHKRGRAHRDENKQPSAWQSDKLQILAYAYLLESALGITVKEGRIRYHADNVL
VHVPLDDAGRSAVREAIQQARTLRESTHRPPVIDNERLCARCSLAPVCLPEEARLAHDREWQPIRLFPEDDER
QVIHVLEPGTSVGRTGEQIKIARRNQPVETIPARQVGQLVLHSFSQISTQALHFCADQGIGVHFISGGGRYLG
SFDTRQGSIQRRIRQYTALSNPDTCLELARKLVVCRAQGQRKFLMRGTRGKAETPESLKAAIAQMKVVLKQVP
QAKSLESLLGLEGNLAALYFSALPCLISKEVPEGMHFAGRNRNPPKDRFNTLLSFGYALLLKDVMNAILTVGL
EPALGFYHQPRSQAAPLALDLIEIFRVPLVDMTVMASVNRNQWDIKADFEVRGEQVWLSEAGKRKFVELYERR
KQESWKHPVTSYSLTYRRLFELEVRLLEKEWSGEGGLFGHLVLR
>PCC10914_01365  Cas_Cas1
MLTDDEKQSVSEAIPKLESGNENIAYVRNIATLHQPFLAEIETELPPQSEEDNSISQKNIWHQEMAAIYLIEQ
GTNIYKDYQRFIIHVSEKPKLEVPIREVQQILVFGNIQLSTPVIQACLQQEIPVLFLSQSGQYHGHLWSEEST
NLDNQLIQIERRNDDSFQFQVSKAIVFGKLMNSKQLLLRFNRKRKVTEVEKAIFGISQDIDSLDLVSNLDALR
GYEGIAAARYFPAFGQLMTNSKFEFSLRNRQPPTDPVNSLLSFGYTLLFNNVLSFLIAEGLSPYLGNFHYGEK
QKPYLAFDLMEEMRSVIVDSLVLNIINHSLLKPQDFDIVPSTGGVYLNQSARRVFLKQFETRMNEEVSHPDLQ
SQITYRHAIQLQVRRYKRCLLSGVPYEPFLRAV
>PCC7417_05178   Cas_Cas1
MSTVYYVIENDAFIGKVDERLTVKAEKKTILDVPLIKINGIVVLGRATVSPAVITELLKNQICLTFLTENGKYL
GRLEPEVTKNLFIRKAQWEAAGESPQAIHLVRGFVRGKLKNYRQVLLRAQREYAHLNLSSQITRLENAIAPIN
QTNKINSLRGLEGAGSAAYFGCFDKLIKTSEFEFSTRNRRPPTDPVNALLSFGYSLLRHDVQSAVNIVGFDPY
LGYLHVERYGRPSLALDVMEEFRPLVVDGMVLSLINKRSLIPDHFITEPLSGAVNLTKEGLRIFLTAYQQKKQ
SSFKHPVMGTKCTYQEAFEIQARFLGKYLMGEVDKYPPLVLK
>WA93_01676 Cas_Cas1
MPIERGYRRLSRYQQELTAIDKLIAARTIIQAKLKNSRTFLQRQQRRNPSQTAEITIKSLEVLAQKAGEAETI
ERLMGLEGAGAAQYFNAFGECLNNPDFVFLARSRRPPGNPVNAMLSFGYQILWNHLLTLIELQGLDPYYACLH
QGTERHAALTSDLIEEFRAPIIDSLVLWLVNTKIMNIDNDFEYHDSGCYLNNSGRPKYIKYFLQRLEEEVQNS
EGEKQPRWDLMTQQIKAFKDFVYQPSKLYKPYQIR
>WA93_02182 Cas_Cas1
MGTLYIVQDDAFIGKVDERINVKYEKKILQDVPFIHLEGVVVLGRATISPAVVDELMQRHIPLSFINKIGHYL
GRLEPEMTGNIFIRKAQWQAAGETPQSIHLVQGFVRGKLKNYRQTLMRYQREFNDIDLSKSIPRIEQVIAAIN
STESINSLRGLEGSGSAAYFGCFDSLIRNSQFSFTKRVRRPPTDPVNSLLSFGYSLLRHDLQSAVNIVGFDPY
LGYLHFDRYNRPSLALDLMEEFRPLVVDAVVLSILNKQLLTPADFVTEPLSGAVSLTPEGRKTFLTLYEKKKI
SEFKHPVMGRKCSYREAFELQARLLAKYLMGTIDKYPPLVLK
>WA93_02450 Cas_Cas1
MQTLYVSEQNCYVCLQKETLLVKQGETVYAEVQLPLLEQILIFGKSQLTTQVIRACLWRDIPIAYLSRMGYCY
GRILPISRGYRQLSRYQQQLSPVEKLITARAMIKGKLKNSRVLLRRQRKKKESELLERVLQSLDYLADQVAQA
DNLERLMGFEGAGAAQYFSAFSECLTNPDFVFSGRSRRPPGNPVNAMLSFGYQVLWNHLLALIEIQGLDPYYA
CLHQAHDGHAALASDLIEEFRAPFIDSLVMWLINRKIVSAESDFEFKNGGCYLNDAGRKKFLRAFLQRMSEEI
QTDEGIKQPKWDLLTKQVRAFKQFVYNPSHHYQPYRID
>PCC6406_05101   Cas_Cas1
VLSLLRVEGVNPYLGNLHRSDRPEAQLAFDLMEEFRSPIVDTLVVQLINQGVFTLEDFTPPNAEGGVYLADPA
RKRFLQGFEGRIMTPSATPTAPNLCPIAR
>PCC6406_05097   Cas_Cas1
MALLYLTQQGTQLRKDYGRFVVQLPDADAIEVPIKEVEHVCLFGNIQLSTSAISTCLYHHIPVVFLSQTGSYK
GHLWSAEDSQQQTQALQYERFSQPQFQWAMAQAIVRGKLWNSRQLLLKLNRRHPLETVAAAAQGINRDRRAAT
NLTLETDSLNSLRGYEGAAAARYFLALGQLLKHPDFTLTQRTRRPPTDPVNSLLSFGYTLLFNNVLSLLRVEG
VNPYLGNLHRSDRPEAQLAFDLMEEFRSPIVDTLVVQLINQGVFTLEDFTPPNAEGGVYLADPARKRFLQGFE
GRIMTPSATPTAPNLCPIAR
>PCC6406_04946   RVT_1+Cas_Cas1
MQLYPEQLATAWHLVQRGSPAAGVDGITVDLFKGIAQEQIRLLHQQMRQERYVASPAKGFYLPKKTGGDRLIG
IPTVKDRIVQRYLLQGIYPHLENTFSEATFAYRPGLSIYTAVAQVMTRYRHQPAWVIKADIQQFFDRLSWPLL
LHQLDQLPLPPVWMRWIEQQLKAGIVIRGHFQRPNQGVLQGSILSGALANLYLNDFDRRCLAADIDLVRYGDD
CVAVCQSYLEATRSLALMQDWIEDLYLSLHPEKTQIIPPGEAFVFLGHRFHNGEVEGPERQKAEGPERSRRAS
KPGARWGPPMVCSIVKPPKRILATSTDEYWRDGMTTLYVTDQGAYLRVQHQQFQVFHQQELRCSVPASRISHV
VLFGCCNVSHGAVRLALQRRIPLLYLSNKGRYFGRLQTTGQAKLDYLTQQVYKAQDPEFIRTQAASIIVGKLH
NSRILLQRLNRRRKTELATQAIDTLAELMQTVPSVESVEAMLGYEGTGASAYFQAYASLLKGEFEFEKRTRRP
```

```
PTDPINSLLSLGYTLLSQNVHAMVEGVGLHTHFGNLHVPRENRPSLVCDLVEEFRAVVVDSLVAYLINSNIFK
ADDFTPPDERGGVYLYPDAMKRFLKHWEEKLQHTVTHPHTGYKVSYRRCFELQVWEYVACLTGEQSVYRPMRM
EK
>PCC6406_04911   Cas_Cas1
MALLYLTQQGTQLRKDYGRFVVQLPDADAIEVPIKEVEHVCLFGNIQLSTSAISTCLYHHIPVVFLSQTGSYK
GHLWSAEDSQQQTQALQYERFSQPQFQWAMAQAIVRGKLWNSRQLLLKLNRRHPLETVAAAAQGINRDRRAAT
NLTLETDSLNSLRGYEGAAAARYFLALGQLLKHPDFTLTQRTRRPPTDPVNSLLSFGYTLLFNNVLSLLRVEG
VNPYLGNLHRSDRPEAQLAFDLMEEFRSPIVDTLVVQLINQGVFTLEDFTPPNAEGGVYLADPARKRFLQGFE
GRIMTPSATPTAPNLCPIAR
>PCC6406_03686   Cas_Cas1
MKQILNTLYVQTQGAYLRLDHDTLQVEIEKQKALQMPLHHLGGITVFGNVLLSPFLIHRCAEDGRSLVWLTEY
GRFKGRLAGNTTGNVLLRRSQHEALNQAAATLDIARYVVAGKLQNTRNVLMRAAREAKAECDRSQLAQAAATH
AEAIRNTEKVADLDKLRGIEGYAAKAYFGAFTAMVRQNRDVFAFTARSRRPPKDPINALLSFVYTLLVSECVA
ACEGVGLDPQMGYLHALRPGRPALALDLMEELRAPLGDRLVLTLINRGQLTPEDFIERPGGAIHLTKEGRATL
LAAYQKRKQEEVFHTVIDAKVPLGLMSHTQARLLARHLRGDVPTYQPFILR
>PCC6406_01802   Cas_Cas1
MGTIYISTDDAFIGKTDERLRVRAQKETLLDVPLIKVDGVVVLGRATVSPAALIELLERKIPLSFMTGTGRFL
GRLEPELNKNIFVRSAQWQAAGETERSIHAVQGFIRGKLKNYRTLLMRSQREGITDLEPGITRLEQAIQPINQ
TAIIDALRGLEGAGSAAYFGHFGRLIRNDDFSFTTRCRRPPTDPVNALLSFGYALLRHDVQGALNIVGFDPYL
GYLHTQRYGRPSLALDLMEEFRPLVVDAMVLAALNRKALSLEDFTPEPISHAVSLTDEGRRKFLRLYEQKKQS
KFKHPVMGRQCTYQEAFEIQTRLLAKYLMAETDQYPPLILK
>PCC7108_00839   Cas_Cas1
MGTLYIVQDDAFVGKVDERINVKFEKKILQDIPFIHLECVVVLGRATISPAVVDELMQRHIPLSFINKIGHYL
GRLQPEMSGNIFVRKAQWQAAGETPQSIHVVQGFVRGKLKNYRQILMRRQRESDDVDLSKNITRIEQIIDAIK
STENINSLRGLEGSGSAAYFGGFNQLIRNPEFSFPKRVRRPPTDPVNSLLSFGYSLLRHDVQSAVNIVGFDPY
LGYLHFDRYNRPSLALDLMEEFRPLVVDAIVLSILNNQLLTPTNFIAEPLSGAVSLTPEGRKTFLTLYEKKKI
SEFKHPVMGRKCSYREAFELQARFLAKYLMGTIDKYPPLFLK
>RoaringCreek_04223   Cas_Cas1
MGTVYVSQDDSFIGKTDERLTVKAEKKQILDVPLIKIDGLVILGRATISPAAVMELLERKIPMSFMTGTGRFL
GRLEPELTKNIFVRRSQWDAAGETPKAIHMVQAFVRGKLKNYRASLMRHLRDYPDNDLQKAIDDLERAIDSLA
AIKAIASLRGVEGHGSAVYWQAFPKLIRADGFSFTTRNRRPPIDPVNAMLSFGYSLLRHDIQGALNIVGFDPY
LGYLHTERYGRPSLALDLMEEFRPLIVDAIVLSAINRKAIAPKDFTFEPLSKAISLSNEARKVFLTLYEQKKQ
SQFKHPVMGRQCTYQEAFELQARLLAKYLMDETDKYPPLVLK
>TJSD091_01108   Cas_Cas1
MTAIYLTEPQSYLEADLRDFAIVYNDRIQSRHSIDTTQQIIAFEGCTLGRSAIAAIYRYRIRLSFIRDNGHIT
AQITPQETPFYTVLQQERNSDRNFRIAFIHSLLRGRWYNATKVLQSLGYSGIRECLEQQLEVLPGLNSWLALQ
TWQVQVFQRYQKALRSVLVSVKDFEPTLELAYALLSRELYVLLLSSGCCAEVGTLHLHCQNHLPLPCDLMEPF
RPVVEAWVFKKAPKFSDRGGLSQSYRATFVQNWESLMATQIFHAYAGQVSLRSALAWQVEEYVRAITETLEYR
PFLLKS
>TJSD091_01107   RVT_1+Cas_Cas1
MGDALLAAWYQVRTRSKTAGIDGISVELFAHFLDEELKRLAHQLQQDIYHPLPALGFYRAKKSGGHRLIGIPT
VRDRIVQRLLLRSLYPALEETFQDCSFAYRPGVGVKHAIERVAEVYSSQTWTVKADISQFFDSLCRTLLLSQL
EELSVDQTVVRYIKGQLEAGIVVGGMPILSGRGVLQGGILSGALANLYLSEFDRRCLNAGASLTRYGDDFVMV
ARSLLEATRFLNLIEDWLSDIYLTLQPEKTHIFAPGEEFVFLGYGFQAGEIITPEYKQPRRNHNKKAVSVASR
PPIACSIVTPTRKSLSTRNFLDSWREGMTTLYVTDQGAYVKVKHQQFQVLLGNDLKVSIPVNVVDYIILFGCC
NLSHGAIGLALRRRIPILFLSYQGRYFGRLQTDGMTRVDYLSRQVHCAEDETFVLRQAKVIVAGKLHNCRILL
RRLNRDRQISQVIEAIEELGVWQEKIAEVELLESLLGYEGFGTRIYFQALGALVQPPFTFEHRTRRPPTDPVN
SLLSLGYTLLHQNIHSLILAVGLHPHYGNLHVPRSNHPALVSDLIEEFRAPVVDSLVIYLVNSGIFTPEDFTP
SDERGGVYLYSDALKKYLKHWQDKLSLKTTHPHTGYKVSYYRCLELQVWEYISCLMGEREVYRPMKLEKW
>TJSD091_01721   Cas_Cas1
MTTLYLTAPRTVVRYHNQTLIIEQSKGSTTCRLAEIDLVVILPGVQLTDRAIAQLLDRGIETIFLRQDGRFRG
RLQGHFHTNPTVRLAQYRALDTTFGMALAQKLILGKVRNQRYLLQRRNRETHGKITQLTEAIDTIASYLPRLK
QTETPLARDELMGVEGLCARIYFQALTHWFPPEWKFTGRNRRPPKDPVNALLSWGYGVLLARVFAATVQAGFD
PYLGFFHAIAPFRPNLVLDLMEEFRPVVVDRVAIAIVNSRLIDPNDFEPAPGEEEGIWLGQTAKKLYLGHLER
QLQDSILYPPQHRKLKLSQILLEQARSLGRCLVETRLDYEAFIFR
>PCC9228_02496   Cas_Cas1
MSVLYVLQPDAILNKQYEAFTVALQQEDGTWNKRSVPVQTVEQIVLMGNPQITGDALTYALELGVPIHYLSGF
GKYLGSALPGYSRNGQLRLAQYEAYFNTTRRLEIAKTIVAGKLRNQQEVLSRSGCTDNSLKALRKKVWLQSTL
DEVRGIEGIAAKEYFGFLSQIIKSPWTFSGRHRRPPTDPVNALLSFAYGLLRVQVTSVAHMAGLDPYIGYLHE
```

```
ANRGQPAMVLDLMEEFRPLVADKLVLAILNNKEIQPQDFEESLGAYRLSDSARKTFLQAFERKMKDDFKHPTF
EYRCTYRRGIELQARLFARHLQEDLPYKALALR
>PCC9228_01207   Cas_Cas1
MGTVYVSQENAFIGKTDECLKVKAETPLEVPLIKIDGVVVLGRATVSPSVVNELLERHIPLSFLTQTGRYLGR
LEPELTKNIFVRQAQWQAAGHTEKALHVVRGFVRGKLKNYRNILMRRERSHEGCDLQNAIAGLEKAIAPIDQT
TSIDALRGLEGSGSAAYFSAFAQLFSNEDFQFESRRRRPPTDPINALLSLGYALLRHDIQSAVNIVGFDPYLG
YLHMERYGRPSLALDVMEEFRPLIVDTMVFSALNKRWLKPADFTSEPLSGAVSLTKDGLKTFLRLYEQKKQSK
FKHPVLKRNCTYQEAFEIQARLLAKYLMAEIEMYPPLVMK
>PCC9228_03691   Cas_Cas1
MTTLYLTEPGTTVGYSNQVLTVSRKEKTTSCRLAEVHLVVVLPGIQLSHSAISAFFDNGIETIFLRRDGRFRG
RLQGQFATNPTVRLAQYQKVNTTFGRALAQKLVFGKVCNQRVLLQQRNRKSRGRISQLTEAIDILSAYQQQLK
ASDTPLQEDEIMGIEGICARTYYQALRHWFPEKWGFTGRNRRPPRDPINALLSWGYGVLLARVFAACVQGGLD
PYLGFFHAIAPYRPNMVLDLMEEFRPVIVDRATIRLLEADLLTWEDFEPSPDGAGIWLGQTAKKLFLAQLEQQ
FQTKYRYPPQNQRRLALNQILLEQARWMGRCLASSELDYEAFVLR
>PCC9228_01055   Cas_Cas4+Cas_Cas1
MKTLNSPLLKSTIRVSALHALAYCPRLFYLEEVEELYTQDAAVFAGRRLHAEIDKQEDEEWEDLYLESEQLGL
RGRVDALRTRDRALIPYEHKRGRCHRDTNRQPQAWESDRLQILAYACLIEFALETTVREGCIRYHADNVLVRV
SVDDEGREAVRAAVAQARELLVSTQRPPVTENERLCTRCSLSPVCLPEEARLSHNQEWQPMRLFPKDDEREIL
HVLEAGTSVGRTGEQLKLVRRGQPVETVSIQQVSQLVLHSFSQISTQALHLCASRDIGVHFISGGGRYVGSVD
RRQGSIQRRIRQYGAFVDSDRCLQLARKLVACRGNGQRKLLMRGKRNLSEVPASLERGIRQMKTLLRQAERVD
SIESLLGLEGNLAAIYFQNLAHLISSKVPPELHFRDRNRRPPKDRFNALLSFGYALLLKDVNNAILTVGLEPA
LGFYHQPRSQAPPLALDLMEIFRVPLVDMVVLGSVNRCQWDCQVDFEVRGQQVWLSEPGRRKFVSLYEQRKAE
TWKHPVTKYSLTYRRLLELEVRLLEKEWMDEEGLFGQLVLR
>ULC007_03669    Cas_Cas1
MGTVYITQDDAFIGKTDERLQVKAQKKTLLDVPMIKVDGVVVLGRATISPAAMIELLERKIPLSFLTGTGRYL
GRLEPPLTKNIFVRSAQWKASGESAQAVHVVRGFIRGKLKNYRNTLLRAQRETVELNLQAAIARLEQTIAPIN
NTEKIDSLRGLEGAGSAAYFGSFAQLIRVKEFCFETRNRRPPTDPVNALLSFGYSLLANDVQSAVNIVGFDPY
LGYLHVERYGRPSLALDLMEEFRPIVVDAMVLSVINRRVLTPDDFTTEPLSGAVSLSDEGRRSFLRAYEQKKQ
SEFKHPVMGRKCSYQEAFEIQARLLGKYLMGETEQYPPLILK
>ULC007_02338    Cas_Cas1
MESLYISQQGCHVSLRQDCILVKQGKTLLQEAQLPMLEQILVFGQSQLTTQLIRSCLFRNIPIAYLSRMGRCY
GRVLPVTSGYRSLARQQYALSATTRLAIARLIVGAKLKNSRVILQRQHRRQASAELLGAITQLAQLAQQVTQA
TSIEQLMGYEGAGAASYFGVFGACVEQSDFKWAGRSKRPPGDPMNAMLSFGYQVLWNHLLSLIELQGLDPYEA
CLHQGSERHPALASDLIEVFRAPIVDSLVLYLVNRRMVNAVDDFEFPATGGCYLNESGRKKLLTALIMRMEEQ
VQVAPGVMAPRWELLNRQVKRFRQAVIQVSSEYESYQIR
>AWQC131C_03408  Cas_Cas1
MGTLYITQDDAFIGKVDERINVKFDKKIIIDVPFIKLDGVVVLGRSTISPAVVSELMIRHIPLSFIDERGHYL
GQLQPEMSGNIFVRKAQWQAAGNTPQSIHVVQGFVRGKLKNYRQLLVRRQRESNHLDLSADITRLENIIAAIN
KTESIDSLRGLEGAGSAIYFGCFNQLILNPNFEFTKRVRRPPTDPVNSLLSFGYSLLRHDIQSAVNIVGFDPF
LGYLHFDRYNRPSLSLDIMEEFRPLVVDSLVLTLLNKQLLKPDDFIREPLSNAVSLTAEGRKIFLTNYEKKKQ
SDFKHPVMGKKCTYRESFELQARLLAKYLMGTTDKYPPLFLK
>AWQC131C_02364  Cas_Cas1
MQTLYVSEQNCYVCLQKETLLLKQGDTVYIEVQLPLLEQILIFGQSQITTQVIRACLWRDIPIAYLSKMGYCY
GRILPISRGYRQLSRYQQQLSPIDRLITARELVKGKLKNSRVLLRRQQKKRESELIERVLESLDYLAAQAATA
DNLERLMGFEGAGAAQYFSAFGECLTNSDFVFSGRSRRPPGNPVNAMLSFGYQVLWNHLLALIEIQGLDPYYA
CLHQGNDGHAALASDLIEEFRAPLVDSLVMWLINRKIMDVKADFEFKNGGCYLNNLGRKKFLRGFLQRMTEEI
KTDQEIKQPKWDLLTQQVKAFKQFVYNPSHHYQCYRID
>UU774_00989     Cas_Cas1
MTTTLYLTEPGTMVRYRNESLVIIRKEQSKNCRLAEVDLLVVLPGVQFTDVAISKLLDKSIETIFLRQDGHFR
GRLQGNYATNPAIRISQYRILESAFGMALAQKLVMGKIRNQRVLLQRRNRASQGQIAELAEAINLINTYSYNL
HNNSTLLNRSELMGVEGICARTYYQALRYWFPPQWNFTGRNRRPPLDPINALLSWGYGVLLARVFSACVQVGL
DPYLGFFHATEPYRPNLVLDAMEEFRPVIVDQAVISIIQSDVLTPEDFHTSPDGVGVWVGDVAKKLFLAELEK
QFRNTVLYPPQNRHLSMSQIILEQARSLGRCLVELSLEYEAFVIK
>UU774_05881     Cas_Cas1
MAAIYLIEQGTNIYKDYQRFIIHVSEKPKLEILIREVQQILVFGNIQLSTPVLQVCLQEQIPVLFLSQNGRYH
GHLWSEESTNLDNQVIQIERRNDDYFQFSVSKAVVLGKLMNSKQLLMRFNRKRKLEDVEKAIYGINQDIDALE
YVDNLDSLRGYEGIAAARYFPVFGQLITNSAFSFSQRFRQPPIDEVNSLLSFGYTLLFNNVLSFIIAEGLSPY
IGNFHYGERQKPYLAFDLMEEFRSPIVDSLVLKVINKSLFKPQDFDVVASTGGVYLSQSARKVFLNQFETRMN
EEVSHPDLISQVTYRHAIQLQVRRYKRCLLSGVPYESFLRAV
```

```
>UU774_03492      RVT_1+Cas_Cas1
MFTLEQLLFAWVQVRAGSRSAGVDGITVDLFAASADQQLRIILRQLQQKSYRASPAKGFYLTKKSGGKRLIGI
STVRDRIVQRLLLEELYLPLEDTFVDCSYAYRPGCNIQQAVQRLFSYYQYHPTWIIKADIAQFFDNLSWALLF
TGLETLHLEAIVLELLEQQIKSGIVLGGKYINFGKGVLQGGILSGALANLYLTAFDRKCLSHGINLVRYGDDF
AVACSSWTEANRILDKITTWLGGLYLTLQPEKTQVFAPHEEFTFLGYRFVNGKVYAPPTPEPRREGEWLVNAS
GTPYFRPKQRPLKFVSRPPKACDISKPINFPRAPISHLWQEFMTTLYVTDQGAYLSLKNQQFQVFYQGQLRIK
VPIVRVTNIVLFGCCNVSHGAVSMALRRRIPIMYLSQKGRYFGRLETEGQAKVEYLARQVECSRDDKFTRQQA
EVIVWAKLHNSRALLLKLNRRRPSKIALSAINLIADLMDNLPQAESMDALRGYEGKAATLYFQALGSLFTGVF
AFDKRTRRPPTDPINSMLSLGYTLLSQNVYSFIQAMGLHTHFGNLHVPRDNHPALVSDLVEQFRAQLVDSLIA
YLVNSNIFTIEDFTNPDERGGVYFQPHALKKFLKHWEEKLQSEITHPHTEHKVAFRRAIELQVREYIACLMGE
VEVYRPMIWKL
>UU774_03490      Cas_Cas1
MSTIYITEQDAMLKIQHHYLKVFHQQKQCISIRISNVSQIIIFGKIPLPREVIKTVLAHQIPVLYLTQDSEYI
GRLENTSQKPAKYLIHQRRRARDVEFNRATAESIIWAKLHNQNIFLQSWTRHYTSDATQRASNYLMLLMDNLA
LATSLDELRQYSEEADNVYYCAIGSLLNFYNGCPQTTAKRIRRFLNLGNQLLHQYIYTLVNTAGLHPDYAILH
YDTDHELPLVWDFMAEFHAPIVDDLVLNFARHLSNSHSNGNGKTQPRSLLQRFLQYWEAKLRTFVLHPYAGEV
SYRQCIDLQVREYLACLLGDIEYYRPLALKFHPNDPSFTNIAETHKPALTLVKK
>Ana90_04894      Cas_Cas1
MGTLYITQDDAFIGKVDERINVKFEKKILQDIPFIHLEGVVVLGRATISPAVVDELMTRHIPLSFINKIGHYL
GRLEPEMTGNIFIRKAQWQAAGETPQSIHAVQGFVRGKLKNYRQTLMRYQREFNDLDLSKSIARIEQVIAAIS
STENINSLRGLEGSGSAAYFGCFDSLIRNSKFSFTKRVRRPPTDPVNSLLSFGYSLLRHDVQSAVNIVGFDPY
LGYLHFDRYNRPSLALDLMEEFRPLVVDAVVLSILNKQLLTPADFVTEPLSGAVSLIPEGRKTFLTLYEKKKL
SEFKHPVMGRKCTYREAFELQARLLSKYLMGTTDKYPPLVLK
>CS508_02369      Cas_Cas1
MQTLYVSEQNCYICLQKEILVVKRGEVTQAQVQLPLLEQILIFGSSQVTTQVVRACLLRDIPIAYLSRMGYCY
GRILPISRGYRQLSRYQQQLSAMEKLVTAREIIKSKIKNSRVLLQRQKKKIASANLERVLESLNYLAEQTASA
DSCERLMGIEGAGAAQYFSAFGECLSNENFIFLGRSRRPPGNQVNAMLSFGYQVLWNHLLALIEIQGLDPYYG
CLHQGHDGHPALASDLIEEFRSPLIDSLVMWLINRNIFDVGADFEFIKGGCYLNNSGRKKFLQAFLQRMTEEV
QTEDGTKQPKWDLLTQQVRAFKQFVYNPAHHYRPYRIH
>CYA406_04973     Cas_Cas1
MPTIYVIHPGSSIDVHSRYLQVFHKQQLCLGVPINRVTQIILFGICNLSRKTQVLIATRQIPVYYINVHGDQL
GSLELTPPRPFKYQTHQLICSQDLEFVRAIAESITWAKLQNEIIVLKQLQPEFYPCNIENTINFLQLLADDLP
SANSIEELWDYHATATSYYYPILASLVPQWSNPHSSFRLDLSHPVLQMLKLGYTFLHQQMALFINDLGLDPEI
SNLHTLNQTSQPLVDDLIEEFRPILVDSLVVKLLQTGTITSSDFVPPNGSRGGYLQPSALQTFINHWEEKLAT
HIEHPFAGTISYRQCFELQAKEYIACLLGDIEFYQPLLMQQEKGKIKKMYAKEYLTQKNVVQALSGKD
>CYA406_04972     Cas_Cas1
LHPSPTDPINSLLSLGYTLLSQNVCSFIELAGLHTHFGNLHVPRDNHPALVSDLMEEFRAQIVDSFVSYLVNS
KILLLEDFTSPDELGGVYLQPDALKKYLKHWEEKLYSEVTHPQTGYKVSLRRCIELQVREYIAALMGEVEVYR
PMIWKN
>CYA406_01986     Cas_Cas1
MSILYITQPQAILGKKYEAFTVATQEESGKWNKQSVPAQTVEQVVLMGNPQVTGDALTYALELGMPVHYLTSF
GKYLGSALPGYSRNGALRLAQFELYQDTERRLELVKTIVSAKIQNQATVLYRQKIADSPLKNHKKLVFKKATL
DEVRGVEGIAAKDYFREWSAILNEPWFFPGRNRRPPTDPVNALLSFAYGLLRVQVTASVHLVGLDPYIGYLHE
VSRGQPAMVLDLMEEFRPLVADSLVFSVLNKREIQPDDFTESLGAYRLKDEGRKRFLQAFERKMNDEFKHPVF
EYRCTYRRAIELQARLLGRYLQEDLVYKPLVLR
>JSC12_04766      Cas_Cas4+Cas_Cas1
MSTTITLAQEQDTLRVNALHALAYCPRLFYLEEVEELYTQDAAVFAGRRLHAEIEKEESEEWQELVLENNELG
LRGKVDALRTRNGQTIPYEHKKGRCYRDASNKPQAWQSDRLQILAYACLIEAALGITVPEGRIRYHADNVLVH
VPLDDQGRQEVREAIQQARQLRLTTERPPITTNERLCARCSLAPVCLPEEVRRAQGVTQQPIRLFPEDDEREV
IHVLEAGTCVGKSGEQLKITKRDKSTQALPARQVGQLVIHSFAQISTQALYFCAEQDIGVHFVSGGGRYLGSF
EPRAGSIQRRVRQYQALSQPDTCLKLARQLVACRGQSQRKFLMRGQHGSETVLELEQTIAQMQRILKQVPQAD
SLSSLLGLEGNLAALYFKGLPQLLSPGAAPEFQFSGRNRRPPQDRFNALLSFGYSLLLKDVMNAILVVGLEPA
LGFYHQPRTQAPPLALDLMEIFRVPLVDLPVMGAINRGQWDAQADFDVRSGQVWLSESGRRKFIALYERRKEE
SWKHPVLGYSLTYRRLLELEVRLLEKEWMGEGGLFGQLVIR
>JSC12_00903      Cas_Cas1
MSILYVTQPDAVIEKTYEAFTVSLKQEDGSWQKKTIAAQTLEQVVLMGNPRITGDALVYALELGLPVHYLSSF
GKYLGGAMPGYSRNGQLRLAQYSTYFDEAKRLEFVKAIVTTKIHNQYGVLYRHGQKENPLKLRKGAVKTQATI
DQVRGVEGLAAREYFGTWSKMLDARWNFAGRNRRPPTDPVNSLLSFAYGLVRVQVTAAVHMAGLDPFIGYLHE
```

```
VSRGQPALVLDLMEEFRPLVADSVVLTVINNREIQPKDFTQSLGAYRLSETGRKTFLQAFERKMMDEFTHPVF
QYRCTYRRAIELQARLLAKYLQEGVKYVPLAVR
>PCC7107_02882   RVT_1+Cas_Cas1
MFTQEHLHFAWLQVCAGSKTAGVDGISVELFESMATEQLQNLVYQLNNETYTASPAKGFYIPKKNGDKRLVGI
PTVRDRIIQRLLLDELYFPLEGTFLDCSYAYRPGHNILQAVQHLYGYYQYQPKWIIKADVADFFDNLSWALLL
TFLEKLSLEPSVLQLIEQQLQSGMIIAGQYRNFGKGVLQGGILSGALANLYLTNFDKKCLSQGINLVRYGDDF
VIACNSWQEANRILDKITVWLGEVYLTLQSEKTQIFTPNDEFTFLGYRFAGGEVYAPPPPKPVLKGEWVINDS
GNPYFRTKPRPKKPVSHPPKACSIDKPINFPRASLSHYWQETMTTLYITDQGAYLSVKNQQFQVYYQGQLRIK
VPVSRVSNVVLFGCCNVSHGAVSMALRRRISIMYLSQKGRYFGRLQVSGDAKVEYLMLQVERCQNPEFTRTLA
EVIVTAKIHNSRILLMRLKRRKSSEFDDNLVKEAIKDLDILMDKLPFAESMDALRGYEGRAATVYFQALGSLF
SGAFKFEKRTKRPPTDPVNSLLSLGYTLLSHNIFSFIQAVGLHSHFGNLHVPRDNHPALVSDLMEEFRAQLVD
SLVAYLINSNIFTEDDFTPPDEKGGVYLQPHALKKFLKHWEEKLQSELTHPNTGYKVSFRRCLELQVREYVAC
LTGEVKIYRPMIWKP
>PCC7107_02880   Cas_Cas1
MSTIYITEQDVTFQIQHQYLKVFYQQTQRICIPIRNLSQFIIFGNINLPKEVIQIVRSHNIPVLYLTQTGEYL
GRLENPSQRQAKYLTYQHRRAQDKEFNRATAESIIWAKLHNQHTFLQNWTRHYTNYIIQRALNYLTLLMDNLS
VAPGIEELHEYSEEADNIYYCALTALLSFYNGSPYINSKRISGLIDLGNQLLHQYIYTLLNTVGLDPNYAILH
RNNHNELPLAWDFTAEFRTLMVDDLVLNFIRNLANTNGNGNGKSHSHKLLQKFLQHWEGKLRTFVLHPYAGEI
SYRQCLDLQVREYLASLLGDVEFYRPLALKFHPANSDFTNTIKPQTVPLKLVKR
>PCC7107_01473   Cas_Cas1
MGTVYISQEDAFIGKIDERLSVKFEKKIILDVPLIKIDGVVILGRATVSPAVISELLQRSIPLTFLTENGRYL
GRIEPEVTKNIFVRKAQWQAAGESPQAIHVVQGFVRGKLKNYRQILTRRQRESSDIDLSKKITQLEQAIAPID
TTHNINSLRGLEGAGSAAYFGCFNELIRNVEFQFSHRLRRPPTDPVNSLLSFGYSLLRHDVQGAINIVGFDPY
LGYLHCDRYGRPSLALDLMEEFRPLVVDAVVLSTLNKQLLTPADFITEPLSGAVSLTPEGRKTFLRLYQQKKQ
SEFKHPVMWRKCTYQEAFELQGRLLAKYLMGETEKYPPLVLK
>CRKS33_01557    Cas_Cas1
MGTLYITQDDAFIGKVDERINVKFDKKIIIDVPFIKLDGVVVLGRSTISPAVVSELMIRHIPLSFIDERGHYL
GQLQPEMSGNIFVRKAQWQAAGNTPQSIHVVQGFVRGKLKNYRQLLVRRQRESNHLDLSADITRLENIIAAIN
KTESIDSLRGLEGAGSAIYFGCFNQLILNPNFEFTKRVRRPPTDPVNSLLSFGYSLLRHDIQSAVNIVGFDPF
LGYLHFDRYNRPSLSLDIMEEFRPLVVDSLVLTLLNKQLLKPDDFIREPLSNAVSLTAEGRKIFLTNYEKKKQ
SDFKHPVMGKKCTYRESFELQARLLAKYLMGTTDKYPPLFLK
>CRKS33_00155    Cas_Cas1
MQTLYVAEQNCYVCLQKETLLVKQGDTVYIEVQLPLLEQILIFGQSQITTQVIRACLWRDIPIAYLSRMGYCY
GRILPISRGYRQLSRYQQQLSPIDRLITARELVKSKLKNSRVLLRRQQKKRDSELIERVLDSLDYLATQAATA
DNLERLMGFEGAGAAQYFSAFSECLTNSDFVFSGRSRRPPGNPVNAMLSFGYQVLWNHLLALIEIQGLDPYYA
CLHQGNDGHAALASDLIEEFRAPLVDSLVMWLINRKIMDVKADFEFKNGGCYLNNLGRKKFLRGFLQRMTEEI
QTNQEIKQPKWDLLTQQVKAFKQFVYNPSHHYKSYRID
>CRKS33_03258    Cas_Cas1
MTTLYIIQPDAILSKDYEAFQVALKQEDGSWKKQKIAAQTVNEVILMGNPQVTGDAFVYALELGMPVHYLSSF
GKYLGSALPKSSRNGQLRLAQYAVYHDPVRRLELVKTIVSAKIHNQYHVLYRHHEKDNPLKERKTLVKTQQNI
DQVRGVEGLAAREYFACWQRMVGKQWTFNGRNRRPPTDPVNSLLSFAYALLQGQVMAGVHVAGLDPYIGYLHE
VHHGQPAMVLDLMEEFRALIADNLVLSLLHKHEIKPEDFNESLGAYRLKDNARKTFLQAFDKKMNDEFKHPVF
DYRCTYRRAIELQARLLSRHLQEGIPYKPLSLR
>OSCR_01298 Cas_Cas4+Cas_Cas1
MTTTTLSSTSTIRVCALHAFAYCPRLFYLEEVEELYTQNQAVFEGRRLHAELERQEEGEWDDFFLTDETLGLR
GRVDALRTRDGQLIPYEHKKGRCYRDANKQAQAWQSDRLQILAYCCLLETIAQHPIQEGRIRYHGDNTLVKIP
FNPETRAEVEAAVAAAQTLRQSQQRPPVTDNERLCASCSLAPVCLPEETRFTDDDTHQPPRLFPKDDDRQIVH
ILEPGTRVGRTGEQLKITRRNQPVETLSVRQVQQVVLHSFSQISTQALHLCTYHDIGVHWISGGGRYVGSLDW
RQGSLQRRCGQYRALTQPEICLDLAQRLVTCRAKGQRQFLMRGKRNLNGSTEAVASVVTSMKRLLKSIERAES
LEQLLGLEGNLAALYFSGLGHLVSAKVAPELQFNGRNRRPPKDRFNALLGFGYSLLLKDVINAILTVGLEPGL
GFYHQLRSQAPPLGLDLMEIFRVSLVDMPILASVNRSQWDVVADFDIRGDRVWLSEAGRRKFIEIYERRKGES
WKHPVLKYSLTYRRLLELEVRLLEKEWSGEANVFARSVVR
>OSCR_02854 Cas_Cas1
MLSLITDIPMTTLYLTRPQTTLTADRTTFTLHYHHLITGRYPISQIQQIICFEGTHLDRRSQREIQRQQIPLL
QLDRTGTICGRWNPPIAPLNRHASPSSHLDLAQWLAQYHLKEAIDLLSNLQGTPALAATRRVLALLHDSLNDA
ETLDLVRGYYATATAYYDRALRESLIAYLLQLNGSRPLMSLSQLLRVSESLLHHHLQLSLSRAYLDLNLGYLH
SPTDCDAPLVCDAVAVLRPAFVDGFVWRWLHRCGGGSRTWLGFCSAWDCQPLLGDRLEELVHSLSRWSDRTT
>OSCR_00822 Cas_Cas1
```

```
MGTMYVIHEDAFLGKLDERLQIKHQGKRLLDVPLLKLDGVVLLGRATVSPYVVDELLERNIPLSFLTRYGKYR
GRLEPEVTRNIFVRRGQWNAAGETPVSVVAVRGFVRGKLKNYRSLLQRKQRDSQDLDVTGAIDQLQGAIAALP
KTPTIDQLRGYEGAGSAAYFGVFNELLQGTPFEFTHRNRRPPKDPVNSLLSLGYALLRHDVQSALNIVGFDPY
LGYLHTERYGRPSLALDVMEEFRPLVVDAMVLSLLKKPNGLSLADFEQEPLNGAVSLTRDGLKQFLRAYEEKK
QSTFKHPVMGRKCTYQEAFEIQGRLLAKFLMGEIESYPPLVLR
>OSCR_02343 Cas_Cas1
MTTLYLTEPGTVVSVKNQSLDLKRGTWQRSCRLAEIELVVVLPGVQLTGAAIALLLDSGVETLFLRRDGRFRG
RLQGQFHSNPTVRLAQYRVVETTFGMGLAQKLVLGKIRNQRSLLQRRNREMEGRVTALVEAVDTMAVYLRQLQ
RLESPLSRDQLMGIEGICAKTYFQGAMALVPGSWGFSGRNRRPPRDPVNALLSWGYGVLLARVFSASVQAGLD
PYLGFFHATEPYRPNLVLDLMEEFRPLVVDWVVWGLLRSSLLEARDFEPSPDGEGVWLGHLAKKLFLEQLERR
FQSKLFYPPQERQLALSQILVEQGRAIGRACISLRLDEFESLEVR
>NIES3756_03630  Cas_Cas1
MSVLYVTQADAVLSKNYEAFNIALKQEDGSWKKQSIPAQTVEQVVLMGNPQVTGDALVYALELGMPVHYLSHF
GKYLGSALPGYSRNGQLRLAQYQLYNDPVRRLELVKAIVTAKIHNQYQVLYRNNERDNPLKERKSLVKSQKTI
DQVRGIEGLAAREYFACWPSMVGKQWTFTGRNRRPPTDPVNSLLSFAYGLLRVQVTAAVHVAGLDPYIGYLHE
VHHGQPSMVLDLMEEFRALIADSVVLTVLHKREIQPQDFTESLGAYRLKEAAKKSFLQAFERKMNDDFKHPVF
EYRCTYRRAIELQARLLSRHLQEGIPYKPLVLR
>PCC7429_00686   RVT_1+Cas_Cas1
MQNFSDKFLSPENFQLAWEKVAAKNGCAGVDRESVVHFAKNSEAYLSQLRRSLASGYYHPMPLRQLFIPKKAG
GWRELGVPTVRDRIVQHALLNILHPLLEPQFEACSFAYRPGRSHLSAVRQIAQWRDRGYEWVLDADVVRYFEN
ILWQRLLDEVAERLAAPEVLSLISAWLSVGVLSKEGLMFPQKGISQGSAISPILANVYLDDFDEIVTATGLKL
VRYADDFVVMSRSQKRIVEAKDEVADLMNGIGLQLHPDKTRIVDFDRGFRFLGHAFAGKVIVKTEGGREKREE
DMRGAVDPCPDGQLVYSDPQIKPTQMQKAMLEALKSSAQPIPPPLFVVLGYSLREVKPVKIESDEAIWTTGMS
TLYLVHQGATLRKEQGRFLVQPLKESALEIPIAEVELVLVFGNIQLTTSAIAACLDAKIPVIFLTQMGEYKGQ
LWNSEFCDLPSEEAQWQRRLDVAFQLETARAIIWGKLMNSKQLLLRLNRKRQLEDVTTAIAGITSDLESVETA
ETLESLRGYEGIAANRYFVALGLLITNEGFSLTGRTRRPPKDPVNSLLSFGYTLLYNNVLSLILAEGLNPYLG
NLHRSDRKETHLAFDLMEEFRSPVVDTLVMNLINKKILRPTDFTYPDKDGGIYLADAARRVFLKHFEERISLQ
IVYADLKEKISYRRVIQHQIWRYRSALLGEAPYESFRRND
>PCC7429_01751   Cas_Cas1
MGTLYITHDDSFIGKTDERLTVKADKKQLIDVPLIKIDGLVILGRATISPAAVIELLERKIPMSFMTGTGRFL
GRLEPELTKNIFVRRSQWGAAGETPKAIHMVQAFVRGKLKNYRASLMRHMRDYPDNDLQKAIGDLERAIDSLA
TIKAIASLRGVEGHGSAVYWQAFPKLIRADGFSFTTRNRRPPVDPVNAMLSFGYSLLRHDIQGALNIVGFDPY
LGYLHTERYGRPSLALDLMEEFRPLIVDAIVLSAINRKAIAPKDFTSEPLSKAISLSNDARKVFLTLYEQKKQ
SKFKHPVMDRQCTYQEAFELQARLLAKYLMDETDKYPPLVLK
>PCC8005_04815   Cas_Cas1
MTAIYLTEPQSYLEADLRDFAIVYNDRIQSRHSIDTTQQIIAFEGCTLGRSAIAAIYRYRIRLSFIGDNGHIT
AQITPQETPFYTALQQERNSDRNFRIAFIHSLLRGRWYNATKVLQSLGYSGIRECLEQQLEVLPGLNSWLALQ
TWQVQVFQGYQKALRSVLVSVKDFEPTLELAYALLSRELYVLLLSSGCCAEVGTLHLHCQNHLPLPCDLMEPF
RPVVEAWVFEKAPKFSDRGGLSQSHRATFVQNWESLMATQIFHAYAGQVSLRSALAWQVEEYVRAITETLEYR
PFLLKS
>PCC8005_04814   RVT_1+Cas_Cas1
MGDALLAAWYQVRTRSKTAGIDGISVELFAHFLDEELKRLAHQLQQDIYHPLPALGFYRAKKSGGHRLIGIPT
VRDRIVQRLLLRSLYPALEETFQDCSFAYRPGVGVKHAIERVAEVYSSQTWTVKADISQFFDSLCRTLLLSQL
EELSVDQTVVRYIKGQLEAGIVVGGMPILSGRGVLQGGILSGALANLYLSEFDRRCLNAGASLTRYGDDFVMV
ARSLLEATRFLNLIEDWLSDIYLTLQPEKTHIFAPGEEFVFLGYGFQAGEIITPEYKQPRRNHNKKAVSVASR
PPIGCSIVTPTRKSLSTRNFLDSWREGMTTLYVTDQGAYVKVKHQQFQVLLGNDLKVSIPVNVVDYIILFGCC
NLSHGAIGLALRRRIPILFLSYQGRYFGRLQTDGMTRVDYLSRQVHCAEDETFVLRQAKVIVAGKLHNCRILL
RRLNRDRQISQVIEAIEELGVWQEKIAEVELLESLLGYEGFGTRIYFQALGALVQPPFTFEHRTRRPPTDPVN
SLLSLGYTLLHQNIHSLILAVGLHPHYGNLHVPRSNHPALVSDLIEEFRAPVVDSLVIYLVNSGIFTPEDFTP
SDERGGVYLYSDALKKYLKHWQDKLSLKTTHPHTGYKVSYYRCLELQVWEYISCLIGEREVYRPMKLEKW
>PCC8005_03720   Cas_Cas1
MTTLYLTAPRTVVRYHNQTLIIEQSKGSTTCRLAEIDLVVILPGVQLTDRAIAQLLDRGIETIFLRQDGRFRG
RLQGHFHTNPTVRLAQYRALDTTFGMALAQKLILGKVRNQRYLLQRRNRETHGKITQLTEAIDTIASYLPRLK
QTETPLDRNGINGS
>PCC8005_01698   Cas_Cas1
VVRYHNQTLIIEQSKGSTTCRLAEIDLVVILPGVQLTDRAIAQLLDRGIETIFLRQDGRFRGRLQGHFHTNPT
VRLAQYRALDTTFGMALAQKLILGKVRNQRYLLQRRNRETHGKITQLTEAIDTIASYLPRLKQTETPLDRNEL
MGVEGICARIYFQALTHWFPSQWKFTGRNRRPPKDPVNALLSWGYGVLLARVFAATVQAGFDPYLGFFHAIAP
```

```
FRPNLVLDLMEEFRPVVVDRVAIAIVNSRLIDPNDFEPAPGEEEGIWLGQTAKKLYLGHLERQLQDSILYPPQ
HRKLKLSQILLEQARSLGRCLVETRLDYEAFIFR
>PCC7104_05900   RVT_1+Cas_Cas1
MNSLQNRFLDPTNFELAWDKVATNQGCAGVDGETIYAFGLHKSRNLTRLLQQVATSTYRPLPLRQFFIPKKSG
GWRELGVPTVRDRIVQQALLQVLHPVFEVEFEPQSYAYRPGRSHRMAVERVAHWRSRGYDWVLDADIVKYFDT
LQHPRLLAEVKERLNQPWVLALLQGWITAGTLTREGILLPTCGVPQGSPISPLLANVYLDDFDELLTQAGHKL
VRYADDFVVLARTQQRLVEAQTYVAQLLEGMGLSLHPNKTQITTFDRGFRFLGHAFAGDVVVPLSPRRPQALT
PNPSPNLGEGSTMKLTYADGPAQSTAIQQALVEALKRSQQPIPPPLFVVLGYQVRADVRVALESQEITWRNGM
SSLYVVEQGAYLHKDQGRLVLKAPRAKDSLEIPLAEVERILLFGNVQLTTAVIAACLQQQIPVIFLSQLGDYK
GHLWSAEITDLTAEAEQFARQHDEAFGCTTARAVVYGKLWNSKIFLLRQNRKRQFAEVRTAIERLDGALTTLA
GDAPLSLEQIRGYEGNGASEYFQTFGPLITNPGFSWAGRNFHPPTDPVNSLLSFGYTLLFNNVFSLLLVEGLN
PYLGHLHGAERQKAYLAFDLMEEFRSPVVDALVMRLINQKIIRPTDFSWPKENNGVYLTDPARRVFLRHFEQR
ITEKVTHPDVKEPVSYRRVIQLQVKRYKRAVLGNQPYKAFERMR
>PCC6306_06056   Cas_Cas1
VKTLYVSHQGCYLSLRQESVLIKRGETILDQVQLPLLEQILIFGQSQVTTQVIRACLQRNIAIAYLSRMGKCY
GRVLSITQGYRQLSRYQQLLSPELQLQTARIIVQAKLKNSRVILMRQQRRRASEQTALAIESLAYLVQQAGEA
VSIARLMGLEGAGAASYFSAFGECLSGAGFVFLARSRRPPGNPVNAMLSFGYQVLWNHLLTLIELQGLDPYWG
CLHQGHDRHAALASDLIEEFRSSIVDSLVLYLVNRGLVSADDDFEYHDGGCYLNQSGRRKFLQAFVQRMEESV
EEQPRWNLLSQQVRAYKQFVYDPSCLYSPYLIR
>PCC6306_00767   Cas_Cas1
MEGLAAREYFGYWSQMLKGNWQFSGRNRRPPTDPVNALLSFAYGL
>PCC6306_00766   Cas_Cas1
MYALELGMPVHYLSGFGKHLGSALPAYSRNEQLRLAQYRRYEDQAQKLELVKAIVTTKIYNQATIFYRHGVKE
HPLKQRKASVAAQKIDQVRGWKGWQRGSILVIGRRC
>C1_03964  Cas_Cas1
MITKRTVIIEQSKGSTTCRLAEIDLVVILPGVQLTDRAIAQLLDRGIETIFLRQDGRFRGRLQGHFHTNPTVR
LAQYRALDTTFGMALAQKLILGKVRNQRYLLQRRNRETHGKITQLTEAIDTIASYLPRLKQTETPLARDELMG
VEGLCARIYFQALTHWFPPEWKFTGRNRRPPKDPVNALLSWGYGVLLARVFAATVQAGFDPYLGFFHAIAPFR
PNLVLDLMEEFRPVVVDRVAIAIVNSRLIDPNDFEPAPGEEEGIWLGQTAKKLYLGHLERQLQDSILYPPQHR
KLKLSQILLEQARSLGRCLVETRLDYEAFIFR
>C1_00162  RVT_1+Cas_Cas1
MGDALLAAWYQVRTRSKTAGIDGISVELFAHFLDEELKRLAHQLQQDIYHPLPALGFYRAKKSGGHRLIGIPT
VRDRIVQRLLLRSLYPALEETFQDCSFAYRPGVGVKHAIERVAEVYSSQTWTVKADISQFFDSLCRTLLLSQL
EELSVDQTVVRYIKGQLEAGIVVGGMPILSGRGVLQGGILSGALANLYLSEFDRRCLNAGASLTRYGDDFVIV
ARSLLEATRFLNLIEDWLSDIYLTLQPEKTHIFAPGEEFVFLGYGFQAGEIITPEYKQPRRNHNKKAVSVASR
PPIACSIVTPTRKSLSTRNFLDSWREGMTTLYVTDQGAYVKVKHQQFQVLLGNDLKVSIPVNVVDYIILFGCC
NLSHGAIGLALRRRIPILFLSYQGRYFGRLQTDGMTRVDYLSRQVHSAEDETFVLRQAKVIVAGKLHNCRILL
RRLNRDRQISQVIEAIEELGVWQEKIAEVELLESLLGYEGFGTRIYFQALGALVQPPFTFEHRTRRPPTDPVN
SLLSLGYTLLHQNIHSLILAVGLHPHYGNLHVPRSNHPALVSDLIEEFRAPVVDSLVIYLVNSGIFTPEDFTP
SDERGGVYLYSDALKKYLKHWQDKLSLKTTHPHTGYKVSYYRCLELQVWEYISCLMGEREVYRPMKLEKW
>C1_00161  Cas_Cas1
MTAIYLTEPQSYLEADLRDFAIVYNDRIQSRHSIDTTQQIIAFEGCTLGRSAIAAIYRYRIPLSFIRDNGHIT
AQITPQETPFYTVLQQERNSDRNFRIAFIHSLLRGRWYNATKVLQSLGYSGIRECLEQQLEVLPGLNSWLALQ
TWQVQVFQRYQKALGSVLVSVKDFEPTLELAYALLSGELYVLLLSSGCCAEVGTLHLHCQNHLPLPCDLMEPF
RPVVEAWVFEKAPKFSDRGGLSQSHRATFVQNWESLMATQIFHAYAGQVSLRSALAWQVEEYVRAITETLEYR
PFLLKS
>KIOST1_03080    Cas_Cas1
VLSLLHLEGLNPYLGNLHRNDRHEAHLAFDLMEEFRSPIVDTLVLTLLNQQVLKPDAFLPPKANGAVYLTDDA
RRRYLQAFEDRIMTTLHHPDARDPVPYFHSAQRSCNLVLKAVTLWGREFAGSSARVAAMLT
>KIOST1_04599    RVT_1+Cas_Cas1
MNLYREQLATAWHLVERGSQAAGVDGMTVDLFRGIAQEQIRLLHQQMRQERYVASPAKGFYLPKKSGGNRLIG
IPTVKDRIVQRYLLQSIYPKLEDAFSDAAFAYRPGLSIYTAVDRVMERYRHSPAWVIKVDIQQFFDTLSWPVL
LHQLDQLGLPPLWVRWIEHQLKAGMVINGHFYRPNQGVLQGSILSGALANLYLNDFDCRCLEADITLVRYGDD
CVAVCQSYLEASRSLALMQDWIEDLYLTLHPEKTQIIPPDQEFVFLGHRFQGGDVEPPERQKVASKKQKTAKA
VWGPPKVCSIVKSPQPIAKSSTDEYWRDGMTTLYVTDQGAYLRVKQQQFQVFHQQELRCSVPASQISHVVLFG
ACTVSHGAVRLALQRRIPLLYLSNKGRYFGRLQTTGQAKLEYLTQQVYKSQDPDFIRTQAASTIVGKLHNSRI
LLLRLNRRRKTERATQAIATLAELMETVPTADSVEAMLGYEGTGANQYFQAYASLLKGKFEFEKRTRRPPTDP
VNSLLSLGYTLLSQNVHAMVEGVGLHTHFGNLHKPQANRPSLVCDLVEEFRAIAVDSLVAYLINSNIFKPDDF
TPPDERGGVYLHPDAMKRFLKHWEEKLQQSVTHPHTGYKVSYRRCLELQVWEYVACLMGEQPVYRPMRMEK
```

```
>KIOST1_04569    Cas_Cas1
MSLLYLTQQGTRLQKEHGRFVLSRPDRTDVAIPMTEVEQVMVFGNISLTTPTMAACLSRQIPVVFLSQSGTYR
GHLWSAEYDDHRAVAAQFAHRDDLTFQWAVAQAIVRGKAWNSKQLLLKLNRRRQLDTVALAISRIDRDLAAID
QLSATDTGLDQLRGHEGAIAAYYFPALGQLIANPDFSLTTRVRRPPTDPVNALLSFGYTLLFNNVLSLLHLEG
LNPYLGNLHRNDRHEAHLAFDLMEEFRSPIVDTLVLTLLNQQVLKPDAFLPPKANGAVYLTDDARRRYLQAFE
DRIMTTLHHPDARDPVPYRRIIQLQIRRYKACLLDDAPYEPFRRET
>CYA98_02936     Cas_Cas1
MSILYITQPQAILGKKYEAFTVATQEESGKWNKQSVPAQTVEQVVLMGNPQVTGDALTYALELGMPVHYLTSF
GKYLGSALPGYSRNGALRLAQFELYQDTERRLELVKTIVSAKIQNQATVLYRQKIADSPLKNHKKLVFKKATL
DEVRGVEGIAAKDYFREWSAILNEPWFFPGRNRRPPTDPVNALLSFAYGLLRVQVTASVHLVGLDPYIGYLHE
VSRGQPAMVLDLMEEFRPLVADSLVFSVLNKREIQPDDFTESLGAYRLKDEGRKRFLQAFERKMNDEFKHPVF
EYRCTYRRAIELQARLLGRYLQEDLVYKPLVLR
>CYA98_04900     Cas_Cas1
MPTIYVIHPGSSIDVHSRYLQVFHKQQLCLGVPINRVTQIILFGICNLSRKTQVLIATRQIPVYYINVHGDQL
GSLELTPPRPFKYQTHQLICSQDLEFVRAIAESITWAKLQNEIIVLKQLQPEFYPCNIENTINFLQLLADDLP
SANSIEELWDYHATATSYYYPILASLVPQWSNPHSSFRLDLSHPVLQMLKLGYTFLHQQMALFINDLGLDPEI
SNLHTLNQTSQPLVDDLIEEFRPILVDSLVVKLLQTGTITSSDFVPPNGSRGGYLQPSALQTFINHWEEKLAT
HIEHPFAGTISYRQCFELQAKEYIACLLGDIEFYQPLLMQQEKGKIKKMYAKEYLTQKNVVQALSGKD
>CYA98_04899     Cas_Cas1
LHPSPTDPINSLLSLGYTLLSQNVCSFIELAGLHTHFGNLHVPRDNHPALVSDLMEEFRAQIVDSFVSYLVNS
KILLLEDFTSPDELGGVYLQPDALKKYLKHWEEKLYSEVTHPQTGYKVSLRRCIELQVREYIAALMGEVEVYR
PMIWKN
>NIES2104_00766  Cas_Cas1
MPVLYVTQPDAILNKSYEAFIVKLKQADGTWKKQTIPAQTVEQVVLMGNPQVTGDALMYALELGMPVHYLSGF
GKHLGSALPAYSRNGQLRLAQYRRYEDQVQKLELVKAIVTTKIHNQATVLYRHGVKEHPLKQRKASVAAQKTI
DQVRGMEGLAAREYFGYWSQMLKGDWQFSGRNRRPPTDPVNALLSFAYGLLRSQVTAAVHLAGLDPYIGYLHE
VTRGQPALVLDLMEEFRALIADNVVLTVLNNREIQQKDFMESLGAYRLSDAGRRTFLQGFERKLSDEFTHPVF
EYRCSYRKAIELQARLLARHLQEDVPYKPLAVR
>FACHB389_03774  Cas_Cas1
MGTVYVTQEDAFIGKVDERLNVKFEKKTILDVPLIKVDGVVVLGRATVSPAAISELLERHIPLTFLTENGRYL
GRIEPEVTKNIFVRKAQWQAAGESAQAIHVVQGFVRGKLKNYRNLLIRRQRESPDLDLSVNVTRLEQAIAPIE
NTQNINSLRGLEGAGSAAYFGCFNQLIRTTDFTFNTRVRRPPTDPVNSLLSFGYSLLRHDIQGAVNIAGFDPY
LGYLHCDRYGRPSLALDLMEEFRPLIVDAVVLSAINKQLLIKADFVTEPLSAAVSLTPEGRKTFLRLYAQKKQ
SEFKHPVLGRKCTYQEAFEIQARLLAKYLMAEIEKYPPLVLK
>FACHB389_01027  Cas_Cas1
MSTLYVTDQGAYLSVKNQQLQVFFQQEIKIKLPANSVSEIVLFGSCNVSHGAVSMCLKRRIPVMYLSQTGRYF
GRLQVDNNARVEYLTRQVESSLNSEFIRKQAEAIVWTKINNSWILLMRLSRRRDFAEISATPVFKAIDDLDTL
MDKLFFAETMDALRGYEGKAATIYFQALGSLFSSPFKFEQRTKRPPTDPINSMLSLGYTLLSQNVFSCVQAEG
LHTHFGNLHVPRDNHPALVSDLVEEWRAQLVDSLVCYLVNYQIFTFDDFTPPDERGGVYFQYHALKKFLKHWE
EKLQSELTHPHTGQKVTFRRCLELQVREYISCLMGEVQLYRPMIWRM
>BDU141951_03289 Cas_Cas1
MPSLRTLPKARDRISFLYFEHCRIEQEGRAIAVFKEKSKFFVPCAAISTLLLGPGTSITHAAIKTLADNVCEV
QWVGEDVMRFYAHGRSGANNARRLIHQATLWADSIQRLAVVRRMYLFRFDEPLDENLTLQQIRGHEGVRVRTL
YQQWSQKTGVEWHGRDYKTDNWDAADPINRALSIANTCLYSVCQAALHSVGYSPALGFIHTGKPLSFVYDIAD
LYKGETTIPAAFEAVADSYFDLDRFVRHRCHRHFADVRLMHRIVKDIDQVLGFNTEEEEEPICFLWDDVLETV
EGGKNWSDDEEAT
>AO1A_00776 Cas_Cas1
MGTLYVTQENAFIGKVDERIHVKVEKQKIADVPLIKVDGVVAFGAVNFSPALAEELLNRRIPFTLLSSTGKYR
GCLEPELRKNIFLRQHQWAADGNTPQAIHLVQAFVRGKLKNYRTTLQRRGRRADIDLSEEIEQLKGIIGAIPS
ESFANATTEQINSLRGLEGAGSAAYFRCFKQLIQNPDFTFTTRNRRPPTDPVNSLLSLGYALLRHDIQSAVNI
VGFDPYLGYLHVKHYGRPALALDLMEEFRPLVVDSMVLEAINRQVFSPSDFTQQPISNETALTREALKTFLSL
YKDKKQSSFKHPVMKTKCTYQEAFEIQARLLGKYLMGETEQYPPLVLQ
>CYA405_00373    Cas_Cas1
MSILYITQPQAILGKKYEAFTVATQEESGKWNKQSVPAQTVEQVVLMGNPQVTGDALTYALELGMPVHYLTSF
GKYLGSALPGYSRNGALRLAQFQLYQDTERRLELVKTIVSAKIQNQATVLYRQKIADSPLKNHKKLVFKKATL
DEVRGVEGIAAKDYFREWSAILNEPWSFPGRNRRPPTDPVNALLSFAYGLLRVQVTASVHLVGLDPYIGYLHE
VSRGQPAMVLDLMEEFRPLVADSLVFSVLNNREIKPDDFTESLGAYRLKDEGRKRFLQAFERKMNDEFKHPVF
EYRCTYRRAIELQARLLGRYLQEDLVYKPLVLR
>CYA405_04793    Cas_Cas1
```

```
MLTLYVSQQGCYLTLKAESLLIKKGKDVLAEAQIPLLEQVLVFGQSQVTTQAIRACLSRNVPIAYLSRMGYCY
GRILPVERGYRHLSRYQQQLTQIDRLIIARRIVQAKLKNSRVILQRQARRQSSPTINLAVQSLNHLMEKVIQA
ESWDRLMGLEGAAAAQYFSAFGECLTNPKFVFTVRSRRPPGNPVNAMLSFGYQVLWNHLLSLIELQGIDPYFG
CLHQGTERHAALASDLIEEFRAPIVDSLVMYLINRNMIDANTDFTYHNGGCFLNDSGRIKYLKAFLQRMEEKL
TTAAGENQPRWDTLNQQVKAFKQFVYNPIEGYKPYEIR
>Paraca_00698    Cas_Cas1
MITKRTVIIQQSKGSTTCRLAEVDLVVILPGVQLTDSTIAQLLDRGIETIFLRQDGRFRGRLQGHFHTNPTVR
LAQYRALDTTFSMAIAQKLILGKVRNQRYLLQRRNRETQGKITQLTEAIDTIASYLPRLKQTETPLDRNELMG
VEGVCARIYFQGLTHWFPPEWKFTGRNRRPPKDPVNALLSWGYGVLLARVFAATVQAGFDPYLGFFHANAPFR
PNLVLDLMEEFRPVVVDRVAIALINSGLIDPNDFEPAPGEEEGIWLGQTAKKLYLGQLERQLQDSILYPPQHR
KLKLSQILLEQGRSLGRCLVETRLDYEAFIFR
>Paraca_03439    Cas_Cas1
MTAIYLTEPQSYLEADLRDFAIVYNDRIQSRHSIDTTQQIIAFEGCTLGRSAIAAIYRYRIRLSFIGDNGHIT
AQITPQETPFYTVLQQERNSDRNFRIAFIHSLLPGRWYNATKVLQSLGYSGIGGCLEQQLEVLPGLNSWLALQ
TWQVKVFQRYQKALRSVLVSVKDFEPTLELAYALLSGELYVLLRSSGCCAEVGTLHLHCQNHLPLPCDLMEPF
RPVVEAWVFKKAPKFSDRGGLSQSYRATFIQNWESLMATQIFHAYAGQVSLRSALAWQVEEYVRAITETLEYR
PFLLKS
>Paraca_03437    RVT_1+Cas_Cas1
VRDRIVQRLLLRSLYPILEETFQDCSFAYRPGVGVKHAIERVAEVYSSQTWTIKADISQFFDSLCRTLLLSQL
EELSVDQTVVRYIKGQLEAGIVVGGMPILSGRGVLQGGILSGALANLYLSEFDRRCLDAGAYLTRYGDDFVIV
ARSLLEATRFLNLIEDWLSDIYLTLQPEKTHIFAPGEEFVFLGYGFQAGEIITPEYKQPRRNHNKKAVSVASR
PPFACSIVTPTRKSLSTRNFLDSWREGMTTLYVTDQGAYVKVKHQQFQVLLGNDLKVSIPVNVVDYIILFGCC
NLSHGAIGLALRRRIPILFLSDQGRYFGRLQTDGMTRVDYLSRQVHCAEDETFVLRQAKVIVAGKLHNCRILL
RRLNRDRQISQVIEAIEELGVWQEKIAEVELLESLLGYEGFGTRIYFQALRALVQPPFTFEHRTRRPPTDPVN
SLLSLGYTLLHQNIHSLILAVGLHPHYGNLHVPRSNHPALVSDLIEEFRAPVVDSLVIYLVNSGIFTPEDFTP
SDERGGVYIYSDALKKYLKHWHDKLSLKTTHPHTGYKVSYYRCLELQVWEYISCLMGEREVYRPMLTKM
>PCC9212_02299   Cas_Cas1
MRTLYVSQQNCYVSLQKETLIVKQGEIVQGQVQLPLLEQILIFGKSQVSTQAIRACLWREIPIVYLSRMGYCY
GRILPIGRGYRQLCRYQQQLTQVERLITARAIVQAKLKNSRVLLRRQRRRQESELLSKVLQNLDHLVSQAATA
DNWERLMGFEGAGAAQYFSVFGECLTNPEFVFVGRSRRPPGNPVNAILSFGYQVLWNHLLALIELQGLDPYYA
CLHEGSERHAALASDLIEEFRAPIVDSLVLWLVNKNVMDAQNDFEYRDGGCYLNDLGRKKYLKAFLQRMEEEV
QIDGGDKQPRWHLLTTQVKAYKQFVYNPAQVYQPYLIR
>PCC9212_00673   Cas_Cas1
MGTVYITQEDAFIGKVDERLNVKFDKKTILDVPLIKVDGVVVLGRATISPATISELLERHIPLTFLTENGRYL
GRLEPEVTKNIFVRKAQWQAAGESAQAIHVVQAFVRGKLKNYRNSLIRRQRECPDLDLSENVTRIEQAIAPID
TTQSINSLRGLEGAGSAAYFGCFNQLIRAKEFTFNSRIRRPPTDPVNSLLSFGYSLLRHDVQGAINIVGFDPY
LGYLHCDRYGRPSLALDVMEEFRPLVVDAVVLSVLNKGLLTPTDFVTEPLSGAVSLTPEGRKTFLRLYAQKKQ
SEFKHPVMGRKCTYQEAFELQARLLAKYLMGETEKYPPLVLK
>BBD1991_03868   RVT_1+Cas_Cas1
MGGVPIAKVRYTPRFPLDPRHLQDAGDLSLPTTQTRTQIALCHTQNVEFRVWMDINQYQLRAAWQQVKRGSRM
AGVDGITVDWFASIYRDKLQQLHQQLRNDSYFPLPAKGFFLRKNNGGKRLVGISTVRDRIVQRWLLGELYLPL
EDCFLDCSYAYRPGRGIQQSAWHYFELYQTQPTWTVKADIAQFFDHLCHGILRTALDNLDGDCAEAVMQQVEA
GISIEGILTRPTQGVVQGNILSGALANLYLTEFDRRCIESGYRLVRYGDDLVLACDSLPEADRVLSDIETWLK
QLYLALQPEKTHIFAPDEPFTFLSYRFEDGTLTEPPPPQLKHGLARPYGATPKRRPARPPSRPPRACSLKPPV
PQRSTAPPAHYFAEPMTTLYVTDQGAYLKVKHYQFQVFFQNQLQMQVPVNRVTHIVLFGCCNLSHGAVRLALT
RGISVTYLSNQGRYFGRLNRQGTAKVEYLARQVERAADGDFVRTQAEAIVRAKLHNSRVLLLRLNRRRRTLEA
MKAIEQLEKLLDNLPLADSIEALLGYEGKGASVYFRGFGSLLKEQFQFEKRNKRPPKDPANSLMSLGYTLLSY
NVYAFVEAIGLHPHFGNLHVTLDHRPGLVCDLMEEFRSPLVDSFVAFLVNSSVLQLDDFTPPDDRGGVYLHPD
ARRKFVKHWEDKLQTEMTHLTTGYRVTYRRCIELQVREYLACLLGDVEAYRPMLRPLT
>BBD1991_03865   Cas_Cas1
MTTLYLTRSKTHLTVDRSGFAIQYCHRTTSRYPISQVKQIVCFEGVRLNRTSQREIQRHQIPLLQLDRTGTIC
GRWYPKSETIDRHLSPSTVDRLDLAEWLLKRHLEDSIDLLTRLQGTPTLQATGRVLLLLHETLSDADSLDMIR
GYYATGTAYFDRALRETLSAYILHLNGSRPPMTVSQLLRVGESLLHHHLEIALTQARFDLSLGSLHVPADCDA
PLVCDGVAVLRPCFVDAFVVRWLHHCCSTSITWLAFCTAWERQPFLHGQIGEFVESLRRWLTA
>BBD1991_03853   Cas_Cas1
MQTLYVSKQGCYICLKQEQLLVKHQQQILAHVRLPLLEQVLIFGKSQMTTQAIRACLARDIPIAYLSRLGYCY
GRILPIERGYRQLARYQQELSAVDRLIVAQNIVHSKLKNSRTFLLRQRRRHSCDKIDLAIESLARLIHQVKQT
YTIESLLGLEGSGAAQYFPAFGECLTHTDFTFLTRTRRPPTNPVNALLSFGYQILWNHLLSLLELQGLDPYHA
```

```
CLHQGSERHAALASDLIEEFRAPLIDSLVIWLINSRTLATNRNFKYRDDGCFLNDSGRTIYLKHFTQRMEESI
QTADGSNQPRWHLLTQQVKRFKQFVYQPSHPYHPYEIH
>BBD1991_00438   Cas_Cas1
MTIPLYITEQGTYLSLKQKQLQVWKGDRLKVSVPLQRLSHIVIFGRCTLSHGVVNVVLKQQIPVMFLTQSGRY
CGRLAAEGTAEIDRLVAQVDRGRDREFVLAQAQQIVYGKLHNCRVLLQRLNRSRRDEEIRACIATLKTWLDRA
LESPTVEGLLGCEGQGTRVYFQGLGRCIQPPFTFDGRTRRPPTDPVNSLLSLGYTLLYQTLFSLVCAVGLHPH
FGNLHVPSSRYASLVSDLVEEFRAPLVDSVVVSLVNRGGLQLDDFYPADIRGAVYLRPTALKSFLKDWQERLQ
TPVSHPQVSEQILYVRVLEIQVWEYVACLVEDRSIYRPFCWKV
>BBD1991_02179   Cas_Cas1
VGTVYVTQEDAFLGKVDERLSVKYQRDRLLDVPLIKIDGVVIFGRATVSPAVVEELLERNIPLSFITRYGRYR
GRLEPEVTRNIFVRRAQWGAAGETQTSIDAVRGFVRGKLKNYRSLLQRKQRENDDLTLTESIEQLNRSIAALP
KVREIDRLRGYEGAGSATYFGCFGQLVRGNDFRFTHRNRRPPLDPINSLLSLGYALLRHDIQSALNIVGFDPY
LGYLHTEQYGRPSLALDVMEEFRPLVVDAVALAMANKRLLTPGDFETEPVSGAVSLTSDGLKVFLRAYEEKKQ
SEFKHPVMGRKCTYRKAFEIQARLLAKFLMGETDKYPPLVLR
>BBD1991_03600   Cas_Cas4+Cas_Cas1
MNAQITAETLRVCALHALAYCPRLFYLEEVEELYTQNQSVFEGRRLHEELDRTEDGEWEDLFLTDETLGLRGR
VDALRTRDGHLIPYEHKKGRSHRDAQKQPQAWESDRLQILAYCCLLETVAKHPIREGRIRYHGDNALVKVPFD
ENSRAEVKAAVDRAKALRESPERPPIAENERLCQNCSLAIVCLPEETRFFQDATHQPSRLFPKDDDRQIVHVL
EPGTRVGRTGEQLKITRRDRSVETLAVRQISQVVLHSFSQISTQALHLCASQDIGIHWVSGGGRYVGSFDGRR
GSIQRRIQQYAALSQPDVCLNLARRLVACRGRGQRQFLMRCQRNLNGTREAVAASVKQMKLLLKQVETAQSLA
QLLGIEGNLAALYFKGLARAISVKVPPELQFEGRNRRPPKDRFNTLLGFGYGLLLKDVTNAILAVGLEPALGF
YHQPRTQAPPLALDLMEIFRVSLVDMTVLASVNRGQWDATADFDPRGNGVWLSDLGRRKFVELYERRKGESWK
HPVLGYSLTYRRLLELEVRLLEKEWLGEGDGFARLVVR
>Ana_01957 Cas_Cas1
MRSLYVSQQGCYLSLQQEKILIKRGQQVLDEVQLPLLEQILVFGQSQVTTQVIRACLQRDIPIVYLSRMGFCY
GRLLPIGRGYRQLARYQQQLSAVERLITAQAIVQAKLKNSRVILLRQQRRQATAPLTLAIESIEYLVQQAATA
ESTERLMGFEGAGAASYFGAFSECLADERFVFIARSRRPPGNEVNAMLSFGYQVLWNHLLSLIELQGLDPYYA
CLHQGSERHAALASDLLEEFRAPIVDSLVLYLINRRMMDIESDFEPKGGGIYLSNLGRKKYLRAFLQRMEERI
QTDVGEQPRWDLLNQQVRAYKLFVYDPVRVYKPHVIR
>Ana_02053 RVT_1+Cas_Cas1
MQVYPEHLQAAWALVQRGSPAAGVDGITTDLFVGVANEQLAQMHRQLRREVYEASPAKGFYVPKKNGGQRLIA
LSTVRDRILQRYLLQSIYPRLEKAFTDSTFAYRPGLSIYGAVDRVMAIYAPQPTWVIKADIQQFFDNLSWGVL
LSQLERLKVAPAQVRLIEQQLKAGLILQGQFYRPNKGVLQGGILSGALANLYLSEFDRLCQEAEIPLVRYGDD
CVAVCHSYLQANRFLAMMQGWLEDIYLTLNPDKTRIVGPDEGFVFLGHMFAKQRVIAPERKWVQAGKKPKVAR
PMYGPPKACSIVPLPKVKTKKSTDEYWRDGMTTLYVTEQGAYLRVRHSQFQVFHEQELRISIPANNITFVVMF
GACTVSHGAVRLALQRRIPLLYLSNKGRYFGRLQTTGQAKIEYLVAQVTKSQDEDFVRRQAMNVIVGKLHNSR
KLLMRLNRRRKTDLATKAVKEIAVLIKQVQQAETVESMLGYEGQGASLYFRAYGSLLKGEFGFDKRTRRPPTD
PTNSLLSLGYTLLSQNVHAMTEAAGLHTHFGNLHVPQSHRPSLVCDLVEEFRAMVVDSLVAYLINSNIYTPED
FTPPDGRGGVYLHTDAMKRFLKHWEEKLQSETVHPHTGYKVSLRRCFELQVWEYAACISGEREEYRPMVWQK
>ESFC1_04790     Cas_Cas1
MRSLYVSQQGCYVSLQQETLQVHHRSQVLQQIQLPHLEQILIFGKSQITTQAIRACLWRNIPIVYLSRMGYCY
GRVMSIERGYRHLQRYQQELVLGDRVTLARAIVAGKLKNSRTLLLRQQRQRHLPQLDPIIPRLEALAQQAQNT
TALDKLMGLEGAGAAQYFSALGQCLTHPDFDFIARSKRPPLNPVNALLSFGYQVVWNHLLSLIEVQGLDPYYA
CLHEGTERHAALASDLIEEFRAPVIDALVLSVINRRWLEVGRDFVYHQGGCYLNDQGRKRFLKAFLQRMEEPI
KTESGLQPKWDILNQQVKAFKQFVYQPSRGYKPYQMR
>ESFC1_03524     Cas_Cas1
MGTVYVLEEDAFIGKVDERLTVKADKKRLLDVPLIKVEGLVILGRATLSPAVVIELLERHISVSYLTRTGRYL
GRLEPEATKNIFVRKAQWLAIGESKQAIHAAGGFIRGKLKNYRNALLRRQRKNPELDLQQPLKRLENAIAPLD
EKSSINSLRGLEGAGSAAYFGSFDNLIQNKSFQFRDRNRRPPTDPVNAMLSLGYALLRHDVQSAVNLVGFDPY
LGYLHVERYGRPSLALDLMEEFRPLVVDAVVLNAINKNHLSPDDFSSEPLSKAVSLTPDGLKIFLRLYEQKKQ
TQFKHPVFKRQCTYQEAFEWQARLLAKYLMAETDKYPPLILK
>ESFC1_02443     Cas_Cas1
MSAVYITQPDATLTKAQEAFKVALKQEDGSWKKHNVPAQTIEQIVLMGHPSITGEALCYALELGIPIHYLSIF
GKHLGVALPSYSRNGQLRLAQYATHTNTEKRLALVKAIVTHKIHNQYQVLYRHDQKDNPLKKHKKLVSQQTTL
NQVRGVEGLAAKDYFACWRNLFKDTWSFNGRNRQPPTDPINSLLSFAYGLLRVQVTSAVHVAGLDPYIGYLHE
TTRGQPAIVLDLMEEFRPLVADNLVLSVISRKQIKPDDFNESLGAYRLSDKARKTFLEAFEQKITSELKHPVF
GYRCTYRRAIELQARLLARHLQEDVPYQPLTIR
>ESFC1_00816     Cas_Cas1
```

```
VGDAHPTTLKAPRPLGEGLGVRAFHYTEIMRTLYVSQQGCYVGLRQECLEVRRKDEVLARVQLPLTEQILIFG
KSQVTTQAIRACLWRDIPIVYLSRLGYCYGRIIAIERGYRHLARYQQEISLVERLLVARAIVRGKLNNSRIIL
LRQQRRYPAEALSLAIQRLGYLIEQIPQTDTIEQLMGYEGAGAASYFTAFGACLQGTGFVFAGRSRRPPGNPV
NALLSFGYQVVWNHLLALIEVQGLDPYQACLHEGSARHAALASDLIEEFRAPIVDSVVLWLLGSRVMDADKDF
EYRDGGCFLNESGRRKYLTAFMQRMEEEVKVAEDERQPRWDLLMQQVKKYRQFVYNPAYGYQPYQIR
>BLJ_01183  Cas_Cas1
MGTIYIIQEDAFIGKTDERLIVKLGKEKLIDVPLIKVEGIIILGRATISPAVVMELLERKISLTFLKKTGKYL
GRLEPELTKNIFVRKSQWAAAGESSQAIHAVQGFVRGKLKNYRNTLLRRSRKYTELNLQDSITKIEQVIAPID
SKNSINSLRGLEGAGSAAYFGCFNQLINSNSGFSFQFRNRRPPTDPVNSLLSFGYALLRHDIQSAINIVGFDP
YLGYLHVERYGRPSLAFDLMEEFRPIIVDAVVLSAINNKLLSCNDFTTEPLSKAVSLTPEGLKIFLRLYEQKK
QSKFKHPVLKRQCTYQESFEIQARLLAKYLMEETDKYPPLVLK
>BLJ_02107  Cas_Cas1
MSVLYVLQPNAILNKVHEAFHVALQQEDGTWKKQAVPAQTVEQVVLMGNPQITGDALCYALELGMPVHYLTSF
GKYMGSALPGYSRNGQLRLAQYDLYHDRDRRLELVKAIVMAKIHNQHLILYRQKVKKNPLKQRKKDVKTQTTL
DQVRGMEGLAAREYFAGLSSILEDQWGFKGRFRRPPTDPVNALLSFGYGLLRVQVTSAVHIAGLDPYIGYLHE
VHHGQPAMVLDLMEEFRALVADSVVLSVINTREIQCNDFTESLGAYRLSDSARKTFIQAFERKMTDEFKHPIF
DYRCTYRRAIELQARLLSRHLLEDIPYKPLMLR
>BLJ_05160  Cas_Cas1
MLTLYVSQQGCSVYKAQETLQIKRNQTVEVEVQLPLLEQILIFGKSQVSTQAVQACLKRNITIGYLSKTGYCY
GHILPVKQGYHYLSHYQQQLSVKDKLLVACQIVQAKLKNSRTILQRQHRRQPSETLALVINSLGYLMDKVAMS
ETWERLMGLEGAGAAAYFKAFGTCIGNSEFVFTERSKRPPGNPTNALLSFGYQVLWNHIYSLIELQGLDPYFG
CLHQASERHAALASDLIENFHHYLPVLSRVQLLN
>PCC6506_05099   Cas_Cas1
MGTLYINQDDVFIGKTDERLNVKLKNQTLQDVPLIHLDGVVIMGRATISPAAIDELMQRKIPLTFLSDTGHFL
GKLEPELSKNIFVRNAQWKAAGETPQAIHVVQGFIRGKLKNYRNSLTQTKRDHSELNLESAIDRISDIIVKID
KTHNIDSLRGLEGAGSAAYFGVFNQLIRAEGFVFKPRHRPATDPVNSLMNFAYTLLRHDVQSAINIVGFDPYL
GYLHVQRYGRVGLAFDLMEEFRPLVADNMIFNAINRRIITINDFTVEPISNVVRLSVEGRKEFLKAYEKKKQS
EFKHPVLGKKCTYQEAFEIQARLLAKYLMGEIDQYPPLVLR
>PCC6506_00395   Cas_Cas4+Cas_Cas1
MQLLEIDYFSPPKQSTIRVSALHAIAYCPRLFYLEEVEELYTQDAAVFAGRRLHVEIEKQEDEDWEDLYLESE
ELGLRGRLDALRTRDGQTIPYEHKRGRCHRDAHNQPQAWESDRIQILAYAYLIESALRIPVPEGRIRYHADNV
MVRVPFDDTGRAAVREAIEQARSLRQSPCRPPVTDNERLCTRCSLAPVCLPEEARLAHNREWQPIRLFPEDDE
RQIVHVLEPGTSVGRCGEQIKISRRNQPTEILPARQIGQIVLHSFSQISTQALYFCTEQAIGIHFVSGGGRYV
GSFDARGGSIQRRIRQYEALSNSQNCLELARKLVICRGQGQRKFLMRGGRGKEAPEELKKAINQMKIVLAQVP
QAKALENLLGLEGNLAALYFGALPYLIGNDVSPLLNFCGRNRRPPKDRFNALLSFGYALLIKDVMNAILTVGL
EPALGFYHQPRSQAAPLALDLMEIFRVPLVDMVVMGSINRGQWQPQTDFEIRGEQVWLNEAGRRKIVDLYERR
KEETWKHPVTGYSLTYRRLLELEVRLLEKEWMGEGGLFAQLIVR
>PCC6506_02824   RVT_1+Cas_Cas1
MEFTPEPLHAAWLQVRKGSRSAGIDNITVDLFAGVARYQLQVLLWQLQQENYFPRPAKGFYLRKASGGKRLIG
IPTVRDRIVQRFLLDELYWPLEDVFLDCSYAYRPGRGIQMAVKHLYSYYQFGQAWVIKADIEKFFDNLCWPLL
LTDLEKLQFEPTLRQLIEQHLASGIVVKGQHFHPNQGVLQGGILSGALANLYLNEFDRLCLSHGFNLVRFGDD
FAVACADSIQANRCLEQINSWLGSFYLKLQPEKTRIFAPDEEFTFLGYLFRNGQVFAPERTQPTTKVVRASGT
SFSRSVTRKVPVFSGKPLACSISVKPIVFPRANSEHFWREPMSTLYVTDQGSYLSIYNQQFQVFYQRELRIKV
PASRVSHIILFGCCNLSHGAVSLALHGRIPIMYLSQRGRYFGRLQAEGMAKVEYLERQVIGSQSSKFVRQQAE
VIVWTKLHNSRALLMKLNRRRPSKIAAIAIEAILQLMEDLHQAASMDALRGYEGQAASLYFQGLGSLFTGSFV
FEKRTKRPPTDPINSLLSLGYTLLSQNVHSFVQAVGLHTHFGNLHTPRDYHPALVSDLMEEFRAQVVDSLVAY
LVNSKIFVLEDFTPPDERGGVYLQPYALKKYLKHWEEKLLSEVTHPHTGYQVSLRRCMELQVREYIACLMGEV
EVYRPMLWKL
>PCC6506_02816   Cas_Cas1
MTAVYITEPGSILSVQQQHFQITHKDNLCVSVPVKQVTLITLFDLCQLSRQAVNLVLSRKIPIYYISRQGQEL
GYCEPDSKRPAKYRSKQLKRQQEHEFVRAFAESITWAKMHNQSVLLQQLQVSESEFESEVYRAFFNLLMDELA
MNSIRRA
>BDU130791_02904 RVT_1+Cas_Cas1
VGCVAAQTHLTLNRTWVEYQKPGVSQTLGFWRLQGCPNKRPTWTVKADIAQFFDHLCHGILQAALDGLDLDSP
SREAVMQQVKAGIVIQGTLSRSSQGVVQGNILSGALANLYLTEFDRRCITSGYRLVRYGDDLVLACDSLSEAE
RVLQHIHNWLKQLYLSLQPDKTQIFAPDQPFSFLSYRFENGTLTEPPPPQLKNGLARPYGATSKRRPVRPPSR
APRACSLKPPTPECSTAPPAHYFADPMTTLYVTDQGAYLKARHSQFQVFHQHQLMQVPVNRVTHVVLFGCCN
LSHGAVRLALSRGISVTYLSNQGRYFGRLNRQGNAKVEYLARQVQCAADADFVRTQAETIVRAKLHNSRVLLL
RLNQRRQTPEAITAIDQLAMLMDHLPLADSVEALLGYEGKGASVYFRGFGSLLKTEFQFTKRSKRPPQDPANS
```

```
LMSLGYTLLSYNLYAFVEAIGLHSHFGNLHVTLDHRPGLVCDLMEEFRSPLVDSFVAYLINSSVLQPDDFTPP
DDRGGVYLHPDALRKFVKHWENKLQTEMTHLTTGYRVTYRRCLELQVREYLACLLGDVEAYRPMLRPIK
>BDU130791_03035 Cas_Cas1
MTIPLYVAEQGTYLSLKQQQLQVWQGDNLLVSLPLQRLSHIVVFGRCTLSHAVVNRVLQQQIPVMFLTQSGRY
CGRLSAEGLPDIERLVAQVERSRDRQFVLTQAQQIVSGKLHNCRVLLQRLNRSRRDAEIRVCIAQLKDWRDRA
LESPSLEGLLGCEGQGTRIYFRGLGRCIQPPFTFSRRSRRPPTDPVNALLSLGYTLLYQTVFSLVRAVGLHPH
FGNLHVPSSRYASLVSDLVEEFRAPLVDSVVVSLLNRGGVQPDDFFPADVRGAVYLRPAGLKRFLSAWQQRLQ
TSVTHPQVGEPILYVRAIELQVWDYVVCLVGERSVYCPFVWKV
>BDU130791_02872 Cas_Cas1
MGTMYVMHEDVFLGKLDERLQIKHQGERLADVPLLKLQGVVVLGRATVSPSVVDVLLERNIPLSFLTRYGKYR
GRLEPEVTRNIFVRRGQWTAAGDTETARTAVRGFVRGKLKNYRSLLQRKQRDGHGVDVTGAISQLEGAIAALP
KTPTIDQLRGYEGAGSAAYFGVFNQLLQGTPFEFTHRNRRPPKDPVNSLLSLGYALLRHDVQSTLNIVGFDPY
LGYLHTERYGRPSLALDVMEEFRPLVVDAMVLSLLKKPNGLSLGDFEQEPLNGAVSLTEDGLKRFLRAYEEKK
QSEFKHPVMGRKCTYQEAFEIQGRLLAKFLMGEIESYPPLVLR
>BDU130791_02467 Cas_Cas1
MTTLYLTEPGTVVSAKNQTLDLKRGTWRRSCRLAEIDLLVVLPGVQLTGAAIALLLDSGVETLFLRRDGRFRG
RLQGQFHSNPSLRLAQYRVVETTFGMGLGQKLVLGKIRNQRSLLQRRNRETGGRVTALVEAVDMMAVYLRKLQ
RLDEPLSRDELMGIEGICAKTYFQAMMTQVPASWGFGGRNRRPPRDPVNALLSWGYGVLLARVFSASVQAGLD
PYLGFFHATEPYRPNLVLDVMEEFRPLVVDWVVWGLVRSSVLEARDFEPSPDGEGVWLGNLAKKLFLEQLERR
FQSKLLYPPQERQLALSQIVLEQGRAIARACVSLRLDEFESFEVR
>BDU130791_03423 Cas_Cas1
MSVLYVTQPDAILSKKYEAFRVGLKQDDQGWVYQSIPAQTVEQVVLMGNPGITGDALTYALELGLSVHYLTRF
GKYLGSALPGHSRNGQLRLAQYRAYCDETQRLAIVQEIVTGKIHNQYGLLYRRGVTENPLKLRKAAVAKQADI
NATRGIEGLAAREYFGCWSDLLKSPWTFTGRHRPATDPVNALLNFAYGLLRVQVTAAVHLAGLDPYVGYLHET
TRGQPAMVFDLMEEFRPLVADSLVLSVVNLGELQLSDFEESLGTYRLSDGGRRCFLEAFERKMTDEFKHPLFG
YRCTYRRAIELQARLFARHLQEGVSYRPLVRR
>WA102_04454     Cas_Cas1
LLVKQGDTVYIEVQLPLLEQILIFGQSQITTQVIRACLWRDIPIAYLSRMGYCYGRILPISRGYRQLSRYQQQ
LSPIDRLITAREIVKGKLKNSRVLLRRQQKKRDLELIERVLDSLDYLATQAATADNLERLRGFEGAGAAQYFS
AFGECLTNSDFVFSGRSRRPPGNPVNAMLSFGYQVLWNHLLALIEIQGLDPYYACLHQGNDGHAALASDLIEE
FRAPLVDSLVMWLINRKIMDVKDDFEFKNGGCYLNNLGRKKFLRGFLQRMTEEIKTDQEIKQPKWDLLTQQVK
AFKQFVYNPSHHYKSYRID
>PCC9006_00152   Cas_Cas1
MRTLYLSRQGCFVSLQQEMLVVKQGDRTLEEVRLPLLDQVLVFGQSQLTTQAIRACLWRNIPIAYLSRMGYCY
GRVVAIEGGYRQLARYQRELTALERLLLAQALVKAKLLNSRVILQRQQRRRSSLMLETAIEQLAYLIHQISDT
DQIDRLMGLEGAAAAQYFQGFGDCLRHADFAFVSRSRRPPGNPVNALLSFGYQVLWNHLLTLVELQGLDPYEG
CLHQGSYRHAALVSDLIEPFRAPLVDSLVLQLINNGELDAQGDFEYRDGGCFLGERGRRKWLGAFVQRMEMPV
QSEEDAAQPSWDLLNRQVRAYKEFVYSPVQGYAPYRIR
>BC1401_04262    Cas_Cas1
MGTLYITQDDAFIGKTDERLTVRADKQTLQDIPLIKLNGVVILGRATVSPAVVIELLERKIPLSFLTKMGRYL
GRLEPELTKNIFVRAAQWKAVAPSKQASHIVQGFVRGKLKNYRNALLRGQREGSDLDFQKALTQLEQVIAPIE
TTTAIDSLRGLEGAGSAAYFGSFDRLIRASEFKFEARRRRPPTDPVNALLSLGYSLLRHDLQSAVNLVGFDPY
LGYLHTLRYGRPSLALDLMEEFRPLIVDAVVLSGLNRRTIVLTDFMTEPLSGAVSLTPEGLRTFLRLYEQKKQ
AKFKHPVLQTQCTYQESFEIQARLLAKYLMEETDKYPPLVLK
>BC1401_04300    Cas_Cas1
MSSLYLVHQGATLRKEQERFVVDEPKPKIKADDLLSEGLLSPVRVPIKEVEQVLVFGNVHLSTAAVSACLEEQ
VPVVFLSQSGRYKGHLWSAERCNLAVESAQFDRRHDQIFQLEMARSMVWGKLMNSRQLLLRLNRKRQVMEVER
AIGGIDRDLESVATVAQIESLRGYEGAAAARYFPALGRLLINPGFSLTDRSRRPPKDAVNSLLSFGYTLLFNN
VLSLLLTEGLNPYLGNLHRSDRHDPHLAFDLMEEFRSPVVDTLVVTLINQQVLKPEDFSDVSEGGIYLTNAAR
RLFLRRFEERMSELVAHPNVQGQVSYRRAMQLQIRQYQRCLLDGVEYESFLRAV
>BC1401_04349    Cas_Cas1
MGTLYITQDDAFIGKTDERLTVRANKQTLQDIPLIKLDGVVILGRATVSPAVVIELLERKIPLSFLTSMGRYL
GRLEPELTKNIFVRAAQWKAVAPSEQASHLVRGFVRGKLKNYRNALMRGQREGSDLDFQKALTQLEQVIAPIE
TTTEIDSLRGLEGAGSAAYFGCFDRLIRADGFRFEARRRRPPTDPVNALLSLGYSLLRHDVQSAINLVGFDPY
LGYLHTQRYGRPSLALDLMEEFRPLIVDAVVLSGLNRRTIVPADFTTEPLSGAVSLTPEGLRTFLRLYEQKKQ
AKFKHPVLQTQCTYQESFEIQARLLAKYLMQETDKYPPLVLK
>CS328_00313     Cas_Cas1
MITKRTVIIEQSKGSTTCRLGEIDLVVILPGVQLTDRAIAQLLDRGIETIFLRQDGRFRGRLQGHFHTNPTVR
LAQYRALDTTFGMALAQKLILGKVRNQRYLLQRRNRETHGKITQLTEAIDTIASYLPRLKQTETPLARDELMG
```

```
VEGLCARIYFQALTHWFPPEWKFTGRNRRPPKDPVNALLSWGYGVLLARVFAATVQAGFDPYLGFFHAIAPFR
PNLVLDLMEEFRPVVVDRVAIAIVNSRLIDPNDFEPAPGEEEGIWLGQTAKKLYLGHLERQLQDSILYPPQHR
KLKLSQILLEQARSLGRCLVETRLDYEAFIFR
>CS328_01847     RVT_1+Cas_Cas1
MGDALLAAWYQVRTRSKTAGIDGISVELFAHFLDEELKRLAHQLQQDIYHPLPALGFYRAKKSGGHRLIGIPT
VRDRIVQRLLLRSLYPALEETFQDCSFAYRPGVRVKHAIERVAEVYSSQTWTVKADISQFFDSLCRTLLLSQL
EELSVDQTVVRYIKGQLEAGIVVGGMPILSGRGVLQGGILSGALANLYLSEFDRRCLNAGASLTRYGDDFVIV
ARSLLEATRFLNLIEDWLSDIYLTLQPEKTHIFAPGEEFVFLGYGFQAGEIITPEYKQPRRNHNKKAVSVASR
PPIACSIVTPTRKSLSTRNFLDSWREGMTTLYVTDQGAYVKVKHQQFQVLLGNDLKVSIPVNVVDYIILFGCC
NLSHGAIGLALRRRIPILFLSYQGRYFGRLQTDGMTRVDYLSRQVHCAEDETFVLRQAKVIVAGKLHNCRILL
RRLNRDRQISQVIEAIEELGVWQEKIAEVELLESLLGYEGFGTRIYFQALGALVQPPFTFEHRTRRPPTDPVN
SLLSLGYTLLHQNIHSLILAVGLHPHYGNLHVPRSNHPALVSDLIEEFRAPVVDSLVIYLVNSGIFTPEDFTP
SDERGGVYLYSDALKKYLKHWQDKLSLKTTHPHTGYKVSYYRCLELQVWEYISCLMGEREVYRPMKLEKW
>CS328_01846     Cas_Cas1
MTAIYLTEPQSYLEADLRDFAIVYNDRIQSRHSIDTTQQIIAFEGCTLGRSAIAAIYRYRIRLSFIRDNGHIT
AQITPQETPFYTVLQQERNSDRNFRIAFIHSLLRGRWYNATKVLQSLGYSGIRECLEQQLEVLPGLNSWLALQ
TWQVQVFQRYQKALGSVLVSVKDFEPTLELAYALLSGELYVLLLSSGCCAEVGTLHLHCQNHLPLPCDLMEPF
RPVVEAWVFEKAPKFSDRGGLSQSHRATFVQNWESLMATQIFHAYAGQVSLRSALAWQVEEYVRAITETLEYR
PFLLKS
>CAUPA1101_00771 Cas_Cas1
VRTLYVSQSGCYLCLHQEQLLIKQGQTVLDQVQIPLLEQILIFGTSQLTTQLIRVCLQRNIPILYLSRMGYCY
GRLIAIERGYRQLSRYQQQLTAIERLMVARAIVQAKLKNSRVILMRHQRKQPSETIAIAVQNLDHLVSQAGQA
DGIERLLGLEGAGAACYFPAFGECLSNRSFVFVARSRRPPGNPVNAMLSFGYQVLWNHLLSLIELQGLDPYHA
CLHQGNERHAALASDLIEEFRAAIVDSLVLYLVNRGVIDADQDFVYADGGCYLNDSGRKKYLRAFLNRMEERL
TIDGTQQPRWDVLMQQVKKFKQFVYDPSHIYQPYQIR
>CAUPA1101_01523 Cas_Cas1
VVLHSYSQISTQALHFCADQGIGVHFISGGGRYLGSFDGRQGSIQRRIRQYDALSNPEMCLHLARKLVICRGQ
GQRKFLMRGQRGRAEVPVLLKQAISQMKAVLKQIPLVMSVESLLGVEGNLAALYFGALPHLIAEGVSPQLKFE
HRNRRPPRDRFNALLSFGYALLLKDVMNAILTVGLEPALGFYHQPRSQAAPLALDLMEIFRVPLVDMVVLAAV
NRGQWDVQADFEERGQQVWLSDTGRRKFVSLYERRRSETWKHPVTEYSLTYRRLIELEVRLLEKEWSDEGGLF
AQLILR
>CAUPA1101_00163 Cas_Cas1
MSVLYVTQPDAVLNKTYEAFTVSLKQEDGSWQKRAIPAQTLEQVVLMGNPQVTGDAFVYALELGMPIHYSIFH
SCEA
>CAUPA1101_04527 [CAS]: Type I-C CRISPR-Cas system spacer integrase Cas1
(CAS-I-C)
MFVMVVDAVVLSGINRRTLSPKDFTSEPMSKAVSLTGEGLRIFLRLYEQKKQSKFRYSVLQTQCTFQKAFEIQ
ARLLAIYLMGETEKYPP
>Ana39865_04022  Cas_Cas1
MGTLYITQDDAFIGKVDERINVKFEKKILQDIPFIHLEGVVVLGRATISPAVVDELMTRHIPLSFINKIGHYL
GRLEPEMTGNIFIRKAQWQAAGETPQSIHAVQGFVRGKLKNYRQTLMRYQREFDDVDLSKNIARIEQVIAAIS
STENINSLRGLEGSGSAAYFGCFDSLIRNSKFSFTKRVRRPPTDPVNSLLSFGYSLLRHDVQSAVNIVGFDPY
LGYLHFDRYNRPSLALDLMEEFRPLVVDAVVLSILNKQLLTPADFVTEPLSGAVSLTPEGRKTFLTLYEKKKL
SEFKHPVMGRKCTYREAFELQARLLAKYLMGTTDKYPPLVLK
>PCC73102_02844  Cas_Cas1
MGTLYVTQADAFIGKVDERLTVKSEQKTIMDIPLIKLEGIVVLGRATISPAVVSELLERHICLTFLTQNGRYL
GRLEPEVTKNIFVRKAQWQAVGESQPAIHLVRGFVRGKLKNYRHSLLRTQREHPDTDLNNNITRLENAIAPIE
KTSSIDSLRGLEGAGSAAYFGCFQQLIKTPEFRFEARRRRPPTDPVNALLSFGYSLLRHDVQSALNIVGFDPY
LGYLHVERYGRPSLALDLMEEFRPLIVDAVVLSLINKRSLTLTDFTTEPLSGAVSLTKEGLHTFLRAYQQKKL
SEFKHPVMGNKCTYQESFEIQARLLAKYLMNEIDKYPPLVLK
>PCC7126_04705   Cas_Cas1
MRTLYVSRQGCYITLKGENLIIKQGETVHGEVQLPFLEQILIFGKSQITTQAIRTCLWRDIPIAYLSRMGYCY
GRLIPITRGYRQLSRYQQQLTADERLLVARKIVQGKLKNSRTFLQRQQRRHPSQTTEIAIKSLEVLVQKAGEA
ETIERLMGLEGAGAAQYFSAFGECLNNPEFIFLARTRRPPGNPVNAMLSFGYQILWNHLLTLIELQGLDPYHA
CLHQGTERHAALASDLIEEFRSPIIDSLVLWLVNTKVINTQADFEYHDGGCYLNNSGRQKYIKYFLQRLEESV
QDSTGEQQPRWDLMTQQIKAFKNFVYQPNQTYQPYQIR
>PCC7126_01383   Cas_Cas1
MGTVYITQDDAFIGKTDERLNVKFEKKTILDIPLIKIDGLVIVGRASISPLAIAELVNQKIPLTFLTNTGKYI
ARLEPEMSKNIFVRSAQWKAAGETTQAIHLTQGFVRGKLKNYRYSLSQAKRRYSELDLDAGINQLSNAIASLD
```

```
KANSINTIRGFEGAGSAAYFGCFNQLIRVDDFNFSTRNRRPPLDPVNSLLSLGYSLLRHDIQGALNIVGFDPY
LGYLHTERYGRPSLALDLMEEFRPLIVDAVVLTAINRRMLTPKDFITEPLSNAVSLTKDGLHIFLRLYQEKKQ
NKFKHPLMQNQHTYQETFEIQARLLAKYLLGEIDKYPPLVMK
>PCC6303_01733   Cas_Cas1
MGTVYVTQDDAFIGKVDERLHVRIEKKTILDVPLIKIDGLVVMGRASISPMAIMELIQNKIPLTFLTINGKYI
ARLEPEMSKNIFLRNAQWKAAGDSSQAIHVTKAFVRGKLKNYRYALLRLQRSYLEHDLSQSLIQLKQSIESLE
KVNNIDSIRGYEGAGSAVYFGCFDKFIRVDNFSFKNRNRRPPVDPVNSLLSLGYSLLRHDIQSALNIVGFDPY
LGYLHVERYGRPSLALDLMEEFRPLIVDAVVINSINKKILKPSDFITEPVSNAVSLTKEGLHTFLKLYQEKKQ
AKFKHPVLGIQKTYQESFEVQARLLGKYLMGEIDKYPPLIMK
>PCC6303_00270   Cas_Cas1
MSTIYITEPDASFKLQNKYLKLLQQDKQRFFIRISNISQFAIFGNIKLPKDVMQIVRLNQIPAIYFTHTGEYV
GRLENPSTVQAKYLNYQRQRLHDREFNRATAESIVWAKLHNQQTFLQNWTRYCTDYTTQRASDYLTLLMDNLS
LAPCVEELHQYIEEADKVYYCAVASILSLYSSNPHTTAKQINRFLNLGNQLLHQYIYNHLKTAGLHPDYSILH
RDTHHELPLAWDFSAEFRAPIVDDLVLNFVRNLAKNNGNGNGNGKKPYTHLQRFLQHWEGKLKTFVLHPSAGE
VSYRQCIDLQVKEYIASLLGDVEYYRPLALKFHPEQTNFINMLETQKPILTLVK
>PCC6303_00269   RVT_1+Cas_Cas1
MFTLEHLNYAWLQVRAGSKSAGVDGISIGFFESMATEQLRNLVSQLQYGTYTASPAKGFYVPKKNGGKRLIGI
PTVRDRIIQRLLLDELYFPLEDTFVDCSYAYRPGRNIQQAVQHLYRYYQYQPKWIIKADIVEFFDNICLALLL
NALEKLRLEPNILQLIEQQIKSGIIINGQYQNAGKGLLQGGTLSGALANLYLTDFDQKCLNQGINLVRYGDDF
VIACSNFAEANRVLDKITGWLGGVYLTLKAEKTEIFSPDDEFTFLGYRFAKGEVFAPPPPEPTQQGEWVINDS
GTPYFRRKPKPEKPVSRPPKACSIDKPVNFPRAAISHYWQDFMSTLYVTDQGAYLSMKNQQFQVFYQGELKIK
IPVVRVSSIVMFGCCNVSHGAVSMALIRRIPIMYLSQKGRYFGHTAVQGNAKVEYLMKQVICCQNPVFTRQQA
ETIVAAKLHNSRILLMRLNRRRETEIATQAIDFIEILIDSLPKAESMDALRGYEGKAATIYFQALGSLFTGFF
AFDKRTKRPPTDPINSLMSLGYTLLSQQVFSFVQSVGLHTHFGNLHVPRDNHPALVSDLMEEFRAQIVDSFVS
YIVNKKILTPEDFTSPDERGGVYLQASALKKYLKHWEEKLQSETTHPNTGYKVAYRRCIELQVREYVACLVGE
VEVYRPMTWKM
>BDU20041_01519  Cas_Cas1
MGTVYITQEDAFLGKVDERLSVKYQRDRLLDIPLIKINGIVVLGRATVSPAVVEELLDRNIPLSFLTRTGRYR
GRLEPEVTRNIFVRRAQWGASGETSISTDAVRGFVRGKLKNYRSLLQRKQRENPELTLGEAIEQLHRAIAALK
KTHNIDRLRGYEGAGSAAYFGCFGQLVRGEGFSFTHRNRRPPLDPINSLLSLGYALLRHDIQSALNIVGFDPY
LGYLHTERYGRPSLALDVMEEFRPLVVDVVALAVTNKRLLTPGDFETEPVSGAVSLTSDGLKVFLRAYEEKKQ
SKFKHPVMGCQCTYQEAFEIQGRLLAKFLMGETDKYPPLVLR
>PCC10802_05487  Cas_Cas1
MSVLYVTQPDAVLNKSYEAFTVALKQEDGSWKKQSVPAQTVEQVVLMGNPQVTGDALCYALELGMPVHYLSSF
GKYLGSALPGYSRNGQLRLAQYAVYNDEKRRLDLVKAIVTGKIHNQHGVLYRHSQPDNPLKLRKKQVSERGTL
DEVRGIEGLAAREYFACWQEMLGDKWFFAGRNRRPPTDPVNSLLSFAYGLLRVQVTAAVHVAGLDPYIGYLHE
VHHGQPAMVLDLMEEFRPLVADNLVLSVLHNQEIKPKDFVESLGAYRLSDGGRKTFLQAFERKMSDEFKHPTF
EYRCTYRRAIELQARLLSRHLQEGVVYKPLMLR
>PCC10802_02639  Cas_Cas4+Cas_Cas1
MQTLDFTEKTAQETIRVSALHAIAYCPRLFYLEEVEELHTQDAAVFAGRRLHAELEKQEDEEWEELFLESEEL
GLRGRVDALRTRDGETIPYEHKRGRCFRDENNQPQAWESDRLQVLAYCCLIEYALKIRVKEGRIRYHAENILV
HVPLDDAGREAVRDAVKQAKALRESAHRPPVIENERLCARCSLAPVCLPEEARLAHNREWQPIRLFPEDDDRQ
ILHILEPGTAVGRTGEQIKISRRNQPVETISARQVGQIVLHSFAQISTQALHFCAGEGIGVHFISGGGRYLGS
FDSRQGSIHRRIRQYAALSNPEICLDLARKLVTCRGESQRKFLMRGQRGVKSVPVALKGAISQMKAVLKQVEK
ADSLSSLLGVEGNLAALYFGALPCLISEAVVPELRFAGRNRRPPKDRFNCLLSFGYALLLKDVMNAILTVGLE
PALGFYHQPRTQAPPLGLDLLEIFRVPLVDMPVMASVNRGQWDVLADFDVRGEQVWLKDAGKEKLIDIYEKRK
EESWKHPATGYSLTYRRLFELEVRLLEKEWSGEGGLFGQLVLR
>PCC10802_01433  Cas_Cas1
MQFGCGVGGCRKGFNPVAGIQLAKNFLLLPPAQIIMAFIWASLHNRHAPPEFTRRQAEAIVRAKLHNSRILLM
RLNRRNQSKLSLGAGVSEAIEDLARLMDCLPVAESLDALRGYEGKAATVYFQALGSLFTGPFTFAKRTKRPPT
DPINSLLSLGYTLLSQNVHSFVQAIGLHTHFGNLHVPRDNHPALVSDLVEEFRAQLVDSLVAYLVNSKIFEPA
DFTPPDEWGGVYLHPDALKKFLKHWEEKLQSQLTHPHTGYKVSFRRCLELQVREYAACLISEVPVYRPMIWSK
>CYA56_00830     Cas_Cas1
MSILYITQPQAILGKKYEAFTVATQEESGKWNKQSVPAQTVEQVVLMGNPQVTGDALTYALELGMPVHYLTSF
GKYLGSALPGYSRNGALRLAQFQLYQDTERRLELVKTIVSAKIQNQATVLYRQKIADSPLKNHKKLVFKKATL
DEVRGVEGIAAKDYFREWSAILNEPWSFPGRNRRPPTDPVNALLSFAYGLLRVQVTASVHLVGLDPYIGYLHE
VSRGQPAMVLDLMEEFRPLVADSLVFSVLNNREIKPDDFTESLGAYRLKDEGRKRFLQAFERKMNDEFKHPVF
EYRCTYRRAIELQARLLGRYLQEDLVYKPLVLR
>PCC9431_05236   Cas_Cas1
```

```
MAGLDPYVGYLHEATRGQPAMVLDLMEEFRPLVADNLILSVISHKEIKPNDFTDHLGAYRLSESGRKQFLQAW
ERKMNDEFKHPTFGYRCSYRRAIELQARLLSRHLQEGVPYKALTPR
>PCC9431_05818   Cas_Cas1
MQTLYVSQQNCYVALQQEKLIVKQGETIHAQVQLPLLEQILIFGNSQVSTQAIRACLWQNVPIAYLSRMGYCY
GRLLPLLRGYRQLSRYQQQLSTVERLITARAIAQTKIKNSRVLLRRQKKRLESEELEKTLQNLDYLAAQAGEA
DSWERLLGFEGAGAAQYFSAFGKCITNPNFVFTGRSRRPPGNPVNAMLSFGYQVLWNHLLALIELQGLDPYYA
CLHQAHDGHAALASDLIEEFRAAIVDSLVLWLINRKIVDAQQDFE
>PCC9431_05758   Cas_Cas1
MAAIYLIEQGTNIYKDYQRFIIHVSEKPKLEILIREVQQILVFGNIQLSTPVLQVCLQEQIPVLFLSQNGRYH
GHLWSEESTNLDNQVIQIERRNNDYFQFSVSKAVVLGKLMNSKQLLMRFNRKRKLEDVEKAIYGINQDIDALE
YVDELNALRGYEGIAAARYFPAFGKLITNSAFSFSQRFRQPPIDEVNSLLSFGYTLLFNNVLSFIIAEGLSPY
IGNFHYGERQKPYLAFDLMEEFRSPIVDSLVLKVINNGEFKPKDFDVVASTGGVYLSQSARKVFLKQFENRMN
EEVSHPDLISQVTYRHAIQLQVRRYKRCLLGGVPYESFLRAG
>PCC9431_05844   Cas_Cas4+Cas_Cas1
MQALELTRATELQISTIRVSALHALAYCPRLFYLEEVEELYTQDAAVFAGRRLHTEIEKQEDEEWEELFLESE
ELGLRGRLDALRTRDGQTIPYEHKRGRAHRDENKQPQAWESDRLQILAYAFLLESALGITVNEGRIRYHADNV
LVHVPFDDAGRSAVREAIEQARTLRQSTHRPPVTNNERLCARCSLAPVCLPEEARLAHDREWQPIRLFPEDDE
RQVIHILEPGTSVGRTGEQIKIARRNQPLETVPARQVAQLVLHSFSQISTQALHFCADQDIGVHFISGGGRYL
GSFDSRNGSIQRRIRQYAALSNPDMCLELARKLVVCRGQGQRKFLMRGTRGKKEVLETLNKAIAQMKATLKQI
PQAKSLESLLGLEGNLAAVYFGVLPCLISKEVATELHFNGRNRRPPKDRFNAMLSFGYALLLKDVMNAILTVG
LEPALGFYHQPRSQAAPLALDLLEIFRVPLVDMTVVASINRGQWDVKADFEVRGAQVWLSQTGRRKFVELYER
RKQESWKHPATGYSLTYRRLFELEVRLLEKEWSGEGGLFGQLIVR
>MRB220_04045    Cas_Cas1
MMSILYVTQPDAILSKVAEAFKVALKQEDGGWEKKSVPAQTVEQVVLMGNPQITGDALVYALELGMPVHYLSG
FGKYLGSALPGYSRNGQLRLAQYAAYYDTAHRLEIVKTIVIGKIQNQHSVLYRHDQKDNPLKSRKQLVKEQTT
LDEVRGVEGLAAREYFAAWSAMLEEQWSFNGRNRRPPTDAINALLGFAYGLLQVQVTAAVHMAGLDPYVGYLH
EATRGQPAMVLDLMEEFRPLVADNLILSVISHKEIKPNDFTDHLGAYRLSESGRKQFLQAWERKMNDEFKHPT
FSYRCSYRRAIELQARLLSRHLQEGVPYKALTPR
>MRB220_00767    Cas_Cas1
MSVLYVTQAEAVLSKVYEAFTVSLKQEDGTWKKKNIPAQTVEQVVLMGNPQVTGDALVYALELGMPVHYLSFF
GKYLGSALPGYSRNGQLRLAQYAVYNDPARRLELVKAIVTTKIHNQYNVLYRHNERDNPLKQRKTMVKTQKTI
DQVRGIEGLAAREYFACWQRMVGKQWTFNGRNRNPPTDPVNALLSFSYALLKTQVTSAVHLAGLDPYIGYLHE
VHHGQPAMVLDLMEEFRALVADNVVLGLLGKREIQPQDFTESLGAYRLSDSGRKCFLEEFERKMNDEFTHPVF
NYRCTYRRAIKLQARLLSRHLQEGIPYKPLTLR
>MRB220_01681    Cas_Cas1
MRTLYVSRQNCYITLKGENLIVKQGETIHGEVQLPLLEQILIFGKSQITTQAIRTCLWRDIPIAYLSRMGYCY
GRLIPIERGYRQLSRYQQQLTAIERLLVARKIVQAKLKNSRTFLQRQQRRHPSQIAEVAIQSLEVLIQKAGDA
ETIERLMGLEGAGAAQYFSAFGECLNNPDFVFLARSRRPPGNPVNAMLSFGYQILWNHLLTLIELQGLDPYHA
CLHQGTERHAALASDLIEEFRVPIIDSLVLWLVNSKAINAQTDFEYHDGGCYLNNSGRQTYIKYFLQRLEEEV
QNTNREKQPRWDLMTQQIKSFKEFVYQPNQIYKPYQIR
>MRB220_01706    Cas_Cas4+Cas_Cas1
MQFFTMQTLDVSILSGQTTLRVSALHALAYCPRLFYLEEVEELYTQNAAVFAGRRLHAEIEKQEDEEWEELFL
ESEELGLRGRLDALRTRDGQTIPYEHKRGRAHRDENKQPQAWESDRLQILAYAFLLESALGITVNEGRIRYHA
DNILVHVPFDDAGRAAVREAIEQARILKQSTYRPPVIDNERLCARCSLAPVCLPEEARLAHDREWQPIRLFPE
DDKRQVIHILEPGTSVGRTGEQIKIARRNQPLETVPARQVAQLVLHSFSQISTQALHFCADQDIGVHFISGGG
RYLGSFDSRNGSIQRRIRQYAALSNPDICLELARKLVVCRGQGQRKFLMRGTRGKKEVLEALNKAIAQMKATL
KQVPQAKSLESLLGLEGNLAAVYFGVLPCLISKEVASELHFNGRNRRPPKDRFNAMLSFGYALLLKDVMNAIL
TVGLEPALGFYHQPRSQAAPLALDLLEIFRVPLVDMTVVASINRGQWDVKADFEVRGEQVLLSQTGRRKFVEL
YERRKQESWKHPATGYSLTYRRLFELEVRLLEKEWSGEGGLFGQLIVR
>CENA21_01280    Cas_Cas4+Cas_Cas1
MQALELSRATEVQVSTIRVSALHALAYCPRLFYLEEVEELYTQDAAVFAGRRLHTQLEKQEDEEWEELFLESE
ELGLRGRLDALRTRDGQTIPYEHKRGRAHRDENNQPQAWESDRLQILAYAYLLESAHGITVQEGRIRYHADNV
VVHVLLDEAGRLKVKQAIQQARILRQSTYRPPVIDNERLCARCSLAPVCLPEEARLAHNREWQPIRLFPEDDE
RQVIHILEPGTSVGRTGEQIKITRRNQPVETVPARQIGQLVLHSFSQISTQALHFCAAQGIGVHFISGGGRYL
GSFDTRQGSIQRRIRQYAALSNPDACLELARKLVICRGQGQRKFLMRGTRSKAKTEKLEKAIAQMKAVLKQVP
QAKSLESLLGYEGNLAALYFSALPDLISPEVPKELHFAGRNRRPPLDRFNAMLSFGYALLLKDVMNAILTVGL
EPALGFYHQPRSQAAPLALDLLEIFRVPLVDMTVMAAVNRGQWDAKADFEIRGKQVWLSELGRRKFVELYESR
KQESWKHPVTGYSLTYRRLFELEVRLLEKEWSGEGGLFGQLILR
>PCC9445_03433   Cas_Cas1
```

```
MKPILNTLYVQTQGAYLHLERDTLKLKAEEQKPFQIPLHHLTGLVVFGNVSVSPFLLHRCAQDGRDFVWLSEY
GKFKARLSGPTTGNVLLRRAQNAAVDDGAKTLDIARYMVAGKIQNARKLLMRSARDAKQEGDRTALTKAAQHH
AESLTRLEAATTVETVRGIEGNAAKAYLEVFDHAIVQNREAFGFKTRSRHPPLDPTNSLLSFLYGLLKNECIA
ACESVGLDPQMGFLHVLRPGRASLAFDLMEELRPIVCDRLVLTLINRQQIQPDDFMERPGGAIYLTESGRKTV
LGAYQKRKQEEITHPLLKTKIPLGLVAHTQARLFARYLRGDLDHYPPFIPR
>PCC9445_03320   Cas_Cas1
MGTVYVTQEDAFIGKTDERLTIKADQKQLLDIPLIKVEGIVALGRATVSPAVVTELLERKIPLSFLKVTGKYL
GRLEPEMTKNIFVRKAQWEARGTDKAIQVVRGFIRGKLKNYKNALLRRDRQNSALNLKPTIQKLEQALNPLPK
NDQIDSLRGLEGAGSAAYFTCFNQLIQNPLFSFKTRNRRPPTDPVNSLLSLGYALLRHDIQSAVNIVGFDPYL
GYLHVPHYGRPALALDLMEEFRPLVVDAVVLASLNQQKLKPDHFSTESLSEAVSLTPEGLKIFLRLYEEKKQA
KFKHPVLGRQCTYQESFEIQGRLLAKYLMAETEQYPPLILK
>PCC9445_01064   Cas_Cas1
MSSLYVSQQGCYIALQQEQLLVKRHNEILTQVQLPHLEQVLIFGKSQVTTQAIRACLWRNIPIVYLSRMGYCY
GRIIAIERGYRALTRYQQELSLGERLGVAQNIVLAKLNNSRTLLMRQKRRHPSPCLTLAIESLDYLQQQVKQA
DSQDRLLGFEGAGAASYFAALGDCLRNPDFVFTARSRRPPGNPVNALLSFGYQVLWNHLLALIEVQGLDPYQA
CLHSSSERHAALASDLIEEFRAPLVDSLVLELVNGRVLKLEHFDYRNGGCFLNELGRKKYLQGFVQRMEEQVQ
IGAESKLRWETLMEQVKGFKAFVYNPSGGYSPYLIR
>PCC9445_00553   Cas_Cas4+Cas_Cas1
MQTIDLESINILQPTPETLRVCSLHAFAYCPRLFYLEEVEELYTQDAAVFAGRRLHEAIELEDGEEWQDLYLE
DELLGLRGRVDALRTREGKTIPYEHKRGRCYRDEQKNPQAWESDKLQVLAYCCLLEAALGVTIVEGRIRYHAD
NVLIHVPLDDSGRQWVREMIQAARELKQSPYRPPVINNEHLCSRCSLAAVCLPEEARLAHNKEWHPVRLFPAD
DEREVIHVLEPGTRVGRTGEQLKISRRNEKEEKVSIQQVSQVVLHSFSQISTQALHFLAAKDVGVHFVSGGGR
YVGSLDCRNGSVQRRIRQYRGLSEGDFCLVLARKLVSCRGEGQRKFLMRGKRRKGTAGELDKAIGQMKAVLKQ
VPKAESLDSLLGIEGKVAALYFGALPGLLGADVPAELRFSGRNRRPPKDRFNALLSFGYALLIKDVMNGILTV
GLEPALGFYHQPRTQAPPLALDLMEVFRVPLVDMTVVTAVNRQQWDVQGDFEVRGEQVWLSEEGRRKFIRLYE
QRKGEMWKHPVTGYSLTYRRLLELEVRLLEKEWSGEGGLFGRLIVR
>hensonii_00338  Cas_Cas4+Cas_Cas1
MVITNDSIAQQATLRVSALHAFAYCPRLFYLEEVEELYTQDAAVFAGRRLHVELEKQEDEAWEELFLESEELG
LRGRVDALRTRDGNTIPYEHKRGRSHRDEHNQPQAWESDRLQILAYACLIETVLQVPVAEGRIRYHADNVLVH
VPLNEAGRQTVQTAIAQARSLRASLYRPPVTENERLCARCSLSSVCLPEEARLAHDREWQPVRLFPQDDDRQI
IHVLEPGSSVGRTGDQLKITRRNQPIEKIPIRQIGQVVLHSFSQISTQALHFCSEQGIGVHFISGGGRYLGSF
DNRHGSIQRRIRQYGALSQPETCLELARQLVICRGQGQRKFLMRGQRGKDPVPEALQQAIHQMKATLKQVNKA
PALETLLGLEGNLAALYFGALPHLIGDQVDPLLQFDGRNRRPPRDRFNALLSFGYALLLKDVMNSILTVGLEP
ALGFYHQPRTQAPPLALDLMEIFRVPLVDMTAMASINRGQWEVQADFEVRGVQVWLSDTGRRKFVDLYERRKA
ESWKHPVTGYSLTYRRLLELEVRLLEKEWSGEGGLFGQLVIR
>hensonii_03184  Cas_Cas1
MRTLYVSQQGCYLSLQKELVLIKRGETILDEVQLPLLEQILIFGQSQVTTQVIRACLQRDIPIAYLSRMGYCY
GRILPVERGYRHLSRYQQQLLEVDRLVVARTIVQAKLRNSRVILQRQQRRRPSDASTFAIQSLEYLIEQAGQA
ATLDKLMGMEGAGAASYFSAFGDCLTGDGLVFIARSRRPPGNPVNAMLSFGYQVLWNHLLSLIELQGLDPYQA
CLHQGSERHAALASDLIEEFRAPVVDSLVLYLANRGIVKADEDFEYHDGGCYLKDMGRKKFLRAFLQRMEEQL
QIDEEEQPRWGLLMQQVKAYKQFVYNPAQGYKPYLIR
>hensonii_02221  [CAS]: CRISPR-Cas system spacer integrase Cas1 (CAS-
I,CAS-II,CAS-III,CAS-V)
MKLVNQKILKPTDFTWPTPEGGVYLNDPARRVFLKHFEERISLKISHPDVSDLVSYHRIIQLQIQRYKQTLKS
GTPYAGFRRVD
>hensonii_02203  Cas_Cas1
MTTLYLTEPGTVVRFQNEALLICKQGQTRTLRVAELTLVVVLPGVQLTSVVIAELLDRGIETIFLRQDGQFRG
RLQGHFATNPTIRLAQYRMVETTFGLALAHKLVEGKVRNQRVLLQRHNRKTQGQVSELVEAIDTIAAYLPRLH
NSSAPLSRDELMGIEGICARTYYQALRHYFPAIWRFSGRNRQPPLDPINALLSWGYGVLLARVFSASVQAGLD
PYLGFFHAIEPYRPNLVLDLMEEFRPIVVDQAVITLVQSDVLDLADFEPSPDGVGIWLGSTAKKLFLGELEKQ
LRQPLLYPPQNRRLSLSQILLEQARCLGRCLIDRGLNYEAFVLK
>hensonii_04279  Cas_Cas1
MTTLRTLPKIRDSISFLYFDRCKVEQESKAIAIFTKQDKYVIPCASLSTLLLGPGTRITHAAIKTLAENACEV
QWVGEDGFRFYSSGRSSSTSVQRLYHQAQTWANPQQRLTIVRKMYQFRFQEELSDQLTLQQIRGLEGVRVRTA
YSRLSKATGVEWKGRNYKQDDWFDADPINRALSSANSYLYALCQAAVVSVGYSPALGFIHIGKPLSFIYDIAD
LYKAETSIPAAFEAVASAREASGMTVRQKCREKFSNYRLMQRIIQDVDKVLGYMSGDTEIPAVSFLWDDQVDW
VEGGQDWAEET
>hensonii_02737  RVT_1+Cas_Cas1
```

```
MLPSLRDRFLSLSNFEQAWDAVASNRGCAGVDGETITDFGSSKADALSTLRRAIAEDCYRAMPLRQIWIPKRS
GGWRELGVPTVRDRLVQQALLQVLYPLMEPQFEPMSFAYRPGRSHRMAVEQVGVWRDRGYTWVLEADIVQYFD
RIQHSRLLAELKERVDPPWVVDLVEAWITTGKLTPEGILLPTAGIPQGSPISPMLANVYLDDLDEAMAQPNWK
LVRYADDFVVLARSQERLQEAWVALNSLLTAMGLELHPDKTGITTFKQGFRFLGHTFAGDVVVPPTPQRRRIK
PESRQGSEMRLIHVDRREAGETVMQQALVEALKAAQQPIPPPLFVVLGYGVRSDSSVKIESPEREWSNGMSSL
YVVEQGTCVNKEEGRFRLSPPRGDSLEVPIREVERILIFGQVQVSTAVIGLCLNQQIPVIFLSQTGEYKGHLW
SAEFEDLTAERYQFLRQAEDPFQICTARAIVWGKLMNSKQLLLRLNRKRKLEVVAEAIQGIQRDIESLSSVTS
LEQLRGYEGITAARYFPALGQLLVNPGFSFRERNRRPPKDPVNSLLSFGYTLLFNNMMSLILAEGLNPYVGNL
HGTERKMPALVLDLMEEFRSPIVDSLVVKLINSKTLQPTDFTWPTSEGGVYLAASARRLFIQKFEQAITQSVK
HPDLQEAVSYRRAMHLQVQRYKAALLKDIPYEAFLRSV
>CENA302_01023   Cas_Cas1
MQTLYISEQNCYVCLQKETLIVKQGDTVHVEVQLPLLEQILIFGRSQITTQVIRACLWRDIPIAYLSRMGYCY
GRILPISRGYRQLSRYQQELSPVDRLITARAIVKGKLKNSRILLRRQRKKRESETLERTLQSLDYLAEQAVQA
DTWERLMGFEGTGAAQYFSVFSECLTNPDFVFSGRSRRPPGNPVNAMLSFGYQVLWNHLLALIEVQGLDPYYA
CLHQANDGHAALASDLIEEFRAPLVDSLVMWLINRNIMDVKEDFQFKNGGCYLNDSGRKKFLRGFLQRMTEEI
QTNDGKQPKWDLLTQQVRNYKQFVYNPSHHYKPYQID
>CENA302_02553   Cas_Cas1
MTILFLLEQDTLVRQTDERLEIFKYDKRLTDVPLCKLEAVVVFGKVILTIPVLKVLNERSIPITYLSEEGRTV
ATLLPEPNPNAILRAKQYQASFDQRKTLDIAREFVRGKLFNQHTVLARFSRQNERSERIGVALKSLKSCQKSL
DQTTSLNELRGYEGQGASSYFGVLGELLTDTPWEFSHRTRRPPTDPVNALLGFGYALLYGDCRAALHAVGLDP
YQGFLHGERYGRANLALDLMEEFRPIFVDGLVLQLLRNNGLEQDCFRNYPGGAVYLNEEGIKKFIQSYEQRKN
TTFQHPVFEIQTDYRRALILQSRLLAKHLNQEIDNYLPLKIR
>CS505_03049     Cas_Cas1
MTILFLLEQDTLVRQIDERLEIFKHDKRLTDVPLCKLESVVVYGRVILTIPALKILNQRSIPITYLSEEGRTV
ATLLPEPNPNAILRSKQYQASFDTHKTLAIAREIIRGKLFNQHTILARFSRQNERSEKVISALKSLKACQKSI
DQTTSLNELRGYEGQGASSYFGVLGELLTGTPWSFSHRTRRPPTDPVNALLGFGYALLYGDCRAALHTVGLDP
YQGFLHGERYGRANLALDLMEEFRPIFVDGLVLQLLRNNSLEKESFINYPGGAVHLNEQGMKKFIQSYEQRKN
TTFQHSVFEIQTDYRRALILQARLLAKHLNQEIEKYLPLKVR
>CS505_01582     Cas_Cas1
MQTLYVSEQNCYICLQKEILVVKRGEVTQAQVQLPLLEQILIFGSSQVTTQVVRACLLRDIPIAYLSRMGYCY
GRILPISRGYRQLSRYQQQLSAMEKLVTAREIIKSKIKNSRVLLQRQKKKIASANLERVLESLNYLAEQTASA
DSCERLMGIEGAGAAQYFSAFGECLSNENFIFLGRSRRPPGNQVNAMLSFGYQVLWNHLLALIEIQGLDPYYG
CLHQGHDGHPALASDLIEEFRSPLIDSLVMWLINRNIFDVGADFEFIKGGCYLNNSGRKKFLQAFLQRMTEEV
QTEDGTKQPKWDLLTQQVRAFKQFVYNPAHHYRPYRIH
>CYA34_04354     Cas_Cas1
MSILYITQPQAILGKKYEAFTVATQEESGKWNKQSVPAQTVEQVVLMGNPQVTGDALTYALELGMPVHYLTSF
GKYLGSALPGYSRNGALRLAQFQLYQDTERRLELVKTIVSAKIQNQATVLYRQKIADSPLKNHKKLVFKKATL
DEVRGVEGIAAKDYFREWSAILNEPWSFPGRNRRPPTDPVNALLSFAYGLLRVQVTASVHLVGLDPYIGYLHE
VSRGQPAMVLDLMEEFRPLVADSLVFSVLNNREIKPDDFTESLGAYRLKDEGRKRFLQAFERKMNDEFKHPVF
EYRCTYRRAIELQARLLGRYLQEDLVYKPLVLR
>CYA34_01952     Cas_Cas1
MLTLYVSQQGCYLTLKAESLLIKKGKDVLAEAQIPLLEQVLVFGQSQVTTQAIRACLSRNVPIAYLSRMGYCY
GRILPVERGYRHLSRYQQQLTQIDRLIIARRIVQAKLKNSRVILQRQARRQSSPTINLAVQSLNHLMEKVIQA
ESWDRLMGLEGAAAAQYFSAFGECLTNPKFVFTVRSRRPPGNPVNAMLSFGYQVLWNHLLSLIELQGIDPYFG
CLHQGTERHAALASDLIEEFRAPIVDSLVMYLINRNMIDANTDFTYHNGGCFLNDSGRIKYLKAFLQRMEEKL
TTAAGENQPRWDTLNQQVKAFKQFVYNPIEGYKPYEIR
>MDT14a_01923    Cas_Cas1
MQTLYVSEQNCYVCLQKETLLVKQGDTVYIEVQLPLLEQILIFGQSQITTQVIRACLWRDIPIAYLSRMGYCY
GRILPISRGYRQLSRYQQQLSPIDRLITAREIVKGKLKNSRVLLRRQQKKRDLELIERVLDSLDYLATQAATA
DNLERLRGFEGAGAAQYFSAFGECLTNSDFVFSGRSRRPPGNPVNAMLSFGYQVLWNHLLALIEIQGLDPYYA
CLHQGNDGHAALASDLIEEFRAPLVDSLVMWLINRKIMDVKDDFEFKNGGCYLNNLGRKKFLRGFLQRMTEEI
KTDQEMKQPKWDLLTQQVKAFKQFVYNPSHHYKSYRID
>PCC7113_06589   Cas_Cas1
MSVLYVTQPEAVLSKKYEAFTVALKQEDGSWQKKSVPAQTVEQVVLMGNPQVTGDALVYALELGMPIHYLSEF
GKYLGSALPGYSRNGPLRLAQYALYHDTKGRLELVKAIVSAKIHNQHLVLYRYKLMGNPLKSHKKLVAEQTTL
DQVRGVEGLAAKDYFGCWDALLKEPWSFPRRNRRPPTDPVNSLLSFAYGLLRVQVTAAVHLAGLDPYIGYLHE
VHHGQPSMVLDLMEEFRPLVADSLVLSVLSHRELQPDDFTESLGAYRLSEDGRKTFLQAFERFLTGMQWASLQ
ISMEKLIKPSEDSVRVYVLDAGALKRTLTYGSQQPRQDTTIIL
>PCC7113_06551   Cas_Cas1
```

```
MNNLYITGTESYLKVQDQHLLVFHPHKPCCKVAINQVNNLVLIGYCKRSDKAIYFALSRCIPVLFLDDKGNYF
GRIETEERRRVKYLAHQLKRAKDLEFTRAIAESIVRAKLHNSCTLLRHFSHCHPTKATQTTLDAMALLLDDLP
LADSVDVLREYGQTATKLYFQSLGCLFAGALNFDQPIQHQPTHPINSLLNLGYTLLNQNIYALLQAVGLHSHF
GNLHIHRDHHPALVADFMEEFSALLVDALVVDLVILQLFTPADFTLPDARGGVYLYPNAINKFIQYWEKTLQS
EVNHLYAGTVNYQQCLEWQVQEYVACLLGEREFYRPFLSKK
>PCC7113_06550   RVT_1+Cas_Cas1
MDFPLEDLIAAWRQVRRGSRSAGIDGITTDLFAGVAKEQLYSLQRQLQQEHYAAHPALGFYLRKTRGGKRLIG
IPVVLDRIVQRLLLEELYLPLEDTFLDCSYAYRPGRGIQMAVQHLESYYQFQPTWVIKADIAQFFDNLCHALL
FTHLEQLQLEPIVLQLIEQQLKAGIVIKGQRLFPQKGVLQGAVLSGALANLYLTEFDRQCLSHGLNLVRYGDD
FVVVAPDWIQANRALEQITTGLAQLYLTLQPEKTKIFAPDEEFTFLGYQFAGGTVTAPPPVVNEDSEAFSESG
SPSRRKTRQPPKLKSRPPKTCNIVKRSQFPIADLDHYWKEPMTTLYITDQGAYLKVEHQQFVVFYQGELRCKV
PVNRVSHIVLFGCCNLSHGAVSLALRRRIPVLYLSQKGRYFGRLQTEGQAEVEYLVRQVQCALDPDFTRRQAK
SIVLGKLHNSRILLMRLNRRVKSEKAAKAIDSLALLMLDVPDAESIEVLLGYEGHGSHLYFQALGTLFKEPFV
FEKRTKRPPTDPVNSLLSLGYTLLSQNMYSMVQTVGLHPHFGNLHVPRNNHPALVSDLMEEFRALLVDSLVAY
LVNSKIFVLEDFTPPDERGGVYLFPDALKKYLKHWEERLQSKMTHPHTGYKVSNRRCLELQVWEYIACLMGEQ
EIYRPMTWKM
>PCC7113_03332   Cas_Cas1
MGTVYITQEDSFIGKTDERLIVKADKKNILDVPFIKVEGVVVLGRATVSPAVVSELMERKIPLSFMMGTGKYL
GRLEPEMTKNIFVRKAQWQAAGETEQSVHVVRGFVRGKLKNYRSILNRNQRKYSELDWQAAIDRVQNAIAPID
KTNSINSLRGLEGAGSAAYFGAFNQLIRVEGFSFAARRRRPPTDPVNALLSFGYSLLRHDVQSAVNIVGFDPY
LGYLHFEHYGRPGLALDLMEEFRPLVVDSMVLNAINHGILKPSDFTSEPLSNAVFLSPEGRREFLRQYEMKKQ
SEFKHPVLGRKCTYQEAFEIQARLLAKYLMGEIDKYPPLFLR
>O77_04775 Cas_Cas1
MSLLYLTQQGAKLQKNQNRLVLTLPESASIHIPVQEVEQVMVFGNVQLTTQALTTLLLRQVPVMFLSQTGSYR
GHLWRDRHDHRAQAAQFARLRDPAFQLALAREIVVGKAWNSKQLLLKLNRRRNLATVDTAIARIDREIETVNH
LDPSPDSLPDIIPNILEQLRGYEGAIAAHYFPAFGQLILPPEFTLTERNRRPPRDPVNAMLSFGYTLLCNNVL
SLLILEGLNPYLGNLHRSDRNEAQLAFDLMEEWRSPIVDVLVIALLNQQIFKLHYFSPPQPNGGIYLNEQGRR
RFLQAFERRIMTPVHHPDAREPVPYRRAIQLQIRRYKRSLLENIPYETFRRVT
>O77_04697 RVT_1+Cas_Cas1
VAIGHLVEQWIGSGVSTASGLILPNKGVPQGAVISPILANVYFDDFDEAIEAAGLKLVRYADDFVILAKSKAR
IERAYNLVASLLHAMGLELHPDKTRVTTFNEGFRFLGHTFVRSLVVQDQPKTRQQWQLEQTAIAQKATHPEAS
PILHHSDPPPQTTQMQQALLAAIQTAEQPIPPPLYVVMGYRVRDEKPVQIESKEWSWKNGMSTLYLVRQGTTL
RKDHGRFVIEGGRFETDEDESLTAQSPPSQPPTPKSKPQNPQSKTPEPFPPLEIPIKEVGRILVLGNVQISTS
ALSECLEHQIPVVFMSRAGDYKGHLWSSEFCDLPTEAAQFGRRHDPGFQVKMAQQILHGKLTNSRHLLLRLNR
KRKVEGLSAKIHRIDQHIAALTKTTDLDAMRGHEGASARLYFQALGQLITNPGFSLTERNRRPPKDPVNSLLS
FGYTLLFNNVLSLILAEGLNPYLGNLHRSDRKEPHLAFDLMEEWRSPIVDSLVMWLINKKAIRPTDFTFPNAE
GGVYLENTARRVFLKHFEDRITETVTHPTVQQPVSYRRAIQLQIQRYKKCLQDSQPYAPFIRSI
>O77_02322 Cas_Cas4+Cas_Cas1
MPESTANTSTASTIAARSVECATEQETLRVNALHALAYCPRLFYLEEVEELYTQDDAVFAGRRLHAELERDEG
DDWEELVLESAELGLRGKVDALRNRDGQTIPYEHKRGRCYRDREGKAQAWESDRLQILAYACLIEAARKIPIQ
EGRIRYHADNVVVKVILTDTGRQEVREAVQRARALRQSTERPPITTNERLCVRCSLSPVCLPEETRYTQDPER
HPVRLFPEDDDREIIHITEVGTMVGKSGEQLKISPREKPPVTLPGRQVGQLVIHSFAQVSTQALYFCAEQDIG
VHFISGGGRYLGSLDTRQGSIQRRVRQYRALSNPEFCLELARKLVLCRGESQRKFLMRGRRGSEAENETLNRT
IDQMQRVLKQVPQASALDSLLGLEGNVAALYWQAVPTMLSQNAPPEFHFVSRNRRPPKDRFNALISFGYALLL
KDIVNAILVVGLEPALGFYHQPRTQAPPLALDLMEIFRVPLVDLAVMGSVNRNQWDVETDFDIRGEQVLLSSS
GRRKFIDLYERRKEESWKHPALGYSLTYRRLLELEVRLLEKEWSGEGGLFGQLVIR
>CYA126_04156    Cas_Cas1
MSILYITQPQAILGKKYEAFTVATQEESGKWNKQSVPAQTVEQVVLMGNPQVTGDALTYALELGMPVHYLTSF
GKYLGSALPGYSRNGALRLAQFELYQDTERRLELVKTIVSAKIQNQATVLYRQKIADTPLKNHRKLLFKKATL
DEVRGVEGIAAKDYFREWSAILNEPWSFPGRNRRPPTDPVNALLSFAYGLLRVQVTASVHLVGLDPYIGYLHE
VSRGQPAMVLDLMEEFRPLVADSLVFSVLNKREIQPDDFTESLGAYRLKDEGRKRFLQAFERKMNDEFKHPVF
EYRCTYRRAIELQARLLGRYLQEDLVYKPLVLR
>CYA126_01557    Cas_Cas1
MLTLYVSQQGCYLTLEAESLLIKKGKEVLAEAQIPLLEQVLIFGKSQVTTQAIRACLIRNVPIAYLSRMGYCY
GRILPVERGYRHLSRYQQQLTRVDRLIIARRLVQAKLKNSRVILQRQARRQSSEVITFAIQSLNHLMEKVIEA
ESWERLMGLEGAAAAQYFSALGECLTNPKFVFTVRSRRPPGNPVNALLSFGYQVLWNHLLSLIELQGLDPYFG
CLHQGTERHAALASDLIEEFRAPIVDSLVMYLINRNMIDANTDFTYHDGGCFLNDSGRIKYLKAFLQRMEEKL
TTATGENQPRWDTLTQQVKAFKQFVYNPIEGYKPYEIR
>VB521301_02857  Cas_Cas4+Cas_Cas1
```

```
MQILQHIELSTKLETIRVSALHALAYCPRLFYLEEVEELYTQDAAVFAGRRLHAELEKQEDEDWEELFLTSEV
LGIRGRVDALRTREGETIPYEHKRGRCHRDENKQPQAWESDKLQILAYCCLLESSLGIHIKEGRIRYHADNVL
VRIPLDDLGRSAVREAIEKARHLKQSIHRPPITDNERLCARCSLSPVCLPEEARLAHDRQWQPIRLFPEDDDR
QVIHVLEPGTSVGRTGEQIKIARRNQPLETIPARQVAQLVLHSYSQISTQALHFCGEQGIGVHFISGGGRYLG
SFDTRNGSVQRRIRQYAALSSPDTCLELARKLVICRAQGQRKFLMRGKGKNKASEKLTKAIAQMKAVLKLVPH
AKSLESLLGYEGNLAALYFDALPDLISPDVPAEMGFDIRNRRPPKDRFNCLLGFGYALLLKDIMNAILIVGLE
PALGFYHQPRSQAAPLVLDLMEIFRVPLVDMIVMSSVNRNQWDIKADFHICGQQVWLSETGKRKFVELYENRK
QETWKHPATGYSLTYRRLLELEVRLLEKEWMGEGGLFAQLIVR
>VB521301_02614  Cas_Cas1
MSVLYVTQPDAVLSKAAEAFKVALKQEDGNWEKKSVPAQTVEQVVLMGNPQITGDALVYALELGMPVHYLSGF
GKYLGSALPAYSRNGQLRLAQYAAYCDAAHRLEIVKTIVTGKIHNQHAVLYRHDQKDNPLKARKQLVKEQTTL
NELRGVEGLAAREYFAAWSKMLEEEWSFHGRNRRPPTDAINALLGFAYGLLQVQVTAAVHMAGLDPYVGYLHE
ATRGQPAMVLDLMEEFRPLIADNLILSVISHKEIKPDDFTDNLSAYRLCESGRKQFLQAWERKMNDEFKHPTF
GYHCSYRRAIELQARLFSRHLQEGIPYKALTLR
>VB521301_02369  Cas_Cas1
MRTLYVSQQNCYVSLQQEMLIVKQGETICAEVQLPLLEQVLIFGKSQVSTQAIRACLWRDIPIVYLSRMGFCY
GRILPIERGYRQLCRYQQQLSPVEKLVTARAIVQAKLKNSRVLLRRQRKRLESQVLEQALQSLDYLVSQAAEA
DSWERLMGFEGSGAAQYFSAFGECLTNSEFVFTGRSRRPPGNPVNAMLSFGYQVLWNHILAMIEVQGLDPYYA
CLHQGSERHAALSSDLIEEFRAPIVDSLVLWLVNTKTVNTQDFEYRDGGCYLNDSGRRKYLKAFLQRMEEEVQ
TDFEEKQPKWDLLTQQIKAYKQFVYNPSHAYQPYLIR
>VB521301_07578  Cas_Cas1
MIFGNIQLSTPVINACLQEKISVLFLNQSGQYHGHLASEESTNLDNHLVQMEKRNSDEFQLQVSKAIVCGKLV
NSKQLLMRFNRKRKIAGVDKAIYGIGQDIAALNLVDDLDSLRGYEGVTAARYFPAFGELITNPNFGFSVRNRQ
PPTDPINSLLSFGYTLLFNNVLSFIIAEGLSPYIGNFHYGERQKPYLAFDLMEEFRSVIVDSLILKIINNYEF
KIQDFDILSSTGGVYLNQSSRRIFLKQFEVRMNEETSHPDLQSQVTYRHAIQLQVRRYKRHLLLGIPYEAFLR
PA
>VB521301_04864  Cas_Cas1
MGTVYITQEDAFIGKVDERLHVKFEKTTILDVPLLKVDGVVVLGRATVSPAAVNELLQRQIPLTFLTDTGSYL
GRLEPEMTKNIFVRKAQWDAAGDTPRAIHVVQGFVRGKLKNYRTILVRRQRECADIDLSTFIHRLEQAIAPID
TTHNIDSLRGLEGAGSAAYFGCFDRMIRNMAFSFTNRVRRPPTDPVNSLLSFGYSLLCHDIQGAVNIVGFDPY
LGYLHCQRYGRPSLALDLMEEFRPVIVDAVVLSALNKQFLKAEDFVQEPISNAVSLTKEGRKTFLQLYAQKKQ
SEFKHPVLGRKCTYQEAFEIQARFLAKYLMGEIEKYPPLVMK
>VB511283_01515  Cas_Cas1
MGTVYITEADAFIGKADERLNVKANKQTLLDVPLIKVDGVVVLGRATVSPAALAELLERHIPLTFLTDTGKYL
GRLEPEVTKNIFVRKAQWQAAENNSQAVHIVKGFVRGKLKNYRNILLRAQRENSQLDLNASITRLEGAIAPID
TTSNIDSLRGLEGAGSAAYFGCFDQLIRAQGFSFEARRRRPPTDPVNSLLSLGYSLLRHDVQSAVNIVGFDPY
LGYLHVERYGRPSLALDLMEEFRPVVVDAVVLSAINKKTLVPADFTTEPLSGAVSLTKEGLRTFLRLYEQKKQ
SKFKHPVMGRQCTYQEAFEIQARLLAKYLMGETDKYPPLVLK
>VB511283_05168  Cas_Cas1
LQKEELHFNGRNRRPPLDRFNALLSFGYALLLKDVMNAILTVGLEPALGFYHQPRSQAAPLALDLLEIFRVPL
VDMTVMASVNRGQWDTKGDFEIRGEQVWLSEAGRRKFVELYERRKQESWKHPITGYSLTYRCDC
>VB511283_05167  Cas_Cas1
VFELCDLSVLCKFCVAQLVLHSFSQISTQALHFCADQGIGVHFISGGGRYLGSFDTRQGSIQRRIRQYAALSN
FDTCLELARKLVVCRGQGQRKFLMRGTRGKASVPETLDKAIAQMKAVLKQVPQAKSLESLLGLEGNLAALYFN
ALPCLISLEVAERRIAL
>ASM187336_00953 Cas_Cas1
MSILYITQPQAILGKKYEAFTVATQEESGKWNKQSVPAQTVEQVVLMGNPQVTGDALTYALELGMPVHYLTSF
GKYLGSALPGYSRNGALRLAQFQLYQNTERRLELVKTIVSAKIQNQATVLYRQKIADSTLKNHKKLVFKKATL
DEVRGVEGIAAKDYFREWSAILNEPWFFPGRNRRPPTDPVNALLSFAYGLLRVQVTASVHLVGLDPYIGYLHE
VSRGQPAMVLDLMEEFRPLVADSLVFSVLNNREIKPDDFTKSLGAYRLKDEGRKRFLQAFERKMNDEFKHPVF
GYRCTYRRAIELQARLLGRYLQEDLVYKPLVLR
>ASM187336_03438 Cas_Cas1
MTTLYLTEPGTVVRYQNQSFLVKYQDQTHSCRLAELTLVVILPGVQLTDEAIAQLLDRGIETLFLRRDGQFRG
RLQGQFQTNSMFRLAQYRIVDTTFGMAIAQLLVMAKIRNQRVLLQQRHRSTKAQISELAEAIDAIAAYMLQLK
NTSTPLNRSELMGIEGICARTYYQALRHYFPPRWNFTGRNRRPPLDPINALLSWGYGVLLARVFAACVQAGLD
PYLGFFHATEPYRPNLVPDLMEEFRPVVVDQAVISMIKSDLLQREDFQPSPDGEGIWLGVTAKKLLLGQLERR
LREPFLYPAQNRRLSLSQILLEQARSLARCLVESRLDYQGLVLK
>IPPASB1200_00243      Cas_Cas1
```

```
MKTVYISQQGCYLCLEKEKIRVTKNKQIINEIQLPLVEQILIFGKSQLTTQLIRKCLWQNIPILYLSRMGYCY
GRIIAIERGYRQLSRYQQQVEFHDKLIIAKTIVSAKIHNSKIILQRQQRRKHNIELQQTIDSLTYLEEQALKV
DSMDKLFGYEGAGAASYFSAFGQCLSNSDFIFYSRSRRPPGNEVNALLSFGYQVLWNHILSLIELQGLDPYYS
CLHQNSERHASLASDLVEEFRAGIVDSLVLYLINKNIINAQEDFMFRNGGCYLNDKGRKKYLQVFITRMEEMV
TNDFGEKEPRWDLLMQQVRKIKHFFYNPVIGYQPYLFR
>IPPASB1200_01204      [CAS]: Type I-C CRISPR-Cas system spacer integrase
Cas1 (CAS-I-C)
VSLTKEGLKVFLRLYEEKKQSKFKHPVLKRQCTYQEAFEIQTRLLAKYLMDETEQYPPLIVKK
>PCC7116_04914   Cas_Cas1
LGYENIRYLHKPYFAETITEKPPQTEEDDSPFQKNIWNAEMAAIYLIEQGTNIYKDYQRFIIHVSEKPKLEVP
IRDVQQIIVFGNIQLSTPVIQACLKEQIPVVFLSQTGTYHGHLWSEKSIHLDNQLVQAERRNDDLFRFSVSRA
VVLGKLLNSKQLLMRFNRRRKIGKVEEAIYGINQDIDALNYVDNLDTLRGYEGIAAARYFPAFGNLITNPKFS
FSQRFRQPPTDEINSLLSFGYTLLFNNVLSFIITEGLSPYIGHFHYGDKQKTYLAFDLMEEFRSPIVDSLVLK
IINKSLFKPQDFDVVASTGGVYLSQTSRRVFLKQFENRMNEEISHPDLISQVTYRHAIQLQVRRYKRCLLSDN
IYESFLRAD
>PCC7116_00230   [CAS]: Type I-C CRISPR-Cas system spacer integrase Cas1
(CAS-I-C)
LKIILKYTINFALVLKGLIEPKDFETEPVSNAVSLTKEARKTFLKLYTQKKQTRFTHPVMKRKCSYQEAFEIQ
ARLLGKYLMGETDKYPPLVRW
>PCC7116_00227   Cas_Cas1
MGTVYITQMDSFVGKVDERLTVKFEKNKILDVPLIKVDGVVVLGRATISPAVVYELLERHIPLTFLTETGRYL
GRLEPEVTKNLFLRKAQWRATENSPTSVGAVQSFVRGKLKNYRNVMVRRQRESAKLDLSKNITHLENAIAPID
TTENIDSLRGLEGAGSAAYFSSFDQMIRTTEFKFTRRVRRPPTDPVNSLLSFGYSLLRHDVQGAVNIVGFDPY
LGYLHCDRYGRPSLALDLMEEFRPLVVDAMVLSTLNKYLCKINCIF
>ATCC29413_04336 Cas_Cas1
MGTVYVTQDDSFISKIDERLNVKFEKKVILDVPLIKIDGLVVMGRASISPAAIFELIDKKIPLTFLTNNGKYL
ASLEPEMGKNIFVRAAQWKAAGESDQAIHVTQGFVRGKLKNYRHSLLEAQRRYEVDLNSNISQLSHAITSIDK
ANSINTIRGLEGAGSAAYFGCFNQLIRVDNFSFHTRNRRPPIDPVNSLLSLGYSLLRHDIQGALNIVGFDPYL
GYLHTERYGRPSLALDLMEEFRPLIVDAVVLTAINRRMLSPKDFVTEPISNAVSLTKEGLHIFLRLYQEKKQT
QFKHPVMQKKYTYQETFEIQARLLAKYLMGELDKYPPLVMR
>PCC7507_02970   Cas_Cas1
MGTVYITQDDAFIGKTDERLNVKFEKKTILDVPLIKIDGLVIMGRASISPLAIAELVNQKIPLTFLTSSGKYI
ARLESEMSKNIFVRSAQWKAAGETAQAIHVTQGFVRGKLKNYRYALSQAQRRYSELDLNAGITQLANAIASLE
KANSINSIRGFEGAGSAAYFGCFNQLIRVDNFNFSTRNRRPPLDPVNSLLSLGYSLLRHDIQGALNIVGFDPY
LGYLHTERYGRPSLALDLMEEFRPLIVDAVVLTAINRRMLAPKDFITEPVSNAVSLTKEGLHIFLRLYQEKKQ
DKFKHPVLQKQYTYQETFEIQARLLAKYLLGEIHQYPPLVMK
>PCC7507_02705   RVT_1+Cas_Cas1
MFTLEHLHFAWLQVRAGSKTAGIDGISVELFESMATEQLQNIANQLYDETYTASPAKGFYIPKKNGSKRLIGI
PTVRDRIIQRLLLDELYFPLEDTFLDCSYAYRPGHNIHQAVQHLYGYYQYQPKWIIKTDIADFFDNLSWALLL
TALDELSLEPIVLCLLEQQLHSGIIIAGQYRNFGKGVLQGGILSGALANLYLTNFDRKCLSQSINLVRYGDDF
VIACNSWQEANRILDKITTWLGEVYLTLQPEKTQIFTPNEEFTFLGYRFAGGEVYAPPPPQPIRQGEWVINDS
GTPYFRTKPKPAKPVSHPPKACSIDKPSNFPQASLSHYWQETMTTLYITDQGAYLSVKNQQFQVYYQGELRIK
VPVSRVSNIVLFGSCNVSHGAVSMALRRRIPIMYLSQKGRYFGRLQAEGEARVEYLMLQVERCQNSEFTRRQA
EAIVKAKIHNSRILLMRLNRRQKSKNVDETIIKKASNELEILMSKLPFADNMDVLRGYEGRAATIYFQALGNL
FSGSFTFEKRTKRPPTDPINSMLSLGYTLLSQNVYSFIQSVGLHTHFGNLHVPRDHHPALVSDLMEEWRAGLV
DSLTVYLVNSEVFTIDDFTLPDERNGVYFQPHALKKFLKHWEEKLQSEVTHPHTGQKVVYRRALELQVREYIS
CLKGEVEVYRPMIWEK
>PCC7507_02702   Cas_Cas1
MSTIYITEQDVSFQIQHHYLKVFHQQNQRISIPIRNISQFIIFGNISLPKEVIKIIHSHHIPILYLTQTGEYL
GRLENPSQLPAKYLTYQRRRARDNEFNRATAESIIWAKLHNQHTFLQSWTRHYTDYTIQRALNYLTLLMDNLP
LAPRIAELQEYSKEADNIYYCAVASLFSFHNGSSFTNAKRTSGLINLGNQLLHQYIYTLLNTAGLNPNYAILH
RDGHHELPLAWDFTAEFHAPIVDDMVLNFVRNLTNTNGNGNGKSHSHKLLQNFLQHWEARLRTFILHPYAGEI
SYRQCLDLQVREYLASLLGDVEFYRPLALKFRPTNSNFANTIKPQTVPLKLVKR
>PCC7507_02570   Cas_Cas1
MTTLYITDQGAYLSVKNQQFQVFYQRELRIKVPVSRVSNVVLFGSCNVSHGAVSMALWRRIPIMYLSQKGSYF
GRLQAEGDARVEYLMLQVERCQNPEFVKKQAEAIVRAKLHNSRILLMRLNRRQKSKNVDGTIIKKASNELEIL
MSNTVFL
>PCC7507_02569   Cas_Cas1
```

```
MDVLRGYEGRAATIYFQALGSLFSGSLTFEKRTKRPPTDPINSMLSSGYTLLSQNVYSFIQSVGLHTHFGNLH
VPRDHHPALVVCQG
>PCC7507_02568   Cas_Cas1
MEEWRAGLVDSLTVYLVNSEVFTIDDFTLPDERNGVYFQPHALKKFLKHWEEKLQSEVTHPHTGQKVVYRRAL
ELQVREYISCLKGEVEVYRPMIWEK
>PCC7507_01241   Cas_Cas1
MSILYVTQPDAVLSKAAEAFKVALKQEDGNWEKKSVPAQTVEQVVLMGNPQITGDALVYALELGMPVHYLSGF
GKYLGSALPGYSRNGQLRLAQYAAYCDAAHRLEIVKTIVTGKIHNQYSVLYRHDQKDNPLKSRKQLVKEKTTL
DEVRGVEGLAAREYFAAWSVMLEEQWSFNGRNRRPPTDAVNALLGFAYGLLQVQVTAAVHLAGLDPYVGYLHE
ATRGQPAMVLDLMEEFRPLVADNLILSVISHKEIKPDDFTDNLGAYRLSEAGRKQFLQAWERKMNDEFKHPSF
GYHCSYRRAIELQARLLSRHLQEGVPYKALTLR
>PCC7507_00921   Cas_Cas1
MTTLYITDQGAYLSIKNQQFQVFCQGQLRIKVPVSRVSNVVLFGSCNVSHGAVSMALRRRIPIMYLSQKGRYF
GRLQAEGDARVEYLMLQVERCQNLEFTRNQAEVIVTAKIHNSRILLMRLNRRKSPEFDDSLVKEAIKNLDVLM
DKLPFAESMDALRGYEGKAATVYFQALGNLFSGAFKFEKRTKRPPTDPVNSLLSLGYTLLSQNVFSFIQSVGL
HTHFGNLHVPRDNHPALVSDLMEEFRAQLVDSLVAYLINSNIFTADDFTLPDEKGGVYLQPHALKKFLKHWEE
KLQSELTHPNTGYKVSFRRCLELQVREYVACLRGEVKVYRPMIWKP
>PCC7507_00865   Cas_Cas4+Cas_Cas1
MLVWYQIIFVQYFSNILENPVSSDEEWEELFLESEELGLRGRLDALRTRDGQTIPYEHKRGRCYRDENKQPFD
SAQGKPQAWESDRLQILAYAYLLESAWGITVKEGRIRYHADNVLVHVPVDDAGRTAVKEAIQQARILRQSTHR
PPVTNNERLCARCSLAPVCLPEEAQLAHDKEWQPIRLFPEDDERQVIHILEPGTCVRRTGEQIKITRRNQPVE
TVPARQVGNLAALYFSALPCLISPEMPKELHFADRNRRPPQDRFNAMLSFGMLCY
>PCC7524_05590   Cas_Cas1
MQTLYVSEQNCYVSLQKETLLVKQGETVYVEVQLPLLEQILIFGRSQVTTQVIRACLWRDIPIAYLSRMGYCY
GRILPISRGYRQLSRYQQQLSPIERLVTARAIVKGKLKNSRVLLRRQRKKKESEILERVLQSLDYLASQAAEA
DTWERLMGFEGAGAAQYFSAFSECLTNPDFVFSGRSRRPPGNPVNAMLSFGYQVLWNHLLALIEIQGLDPYYA
CLHQGHDGHAALASDLIEEFRAPLVDSLVMWLINRKIVSAEADFDFKNGGCYLNDSGRKKFLRAFVQRMTEQI
QTEDGTQPKWDLLTQQVKAFKQFVYNPSHHYQPYKID
>PCC7524_00921   Cas_Cas1
MGTVYITQEDAFIGKIDERLSVKFDKKVILDVPLIKIDGVVILGRATVSPAVISELLERHIPLTFLTENGRYL
GRIEPEVTKNIFIRKAQWQAVGESPAAIHVVQGFVRGKLKNYRNILTRHQRECAEIDLSKHITHLEQAIAPID
NTHNINSLRGLEGAGSAAYFGCFNQLIRHPDFEFSHRLRRPPTDPVNSLLSFGYSLLRHDVQGAINIAGFDPY
LGYLHCDRYGRPSLALDLMEEFRPLIVDAVVLSTLNKQLLTPTDFVTEPVSGAVSLTPEGRKTFLSLYQKKKQ
SEFKHPVMGRKCTYQEAFELQARLLAKYLMGETEKYPPLVLK
>PCC7367_01763   Cas_Cas1
VKVLLNTLYVQTQGSYLRLDHETLKIEVERELTFQIPLHHLGAIVTFGNVLISPFLIHRCAEDGRALIWLSEY
GRFRARANGSTSGNVLLRQAQYSALESTDAVLLIARYIVAGKLQNARLILMRSARDAKNESDRTALSAAAATH
AEAIKSAEKAESVEKLRGIEGYAAKAYFAAFSHMIRLNRDAFALTERSRRPPRDPINAMLSFVYTLLVNDCVS
ACEGVGLDPQVGFLHAIRPGRPSLALDLMEELRAPIADRLVLTLINRGQIKPEHFLERPGGAIFMTDDARKTV
LAEFQKRKKVEVNHPILATKVPLGLVNHTQARLLARHLRGDVPTYQPFILR
>PCC9605_00004   Cas_Cas1
MGTVYITQEDAFIGKVDERLNVKFDKKTILDVPLIKIDGVVVLGRATVSPAVVNELLERQIPLTFLTETGRYL
GRLEPEVTKNIFVRKAQWQAAGESNQAIHAVQGFVRGKLKNYRNILNRRQRESTDLDLSASIQRLEQVIAPID
TTQSINSLRGLEGAGSAAYFGCFDQMIRNTEFKFTNRVRRPPTDPVNSLLSFGYALLRHDVQGAVNVVGFDPY
LGYLHCDRYGRPSLALDLMEEFRPLVVDAIVLSTLNKLLLTPADFVTEPLSGAVSLTNEGRKTFLRLYAQKKQ
SEFKHPVLGRKCTYQEAFELQARLLAKYLMGESEKYPPLVLK
>PCC9605_05970   Cas_Cas1
MEFRSNYKIPKPPKLLEHEPLLPTLYLESDTTINLAEDPSYLYLENPEHSYSQSIFPLNAARAILAFEPVEIK
PRALSALLRSQIPLILFSQSGEYLGRIEPTYTTRPDLIKAQALMQDLMQMLVWSVLRRCRRCLLLYQRHQRYH
IR
>PCC9605_05463   Cas_Cas1
MSTIYITEQDVSFKIQHHYLKVFHQQKQCICIRISNISQIILFGNIPLPKEVVKTVISHQIPVLYLTQDSEYI
GRLENTSQRQAKYLIYQRRRARDVEFNRATAESIIWAKLHNQNTFLQSWTRHYTSPATQRALNYLMLLMDNLG
LATSLDELRQYSEEADNVYYCAIGSLLNFYNRCPQTTAKRIRRFLNLGNQLLHQYIYTLLNTAGLHPDYAILH
HDSDRELPLVWDFMAEFHAPIVDDLVLTFVRNFPNLNGNRKNQPRSLLQRFLQYWEGKLRTFVLHPYAGEVSY
RQCIDLQVREYLTCLLGDVEYYRPLALKFNPASEVKS
>PCC9605_05461   Cas_Cas1
MALRRRIPVMYLSQKGRYFGRLETEGQAKVEYLARQVECSRDDKFTRQQAEAIVWAKLHNSRNLLMRLNRRRN
SKEVDGTIVVEAIKNLEILMDKLFFAESMDALRGYEGKAATIYFQALGTLFTGPFKFEKRTKRPPTDPINSML
```

```
SLGYTLLSQNVYSFVQTIGLHTHFGNLHVPRDNHPALVSDLVEEWRAQLVDSLVAYLVNSQIFTIDDFTPPDE
RGGVYFQPHALKKFLKQWEEKLQSELTHSHTGQKVPYRRCIELQVREYIACLMGEVKVYRPMIWKM
>PCC7420_03004   Cas_Cas1
MSKPVNYPALVSDLVEEFRALLVDSLVAYLVNSKIFVLEDFTPPDERGGVYLYPDSLKKFLKHWEEKLQARIT
HPHTGYKVNNRRCLELQVWEYVACLMGEQQVYRPVRWNKW
>PCC7420_02949   Cas_Cas1
LINGVIDAVEEREGELIPLEYKKGRMARHLNDHFQLCGAALCLEERTPFVVSALALMALKCRLRTQYDVHRDD
ARRLELVKAIITTKIHNQHNVLYRYGQKNSPLKQRKLLVQEQTTLDQVWGRIGFRIRDENPF
>PCC7420_03260   Cas_Cas4+Cas_Cas1
MQTLDSTQDTIRVSALHALAYCPRLFYLEEVEELYTQDAAVFAGRRLHVELEKQEDEDWEDLFLESPELGLRG
RVDALRTRDGQTIPYEHKRGRCYRDENNQPQAWNSDKLQILAYAYLIESALGIPISEGRIRYHADNVLVHVPL
DDLGRATVRDAIQQARILRQSSHRPPVTDNERMCVRCSLSPVCLPEEARLAHNREWQPMRLFPPDDERQVIHV
LEPGTAVGRTGEQLKLSRRGQPVETVPVRQVGQLVLHSFSQISTQALHFCVTQGIGVHFISGGGRYIGSFDTR
QGSIQRRIHQYQALSQPDTCLELARKLVTCRGQGQRKFLMRGKRGSQGKSAKLKRAIAQMKAVLKNVSSAPSR
ESLLGLEGNLAASYFAALPCLIAATVPPELRFDHRNRRPPKDRFNALLSFGYALLIKDVMNAILTVGLEPALG
FYHQPRSQASPLALDLMEIFRVPLVDMVVMASINKGQWDVQADFDVRGVQVWLSDTGRRKFVGLYERRKQETW
KHPVTGYSLTYRRLLELEVRLLEKEWFGEGGLFGQLIVR
>PCC7420_00245   Cas_Cas1
MAAIYLIEQGTTIYKEYQRFIIYVSEKPKLEVPIREVQQILVFGNIQLSTPVMQVCLREQIAVVFLSQSGRYH
GHLWSSEFRDLDQELVQVRRWGDAAFQFQVSQAIVYGKLMNSKQLLLRFNRKRKLPDVERAIIGINQDIEALE
FSESLDRLRGYEGIGAARYFPALGQLITNSRFEFSLRNRQPPTDPVNSLLSFGYTLLFNNVLGFIIAEGLSPY
LGNFHYGERQKPYLAFDLMEEMRSVVVDSLVLNIVNHSLFKPQDFDTVPSTGGVYLNQSARRVFLKQFETRMN
EEVSHPDLQSKVTYRQAIQLQVRRYKQSLLSGVSYEAFLRAM
>D9_00392  Cas_Cas4+Cas_Cas1
MEVMDLESEGVKKTSPQDISKTIRVSALHALAYCRRLFYLEEVEELYTQDAAIFAGRRLHTELEKQEDEVWEE
LFLESEELGLRGRLDALRTRDGQTIPYEHKRGYCHRDENKQPQAWESDKLQILAYALLLEFALGITVKEGRIR
YHTDNILIHVPLDHSGRESVHQAIKEARNLRESAYRPPVTENERLCARCSLAPVCLPEEARLTHDKEWQPIRL
FPEDDERLIIHILEPGALVGKTGSQIKINRRNQPIENIPASQIGQLVLHSFSQISTQALYFCAHQDIGIHFIS
GGGRYVGSFDNRRGSIQRRIRQYTALSDPAFCLELARKLVLCRGQGQRKFLMRGTRGRTQVPENLHKSIAQIK
TMLKQVPQAKSLESLLGLEGNIAASYFSALPHLIVADVVDELYFNGRNRRPPKDRFNALLSFGYALVLKDVMN
AILTVGLEGALGFYHQPRSQAAPLALDLLEIFRVPLVDMTVMSSINRGQWHPKEDFEIRREQVWLSEPGKRKF
VELYERRKQETWKHPVTGYSLTYRRLFELEVRLLEKEWSGEPGLFGNLILR
>D9_02760  Cas_Cas1
MQTLYISEQNCYVCLQKETLIVKQGDTVHVEVQLPLLEQILIFGRSQITTQVIRACLWRDIPIAYLSRMGYCY
GRILPISRGYRQLSRYQQELSPVDRLITARAIVKGKLKIVEFY
>D9_02759  Cas_Cas1
MGFEGTGAAQYFSVFSECLTNPDFVFSGRSRRPPGNPVNAMLSFGYQVLWNHLLALIEVQGLDPYYACLHQAN
DGHAALASDLIEEFRAPLVDSLVMWLINRNIMDVKGDFQFKNGGCYLNDSGRKKFLRGFLQRMTEEIQTNDGK
QPKWDLLTQQVRNYKQFVYNPSHHYKPYQID
>PCC11201_04660  Cas_Cas1
MLTLYVSQQGCYLTLEAENLLIKKGKAVLAEAQIPLLEQVLVFGQSQITTQAIRACLSRNVPIAYLSRMGYCY
GRILPVERGYRHLSRYQQQLTRIDRLIIARCLVQAKLKNSRVILQRQARRQSSEVITFAIQSLNHLMEKVREA
ESWERLMGLEGAAAAQYFSAFGECLNHPEFVFTVRSRRPPGNPVNAMLSFGYQLLWNHLLSLIELQGLDPYFG
CLHQGTERHAALASDLIEEFRAPIVDSLVMYLINRNMIDANTDFTYHDSGCFLNDSGRIKYLKAFLQRMEEKL
TTAAGENQPRWDTLTQQVKAFKQFVYNPIEGYKPYEIR
>PCC11201_04582  Cas_Cas1
MTTLYLTEPGTVVRYQNQSFSVKYQDKTHSCRLGELTLVVILPGVQLTDEAIAQLLDRGIETLFLRRDGQFRG
RLQGQFQTNPMFRLAQYRIVDTTFGMALAQLLVMAKIRNQRVLLQQRHRSTKAQISELAEAIDAIAAYMIQLK
NTSTPLNRSELMGIEGICARTYYQALRHYFPPQWNFTGRNRRPPLDPINALLSWGYGVLLARVFAACVQAGLD
PYLGFFHATEPYRPNLVLDLMEEFRPVVVDQAVISIIKSDLLQREDFQPSPDGEGIWLGVTAKKLLLGQLERR
LREPFLYPAQNRRLSLSQILLEQARSLARCLVESRLDYQGFVLK
>PCC11201_04364  Cas_Cas1
MGTVYVTVDDAFIGKTDERLMVKADKKNILDVPLIKVEGIVILGRATVSPAVVQELLERHIPLTFLTGTGRYL
GRLEPEMTKNIFVRKAQWQAAGESEQALHLVRGFVRGKLKNYRTLLLRQGRKYSDLQLDDSVTRLDNAIAPIS
STNNINSLRGLEGAGSAAYFGAFDQLIRGDGFSFQTRNRRPPTDPVNALLSFGYSLLCHDVQSAVNLVGFDPY
LGYLHVQHYGRPSLALDLMEEFRPLVVDSMVLSAINLRQLTPDDFTTEAISNAVFLSDEGRRTFLKLYEQKKQ
SKFKHPVLGKQCTYQEAFELQVRLLAKYLMGDIEKYPPLVLG
>PCC11201_03307  Cas_Cas1
```

```
MATLYLIEQGTTIYKDNQRFVIWRPDQQKTEIPVRDVERILLFGNIQMSTQAINTCLDEKVLILFLTLSGQYK
GHIWSLESTHFGCETVQFSKWKDPRFQLLVSQGIVRGKLWNSKQLLVRLNRKRKIPEVAKAIAGITSDINALN
TVNNLDSLRGYEGISAARYFPAFGKLIANSAFHFSLRNRQPPTDSVNSLLSFGYTLLFNNVLSLIIAEGLSPY
MGNFHYGEDKKPYLAFDLMEEFRSPIVDSFVLKLINSSIVKPQDFDRVASTGGVYLNKQARRVFLEQFEKRMN
EEISHPDLQSSVSYRQAIQLQVRRYKSSLLQGSAYQAFLRSV
>WA102_04123     Cas_Cas1
MQSLYVSEQNCYVCLQKETLLVKQGETVYAEVQLPLLEQILIFGKSQLTTQVIRACLWRDIPIAYLSRMGYCY
GRILPISRGYRQLSRYQQQLSPVEKLITARAMIKGKLKNSRVLLRRQKKKKESELLERVLQSLDYLADQVAQA
DNLERLMGFEGAGAAQYFSAFSECLTNPDFVFSGRSRRPPGNPVNAMLSFGYQVLWNHLLALIEIQGLDPYYA
CLHQAHDGHAALASDLIEEFRAPFIDSLVMWLINRKIVSAEADFEFKNGGCYLNNAGRKKFLRAFLQRMSEEI
QTDEGIKQPKWDLLTKQVRAFKQFVYNPSHHYQPYRID
>WA102_03338     Cas_Cas1
MTTLYIIQPDAVLSKDYEAFQVALKQEDGSWKKQKVAAQTVDEVILMGNPQVTGDAFVYALELGMPVHYLSSF
GKYLGSALPKSSRNGQLRLAQYAVYHDPVRRLELVKAIVTAKIHNQYNVLYRHNEKDNPLKERKVLVKTQQNI
DQVRGVEGLAAKEYFACWQRMIGKQWTFNGRNRRPPTDPVNSLLSFAYALLQGQVMAGVHVAGLDPYIGYLHE
VHHGQPAMVLDLMEEFRALIADNLVLSLLHKHEIKPKDFNESLGAYRLKDNARKTFLKAFDQKMNDEFKHPVF
DYRCTYRRAIELQARLLSRHLQEGVPYKPLSLR
>WA102_03209     Cas_Cas1
MGTLYIVQDDAFIGKVDERINVKYEKKILQDVPFIHLEGVVVLGRATISPAVVDELMQRHIPLSFINKIGHYL
GRLEPEMTGNIFIRKAQWQAAGETPQSIHLVQGFVRGKLKNYRQTLMRYQREFNDIDLSKSIPRIEQVIAAIN
STESINSLRGLEGSGSAAYFGCFDSLIRNSQFSFTKRVRRPATDPVNSLLSFGYSLLRHDLQSAVNIVGFDPY
LGYLHFDRYNRPSLALDLMEEFRPLVVDAVVLSILNKQLLTPADFVTEPLSGAVSLTPEGRKTFLTLYEKKKI
SEFKHPVMGRKCSYREAFELQARLLAKYLMGTIDKYPPLVLK
>WA102_00742     Cas_Cas1
MRTLYVSRQGCYISLKGETLIIKQGEIVHGEVQLPLLEQILIFGKSQITTQAIRTCLWRDIPIAYLSRMGYCY
GRLMPIERGYRRLSRYQQELTAIDKLIAARTIIQAKLKNSRTFLQRQQRRNPSQTAEITIKSLEVLAQKAGEA
ETIERLMGLEGAGAAQYFNAFGECLNNPDFVFLARSRRPPGNPVNAMLSFGYQILWNHLLTLIELQGLDPYYA
CLHQGTERHAALTSDLIEEFRAPIIDSLVLWLVNTKIMNIDNDFEYHDSGCYLNNSGRPKYIKYFLQRLEEEV
QNSEGEKQPRWDLMTQQIKAFKDFVYQPSKLYKPYQIR
>CCY9414_03059   Cas_Cas1
MTTLYLTEPGTVIRYRNESLTITKQEKSQNCRIAEITLVVVLPGVQLTDVVIAQLLDKGIETIFLRQDGQFRG
RLQGQFATNPIIRLAQYRTVETTFGMAIAQKLVIGKIRNQRVLLQRRNRATKGEISQLAEAIDLISSYSSRLS
DFTTPLNRDELMGVEGICARTYYQGLKHWFPPQWNFSGRNRRPPLDPINALLSWGYGVLLARVFSACVQAGLD
PYLGFFHAIEPYRPNLVLDLMEEFRPIIVDQSVISIIQSDLLQSEDFQPSPDGEGIWLGTLAKKLFLGELERR
FHTPILYPAQNRRLSLSQIVLEQARSLGRCLVNSRLEYEAFVIK
>NIES3708_01940  Cas_Cas1
MGTIYITQDDAFIGKTDERLQVKADKKQLLDVPFIKIDGLVILGRANFSPAIVSELLTRKIPLSFLTSTGKYL
GRLEPELTKNIFIRKAQWAVTENSPKAIHVVNGFIRGKLKNYRNLMMLRNRQEYHEKLESSVEQIQQIIKSLD
TKNNINSLRGLEGAGSAIYFSCFNDFIKNNNFSFKTRNRRPPTDPVNSLLSLGYALLRHDIQSAVNIVGFDPY
LGYLHVAHYGRPALALDLMEEFRPIITDAMVLNVINNKKLTPQDFITEPISNAVSLTKEGLKIFLRSYEEKKQ
SKFKHPVLNKQCTYQEAFEIQARLLGKYLMSETEQYTPLITKK
>NIES3708_01244  Cas_Cas4+Cas_Cas1
MFATDLNYHQETEATLRVSALHAFAYCPRLFYLEEVEELYTQDASVFAGRRLHEELEEKEEGEWQDLYLEDEY
LGLRGRIDALKTIHGDIIPYEHKRGKCYRDEKKQPQPWESDRLQILAYCCLIEASLGIVVKEGRIRYHADNVL
IHLPVDDEARQWVKDTIQQARELRKSVYRPPVTNNEYLCSRCSLAPVCLPEEARLAHNKEWQPIRLFPQDDDR
DVIHILEAGTRVGRTGEQLKISRRNEPDEKVSIQQVGQVVLHSFSQISTQALHFLSDKEVGVHFISGGGRYIG
SVDTRNGSIQRRINQYQALTKPEFCLELGKKLVNCRGEGQRKLLMRGKRSLVSQCDELDYAIEQMKKVLRQIP
KIDSLSSLLGIEGNLASLYFGALSHIIGKNVPAELHFQTRNRRPPKDRFNALLSFGYSLLIKDVVNSILTVGL
EPAFGFYHQPRTQAPPLALDLMEIFRIPLVDMVVLASVNRLQWDIKEDFDIRGKQVWLSDVGRKKFIALYEKR
KTESWKHPVTGYSLTYRRLLELEVRLLEKEWCGETGLFGHLVLR
>UTEX2349_04040  Cas_Cas1
MGTVYISQNDAFIGKIDERLSVKFDKKTILDIPLLKIDGIVVLGRATVSPAVVNELLARQIPLTFLTETGRYL
GRLEPEVTKNIFVRKAQWQAAGESLQAIHVVQAFVRGKLKNYRNTLIRRQRDTDLDLSANISRIEHAIAPIET
TLTVNSLRGLEGAGSAAYFGCFNQLIRNGEFEFKTRVRRPPTDPVNSLLSFGYSLLRHDVQGAVNIVGFDPYL
GYLHCDRYGRPSLALDLMEEFRPLVVDAVVLLAINKRHLTPADFLTEPLSGAVSLTPEGRKTFLRLYAQKKQS
EFKHPVLGRKCMYQEAFELQARLLAKYLMGEVEKYPPLVLK
>UTEX2349_00166  Cas_Cas1
MAAIYLIEQGTSIYKDYQRFIIHVSEKPKLEVPIREVQQILVFGNIQLSTPVINSCLQDKIAVLFLSQSGQYH
GHLWSEESTNLDNQLIQIERRNDDYFQFSVSKAVVLGKLMNSKQLLMRFNRKRKLEDVEKAIYGINQDIDALE
```

```
YVDNLDSLRGYEGIAAARYFPAFGKLITNSAFSFSQRFRQPPIDEVNSLLSFGYTLLFNNVLSFIIAEGLSPY
IGNFHYGEKQKPYLAFDLMEEFRSPIVDSLVLKVINKSLFKPNDFDLVASTGGVYLSQSARKVFLKQFETRMN
EEVSHPDLISQVTYRHAIQLQVRRYKRCLLSGVPYESFLRAG
>UTEX2349_00401  RVT_1+Cas_Cas1
MRFPLELLNAAWFQVRAGTKSAGVDGISLDLFESVAAEQLRNIEYQLHHETYTASPAKGFYVPKKNGDKRLIG
IPTVRDRIVQRLLLEELYFPLEDTFLDCSYAYRPGRNIQQAVQHLYSYYQLQPKWVIKADIAEFFDNLCWALL
LTALEDLQLESIVLQLLEGQLKSGIVIAGKPVYPGKGVLQGGVLSGALANLYLTNFDRKCLSHGINLVRYGDD
FAIACTSFHEANRILDKITTWLGELYLQLQPEKTQIYAPDDEFIFLGYRFKKGEVYAPPPPAPVRRGEWVIND
SGTPYFRTKPRPTKPVSRPPKACSIDKPSKYPTAPISHLWQDFMTTLYVTDQGAYLSVKNQQFQVYYQGELRI
KVPATRVTNIVLFGCCNVSHGAVSYALRRRIPIMYLSQKGRYFGRLQTEGNAKVEYLMRQVICSQNPEFTRKQ
AENIVWAKLHNSRALLLKFNRRRPSKMASSAIELIADLMDKLPQAESMDALRGYEGKAATLYFQALGSLFTGV
FEFDRRTKRPPTDPINSLMSLGYTLLSQNVFSFVETVGLHTHFGNLHVPRDNHPALVSDLMEEFRAQLVDSLV
AYVVNSKILTLEDFTPPDEKGGVYLQPHALKKFLKHWEEKLQSELTHPHTGYQVSFRRCVELQVREYVACLMG
EVELYRPMIWKL
>UTEX2349_00397  Cas_Cas1
MPNIYITESDASFQMKHNYFQVFHQQKQCISLPIRHVNQIIIFGNVNLPNDVIKILRLHHIPVLYLTQDGEYL
GRLENPSQVQPKYMFHQRKRLRDIEFNRATAESMIWAKLHNQHIFLQSWTRYHANHATQRASNYLTLLMDNLA
IAKSLDELCEYNDEADNVYYTAINSLLSFYSIYPQTRVRQISGLFNLGYQLLHQYIYTLLHTVELHPDYAILH
RHGHHELPLAWDFTVEFRAPIVDDLVLNFVRNLPSSNGNGNGNGKTQPKMILQNFLQHWEAKLRTFVLHPHAG
EVSYRQCMDLQVREYLACLLGDVEYYRPLALKFHPNHPHFTNIAEPEKTPLTLVKR
>PCC6304_05820   Cas_Cas1
VGTVYITQDDCFIGKTDERLNIRVNKQKLLDVPLINIEGVVIMGRATVSPAVVTELLNRHIPLSFLTTTGRYL
GRLEPELTKNIFVRKAQWQAAGETEKALHLVRGFIRGKLKNYRTALLRKQRQGSDVNFQTATDRLKQILSSLT
STATIDSLRGLEGAGSAAYFQCFDALIRSEGFSFEARRRRPPTDPVNALLSLGYSLLSHDIQSAVNIVGFDPY
LGYLHVERYGRPSLALDLMEEFRPLVVDAMVLNAINNRIITPSDFTTEPISNAVSLTPEALRTFLRLYEQKKQ
SDFTHPVMGKKCTYQESFEIQGRLMAKYLMEEIDQYPPLVLK
>CS505_01270     Cas_Cas1
MQTLYVSEQNCYICLQKEILVVKRGEVTQAQVQLPLLEQILIFGSSQVTTQVVRACLLRDIPIAYLSRMGYCY
GRILPISRGYRQLSRYQQQLSAMEKLVTAREIIKSKIKNSRVLLQRQKKKIASANLERVLESLNYLAEQTASA
DSCERLMGIEGAGAAQYFSAFGECLSNENFIFLGRSRRPPGNQVNAMLSFGYQVLWNHLLALIEIQGLDPYYG
CLHQGHDGHPALASDLIEEFRSPLIDSLVMWLINRNIFDVGADFEFIKGGCYLNNSGRKKFLQAFLQRMTEEV
QTEDGTKQPKWDLLTQQVRAFKQFVYNPAHHYRPYRIH
>CS505_00652     Cas_Cas1
MTILFLLEQDTLVRQIDERLEIFKHDKRLTDVPLCKLESVVVYGRVILTIPALKILNQRSIPITYLSEEGRTV
ATLLPEPNPNAILRSKQYQASFDTHKTLAIAREIIRGKLFNQHTILARFSRQNERSEKVISALKSLKACQKSI
DQTTSLNELRGYEGQGASSYFGVLGELLTGTPWSFSHRTRRPPTDPVNALLGFGYALLYGDCRAALHTVGLDP
YQGFLHGERYGRANLALDLMEEFRPIFVDGLVLQLLRNNSLEKESFINYPGGAVHLNEQGMKKFIQSYEQRKN
TTFQHSVFEIQTDYRRALILQARLLAKHLNQEIEKYLPLKVR
>PCC7521_01324   Cas_Cas1
MGTVYITQEDAFIGKVDERLNVKYEKKTILDVPLIKIDGLVVMGRASISPATIAELVNQKIPLTFLTNNGKYI
ARLEPEMSKNIFLRSAQWKAKGESPQAIHASKGFVRGKLKNYRYSLLKSQRSYPDLDLNGGLNQLATAIANLD
QATSIDSIRGLEGAGSAAYFSCFNQLIRVDNFQFHTRNRRPPTDPVNSLLSLGYALLRHDVQGAINIVGFDPY
LGYLHTERYGRPSLALDLMEEFRPVIVDAVVLTAINRRILVPTDFITEPLSNAVSLTKEGLHKFLRLYQEKKQ
TKFKHPVMQKQYTYQESFELQARLLAKYLMGEIEKYPPLVMK
>PCC7521_01907   Cas_Cas1
MDKLPFAESMDALRGYEGRAATVYFQALGSLFSSVFKFEKRTKRPPTDPVNSLLSLGYTLLSQNVFSFIGT
>PCC7122_01534   Cas_Cas1
MGTLYVTQNDSFVGKVDERIHVKFEKKVLQDIPFIHLKDVVVLGRATVSPAVVFELMSRHIPLSFIDERGHYL
GRLEPEMTGNIFVRKAQWQAAGESLQAVHVVQGFIRGKLKNYRQVLMRCQRESDVDLSKYITRIEQTISPIDS
THNIKSLRGLEGAGSAAYFGCFNQLIRNPDFTFTKRVRRPATDPVNSLLSFGYSLLRHDVQGAVNIVGFDPYL
GYLHCDRYNRPSLPLDLMEEFRPLVVDAVVLSILNKQLLTPADFISEPLSGAVSLTPKGRKIFLTKYEKKKQS
EFKHPVMGRKCTYREAFEWQARLLAKYLMGETDKYPPLVLK
>PCC7122_06044   Cas_Cas1
MQTLYVSEQNCYVCLQKETLLVKQGETVYAEVQLPLLEQILIFGRSQITTQVIRACLWRDIPIAYLSRMGYCY
GRILPISRGYRQLSRYQQQLSPVERLITARAMIQGKLKNSRVLLRRQRKKRESEVLERVLQSLDYLADQTAQA
DTWERLMGFEGAGAAQYFSAFSECLTNPDFVFSGRSRRPPGNPVNAMLSFGYQVLWNHLLALIEIQGLDPYYA
CLHQANDGHAALASDLIEEFRAPFIDSLVMWLINRKIVNAEADFEFKNGGCYLNDEGRKKFLRAFLQRMSEEI
QTDEGIKQPKWDLLTQQVRAFKQFVYNPSHHYKPYRID
>PCC7122_05839   Cas_Cas1
```

```
MSTIYITEQDANLQIQHHYLKVFHQQKQCVSLRISNVSQIILFGNITLPKEIIKVVVAHQIPVLYITPFGEII
GRLENTSQRQYKYLTCQRQWIRNRELKRATAESMIWAKLHNQHTFLQSWTRHHANYITQRALNYLTLLIDNLP
IATFLNELREYSEEADKVYESAVTSIFSFYNVCSHINQKPINGFLNLGYQLLNQYIYTLLDSAGLHPNYAILS
RNSDHELPLAWDLAAEFRAPIVDDCVLNFVRNFSHTNGNGKSQSRTILQRFLQHWEAQLRTFVIHPYAGEVSY
RQCMDLQVREYIACLLGDAEYYRPLAFKHHPEEVKSQKIKVKREESEEVKSQKVKVLR
>PCC7122_05838   RVT_1+Cas_Cas1
MFTIEQLYFAWSQVRVGSRSAGVDGISVDLFAASVDEQLTILLRQLQQESYHPSPAKGFYLTKKTGGKRLVGI
PTVQDRIVQRLLLEELYFPLEETFVDCSYAYRPGRNIQQAVQQLFSYYQYHPTWIIKADIAQFFDNLCWALLL
TNLEALQLESRILQLLEQQLKAGIIIAGKHINFGKGVLQGGIISGALANLYLTIFDRKCLSNGINLVRYGDDF
AVACSSWKEANRILDKIIAWLGELYLTLQPEKTQIFAPNEELKFLGYRFANGKVYAPPPPEPKLEGEWLANAS
GTPYFRPKERPLKFVSRPPKACDISKPVSFPRAPISHLWQESMTTLYITDQGAYLSAKNQQFQVYYQQELRIK
IPVVRVTNIVLFGCCNVSHGAVSVALHRRIPVMYLSQKGRYFGRLQTEGMAKVEYLARQVECSQNFEFTKKQA
EAIVRAKLHNSRIMLMRLNRRVKSEKATQAIAQIELFMDKLPFSEDVNMLRGYEGIAATTYFQALGSLVTGKF
TFEKRTKRPPTDPINSMLSLGYTLLSQNVHSLIQSVGLHTHFGNLHVPRDNHPALVSDLMEEFRALVVDSLVI
YLVNRNLFTIDDFTSPDERNGVYFQPHALKKFLKHWEEKLQTEVTHPHTDCKLPYRRCIELQVREYIACLMGE
VEVYRPMIWDK
>PCC7122_05748   Cas_Cas1
MTTLYVIQPDAVLSKDYEAFQVALRQEDGSWKKRKVAAQTVNEVILMGNPQVTGDAFVYALELGMPVHYLSSF
GKYLGSALPKSSRNGQLRLAQYAVYHDSVRRLELVKAIVSAKIHNQYNVLYRHNEKDNPLKERKNLVKTQQTL
DQVRGVEGLAAREYFACWQRMVGKQWTFNGRNRRPSTDPVNSLLSFAYALLQGQVMAGVHIAGLDPYIGYLHE
VHHGQPAMVLDLMEEFRALIADNLVLSLLHKHEIKPEDFHESLGAYRLKDNARKTFLQAFDKKMNDEFKHPVF
DYRCTYRRAIELQARLLSRHLQEGIPYKPLSLR
>NIES3754_04306  RVT_1+Cas_Cas1
MFTLEQITSAWLLVRAGTKAAGVDDISLDLFASAINDQLPILLRQLQQESYRASPAKGFYLSKSSGGKRLVGI
PTVRDRIVQRLLLEELYFPLEDTFLDCSYAYRPGRNIQQAVQHLYSYYQFQPKSIIKADIAEFFDSLCWALLL
TALEELSLEPIVVQLLEGQLKSGIVIAGKLIYPGQGVLQGGILSGALANLYLTAFDKKCLSHGINLVRYGDDF
AIACNSWLEANRTLDKVTTWLGELYLHLQPDKTQIFAPGDEFTFLGYRFAGGKVYAPPSPVVTREGEWVTNGS
GLPYFRPKQRPTKSVSRPPKACSIDKPIKFPTAPISHLWQDFMSTLYVTDQGAYLSVKNQQFQVFYQGELRIK
IPVSRVSNVVLFGCCNVSHGAASLALRRRIPIMYLSQKGRYFGRLQVEGDAKVEYLMRQVECCQNPEFTRRQA
EAIVTAKIHNSRILLMRLKRRKSSEFDDNLVKEAIKDLDTLMDKLPFAESMDALRGYEGRAATVYFQALGSLF
SSVFKFEKRTKRPPTDPVNSLLSLGYTLLSQNVFSFIQAIGLHGHFGNLHVPRDNHPALFS
>NIES3754_04305  Cas_Cas1
MEEFRAQLVDSLVAYLINSNIFTEDDFTPPDEKGGVYLQPHALKKFLKHWEEKLQSELTHPNTGYKVSFRRCL
ELQVREYVACLMGEVKIYRPMIWKP
>NIES3754_04304  Cas_Cas1
MSTIYITEQDAMLKIQHHYLKVFHQQKQCICIRISNVSQIIIFGNIPLPREVIKTVLAHQIPVLYLTQDSEYL
GRLENTSQQPAKYLIHQRRRARDVEFNRATAESIIWAKLHNQNVFLQSWTRHYTSDAIQRASNYLMLLMDNLA
LATSIDELRQYSEEADNVYYCAIASLLNFYNGCPQTTAKRIRRFLNLGNQLLHQYIYTLVNTVGLHSDYAILH
HDADYELPLVWDFMAEFHAPIVDDLVLNFVRHLPHSNGNGKSQPRTLLQCFLQHWEAKLRTFVLHPYAGEVSY
RQCIDLQVREYLACLLGDVEYYRPLALKFNPARSEFADVIAPQQVPLTLVKR
>NIES3754_01068  Cas_Cas1
MGTVYITQEDAFIGKVDERLNVKYEKKTILDVPLIKIDGLVVMGRASISPAIIAELVNQKIPLTFLTNNGKYI
ARLEPEMSKNIFLRSAQWKARGESPQAIHASKGFVRGKLKNYRQSLLQAQRRYTELNLSAGISQLGSAIANLD
QATSIDSIRGLEGAGSAAYFSCFNQLIRVDNFQFHTRNRRPPTDPVNSLLSLGYALLRHDVQGAINIVGFDPY
LGYLHTERYGRPSLALDLMEEFRPIIVDAVVLTAINRRILVPTDFITEPLSNAVSLTKEGLHKFLRLYQEKKQ
TKFKHPVMQKQYTYQESFELQARLLAKYLMGETEKYPPLIMK
>PCC6407_02052   Cas_Cas1
MGTLYINQDDVFIGKTDERLNVKLKNQTLQDVPLIHLDGVVIMGRATISPAAIDELMQRKIPLTFLSDTGHFL
GKLEPELSKNIFVRNAQWKAAGETPQAIHVVQGFIRGKLKNYRNSLTQTKRDHSELNLESAIDRISDIIVKID
KTHNIDSLRGLEGAGSAAYFGVFNQLIRAEGFVFKPRHRPATDPVNSLMNFAYTLLRHDVQSAINIVGFDPYL
GYLHVQRYGRVGLAFDLMEEFRPLVADNMIFNAINRRIITINDFTVEPISNVVRLSVEGRKEFLKAYEKKKQS
EFKHPVLGKKCTYQEAFEIQARLLAKYLMGEIDQYPPLVLR
>PCC6407_05742   RVT_1+Cas_Cas1
MEFTPEPLHAAWLQVRKGSRSAGIDNITVDLFAGVARYQLQVLLWQLQQENYFPRPAKGFYLRKASGGKRLIG
IPTVRDRIVQRFLLDELYWPLEDVFLDCSYAYRPGRGIQMAVKHLYSYYQFGQAWVIKADIEKFFDNLCWPLL
LTDLEKLQFEPTLRQLIEQHLASGIVVKGQHFHPNQGVLQGGILSGALANLYLNEFDRLCLSHGFNLVRFGDD
FAVACADSIQANRCLEQINSWLGSFYLKLQPEKTRIFAPDEEFTFLGYLFRNGQVFAPERTQPTTKVVRASGT
SFSRSVTRKVPVFSGKPLACSISVKPIVFPRANSEHFWREPMSTLYVTDQGSYLSIYNQQFQVFYQRELRIKV
PASRVSHIILFGCCNLSHGAVSLALHGRIPIMYLSQRGRYFGRLQAEGMAKVEYLERQVIGSQSSKFVRQQAE
```

```
VIVWTKLHNSRALLMKLNRRRPSKIAAIAIEAILQLMEDLHQAASMDALRGYEGQAASLYFQGLGSLFTGSFV
FEKRTKRPPTDPINSLLSLGYTLLSQNVHSFVQAVGLHTHFGNLHTPRDYHPALVSDLMEEFRAQVVDSLVAY
LVNSKIFVLEDFTPPDERGGVYLQPYALKKYLKHWEEKLLSEVTHPHTGYQVSLRRCMELQVREYIACLMGEV
EVYRPMLWKL
>HK06_02992 Cas_Cas1
MGTLYITQEDSFIGKIDERLTVKFDKKILQDIPLIKIDGIVILGRATISPATVSELLERHIPLTFLTENGRYL
GRLEPEVTKNIFVRKAQWAAAGESPQAIHIVKGFVRGKLKNYRNVMFRRQREANLDLSKSIERIEQVITSIDT
THNINSLRGLEGAGSAVYFGCLNQMIRNVEFQFDYRNRRPPKDPVNSLLSFGYSLLRHDIQGAVNIAGFDPYL
GYLHFERYGRASLAFDLMEEFRPLIVDALVLSTINKSLLTPSDFFTEPLSGAVYMTPEGRKAYLKLYTQKKQS
EFKHPVMGRKCTYQEAFEIQARLLAKYLMGEIDKYPPLILK
>KM1D3_04105      Cas_Cas1
MAGVHVAGLDPYIGYLHEVHHGQPAMVLDLMEEFRALIADNLVLSLLHKHEIKPEDFNESLGAYRLKDNAKKT
FLQAFDKKMNDEFKHPLFDYRCTYRRAIELQARLLSRHLQEGIPYKPLSLR
>KM1D3_02980      Cas_Cas1
MRTLYVSRQGCYISLKAETLIVKQGEIVHGEVQLPLLEQILIFGKSQITTQAIRTCLWRDIPIAYLSRMGYCY
GRLMPIERGYRRLSRYQQELTAIDKLIAARTIIQAKLKNSRTFLQRQQRRNPSQTAEITIKSLEVLAQKAGEA
ETIERLMGLEGAGAAQYFNAFGECLNNPDFVFLARSRRPPGNPVNAMLSFGYQILWNHLLTLIELQGLDPYYA
CLHQGTERHAALTSDLIEEFRAPIIDSLVLWLVNTKIMNIDNDFEYHDSGCYLNNSGRPKYIKYFLQRLEEEV
QNSEGEKQPRWDLMTQQIKAFKDFVYQPSKLYKPYQIR
>KM1D3_05067      Cas_Cas1
MQTLYVSEQNCYVCLQKETLLVKQGETVYAEVQLPLLEQILIFGKSQLTTQVIRACLWRDIPIAYLSRMGYCY
GRILPISRGYRQLSRYQQQLSPVEKLITARAMIKGKLKNSRVLLRRQKKKKESELLERVLQSLDYLADQVAQA
DNLERLMGFEGAGAAQYFSAFSECLTNPDFVFSGRSRRPPGNPVNAMLSFGYQVLWNHLLALIEIQGLDPYYA
CLHQAHDGHAALASDLIEEFRAPFIDSLVMWLINRKIVSAESDFEFKNGGCYLNDAGRKKFLRAFLQRMSEEI
QTDEGIKQPKWDLLTKQVRAFKQFVYNPSHHYQPYRID
>KM1D3_04654      Cas_Cas1
MGTLYISQADSFIGKVDERINVKFDKKIIMDVPFIKLDGVVVLGRSTLSPAVVYELINRHIPLSFIDEMGHYL
GQLQPEMSGNIFIRKAQWQAAGETPQSIHLVQGFVRGKLKNYRQTLMRYQREFNDIDLSESIPRIEQVIAAIN
STESINSLRGLEGSGSAAYFGCFDSLIRNSQFSFTKRVRRPPTDPVNSLLSFGYSLLRHDLQSAVNIVGFDPY
LGYLHFDRYNRPSLALDLMEEFRPLVVDAVVLSILNKQLLTPADFVTEPLSGAVSLTPEGRKTFFNFVRKEKN
IGV
>KM1D3_05752      RVT_1+Cas_Cas1
VNFWLCTIFNALGKKLQKKRGCAGIDRETIDDFARNQALNIYELRNTVANGSYEPLPCKQVVIPKRNGNVREL
KIPTVRDRIVQQALLNVFYPVLEVKFSSASFAYRPNLSYLNAVEKIADLRDEGYVWVFDADIVKFFDNIEHSR
LLQQIRLHLDHPGILCLIKSWVSVGVQTDAGLILSQKGIPQGAVISPILANIYLHEFDEIITDSNLQLVRYAD
DFVVLARTEAEILAAKTEIINLLAGMGLEIHPEKTRTTNFEHGFCFLGHGFIDNAIFPIDTNRTKSTNNSVPQ
AQSKQRKKKTPPRQLAPQIEGLQLDIETNQELEILEQELEILEQEQQHQVAFLHQPYLGETDITTDREIISDD
GDDKNSPLQKNIWNQEMAAIYLLEQGTSIYKDHLRFIIHVSEKPKLEIPIREVEQIMIFGNIQLSTPVINACL
KEKIGVLFLNQLGQYHGHLASEESTNIDNHLTQMRRRDDNYFQFQVSRAIVYGKLTNSKQLLMRFNRKRKIAA
VDQAIYGISKDIESLELVDNLDSLRGYEGITAARYFPAFGQLITNPNFQFSLRNRQPPTDPINSLLSFGYTLL
LNNVMSFIVAEGLSPYIGNFHYGERQKSYLAFDLMEEFRSVIVDSLALKIINNSVFKPQDFDILDSTGGVYLN
QSARRIFLQKFEARMNEETSHPDLQSQVTYRHAIQLQVRRYKRCLLSGIPYTAFLRPA
>BCCUSP55_01905  Cas_Cas1
MGTLYISQADSFIGKVDERIHVKFEKKTIIDVAFIKLDGVVVLGRSTVSPAVVSELMERHIPLSFIDERGHYL
GQLQPEMSGNIFVRKAQWQAAGNTTQSVHVVQGFVRGKLKNYRQLLVRRQREADNLDLSADVTRLENVIAAID
KTESIDSLRGLEGAGSAFYFGCFNQLIRNSNFEFTKRVRRPPTDPVNSLLSFGYSLLRHDIQAAVNLVGFDPY
LGYLHFDRYNRPSLALDLMEEFRPLVVDSLVLTLLNKQFLKPEDFISEPLSNAVSLTPEGRKIFLTNYEKKKQ
SEFKHPVMGKKCTYRESFELQARLLAKYLMGITEKYPPLFIR
>CYA407_02530    Cas_Cas1
MSILYITQPQAILGKKYEAFTVATQEESGKWNKQSVPAQTVEQVVLMGNPQVTGDALTYALELGMPVHYLTSF
GKYLGSALPGYSRNGALRLAQFELYQDTERRLELVKTIVSAKIQNQATVLYRQKIADSPLKNHKKLVFKKATL
DEVRGVEGIAAKDYFREWSAILNEPWFFPGRNRRPPTDPVNALLSFAYGLLRVQVTASVHLVGLDPYIGYLHE
VSRGQPAMVLDLMEEFRPLVADSLVFSVLNKREIQPDDFTESLGAYRLKDEGRKRFLQAFERKMNDEFKHPVF
EYRCTYRRAIELQARLLGRYLQEDLVYKPLVLR
>CYA407_00526    Cas_Cas1
MLTLYVSQQGCYLTLEAESLLIKKGKEVLAEAQIPLLEQVLVFGQSQITTQAIRACLSRNVPIAYLSRMGYCY
GRILPVERGYRHLSRYQQQLTRVDRLIVARRIVQAKLKNSRVILQRQARRQSSEVITFAIQSLNHLMEKVIQA
ESWERLMGLEGAAAAQYFSAFGECLTNPKFVFTVRSRRPPGNPVNAMLSFGYQVLWNHLLSLIELQGLDPYFG
```

```
CLHQGTERHAALASDLIEEFRAPIVDSLVMYLINRNMIDANTDFTYHNGGCFLNDSGRIKYLKAFLQRMEEKL
TTAAGENQPRWDTLTQQVKAFKQFVYNPIEGYKPYEIR
>JSC11_03052     Cas_Cas1
MDKLPFAESMDALRGYEGRAATVYFQALGSLFSSVFKFEKRTKRPPTDPVNSLLSLGYTLLSQNVFSFIGT
>JSC11_00912     Cas_Cas1
LEKTLQNLDYLAAQAGEADNWERLLGFEGAGAAQYFSAFGKCITNPNFVFTGRSRRPPGNPVNAMLSFGYQVL
WNHLLALIELQGLDPYYACLHQAHDGHAALASDLIEEFRAPIVDSLVLWLINRKIIDAQQDFEYHNGGCYLND
SGRRKYLKAFVQRMAEEIQTDAGEKQPKWDLLNQQIKAFKQFVYNPSHHYQPYRID
>JSC11_00911     Cas_Cas1
MQTLYVSQQNCYVALQQEKLIIKQGETIYAQVQLPLLEQILIFGNSQVSTQAIRACLWQNVPIAYLSRMGYCY
GRLLPLERGYRQLSRYQQQLSTVERLITARAIAQTKIKNSRVLLRRQKKRLES
>JSC11_01201     Cas_Cas1
MGTVYITQEDAFIGKVDERLNVKYEKKTILDVPLIKIDGLVVMGRASISPATIAELVNQKIPLTFLTNNGKYI
ARLEPEMSKNIFLRSAQWKAKGESPQAIHASKGFVRGKLKNYRYSLLKSQRSYPDLDLNGGLNQLATAIANLD
QATSIDSIRGLEGAGSAAYFSCFNQLIRVDNFQFHTRNRRPPTDPVNSLLSLGYALLRHDVQGAINIVGFDPY
LGYLHTERYGRPSLALDLMEEFRPVIVDAVVLTAINRRILVPTDFITEPLSNAVSLTKEGLHKFLRLYQEKKQ
TKFKHPVMQKQYTYQESFELQARLLAKYLMGEIEKYPPLVMK
>PCC7375_04356   Cas_Cas1
MGTVYISTDDAFIGKTDERLKVRAQKETLLDVPLIKVDGVVVLGRATVSPAAIMELLERKIPLSFMTGTGRFL
GRLEPELNKNIFVRSAQWQAAGETPRSVHVVQGFIRGKLKNYRTMLLRAQREGTTSLQSGITRLEQAIKPIHQ
TSIINALRGLEGAGSAAYFGLFGRLLRNEAFTFKTRNRRPPTDPVNALLSFGYALLLHDIQGAINIVGFDPYL
GYLHTQRYGRPSLALDLMEEFRPLVVDAMVLSAINRKALILDDFSDEALSHAVSLTDEGRRKFLRLYEQKKQS
KFKHPVMGRQCTYQEAFEIQARLLAKYLMAETDQYPPLILK
>PCC7375_03275   RVT_1+Cas_Cas1
LRVPTVRDRIVQQALLNVLHPVLEPQFEPVSFAYRPGRSHKLAVEKVSAWHRRGYDWLLDGDIVSYFDQVEHS
RLLSEVDERLGASDFETLALRLIEQWNTVGTLTSAGLVLPERGIPQGSVVSPILANVYLDDFDEALQASRFKL
VRFADDFVVMGRSQRQAEQAQAKVAELLTTMGLQLHPDKTQITNFDRGFRFLGHAFVGDVVVPVKRSRSKAPR
AKVDQSPLRIFHADALAKPSQMQQALLASLKKSQQPIPPPLFVVLGYDVRREQRVEIKSSEVIWNKGMSTLYL
VQQGTTVHKEQGRFLIRAPQTKTVEVPIREVEQILIFGNCQLTTQVISVCLDQRITVVYLTQMGDYKGHLWSA
EAEGMNAIMAQFERHQDDAFCLETARVIVARKLQNSKQLLLRLNRKRKLEAVKISIEGLYTDLDAVAMATEIN
MLRGHEGAAAARYFKALGKLINNPGFVFNGRNRRPPTDPTNSLLSFGYTLLHNHVLSLILAEGLNPYLGNLHG
SERKQTFLAFDLVEEFRSVIVDSLVMRLINQKAVRPTDFSWPNKQGGVYLQGTARRLFLKRFEERLSLPVSHP
DVSDTVSYRRILQLQVRRYKQAVLNGVIYEAYRKIN
>MDT13_03258     Cas_Cas1
MQTLYVSEQNCYVCLQKETLLVKQGDTVYIEVQLPLLEQILIFGQSQITTQVIRACLWRDIPIAYLSRMGYCY
GRILPISRGYRQLSRYQQQLSPIDRLITAREIVKGKLKNSRVLLRRQQKKRDLELIERVLDSLDYLATQAATA
DNLERLRGFEGAGAAQYFSAFGECLTNSDFVFSGRSRRPPGNPVNAMLSFGYQVLWNHLLALIEIQGLDPYYA
CLHQGNDGHAALASDLIEEFRAPLVDSLVMWLINRKIMDVKDDFEFKNGGCYLNNLGRKKFLRGFLQRMTEEI
KTDQEIKQPKWDLLTQQVKAFKQFVYNPSHHYKSYRID
>PCC9339_00676   Cas_Cas1
LSQFPEQQSQVAFLHQPYLVETDSKTDGEIISDDHDNNSPLQKNIWNQEMAAIYLLEQGTSIYKDHLRFIIHV
SEKPKLEIPIREVEQIMIFGNIQLSTPVINACLKEKIAVLFLNQLGQYHGHLASEESTNIDNHLIQMQRRNDD
YFQFQVSKAIVYGKLTNSKQLLMRFNRKRKITAVEQAIYGISKDIDSLELVDDLDSLRGYEGITAARYFPAFG
QLITNSNFEFSLRNRQPPTDPINSLLSFGYTLLLNNVMSFIVAEGLSPYIGNFHYGERQKSYLAFDLMEEFRS
VIVDSLVLKIINNLVFKPQDFDILNSTGGVYLNQSSRRIFLQKFEARMNEETSHPDLQSQVTYRHAIQLQVRR
YKRCLLSNIPYAAFLRPA
>PCC9339_04465   Cas_Cas1
MTTTLYLTEPGTMVRYRNESLVIIRKEQSKNCRLAEVDLLVVLPGVQFTDVAISKLLDKSIETIFLRQDGHFR
GRLQGNYATNPAIRISQYRILESAFGMALAQKLVMGKIRNQRVLLQRRNRASQGQIAELAEAINLINTYSYNL
HNNSTLLNRSELMGVEGICARTYYQALRYWFPPQWNFTGRNRRPPLDPINALLSWGYGVLLARVFSACVQVGL
DPYLGFFHATEPYRPNLVLDAMEEFRPVIVDQAVISIIQSDVLTPEDFHTSPDGVGVWAGDMAKKLFLAELEK
QFRNTVLYPPQNRHLSMSQIILEQARSLGRCLVELSLEYEAFVIK
>PCC9339_04324   Cas_Cas1
MGTVYITQEDAFIGKVDERLNVKHEKKTILDVPLIKIDGLVVMGRASISPATVAELVHHKIPLTFLTNNGKYI
ARLEPEMSKNIFLRSAQWKAVGESTQAIHASKGFVRGKLKNYRYSLLKSQRSYSDIDLNAGLNQLATAIANLE
QATSINSIRGLEGAGSAAYFGCLNQLIRVDNFQFHTRNRRPPTDPVNSLLSLGYALLRHDVQSAINIVGFDPY
LGYLHTERYGRPSLALDLMEEFRPLIVDPVVLAAINRRILSPTDFMTEPVSNAVSLTKEGLHKFLRLYQEKKQ
IKFKHPVMQKQYTYQESFELQARLLAKYLMGEIEKYPPLVMK
>USR001_03037    Cas_Cas1
```

```
MTVLYVTQPDAVLSKKYEAFVVAMKQPDDTWKNYSVPAQTVEQIVLMGNPQVTGDALTYSLELGIPVHYLSIF
GKYLGCALPGYSRNGALRLAQFELYKDTERRLQLVKAIVTAKIQNQFTVLYRQKISDSPLKSHKKLVAQKQTL
DEVRGVEGIAAKDYFADWSAILKEPWFFPGRNRRPPTDPVNSLLSFAYGLLRVQVTSAVHIVGLDPYIGYLHE
VSRGQPSMVLDLMEEFRPLVGDSLVFSVLNNRQIKPQDFTESLGAYRLADDARKRFLQAFEDKMNDEFTHPVF
DYRCTYRRAIELQARLLGRYLQEGVNYKPLVLR
>USR001_02445    Cas_Cas1
MGTLYINQDDVFIGKTDERLNVKHKNQTLQDVPLIHLDGVVIMGRATISPAAIDELMQRKIPLTFLSDTGHFL
GKLEPELSKNIFVRNAQWKAAGETPQAIHVVQGFIRGKLKNYRNSLTQTKRDHSELNLESAITQISNIIPKLD
KTHNIDSLRGLEGAGSAAYFGVFNQLIRAEGFIFKPRHRPATDPVNSLMNFAYTLLRHDVQSAINIVGFDPYL
GYLHVQRYGRVGLAFDLMEEFRPLVADNMILNAINRRIIIVNDFTVEPISNVVRLSLEGRKEFLKLYQKKKQS
EFKHPVLGRKCTYQEAFEIQARLLAKYLMGEIEQYPPLVLR
>USR001_05111    RVT_1+Cas_Cas1
MEFTPEPLHAAWLQVRKGSRSAGIDNITVDLFAGVARYQLQVLLWQLQQENYFPRPAKGFYLRKASGGQRLIG
IPTVRDRIVQRFLLDELYWPLEDVFLDCSYAYRPGRGIQMAVKHLYSYYQLGQAWVIKADIEKFFDNLCWPLL
LTDLEKLQFEPILRQLIEQHLASGIVVKGQHFHPNQGVLQGGILSGALANLYLNEFDRLCLSHGFNLVRFGDD
FAIACADSIQANRCLEQISSWLGSFYLKLQPEKTRIFAPDEEFTFLGYLFRNGEVFAPEKAQPTTQVIRASGT
SFSRPATRKVPVFSGKPLACSIDVKPIVLPRANSEHFWREPMSTLYVTNQGSYLSIYNQQFQVFYQRELEIKV
PASRVSHIILFGCCNLSHGAVSCALHRRIPILYLSQRGRYFGRLQSDGRAKVEYLTRQIICSLNPEFVRLQAE
AIVRAKLHNSRILLMRLNRYRKSKNADEMSVLQAVEMLEILMDSLHKADSMDALRGYEGKAATVYFQALGSLF
SGPFAFEIRTKRPPTDPINSLLSLGYTLLSQNVSSFVEAMGLHTHFGNLHVPRDNHPALVSDLMEEFRAQVVD
SLVAYLINSQIFIADDFTPPDERGGVFLQPHALKKFLKHWDEKLLSEVTHPHTGYKVAYRRCLELQVREYVSA
LMGEVEVYRPMMWKI
>USR001_05109    Cas_Cas1
MTAVYITEPGSILSVKNHHFQVMHKEKLCESIPVKQVTQIILFCLCQLSRLAVNLASSRQIPIYFISQQGQEL
GNCEPQSKRPAKYRVRQLKRLKDPEFVRAFAESITWAKMHNQSVLLQQLRVSNSAFEPEVYRAFFNLLMDDLP
MAYSLDELRGYDVTSDTFYYPALGSLLPWNLGFRWRDLQQPTDPINRMLNLGNTLLHQQTSAFVQVLGLDADI
SHLHVPSQHQKPLVCDLMAEFRPMLVDELAVNLAKSRIITARDIVPKNGSGALLSPRALKTFIQHWENKLQTF
VSHPQAGMVSYRHCLELQVREYIACVLGDVQHYRPLVWSLPKNDVAANTNELEAEYPILAKM
>VB511288_08254  Cas_Cas1
MPNIYITEQDALLKTQNQNFQVFHQQAQCISVPIRHVNQIIIFGNLHLPRDVMKTMRLHHIPVLYLTQQGEYL
GRLENTSQLQPKYLAYQRKRSRDVEFTRATAESIIWAKLHNQHLFLQTWTRYQKHHASQMALNYLSLLMDDLP
MAKSLNELREYDEEANNLYYPAFSNLLNFHHGCSRTTKNLINTLLNLGYQLLHQHIYTLLHTSKLHPDCAIFH
HQSHHEFPLACDLMAEFRAPIVDDLVMNFVKNRFLADANGTTSNNGNGINSWRTLLKTFLQHWEAKLRTVVLH
PSAGEVTYRQCMELQIREYLACVLGEIEYYRPLALKFHPVHPDFTNTVEPQKVPLGLVKQ
>VB511288_08203  RVT_1+Cas_Cas1
MFTIEQITSAWLLVRAGSRGAGVDGMTVDLFAAGVNEQLRILLRQLQQESYRASPAKGFFVAKKSGGKRLIGI
PTVRDRIVQRLLLEELYFPLEDTFLDCSYAYRPGRNIQQAVQHLYSYYQYQPKWIIKADIAEFFDNLCWALLF
TALEDLQLEPILLQLLEQQLKSGIVIAGKPIYPGKGVLQGGVLSGALANLYLTSFERKCLSYGINLVRYGDDF
AIACSSWLEANRILDKITTWLGELYLNLQPEKTQIFAPDDEFTFLGYRFAGGEVYAPPPPVMTREGEWVTNES
GLPYFRPKSRPVKFVSRPPKACSIASPIKFPTAPISHLWQEFMTTLYVTDQGAYLSVKNQQFQVFYQGELKIK
VPATRVNNIVMFGCCNVSHGAVSMALRRRIPIMYLSQKGRYFGHTAVQGDARVEYLMQQVKCCENTQFTRQQA
EAIVAAKLHNSRILLMRLNRRRPTEIATQAIDLIEILIDSLPQAESLDALRGYEGKAATVYFQALGSLFTGFF
AFDKRTKRPPTDPINSLMSLGYTLLSQQVFSFIQSVGLHTHFGNLHTPRDNHPALVSDLMEEFRAQIVDSFVA
YLVNKKILTPEDFTPPDERGGVYLQASALKKYLKHWEEKLQTETTHPHTGYKVAYRRCIELQVREYIACLVGE
VEVYRPMVWKL
>VB511288_05518  Cas_Cas1
MTTLYLTEPGTIVRYRNESLVIIKQKQSQNSRLAEVSLVIVLPGVQLTDVAIAQLLDRGIETIFLRQDGQFRG
RLQGHFSTNPTIRLAQYRLVETTFGMALAQKLVMGKVRNQRVLLQRRNRATQGQIAALTEAIDLIGNYSSYLA
NITTPLNRNELMGIEGICARTYYQGLRHWFPIEWNFSGRNRRPPLDPINALMSWGYGVLSARVFSACLQAGLD
PYLGFFHATEPYRPNLVLDLMEEFRPVVVDQTVISMIQSQMLSPEDFQSSPDGVGVWLGSLAKKLFLGELEKQ
FRNPVLYPRQNRRLSLSQIILEQARSLGRCLMELSLDYEAFIIK
>YZ_03330  Cas_Cas1
MITKRTVIIQQSKGSTTCRLAEVDLVVILPGVQLTDSTIAQLLDRGIETIFLRQDGRFRGRLQGHFHTNPTVR
LAQYRALDTTFSMAIAQKLILGKVRNQRYLLQRRNRETQGKITQLTEAIDTIASYLPRLKQTETPLDRNELMG
VEGVCARIYFQGLTHWFPPEWKFTGRNRRPPKDPVNALLSWGYGVLLARVFAATVQAGFDPYLGFFHANAPFR
PNLVLDLMEEFRPVVVDRVAIALINSGLIDPNDFEPAPGEEEGIWLGQTAKKLYLGQLERQLQDSILYPPQHR
KLKLSQILLEQGRSLGRCLVETRLDYEAFIFR
>YZ_02676  Cas_Cas1
```

```
MTAIYLTEPQSYLEADLRDFAIVYNDRIQSRHSIDTTQQIIAFEGCTLGRSAIAAIYRYRIRLSFIGDNGHIT
AQITPQETPFYTVLQQERNSDRNFRIAFIHSLLPGRWYNATKVLQSLGYSGIGGCLEQQLEVLPGLNSWLALQ
TWQVKVFQRYQKALRSVLVSVKDFEPTLELAYALLSGELYVLLRSSGCCAEVGTLHLHCQNHLPLPCDLMEPF
RPVVEAWVFKKAPKFSDRGGLSQSYRATFIQNWESLMATQIFHAYAGQVSLRSALAWQVEEYVRAITETLEYR
PFLLKS
>YZ_02674   RVT_1+Cas_Cas1
VRDRIVQRLLLRSLYPILEETFQDCSFAYRPGVGVKHAIERVAEVYSSQTWTIKADISQFFDSLCRTLLLSQL
EELSVDQTVVRYIKGQLEAGIVVGGMPILSGRGVLQGGILSGALANLYLSEFDRRCLDAGAYLTRYGDDFVIV
ARSLLEATRFLNLIEDWLSDIYLTLQPEKTHIFAPGEEFVFLGYGFQAGEIITPEYKQPRRNHNKKAVSVASR
PPFACSIVTPTRKSLSTRNFLDSWREGMTTLYVTDQGAYVKVKHQQFQVLLGNDLKVSIPVNVVDYIILFGCC
NLSHGAIGLALRRRIPILFLSDQGRYFGRLQTDGMTRVDYLSRQVHCAEDETFVLRQAKVIVAGKLHNCRILL
RRLNRDRQISQVIEAIEELGVWQEKIAEVELLESLLGYEGFGTRIYFQALRALVQPPFTFEHRTRRPPTDPVN
SLLSLGYTLLHQNIHSLILAVGLHPHYGNLHVPRSNHPALVSDLIEEFRAPVVDSLVIYLVNSGIFTPEDFTP
SDERGGVYIYSDALKKYLKHWHDKLSLKTTHPHTGYKVSYYRCLELQVWEYISCLMGEREVYRPMLTKM
>NIES1031_00870  Cas_Cas4+Cas_Cas1
MQSIDLAQDTLTEAFPNTLRVCALHAFAYCPRLFYLEEVEELYTQDAAVFAGRQLHEKIEQQEDEEWLDLYLE
DEILGLRGRVDALRTRDGQTIPYEHKRGRCYRDDNKPQAWDSDKLQILAYCCLLEAALGIVVSEGRIRYHADN
VLIRVPLDEQGRQWVKDSIQQAKQLRKSAYRPPVTSNEHLCTRCSLSSVCLPEEARLAHNKEWHPLRLFPQDD
EREVLHILEPGTRVGRTGEQLKISRRDQADEKVSVGQVSQVVLHSFAQISTQAVHFLAYKDVGIHFVSGGGCY
IGSVDTRVGSIQRRIRQYQALSQSDFCLELARKLVNCRGEGQRKFLMRGKQKSKSDKVELEQAIAQMKAVLKQ
VPQVQSLESLLGIEGNLAALYFGVLSQLLAPDTPDLLKFSGRNRRPPKDRFNALLSFGYSLLIKDVMNAILTV
GLEPALGFYHQPRTQAPPLALDLMEIFRVPLVDLIVVTSINRNQWDVDADFEVRGQQVWLSDRGRRKFISIYE
QRKSESWKHPVTGYSLTYRRLLELEVRLLEKEWSGEGGLFGQLIVR
>SR001_05650     Cas_Cas1
MGTVYITQEDALIGKTDERLNVKFQKEKLLDVPLIKIDGVVVLGRATVSPAVVTELLTRKIPLSFLTKNGKYL
GRLEPEVTKNIFVRKAQWQAAGESEQAVHLVRAFVRGKLKNYRSSLMRCQRENLGVDLSSSIAQIKQSIESLE
NTELIDSLRGYEGAGSAAYFGCFDRLIRNPEFTFKTRNRRPPTDPVNSLLSFGYALLRHDVQSAVNIVGFDPY
LGYLHVERYGRPSLALDLMEEFRPLVVDAMVLAAINKGALSPADFVSEPISNAVSLTPDGLKIFLRLYEQKKQ
SKFKHPVMGRQCTYQEAFEWQARLLAKYLMGETDKYPPLVLK
>SR001_03378     Cas_Cas1
MQTLYVSQQGCYVCLDQEMLVVKQGDVVWGQVQLPLLEQVLIFGKSQITSQAIRACLQRNIPIVYLSRMGYCY
GRTLPIERGYRQLARYQQQLGFAERLQVAQKIVQVKLNNSRTLLQRQQRRQNSEVVALAIKTLQYCREQAGKV
ESLDRLMGFEGAGAAQYFSAFEECLDNPDFVFVSRSKRPPGNPVNAMLSFGYQVLWNHLLSLIELQGLDPYSA
CLHQGTERHAALASDLIEEFRAPMVDSLVLYLANRQMIDSRDDFLYKNSGCYLNDSGRRKFLKAFVQRMEETI
TNEAGEVQPKWDLLTQQVKKYKQFVYQPTELYSPYQIR
>SR001_00584     Cas_Cas4+Cas_Cas1
MLQLPNNLELSQSDLTKSHTIRVAALHALAYCPRLFYLEEVEELYTQDAAVFAGRRLHEELEKEEEEWEDLFL
ESEELGLRGRMDALRTRDGQTIPYEHKRGRPQRDENKQPQAWESDRLQILAYCCLLESALGIAVSEGRIRYHA
ENVLIHVPLDEMGRQAVQDAIQQAKKLRSSTQRPPVCSNERLCTRCSLAPVCLPEEARLAHNPEWEPVRLFPE
DDRRQVLHILEPGTAVGRTGEQIKISRQGQPVEKIPAQQVSQVVLHSFSQISTQALHFCAGEDITVHWISGGG
RYMGSFDSRQGSVQRRIRQYAALTSSETCLDLAKKLVNCRGKGQRQLLMRGKRGLKPVPENLTEAISQMQKVL
KQVPRAESLASLLGLEGNLAALYFGALPNLISAQVPPEMHFAGRNRRPPKDRFNCLLSFGYALLLKDVMSAIL
TVGLESALGFYHQPRSQAAPLALDLMEIFRIPLVDLVVLGSVNRGQWDIQADFEVRGCQVWLSDSGRRKLVEL
YERRKQESWKHPVTGYSLTYQRLLELEVRLLEKEWSGEGGLFGHLVLR
>CYA15_02595     Cas_Cas1
MSILYITQPQAILGKKYEAFTVATQEESGKWNKQSVPAQTVEQVVLMGNPQVTGDALTYALELGMPVHYLTSF
GKYLGSALPGYSRNGALRLAQFELYQDTERRLELVKTIVSAKIQNQATVLYRQKIADSPLKNHKKLVFKKATL
DEVRGVEGIAAKDYFREWSAILNEPWSFPGRNRRPPTDPVNALLSFAYGLLRVQVTASVHLVGLDPYIGYLHE
VSRGQPAMVLDLMEEFRPLVADSLVFSVLNNREIKPDDFTESLGAYRLKDEGRKRFLQAFERKMNDEFKHPVF
EYRCTYRRAIELQARLLGRYLQEDLVYKPLVLR
>CYA15_00237     Cas_Cas1
MEKVIQAESWDRLMGLEGAAAAQYFSAFGECLTNPKFVFTVRSRRPPGNPVNAMLSFGYQVLWNHLLSLIELQ
GIDPYFGCLHQGTERHAALASDLIEEFRAPIVDSLVMYLINRNMIDANTDFTYHNGGCFLNDSGRIKYLKAFL
QRMEEKLTTAAGENQPRWDTLNQQVKAFKQFVYNPIEGYKPYEIR
>IPPASB1220_03504      Cas_Cas1
MGTVYITHEDAFIGKTDERITVKANKQLLLDVPLIKVDGVVVLGRATVSPAVVMELLEKRIPLSFLMATGRYL
GRLEPEMTKNIFVRAAQWEASGESERAVGAVRGFVRGKLKNYRIALLRRQREAPELDLQRGIQRLEDAIAPID
TTHSINSLRGLEGAGSAAYFGCFPQSIKNALFHFPGRRRRPPTDPVNSLLSFGYALLRHDIQGAINIVGFDPY
```

```
LGYLHVERYGRPGLALDLMEEFRPLVVDAVVFNAINNQKLTPSDFTAEPISGAVSLTPDGLKIFLRMYEEKKQ
SKFKHPVMGRQCTYQEAFEIQARLLAKYLMGEIDQYPPLVKK
>Ana43_04548     Cas_Cas1
VTTLYLTEAGTIVRYRNESLVIVKQQQSTNCRLAELTLVVVLPGVQLTDVVISLLLDRGIETIFLRQDGQFRG
RLQGQFSTNPTIRLAQYRTVETTFGMALAQKLVMGKVRNQRVLLQRRNRATKGQISELTEAIDLISAYASQLS
NTTTPLNRNEIMGVEGICARTYYQALKHWFPTEWNFCGRNRRPPLDPINALLSWGYGVLLARVFAACVQAGLD
PYLGFFHAIEPYRPNLVLDLMEEFRPVAVDQSVISLIQSDLLTPEDFQPSPDGVGIWLGTMAKKLLLGELERR
FRTSVLYPPQNRQLSLSQILLEQARSLGRCLVESSLDYEPFVIK
>Ana43_03901     Cas_Cas1
MGTVYITQEDAFIGKIDERLSVKHEKKTILDVPLIKIDGVVVLGRATVSPAVISELLERKIPVTFLTENGRYL
GRIEPEANKNIFMRKAQWQAAGDSAQAIHVVQGFVRGKLKNYRNILIRRQRECPDIDLSKKITHLENAISPID
STDNINSLRGLEGAGSAAYFGCFNQMIRNPAFEFSHRLRRPPTDPVNSLLSFGYSLLRHDVQGAINIVGFDPY
LGYLHCDRYGRPSLALDLMEEFRPLVVDAVVLSTINKQLLSTSDFVTEPLSGAVYLTPEGRKTFLRLYQQKKQ
SEFKHPVMGRKCTYQEAFEMQARLLGKYLMGETAKYPPLVLK
>PCC6802_02542   Cas_Cas4+Cas_Cas1
MFAPAIDIAQDTIRVSALHALAYCPRLFYLEEVEELYTQDAAVFAGRRLHVELEKQEDEDWEELFLESAELGL
RGKVDALRTRDRQTIPYEHKRGKAARDDRKQPQAWESDRLQILAYACLMESALGIPIQEGRIRYHADNVLVRV
PLDDAGRAEVRAAIARARELRISPYRPPVTANERLCARCSLAPVCLPEEARLAHDRAWQPVRLFPKDDDRDIV
HVLEAGTSVGRTGEQLKISRRDRPVEIVPARQIAQVVLHSFSQISTQALHFCAHQGIGVHFVSGGGRYVGSFD
SRQGSIQRRIRQYGALSDSAKCLELARKLVACRGQGQRKFLMRGQRGMARVPEGLEKAIAHMKTLLKQVTQAP
SLESLLGLEGNLAAAYFGAIPYLLGADVPPEMRFDGRNRRPPKDRFNALLSFGYALLLKDVMSAILTVGLEPA
LGFYHQPRTQAPPLALDLLEIFRVPLVDLPIVGSINRGQWNVREDFDVRGEQVWLSDAGRRKSIDLYEKRKEE
SWKHPALGYSLTYRRLFELEVRLLEKEWCGEEGLFAQLILR
>PCC6802_00675   Cas_Cas1
MRTLYISQQGCYVSLRQETLVIKRGEFVLDEVQLPLLEQVLVFGKSQITTQAIRECLRRDIPIAYLSRMGYCY
GRVMAIERGYRQLARYQQELSSLERLLIARQIVRCKLLNSRVLLQRQHRKRPSDVLDMTIKNLGYLSDRALQV
SDVDELLGIEGAGAAAYFPAFGECLDNPDFAFVARSRRPPGNPVNAMLSFGYQVLWNHLLTLIELQGLDPYHA
CLHQGSERHAALASDLIEEFRAAIVDSLILYLVNRGIVKGSDDFEYRDGGCFLNDTGRKKYLKAFLQRMEEPI
QTESGEQPRWDTLMQQVKAFKRSVYDPSIPYKPYTIR
>NIES3755_03371  Cas_Cas1
MSVLYVTQPDAILNKSYEAFTVKLKQADGTWKKQTIPPQTVDQVVLMGNPQVTGDALMYALELGMSVHYLSSF
GKHLGSALPVESRNGQLRLAQYRRYEDQARKLELVKAIVIAKIQNQAAVLYRHNVKEEKLKQRKASVTSQKTI
DQVRGVEGLAAREYFACWSQMLKEPWTFAGRNRRPPTDPVNALLSFAYGLLRSQVTAAIHLAGLDPYIGYLHE
VTRGQPAMVLDLMEEFRALVADNVVLSVLNNREIQTKDFTESLGAYLLSDAGRRTFLQAFERKLSDEFKHPVF
EYRVSYRRAIELQARLLARHLQEGVSYKPLVLR
>BC008_04014     Cas_Cas1
MSVLYVTQPEAVLSKTHEAFKVALKQSDGSWKKHSVPAQTVEEVILMGNPQVTGDALVYALELGMPVHYLSYF
GKHLGSTLPAYSRNGQLRLAQYAAHCDRSRKLEIAQTIVTGKIHNQHSVLYRYDQKDNPLKNRKKLVPEQTTL
DEVRGVEGLAAREYFACWSAILPDKWSFTGRNRRPPTDPINALLGFAYGLLRVQVTAAVHVAGLDPYIGCLHE
TTRGQPAMVLDLMEEFRPLVADSLIFAVIKHKEIKPDDFTESLGAYRLSDKGRKNFLQAWERKMNDEFKHPNF
GYRCNYRRAIELQARLFSRHLQEGLPYKPFIRR
>BC008_02846     Cas_Cas1
MGTVYVTQEDTFIGKVDERLNVRFEKKTILDVPLLKVDGVVVMGRASISPAAIAELVHNKIPLTFLTLNGKYI
ARLEPEMNKNIFLRSAQWKAAGDSTQAIHVTKAFVRGKLKNYRKSLLKTQRNYPDINVSKGISQLENSLVSLH
KATTIDSIRGCEGAASAAYFSCFDNFIRVDDFSFNSRNRRPPTDPVNSLLSLGYSLLRHDIQGALNIVGFDPY
LGYLHTERYGRPSLALDLMEEFRPLIVDSLVLGAINRKIFKPSDFQKEPISNAVSLTKEKLHEFLRLYQEKKL
TKFTHPVLKIKKTYQEAFEIQARLLGKYLMGETDKYPPLIVK
>NIES2119_02204  Cas_Cas1
MSVLYVTQPEAVLNKNYEAFTVSLKQEDDTWKKQSVPAQTVEQVVLMGNPQVTGDALTYALELGIPIHYLSGY
GKYLGSALPGYSRNGSLRLAQFELYNDSPRKLELVKTIVTGKIQNQYTVLYRQKIANSSLKTYKKLVAQKQTL
DEVRGVEGIAAKDYFAEWTAMLKEPWSFPGRNRRPPTDPVNSLLSFGYGLLRVQVTAAVHIVGLDPYVGYLHE
VSRGQPAIVLDLMEEFRPLVADSLVLSVLNKREIQPDDFTESLGAYRLSDDGRKRFLQAFERKMNDEFTHPVF
DYRCTYRRAIELQARLLSRYLQEGITYKPLVLR
>NIES2119_02752  RVT_1+Cas_Cas1
MEFLITLEQLKAAWGQVRLGTKSAGIDGITVDWFATLGDLQLTKIQHQLHQENYNASPALGFYLPKKDGGFRL
IGIPIVADRIIQRMLLEEMYFPLENTFLDCSYAYRPGRNIQMAVQHLYSYYQFQPCWIIKADIMQFFDNICWA
LLLTQIEQLQIETMLLQLLEQQLKSGIVVKGKRLNPGKGVLQGATLSGALANLYLTEFDRRCLQAGLNLVRYG
DDFAIACHSFIEANRILSQVINWLGELYLKLKPEKTHIYSPDDEFIFLGYRFFQGKVYPPPPPETKAGEWVIS
DSGMPYFRRKVRPAKLISRPPKGCALVKPPLIKSSLAQYWQEQMTTLYVTDQGAYINVHHQFQVFYQRELRI
```

```
KVPVNRVSHIVLFGCCNLSHGAVSLALRRRIPIMYLSQKGRYFGRLETEGVAKVEYLSQQVICAQNSQFVRQN
AEAIVRAKLHNSRVLLLRLNQRRKLAEVEKAIDSLSMLMEKLPDADSMDALRGYEGKGANFHFQALGKLFNET
FAFEKRTKRPPTDPVNSLLSLGYTLLSQNVNSFIQAVGLHTHFGNLHVPRDNHPALVSDLMEEFRALIVDSLV
VYVVNSKILTPEDFTPPDFQGGVYLLPDALKKFLKHWEEKLQTEVLHPNTGYKATYRRCLELQVREYIACLVG
DVEVYRPMTWSK
>NIES2119_02750  Cas_Cas1
MSTIYITETQIKVKVKSPYLLLCDRQEVRQKFLINNVSQIILWGRCYLSREAASLATFRKIPILFIGDRGEDL
GRLQHSHKRQPLCLKYQKRRSLDNEFVLAIAESIIRAKLHNCYIVLQRLQNSKFTPILQIASEFLLLLIDDLA
MANSLEEMREYVAIAASFYYPALASLLPKGFHCKGRRKQPPTDGINCLLNLGYTLLHQIIEMFLEELGLHPDW
GNLYTNSRHQSPLACDLMAEFRAPIIDELVAFLAISEIITPSDFLPTEKPNEVSLHPGILKVFFQHWENQLQT
QIVHPYAGTVSYRHCLQLQIKEYINFLLGNSDRYRPLLLQVDSVPANVQSTEKPETKQLYLVKR
>NIES2119_02736  Cas_Cas1
MTTLYVSKQGCYVFRQQETLVIKQKETILGEVQLPHLEQVLIFGKSQVTTDAIRACLARNIPIAYLSRMGYCY
GRIMPIERGYRQLSRYQKQLTAIERLQVARLIVQAKLKNSRVFLMRQQRRKPSEVIGLAIQSLEHFIQQAGEA
NAIQQLMGYEGAGAAQYFSAFGECITNSHFVFIARSRCPPGNPVNAMLSFGYQLLWNHLLALIELQGLDPYHA
CLHQGTERHAALASDLIEEFRAPLIDSLVLWLVNTKVMDAETDFDYYDSGCYLNNLGRPKYIKFFLQRLEEDK
VTESGEKQPRWELLMQQVKAFKEFVYHPVKVYQPYQIR
>NIES2119_02726  Cas_Cas1
MATLYVSRQGCYVILDQEMLVVKQGETVFGEVQLPLLEQVLIFGKSQVTTQAIRACLWRNIPIAYLSRMGYCY
GRILPIERGYRQLSRYQQQLTNLERLLVAQKIVEAKLKNCRVILQRQQRRQSSPTVGFAVENLEYLIGKAGNA
DTIERLMGFEGAGAAQYFSAFGECINNSEFVFAGRSRRPPGNPINAMLSFGYQVVWNHLLSLIELQGLDPYSG
CLHQGNERHAALASDLIEEFRAPLVDALVLYLVNRRMVSVNQDFVYHDGGCFLNDSGRGKFLQAFVQRMEERV
ESVTGEKQPKWDLLTQQVKFFKQFVYNSSLGYRPYQIR
>NIES2119_01094  Cas_Cas1
MGTIYITQEESFIGKTDERLQVKAEKKTLLDVPLIKVDGVVVLGRSTISPAALMELLQRQIPLSFLTANGKYL
GRLEPQLTKNIFVRSAQWKVMQEPSKATHVVRGFVRGKLKNYRNSLLRAQRENETLNLQDAITRLEQVIAPID
KTESIDSLRGLEGAGSAAYFGSFDKLIRAEGFSFESRRRRPPTDPINALLSFGYALLAHDLQSAVNIVGFDPY
LGYLHTERYGRPSLALDLMEEFRPLVVDAVVLSAVNKRSLFPKDFTTEPLSNAVRLTDEGLRTFLRLYEQKKQ
SKFKHPVLQRQCTYQEAFEIQARLLAKYLMGETELYPPLVLR
>NIES2119_04868  Cas_Cas1
MTAIFITQPGSYLKAKYHYFQVFQQNKLCLQIHSSQVSQITLFDNCRLQSQAASLAESLAIPVLFLDNQGNYL
GRLQLESQHSAKYFQQQIKRTQNLNFTKTIAESIVRAYLHNCCVVLFHLTANHSTPKVEIAMEMLDLLIDDLP
MADSIHALREYSTTGASLYYQSITELLAKAFNFQCTCRPHFDSIKSVLRLGNALLNQTIYGFVQSAGLNPNFG
SLHLYCQDELSLIEDFALEFRPFLVDELVATLAISQILTPDDFTQPTKQGVISLHPAALKKFINHWQAKLQTE
VTHLYAEKLSYRCCLEWQVREYIAYLLGDTEFYRPMLLKWNADTAISPISKQPKEQRVLVKL
>NIES2119_04867  Cas_Cas1
VRYGDDFVIASGDWIQAIRAVEQINKWLGDIYLKLQPEKTQIFAPDEEFTFLGYRFLNGEVIPPPPPKVGELT
QRDSGTPYQSKVRPTPVFSQPPKACSLLKPVKFPVANCEHFWKEPMTTLYITDQGAYLSIRNQQFQVFDRGEL
RIKVPANRVSHIVLFGCCNLSHGAVSCALRRRIPVLYLSQKGRYFGRLQTDGQARVEYLGRQVECSQNSEFTR
RQAEAIVWAKLHNSRALLLKLNRRRPSKIAGEALESIANLMDKLPYAESMDALRGYEGKAATLYFQALGSLFT
GSFAFEKRTKRPPTDPINSLMSLGYTLLSQNVCSFIELMGLHTHFGNLHVPRDNHPALVSDLMEEFRAQLVDS
LVSYLVNSKILNSEDFTLPDERGGVFLQPHALKKFLKHWEEKLQSELTHPHTGYKVSFRRCLELQVREYIACL
MGEVDTYRPMIWNL
>PCC73103_04853  Cas_Cas4+Cas_Cas1
MNNTEYIPIAALNQYAYCPHRCWRMFCAEEFIDNQYTIEGTSLHERVHTLGEGNREETWQVRAIYLKSDKYQL
IGKSDLIELENGEWYPVEYKRGRKGEWDNDELQVCAQALCYSLLKSQRSYSNLDLNGRLNQLATAIANLDQAT
SINSIRGLEGAGSAAYFGCFNQLIRVDNFQFHTRNRRPPTDPVNSLLSLGYALLRHDVQSAINIVGFDPYLGY
LHTERYGRPSLALDLMEEFRPVIVDAIVLAAINRRILSPTDFITEPVSNAVSLTKEGLHKFLRLYQEKKQTKF
KHPVMQKQYTYQESFELQARLLAKYLMGEIEKYPPLVMK
>PCC73103_01484  Cas_Cas1
MTTTLYLTEPGTMVRYRNESLIIIRKEESKNCRLAELDLLVVSPGVQFTDVVISKLLDKGIETIFIRQDGNFR
GRLQGNYATNPAIRISQYRILESAFGMALAQKLVMGKIRNQRVLLQRRNRASQGQIAELAEAINLINTYSYNL
HTTSTLLNRSELMGVEGICARTYYQALRYWFPPQWNFTGRNRRPPLDPINALLSWGYGVLLARVFSACVQVGL
DPYLGFFHATEPYRPNLVLDAMEEFRPVIVDQAVISIIQSDILTPEDFHTST
>ATCC33047_03492 Cas_Cas1
VTTLYLTEAGTIVRYRNESLVIVKQQQSTNCRLAELTLVVVLPGVQLTDVVISLLLDRGIETIFLRQDGQFRG
RLQGQFSTNPTIRLAQYRTVETTFGMALAQKLVMGKVRNQRVLLQRRNRATKGQISELTEAIDLISAYASQLS
NTTTPLNRNEIMGVEGICARTYYQALKHWFPTEWNFCGRNRRPPLDPINALLSWGYGVLLARVFAACVQAGLD
```

```
PYLGFFHAIEPYRPNLVLDLMEEFRPVAVDQSVISLIQSDLLTPEDFQPSPDGVGIWLGTMAKKLLLGELERR
FRTSVLYPPQNRQLSLSQILLEQARSLGRCLVESSLDYEPFVIK
>ATCC33047_00989 Cas_Cas1
MGTVYITQEDAFIGKIDERLSVKHEKKTILDVPLIKIDGVVVLGRATVSPAVISELLERKIPVTFLTENGRYL
GRIEPEANKNIFMRKAQWQAAGDSAQAIHVVQGFVRGKLKNYRNILIRRQRECPEIDLSKKITHLENAISPID
STDNINSLRGLEGAGSAAYFGCFNQMIRNPAFEFSHRLRRPPTDPVNSLLSFGYSLLRHDVQGAINIVGFDPY
LGYLHCDRYGRPSLALDLMEEFRPLVVDAVVLSTINKQLLSTSDFVTEPLSGAVYLTPEGRKTFLRLYQQKKQ
SEFKHPVMGRKCTYQEAFEMQARLLGKYLMGETAKYPPLVLK
>IAMM220_03691   RVT_1+Cas_Cas1
VDSQLSAFLALANFQLAWEKVAANNGCAGVDGQSVAQFSRHESRNLARLRQSVSNGTYRALPLRQIFIPKKNK
SWRSLRVPAVRDRIVQQALLNVLHPVFEPQFESCSFAYRPGRSHHMAVDQISAWQRRGYRWVLDADIVQYFDN
VEHQRLLAEVAERLPRKIVSATFTELVLHLVEQWISVGVLTPEGLLLPQKGIPQGSVVSPILANVYLDDFDEA
FLDRSLKLVRYADDFVILGRKQKEIINVRREVSDLLGEMGLQLHQEKTHLTDFKHGFRFLGHAFVGDLVVPVQ
KNKQRHKQSFLGKGGLRRSGVSAGYATHQKAQVPLPKGDLGGSYVPSIADLRIVHADPVSKSSPLQGALVDAL
KKSQQPIPPPLFVVLGYRVRQEKRVEIISQESIWREGMSTLYLVNQGTTVKKEHGRFLVKEPKETLFEIPSRE
VKRLLVFGNVQLTTAVITECLDLQIPVLFLSQLGQYKGHLWSAEYDDISVEMAQYQRQGDESFRLETARNIVI
GKLANSQRLLWRLNRKRQNEVVQKAIEGLDNDMATAEQVTALDALRGCEGAGAARYFKALGQLITNEAFTFEQ
RNRRPPKDPVNSLLSFGYTLLFNNVLSLLLAEGLNPYLGHLHGSEKKKMFLAFDLVEEFRSPIVDSLVMRLIN
QKFFKPTDFTWPNKEGGIYLNDVARRPFLKHFEKRLSLELSHPDVQSPVSYRRIIQLQIQRYKQALLSDIPYV
PYRKVY
>IAMM220_00640   Cas_Cas1
MTTLYLTQPDAVLSKKHEAFNIALKQEDGNWKKQSIPAQILEDIVLLGYPSVTGEALGYALELGLPVHYLSRF
GKYVGSALPNESRNGQLRLAQFRAHEDPNQRLEIVKSIVRGKIHNQYNLLYRRGQKENPLKGREKLVLQKHSI
DEVRGVEGLAAREYFACWQDMLGKDWEFTGRFRRPPTDPVNALLSFGYGLLRTQLTAAAHIAGLDPYIGFLHE
TTRGQPAMVLDLMEEFRSLVADSVVLTVLKQKEIKPKDFTESLGAYRLKDDACKVFLSAFDRKLSSEFKHPTF
KYKCTYRRAIELQARLLARHLQEGIIYEPLIIR
>IAMM220_01861   Cas_Cas1
MKTLYVSEQGCYVRLKHEFLEVWQRETCLAQVQLPLLEQVLIFGQSQVSIQAMRACLWRDIPIVFLSRLGYCY
GRVMAVERGYRQLARYQQGLSVTEKLSVARSLVAAKVHNCRVLLLRQNRKYPTAILTLSISTLQHLRSQILAA
ETTERLMGLEGAAAASYFSAFGECLRNSDFCFVSRSRRPPLNPVNALLSFGYQVLWNHLLALVEIRGLDPYQA
CLHQGSERHAALVSDLIEEFRAPLVDGLVLYLINKKVISSEDFVYRDGGCFLNSAGRKIYLSAFVRQMEEVIN
VGDEREMRWHLLTTQVRAYQNFIYQPVNGYAPYQIR
>CYA540_02756    Cas_Cas1
MLTLYVSQQGCYLTLKAESLLIKKGKDVLAEAQIPLLEQVLVFGQSQVTTQAIRACLSRNVPIAYLSRMGYCY
GRILPVERGYRHLSRYQQQLTQIDRLIIARRIVQAKLKNSRVILQRQARRQSSPTINLAVQSLNHLMEKVIQA
ESWDRLMGLEGAAAAQYFSAFGECLTNPKFVFTVRSRRPPGNPVNAMLSFGYQVLWNHLLSLIELQGIDPYFG
CLHQGTERHAALASDLIEEFRAPIVDSLVMYLINRNMIDANTDFTYHNGGCFLNDSGRIKYLKAFLQRMEEKL
TTAAGENQPRWDTLNQQVKAFKQFVYNPIEGYKPYEIR
>CYA540_00502    Cas_Cas1
MSILYITQPQAILGKKYEAFTVATQEESGKWNKQSVPAQTVEQVVLMGNPQVTGDALTYALELGMPVHYLTSF
GKYLGSALPGYSRNGALRLAQFQLYQDTERRLELVKTIVSAKIQNQATVLYRQKIADSPLKNHKKLVFKKATL
DEVRGVEGIAAKDYFREWSAILNEPWSFPGRNRRPPTDPVNALLSFAYGLLRVQVTASVHLVGLDPYIGYLHE
VSRGQPAMVLDLMEEFRPLVADSLVFSVLNNREIKPDDFTESLGAYRLKDEGRKRFLQAFERKMNDEFKHPVF
EYRCTYRRAIELQARLLGRYLQEDLVYKPLVLR
>PCC7112_05650   Cas_Cas1
VGTLYINQDDVFIGKTDERLNVKAKNQMLQDVPLIHLDGVVILGRATISPAAIAELLQRKIPLTFITDTGHYL
GKLEPELSKNIFVRNAQWKAAGETPQAIHVVRGFIRGKMKNYRSSLSQTQREHSELNLASAIMQISNAIAKLD
KTDNIDSLRAQEGAGSAAYFRVFNQLIRAEGFVFQPRHRPATDPVNSLMNFAYTLLRHDVQSAINIVGFDPYL
GYLHMQRYGRAGLAFDLMEEFRPLVADSIILNAINRRILTPNDFTVEPLSNVVRLSETGRKEFLKLYEKKKQS
EFKHPVLKRKCTYQEAFEIQGRMLAKYLMGEIDKYPPLVLY
>PCC8106_01382   Cas_Cas1
MGTIYIIQEDAFIGKTDERLIVKLGKEKLIDVPLIKVEGIIILGRATISPAVVMELLERKISLTFLKKTGKYL
GRLEPELTKNIFVRKSQWAAAGESSQAIHAVQGFVRGKLKNYRNTLLRRSRKYTELDLQDSITKIEQVIAPID
SKNSINSLRGLEGAGSAAYFGCFNQLINSNSGFSFQFRNRRPPTDPVNSLLSFGYALLRHDIQSAINIVGFDP
YLGYLHVERYGRPSLALDLMEEFRPIIVDAVVLSAINNKLLSCNDFITEPLSNAVSLTPEGLKIFLRLYEQKK
QSKFKHPVLKRQCTYQESFEIQARLLAKYLMEETDKYPPLVLK
>PCC8106_01838   Cas_Cas1
MLTLYVSQPGCSVYKAQETLQIKRNQTVEVEVQLPLVEQVLIFGKSQVTTQAIQACLKRNIPIGYLSKTGRCY
GRILPVQQGYRYLSRHQQQLSAKHKLLVARQIVQAKLKNSRTLLQRQYRRKPSEILDLAIKSLNYLTEKARNS
```

```
ETWERLMGLEGAGAAVYFKAFGTSISNSEFVFTERSKRPPGNPTNALLSFGYQVLWNHIYSLIELQGLDPYFG
CLHQASERHAALASDLIEEFRAPIVDSLVLYLINRKIVNAEADFEYRNGGCYLNNSGRVKYLKAFVLRMEEKV
SSNSGDSQPRWDLINQQVKVFKRFVYDSANQYQPYLNR
>PCC8106_02852   Cas_Cas1
VKDKLLVACQIVQAKLKNSRTLLQRQYRRKPSEILDLAIKSLNYLTEKARNSETWERLMGLEGAGAAVYFKAF
GTSISNSEFVFTERSKRPPGNPTNALLSFGYQVLWNHIYSLIELQGLDPYFGCLHQASERHAALASDLIEEFR
APIVDSLVLYLINRKIVNAEADFEYRNGGCYLNNSGRVKYLKAFVLRMEEKVSSNSGDSQPRWDLINQQVKVF
KRFVYDSANQYQPYLNR
>PCC7103_08488   Cas_Cas1
MGLAINVEKTRITNFERGFCFLGHGFLDNAIFPVDTQEPKLKHETEKQPKGNQRKKKTTHQDNVTILSQALVA
ETNSEAISEDSDDDNTLLQHNIWNQEMAAIYLIEQGTSIYKDHLRFIIHVNEKPKLEIPIREVEQIMIFGNIQ
LSTPVINACLKEQIAVLFLNQSGQYHGHLSSEESTNIDNHLAQLERRNDYVFQLNVSKAIVFGKLVNSKQLLA
KFNRKRKITVIEKAIFGINQDIASLELVDNLDSLRGYEGITAARYFPAFGQLITNTNFEFSLRNRQPPTDPIN
SLLSFGYTLLFNNVMSFIVAEGLSPYIGNFHYGEKQKPYLAFDLMEEFRSIIVDSLVLNIINHSLLKPTDFEI
IPSTGGVYLHQSARRIFLKQFETRMNEEVSHPDLQSQVTYRHAIQLQVRRYKRCLLSGVPYEAFLRNT
>PCC7103_02155   Cas_Cas1
MGTLYITQEDSFIGKIDERLTVKFDKKILQDIPLIKIDGIVILGRATISPATVSELLERNIPLTFLTEYGRYL
GRLEPEVTKNIFVRKAQWAAAGESPQAIHIVKGFVRGKLKNYRNVMFRRQREANLDLSKSIERIEQVITSIDT
THNINSLRGLEGAGSAVYFGCLNQMIRNVEFQFDYRNRRPPKDPVNSLLSFGYSLLRHDIQGAVNIAGFDPYL
GYLHFERYGRASLAFDLMEEFRPLIVDALVLSTINKSLLTPSDFFTEPLSGAVYMTPEGRKAYLKLYTQKKQS
EFKHPVMGRKCTYQEAFEIQARLLAKYLMGEIEKYPPLILK
>PCC7103_01168   Cas_Cas1
MQTLYVSQQNCYVALQKESLIIKQGDIILGEVQLPLLEQILIFGQSQLTTQAIRACLWRDIPIAYLSRMGYCY
GRILPIARGYRQLSRYQQQLSITEQLITARAIIQGKLKNSRTLLRRQRKRLESANLETVLQSLDYLASQASQA
DNRERLMGFEGAGAAQYFSAFGDCFTNSDFVFTARSRRPPGNPVNAMLSFGYQVLWNHLFALIELQGLDPYYA
SLHQAHDGHAALASDLIEEFRAPIVDSLVLWLVNSRVVDARDDFDFRDGGCYLNDTGRRKYLQAFVQRMSEEI
QDDSGEKQPRWDILMQQVKAYKQFVYNPSHQYQPYRIR
>ITEPA1_02636    Cas_Cas1
VPLDHSGRESVHQAIKEARNLRESAYRPPVTENERLCARFSLAPVCLPEEARLTYDKEWQPIRLFPEDDERLI
IHILEPGALVGKTGSQIKINRRNQPIENIPASQIGQLVLHSFSQISTQALYFCAHQDIGIHFISGGGRYVGSF
DVVFNVVFVNILH
>ITEPA1_02635    Cas_Cas1
LSDPAFCLELARKLVLCRGQGQRKFLMRGTRGRTQVPENLHKSIAQIKTMLKQVPQAKSLESLLGLEGNIAAS
YFSALPHLIVADVVDELYFNGRNRRPPKDRFNALLSFGYALVLKDVMNAILTVGLEGALGFYHQPRSQAAPLA
LDLLEIFRVPLVDMTVMSSINRGQWHPKEDFEIRREQVWLSEPGKRKFVELYERRKQETWKHPVTGYSLTYRR
LFELEVRLLEKEWSGEPGLFGNLILR
>ITEPA1_02813    Cas_Cas1
MQTLYISEQNCYVCLQKETLIVKQGDTVHVEVQLPLLEQILIFGRSQITTQVIRACLWRDIPIAYLSRMGYCY
GRILPISRGYRQLSRYQQELSPVDRLITARAIVKGKLKNSRILLRRQRKKRESETLERTLQSLDYLAEQAVQA
DTWERLMGFEGTGAAQYFSVFSECLTNPDFVFSGRSRRPPGNPVNAMLSFGYQVLWNHLLALIEVQGLDPYYA
CLHQANDGHAALASDLIEEFRAALVDSLVMWLINRNIMDVKEDFQFKNGGCYLNDSGRKKFLRGFLQRMTEEI
QTNDGKQPKWDLLTQQVRNYKQFVYNPSHHYKPYQID
>ITEPA1_02648    Cas_Cas1
VPLDHSGRESVHQAIKEARNLRESAYRPPVTENERLCARFSLAPVCLPEEARLTYDKEWQPIRLFPEDDERLI
IHILEPGALVGKTGSQIKINRRNQPIENIPASQIGQLVLHSFSQISTQALYFCAHQDIGIHFISGGGRYVGSF
DNRRGSIQRRIRQ
>CACIAM69d_02034 Cas_Cas4+Cas_Cas1
MEIDTDPWLEPTSTTPASDSSELTATIRVSALHALAYCPRLFYLEEVEELYVQDAAVFAGRKLHQELAEVAEG
EWVELTLESEALGLRGKLDALRSRDGQLIPYEHKRGRCRRGANKQPEAWPSDRLQLLAYACLLEAALSVQVQE
GRIRYHAENVLVRVPIDPGGRAEVAAAIAQARELRRSTQRPPVTDNDRLCTRCSLAPVCLPEEARLVHDRGWE
PVRLFPADDDREVVHVLEAGTRVGRKGEQLKLTQRDNTVNLMPVERVGQVVLHSYAQISTQALHFCHDRAIGV
HFVSGGGRYMGGFDYRAGSIQRRVRQYCALSDPDTSLQLARKLVIAKTKSQRQFLMRVRRTRSAVLPLVDSAI
IQMKPLIDRLERAESLAGLLGFEGKIAALYFSALPSLLVETVPPELQFSDRNRRPPRDRFNCLLGFGYGLLLK
DITNALIAIGLEPALGFYHQPRSQAPPLALDLMELFRLPLVELLVLASVNRGMWQPEEDFAICTVGRGKAQQQ
VQQVWLSDQGRRKFIQLYERRKQDVWKHPVTGYSLTYRRLFELEARLLEKEWSGEPGLFAAWSVR
>UTEXB481_03637  Cas_Cas1
MGTVYITQEDAFIGKVDERLNVKFEKKTILDVPLLKIDGVVVLGRATVSPAVISELLERHIPLTFLTHTGGYL
GRLEPEVTKNIFVRKAQWQAEGESTQAIHLVQGFVRGKLKNYRQVLLRRQRESTDLDLSKNITRIEQAIAPIE
TTHNINSLRGLEGAGSAAYFGCFSELILVPEFDFNGRVRRPPTDPVNSLLSFGYALLRHDVQGALNIVGFDPY
```

```
LGYLHCERYGRPSLALDLMEEFRPLVVDTVVLSALNKRLLTPADFVTEPLSGAVSLTASGRKTFLQLYAQKKL
SEFKHPVFGRKSTYQEAFELQARLLAKYLMGEIEKYPPLILK
>CCY9414_01346    Cas_Cas1
MTTLYLTEPGTVIRYRNESLTITKQEKSQNCRIAEITLVVVLPGVQLTDVVIAQLLDKGIETIFLRQDGQFRG
RLQGQFATNPIIRLAQYRTVETTFGMAIAQKLVIGKIRNQRVLLQRRNRATKGEISQLAEAIDLISSYSSRLS
DFTTPLNRDELMGVEGICARTYYQGLKHWFPPQWNFSGRNRRPPLDPINALLSWGYGVLLARVFSACVQAGLD
PYLGFFHAIEPYRPNLVLDLMEEFRPIIVDQSVISIIQSDLLQSEDFQPSPDGEGIWLGTLAKKLFLGELERR
FHTPILYPAQNRRLSLSQIVLEQARSLGRCLVNSRLEYEAFVIK
>BC008_03921      Cas_Cas1
MSVLYVTQPEAVLSKTHEAFKVALKQSDGSWKKHSVPAQTVEEVILMGNPQVTGDALVYALELGMPVHYLSYF
GKHLGSTLPAYSRNGQLRLAQYAAHCDRSRKLEIAQTIVTGKIHNQHSVLYRYDQKDNPLKNRKKLVPEQTTL
DEVRGVEGLAAREYFACWSAILPDKWSFTGRNRRPPTDPINALLGFAYGLLRVQVTAAVHVAGLDPYIGCLHE
TTRGQPAMVLDLMEEFRPLVADSLIFAVIKHKEIKPDDFTESLGAYRLSDKGRKNFLQAWERKMNDEFKHPNF
GYRCNYRRAIELQARLFSRHLQEGLPYKPFIRR
>BC008_04346      Cas_Cas1
MAFDSKQDTDVSEGDEWAVRAEYWTAKSRAAEPKRLKRSRRANPLILSGHNISLRIENGTLFIRDGFTHHPQK
RDEYRYFPGSLDLPPRIILIDGSGSLSLAVLDWLREQKLPLIRVSWDGKSASTFAENGMPTDPLKVKWQIEQA
NDERARVKFAVSFIRKKLIASRQTLRNLLKDQKGFEKADQVLSENIELLKSDPPTSIDKLYAVEGPCAYFYFR
AWEGLSLKWKGLNRHPLPPTWERFEQRSSVLTGRKPKNWRATHPINAMLNYAYAVLEADVRLKVLADGYDPRF
GIMHRVYKDASDAYVFDLMEPLRPKVDAAVLTFALGERFSPKDFVLRKDGVCRLSPALAKRLCLITDGNLFWN
LVSNCGCSLQERGD
>BC008_03518      Cas_Cas1
MGTVYVTQEDTFIGKVDERLNVRFEKKTILDVPLLKVDGVVVMGRASISPAAIAELVHNKIPLTFLTLNGKYI
ARLEPEMNKNIFLRSAQWKAAGDSTQAIHVTKAFVRGKLKNYRKSLLKTQRNYPDINVSKGISQLENSLVSLH
KATTIDSIRGCEGAASAAYFSCFDNFIRVDDFSFNSRNRRPPTDPVNSLLSLGYSLLRHDIQGALNIVGFDPY
LGYLHTERYGRPSLALDLMEEFRPLIVDSLVLGAINRKIFKPSDFQKEPISNAVSLTKEKLHEFLRLYQEKKL
TKFTHPVLKIKKTYQEAFEIQARLLGKYLMGETDKYPPLIVK
>MDT14b_02324     Cas_Cas1
MGTLYITQDDAFIGKVDERINVKFEKKILQDIPFIHIEGVVVLGRATISPAVVDELMTRHIPLSFINKIGHYL
GRLEPEMTGNIFIRKAQWQAAGETPQSIHAVQGFVRGKLKNYRQTLMRYQREFDDVDLSKNIARIEQVIAAIS
STENINSLRGLEGSGSAAYFGCFDSLIRNSKFSFTKRVRRPPTDPVNSLLSFGYSLLRHDVQNAVNIVGFDPY
LGYLHFDRYNRPSLALDLMEEFRPLVVDAVVLSILNKQLLTPADFVTEPLSGAVSLIPEGRKTFLTLYEKKKL
SEFKHPVMGRKCTYREAFELQARLLAKYLMGTTDKYPPLVLK
>PCC9214_00217    Cas_Cas1
MLTLYVSQQGCYLTLKAESLLIKKGKEVLAEAQIPLLEQVLIFGKSQVTTQAIRACLTRNVPIAYLSRMGYCY
GRILPVERGYRHLSRYQQQLTRVDRLIVARRIVKAKLKNSRTILQRQARRRSSETINLAIQSLNHLMEKVIEA
ESWERLMGLEGAAASQYFSALGECLTNPKFVFTVRSRRPPGNPVNALLSFGYQVLWNHLLSLIELQGLDPYFG
CLHQGSERHAALASDLIEEFRAPIVDSLVMYLINRNMIDADTDFTYHDGGCFLNDSGRVKYLKAFLQRMEEKL
TTADGEEQPRWDTLNQQVKAFKQFVYNPIEGYKPYEIR
>PCC9214_02280    Cas_Cas1
MGTVYVTVNDSFIGKTDERLTVKADKKNVLDVPLIKVDGIVILGRSTVSPAVVQELLERHIPLTFLTDTGRYL
GRLEPEMTKNIFVRKAQWEAAGESEKALHLVRGFVRGKLKNYRMILLRQGRKYSELDLEAALTRLDNAIAPIS
YTNNINSLRGLEGAGSAAYFGAFDQLIRGDGFSFNCRNRRPPTDPVNSLLSFGYSLLCHDVQSAVNLVGFDPY
LGYLHVQHYGRPSLALDLMEEFRPLIVDSMVLSAINLKQLTPDDFTTEPISNAVLMSPEGRRTFLKLYEQKKQ
SKFKHPVLGKQCTYQEAFELQARLLAKYLMGDIEKYPPLVLK
>PCC9214_05455    RVT_1+Cas_Cas1
MDFSLERLWAAWRIVRMRSRAAGVDGMTVDKFASVAEQQLRQLQEQLQQECYVARPALGFYLRKKSGGRRLIG
IPTVKDRIVQRLLLEEMYLPLEDTFLDCSYAYRIGRNIHDAVQHFCLYYQQQPAWVLKGDIAEFYDHLCWALL
LTALQQLELEPIVVRLVEQQFKSGIFLRGQRINPGQGVIQGGVLSGALANLYLTEFDRRCLQLGLNLVRYGDD
FAICSTSQAAANQALEQVRLGLGEIYLSLQSDKTRIYSPQEEFVFLGYRFVAGEVYEPEPPERTVAGEWGISA
SGVPYFKPKRRPAAKVSRPPKGCSLNQEVKPKTNLAHYWIEVMTTLYVTDQGAYLNVYRQQFQVFFEKELRIK
VPVNRVSHIVLFGCCNLSHGAVSLALRRRIPILYLSQRGRYFGRLQTDGHAKVEYLAQQVERALNPEFVRQQA
EAIVLAKLHNSRGLLMKLNRRRKNKVATEAIADLAVLMEKLLFAESLDGLRGYEGKAATIYFQALGSLFTGSF
TFEKRTKRPPTDPINSLLSLGYTLLSQNVCSFIELAGLHTHFGNLHVPRDNHPALVSDLMEEFRAQIVDSFVS
YLVNSKILLLEDFTSPDELGGVYLQPDALKKYLKHWEEKLYSEVTHPQTGYKVSLRRCIELQVREYIAALMGE
VEVYRPMIWKN
>PCC9214_05453    Cas_Cas1
MPTIYVIHPGSSIEVRSRYLQVFHKQQLCLGVPINRVTQIILFGICNLSRQTQVLIATRQIPVYYINVHGDQL
GSLELTPQRSFKYQTHQLICSQDLEFVRAIAESITWAKLQNEIIVLKQLQPKSDHCNIENTINFLQLLADDLP
```

```
SANSIEELWDYHATATSYYYPILASLVPQWSNRHSSFRLDLSHPVLQMLKLGYTFLHQQMALFIHDLGLDPEI
SNLHTLSQTNQPLVDDLIEEFRPILVDSLVVKLLQTGTITSSDFVPPNGSRGGYLQPSALQIFINHWEEKLAT
HIQHPFAGTISYRQCFELQVKEYIACLLGDIEFYQPLLMQKEKGKIKKMYAKEYLTQKNVVQVFSGKD
>MTP1_03881 Cas_Cas1
MIDPLDSVSAEVRRSQIRKLDCTEVCLELARALVAAKIVNQRTLLLRNHSSLPAGVAADLAKEAKNAAQADSV
DAVRGHEGQAAAIYFGHFSGMFKGPLSLEFDANGRKRRPPPDPINSCLSLAYSMLTHECVAALRLARLEPSIG
GFHVSRPGRPAFALDFMEPFRPLIADSVAITCFNRGELTEGHFLRTASGCSLTDAGRRSFFNALGRRMDTEVT
HPVFEYRLSYRRTIVLHARMIAAWLTGDIPSLAFLTTR
>MTP1_04866 Cas_Cas1
MARLRTLPKVRDRISFVYFEHCRIEQDGQSILVFKENSKFFIPCAAISTLLLGPGTSITHAAIKALADNICEV
QWVGEDLMRFYAHGRSGASNTKRLIHQATSWADSLQRLAVVRRMYTFRFDEPLDENLSLQQIRGYEGVRVRTA
YQQWSQKTGVEWHGRNYKQEDWDAANPINQALSIANTCLYAVCQAALYSVGYSPALGFIHTGKPLSFVYDIAD
MYKAQTTIPAAFEAVSETHFDLDSLVRRKCRERFNDYRLMHRVIDDVDEVLGFKSNDDGEPICFLWDDVLEWV
EGGKNWSEEN
>MTP1_00941 Cas_Cas1
MSTLYLVHQGTTVKKEQGRYVVKLPKEEALEIPVREVERVLVFGNIQLTTAVIAECLENQVPVVFMSQLGDYR
GHLWSAEYTDIGTELVQYQHQGNEGFKLATARAIVAGKLANSRQLLLRLNRKRQLPAVAEAIATLAELLESSL
QADSLNALRGYEGLAAAKYFRALGALIVNPAFTFTERNRRPPKDPVNSLLSFGYTLLFNNVLSLLLAEGLNPY
LGNLHGSEKKQTFLAFDLVEEFRSPVVDSLVMRLINQKFLKPTDFTWPNAEGGIYLNDAARRLFLNQFEKRLS
LKVSHPDVAEPVSYRRVMQLQVQRYKRALLEGVPYEAFRKVN
>MTP1_00384 Cas_Cas4+Cas_Cas1
MGTPDDAVPLTRVMALHALAYCERLFYLEEVEEIRLANESVYEGRRIHLQDISEGIKTLTYEMESERWGIRGK
ADAVRYRDGSLVAYEHKKGKSNRRAAWPSDRLQVIAYCALLEEATGKTILEGRIRYHADNVTVRVPFDEAARH
ELCAAVARAAELRQHIERPPVTSDARLCRHCALAPVCLPEEDRLAAMTVNVAPDAVGLDDAKVRRLFPERDER
RILHITEPGTRVERSAEELVIEIPGRAPVRFPGKDVGAVVLHGAVQITTQAIHYCAANDVAIHWLTGGGNVVG
SFRGGTPGVQRRHRQYAAMSDPDFCLGLARRLAMAKVENQIRYVLRATRPMATEAGREPALEKHLTQMRHALA
QISRATEPDEIRGYEGSAGQAYFAAFASLIHTSPDERGGLAGFEFAGRNRRPPRDRVNALLSFFYALVYKDCL
AALIAVGLEPSFGFLHTPRSAAHPLALDLMELFRVALADMPVVASINRRQWDATEDFVVTPHKVWLSLEGKRK
AIALYEERKQESWQHPVLRYSLSYQRTIELEARLLEKEWTGAPGLFARMRLR
>MTP1_04714 Cas_Cas4+Cas_Cas1
MKENFLATPDLIPARMLNEYVYCPRLFYLEYVQREWAENVDVLEGRFVHRRVDRPGGPAPVPAEITAESKIHA
RSVEVGCPELGAVAVIDLLESDDGQVVPVDYKRGAAPNLPEGAWEPERIQVCLQGLLLRANGYRSGYGIIYYA
ADKTRVRVDFTLVLIARTKQLLAAARALAVSGATPPPLVNSNKCPRCSLVGICLPDEINFLRGDDRADDRASE
ATAATALPSATKAAGGALPVPTGREVRRLVPARDDNKAVYVQGQGYALGLNGAVLEIRDKGQAVDAVRLVEIS
QVNLFGNVQISAQALRALVNREVPVLHLSYGGWLQAVTTAPPHKNIELRQRQFQAAGDPGFCLRLAKAFVSGK
IRNSRLLLRRNGRQLDPDVLSQLAALRGAVERCEELSSLLGIEGNAAREYFRHFPAMFKPGNEPPFDFTGRNR
RPPKDPVNALLSFAYSLLLKEMMVAVIGVGFDPHLGFYHQPRYGRPALALDLMEEFRSLVADSVVITVINNGE
IGPSDFIQRGEAVALTPRGRKTFIEAFERRLDQLVTHPVFNYQVSYRRIFEIQARLLARTLMGEIPRYIPFVT
R
>KVJ20_03003     Cas_Cas1
MGTLYVTQTDAFIGKVDERLTVKAEQKTIMDIPLIKLEGIVVLGRATISPAVVSELLDRHICLTFLTENGRYL
GRLEPEATKNIFVRKAQWQAVGESEPAIHLVRGFVRGKLKNYRHSLLRTQREHPDTDLNNNITRLENAIAPIE
KTSSIDSLRGLEGAGSAAYFGCFQQLIKTSEFQFEARRRRPPTDPVNALLSFGYSLLRHDVQSALNIVGFDPY
LGYLHVERYGRPSLALDLMEEFRPLIVDAVVLSLINKRSLTLTDFTTEPLSGAVNLTKEGLHTFLRAYQQKKL
SSFKHPVMGTKCTYQESFEIQARLLSKYLMNEIDKYPPLVLK
>PCC6912_04864   Cas_Cas1
MRTLYVSQQNCYVSLQKETLIVKQGEIVQGQVQLPLLEQILIFGKSQVSTQAIRACLWREIPIVYLSRMGYCY
GRILPIGRGYRQLCRYQQQLTQVERLITARAIVQAKLKNSRVLLRRQRRRQESELLSKVLQNLDHLVSQAATA
DNWERLMGFEGAGAAQYFSVFGECLTNPEFVFVGRSRRPPGNPVNAILSFGYQVLWNHLLALIELQGLDPYYA
CLHEGSERHAALASDLIEEFRAPIVDSLVLWLVNKNVMDAQNDFEYRDGGCYLNDLGRKKYLKAFLQRMEEEV
QIDGGDKQPRWHLLTTQVKAYKQFVYNPAQVYQPYLIR
>PCC6912_00380   Cas_Cas1
MGTVYITQEDAFIGKVDERLNVKFDKKTILDVPLIKVDGVVVLGRATISPATISELLERHIPLTFLTENGRYL
GRLEPEVTKNIFVRKAQWQAAGESAQAIHVVQAFVRGKLKNYRNSLIRRQRECPDLDLSENVTRIEQAIAPID
TTQSINSLRGLEGAGSAAYFGCFNQLIRAKEFTFNSRIRRPPTDPVNSLLSFGYSLLRHDVQGAINIVGFDPY
LGYLHCDRYGRPSLALDVMEEFRPLVVDAVVLSVLNKGLLTPTDFVTEPLSGAVSLTPEGRKTFLRLYAQKKQ
SEFKHPVMGRKCTYQEAFELQARLLAKYLMGETEKYPPLVLK
>PCC7110_07685   Cas_Cas1
```

```
MAAIYLIEQGTSIYKDHQRFIIHVSEKPKLEIPIREVEQIMIFGNIQLSTPVINACLHEKIAVLFLNQSGQYH
GHLASEESTNLDNHLVQMERRNHDEFQFHVSKAVVYGKLINSKQLLIRFNRKRKIAAVEKAIYGISQDIDSLD
LVDDLDSLRGYEGITAARYFPAFGELITNPNFEFSLRNRQPPTDPINSLLSFGYTLLFNNVMSFIIAEGLSPY
IGNFHYGERQKPYLAFDLMEEFRSVIVDSLVLKIINNFVFKTQDFDVLSSTGGVYLNQSSRRVFLKQFETRMN
EETSHPDLQSQVTYRHAIQLQVRRYKRCLLSGTPYEAFLRPT
>PCC7110_06781   Cas_Cas1
MGTVYITQEDAFIGKVDERLHVKFEKTTILDVPLLKVDGVVVLGRATVSPAAVNELLQRQIPLTFLTDTGRYL
GRLEPEMTKNIFVRKAQWQAAGDTPQAIHVVQGFVRGKLKNYRTTLVRRQRECSDIDLSSFIHRLEQTIAPID
TTHNIDSLRGLEGAGSAAYFGCFDQMIRNMAFSFTNRVRRPPTDPVNSLLSFGYSLLCHDIQGAVNIVGFDPY
LGYLHCQRYGRPSLALDLMEEFRPVIVDAVVLSALNKQFLKAEDFVQEPVSNAVSLTKEGRKTFLQLYAQKKQ
SEFKHPVLGRKCTYQEAFEIQARFLAKYLMGEIEKYPPLVVK
>PCC7110_05612   Cas_Cas1
MADIATLHQPFLAEIETELPLQSEEDNSPFQKNIWNREMAAIYLIEQGTSIYKDYQRFVIHVSEKPKLEVPVR
EVQQILVFGNIQLSTPVIQACLQEQIPVLFLSQSGQYHGHLWSEESTHLDNQLIQVERRNDDYFQYSVSKAIV
YGKLMNSKQLLLRLNRKRKIAEVEKAIFGISQDIDSLDSVDNLDTLRGYEGIAAARYFPAFGKLITNPNFEFS
LRNRQPPTDPVNSLLSFGYTLLFNNVLSFIIVEELSPYLGNFHYGERQKPYLAFDLMEEMRSVIVDSLVLKII
NKSLLKPQDFDIAPSTGGVYLNQSARRVFLKQFEARMNEETTHPDLQSVVTYRHAIQLQVRRYKRSLLSGVLY
EPFLRAV
>PCC7110_04957   Cas_Cas4+Cas_Cas1
MQILEDIELSTKLETIRVSALHALAYCPRLFYLEEVEELYTQDAAVFAGRRLHAELEKQEDEDWEELFLISEE
LGIRGRVDALRTREGETIPYEHKRGRCHRDENKQPQAWESDKLQILAYCCLLESSLGIPIQEGRIRYHADNVL
VRIPLDDLGRAAVRGAIEKARNLRQSTHRPPVTDNERLCTRCSLAPVCLPEEARLAHDREWQPIRLFPEDDNR
QVIHVLEPGTNVGRTGEQIKIARRNQPLETIPARQVAQLVLHSYSQISTQALHFCGEQGIGVHFVSGGGRYLG
SFDTRNGSIQRRIRQYTALSSPDTCLELARKLVICRAQGQRKFLMRGKGKNKASEKLTKAIDQMKAVLKLIPH
AKSLESLLGYEGNLAALYFNAFPDLISPDVPKEMGFDIRNRRPPKDRFNCLLGFGYALLLKDIMNAILIVGLE
PALGFYHQPRSQAAPLALDLMEIFRVPLVDMIVMSSVNRNQWDIKTDFDIRGQQVWLSEAGKRKFVELYENRK
QETWKHPATGYSLTYRRLLELEVRLLEKEWMGEGGLFGQLIVR
>PCC7110_02650   Cas_Cas1
MSILYVTQPDAVLSKAAEAFKVALKQEDGNWEKKSVPAQTVEQVVLMGNPQITGDALVYALELGMPVHYLSGF
GKYLGSALPAYSRNGQLRLAQYAAYYDADHRLEIVKTIVTGKIHNQHAVLYRHDQKDNPLKSRKQLVKEQTTL
NELRGVEGLAAREYFAAWSKMLEEEWSFNGRNRRPPTDAVNALLGFAYGLLQVQVTAAVHMAGLDPYVGYLHE
ATRGQPAMVLDLMEEFRPLIADNLILSVISHKEIKPDDFTDNLGAYRLSDSGRKQFLQAWERKMNDEFKHPTF
GYHCSYRRAIELQARLLSRHLQEGVPYKALSLR
>PCC7110_02015   RVT_1+Cas_Cas1
MISTVRDRIVQRSLLDDLYFPLEDTFLDCSYAYRPGHNIQQAVQHLYGYYKYQPKWVIKTDIADFFDSLSWAL
LLTALEELRLEAIILQLLEQHLKAGILINGQYRNFGKGVLQGGILSGALANLYLTDFDRKCLNQGMNLVRYGD
DLVVACSTFSEANRILDKIIAWLGEIYLTLQPEKTQIFSPDDEFTFLGYRFAKGEVYAPPPPEPVREGEWVIN
DSGTPYFRRKPRPTKPVSRPPKACSIGKPINFPRATLSHYWQDFMSTLYITDQGAYLSVKNQQFQVFYQGELK
IKVPVVRVSNIVMFGCCNVSHGAVSMALIRRIPIMYLSQKGRYFGHTAVQGSAKVEYLTRQVQCCENPVFTRQ
QAEAIVAAKLHNSRILLMRLNRRRETELATQAIDLMDVLIDSLPKADSMDALRGYEGKAATIYFQALGSLFTG
FFAFDKRTKRPPTDPINSLMSLGYTLLSQQVFSFVQSVGLHTHFGNLHTPRDNHPALVSDLMEEFRAQVVDSF
VSYIVNKKILTPEDFTPPDERGGVYLQASALKKYLKHWEEKLQSETTHPNTGYKVAYRRCIELQVREYVACLT
GEVEVYRPMTWKL
>PCC7110_02014   Cas_Cas1
MSTIYITDPNASFKLQNKYLKVFQQERQRFLIRISNISQFIIFGNIKLPKDIVQIIRLNQIPVLYLTQTGEYI
GRVENPSPVQTQYLTYQRQRLRDTEFNRATAESIIWAKLHNQHAFLQNWTRQYTFQNTDPTIKRTLNYLELLM
DNLSVAPSINELHEYIHEADSIHYSAIAYLLAFHNQCPPITTKQINKFLTLGHQLLHQYVYTHLQTAGLHPDY
GILYCDRDRELPLAWDLSAEFRAPIVDDMVLHFTRHLANTNGNGKKSYTLIQRFLQHWEGKLKTFVLHPYAGE
VTYRQCIDLQVREYISSLIGDVEYYRPLALKFHPNPLNFTNIAETQKPILTLVK
>PCC7110_01543   Cas_Cas1
VRGIEGLAAREYFACWQKMVGKQWTFNGRNRNPPTDPVNALLSFSYALLKTQVTAAVHLAGLDPYIGYLHEVH
HGQPAMVLDLMEEFRALISDNIVLGVLSKREIQPQDFTESLSAYRLSDAGRKTFLDSGRFCV
>PCC7110_00421   Cas_Cas1
MSTIYITEQDVSFKIQHRYLKVFHQEEQRISLPIRNLSQLIIFGNINLPQDIIKLLRSHPIPVLYLTQSGEYL
GRLENPYKLQPKYLTYQRKRVRDFEFLRATAESIIWAKLHNQHVFLQSWTRHHVNHVIQQALNYLTLLMDNLP
IAQSLSELREYSEEADKVYYWAIASLLNVYSHCPQTTTKRMSQFCHLGNQLLHQYIYTLLNTRELHPGYAILH
RDTHYDLPLAWDLTAEFRAPIVDDLVLNFVRNLSTSNGNGKHQPRSLLQSFLQQWEAKLRTFVLHPYSGEVSY
RQCLDLQVREYLACLLGEVDFYRPLALKFHPTYSTFNVAKPEKAPLALVK
>PCC7110_00419   Cas_Cas1
```

```
LGNQLLHQYIYTLLNTRELHPDYAILHRDTHYDLPLAWDLTAEFRAPIVDDLVLNFVRNLSTSNGNGKHQPRS
LLQSFLQQWEAKLRTFVLHPYPGEVSYRQCLDLQVRDNCFYIIIMVVNSLPLTENKEGGRRKAEVGKNFPSAL
SLFLLPKVEAPQFIDGKVRFCSCHFKSPQFIDGDTTSVFPLPSYER
>PCC7110_00415   RVT_1+Cas_Cas1
MMFTLEQMTSAWLLVRAGTKSAGVDGIDVDLFASAVNDQLRILLRQLQQESYCASPAKGFYLAKSSGGKRLVG
IPTVRDRIVQRLLLEELYFPLEDTFLDCSYAYRPGRNIQQAVQHLYSYYHLRPKWIIKADIAEFFDSLSWALL
LTALEKLPLEPIVVQLLEGQLRSGIVINGKPIYPGKGVLQGGVLSGALANLYLNEFDKKCLHQGINLVRYGDD
FAIACSNWREATRTLDKVAAWLGELYLNLQPEKTQIFAPDDEFTFLGYRFAKGEVYAPPPPVITREGEWVTNE
SGLPYFRPKKRPVKFVSRPPKACSTDKPIKFPSAPISHLWQESMSTLYVTDQGAYLSVKNQQFQVFYQGELRI
KLPVSRVSNVVLFGCCNVSHGAVSMALHRRIPIMYLSQKGRYFGRLQVEGDAKVEYLMRQVECCQNPDFTRRQ
AEAIVRAKLHNSRILLMRLNRRRKSQDVDETVIKTAAHDLEVLMSKLPFAENMDMLRGYEGKAATIYFQALGS
LFSGSFKFEKRTKRPPTDPINSMFSLGYTLLSQNVFSFVQTIGLHTHFGNLHVPRDNHPALVSDLVEEWRASV
DSLVAYLVNSQIFTIDDFTLPDERGGVYFQPHALKKFLKHWEEKLQSELTHPHTGQKVVFRRAIELQVREYIS
CLKGEVEVYRPMVWEK
>PCC7110_00328   Cas_Cas4+Cas_Cas1
MELLDYITTKPDTIRVSALHALAYCQRLFYLEEVEELYTQDAAVFAGRRLHVEIEKAEDEEWEELFLTSEELG
LRGRLDALRTRDGQTIPYEHKRGRCHRDENKQPQAWESDRLQILAYAYLLESALGITVQEGRIRYHADNVLVR
VPLDDSGRSAVREAIQQARTLRQSTHRPPVTDNERLCARCSLAPVCLPEEARLAHDREWQPIRLFPEDDDRQI
IHILEPGTTVGRTGEQIKITRRGQAVETVPARHVGQLVLHSFSQISTQALHFCAEQGIGVHFISGGGRYIGSF
DTRQGSIQRRIRQYAALSNPDTCLELARKLVNCRGQGQRKFLMRGTRGKTEVSDTLKKAINQMQILLKQVPKA
KSLESLLGLEGNLARLYISALPCLISQDIPKELHFSNRNRRPPKDRFNSLLGFGYALLIKDVMNAILTVGLEP
ALGFYHQPRSQAAPLVLDLIEIFRVPLVDMSIMASINRGQWEVQTDFDIRGEQVWLSETGKRKFVELYERRKQ
ETWKHPVTGYSLTYRRLMELEVRLLEKEWSGEAGLFAQLILR
>74_05537  Cas_Cas1
MAAIYLIEQGTNIYKDYQRFIIHVSEKPKLEILIREVQQILVFGNIQLSTPVLQVCLQEQIPVLFLSQNGRYH
GHLWSEESTNLDNQVIQIERRNDDYFQFSVSKAVVLGKLMNSKQLLMRFNRKRKLEDVEKAIYGINQDIDALE
YVDNLDSLRGYEGIAAARYFPVFGQLITNSAFSFSQRFRQPPIDEVNSLLSFGYTLLFNNVLSFIIAEGLSPY
IGNFHYGERQKPYLAFDLMEEFRSPIVDSLVLKVINKSLFKPQDFDVVASTGGVYLSQSARKVFLNQFETRMN
EEVSHPDLISQVTYRHAIQLQVRRYKRCLLSGVPYESFLRAV
>74_01035  Cas_Cas1
MTTTLYLTEPGTMVRYRNESLVIIRKEQSKNCRLAEVDLLVVLPGVQFTDVAISKLLDKSIETIFLRQDGHFR
GRLQGNYATNPAIRISQYRILESAFGMALAQKLVMGKIRNQRVLLQRRNRASQGQIAELAEAINLINTYSYNL
HNNSTLLNRSELMGVEGICARTYYQALRYWFPPQWNFTGRNRRPPLDPINALLSWGYGVLLARVFSACVQVGL
DPYLGFFHATEPYRPNLVLDAMEEFRPVIVDQAVISIIQSDVLTPEDFHTSPDGVGVWVGDVAKKLFLAELEK
QFRNTVLYPPQNRHLSMSQIILEQARSLGRCLVELSLEYEAFVIK
>74_03031  RVT_1+Cas_Cas1
MFTLEQLLFAWVQVRAGSRSAGVDGITVDLFAASADQQLRIILRQLQQKSYRASPAKGFYLTKKSGGKRLIGI
STVRDRIVQRLLLEELYLPLEDTFVDCSYAYRPGCNIQQAVQRLFSYYQYHPTWIIKADIAQFFDNLSWALLF
TGLETLHLEAIVLELLEQQIKSGIVLGGKYINFGKGVLQGGILSGALANLYLTAFDRKCLSHGINLVRYGDDF
AVACSSWTEANRILDKITTWLGGLYLTLQPEKTQVFAPHEEFTFLGYRFVNGKVYAPPTPEPRREGEWLVNAS
GTPYFRPKQRPLKFVSRPPKACDISKPINFPRAPISHLWQEFMTTLYVTDQGAYLSLKNQQFQVFYQGQLRIK
VPIVRVTNIVLFGCCNVSHGAVSMALRRRIPIMYLSQKGRYFGRLETEGQAKVEYLARQVECSRDDKFTRQQA
EVIVWAKLHNSRALLLKLNRRRPSKIALSAINLIADLMDNLPQAESMDALRGYEGKAATLYFQALGSLFTGVF
AFDKRTRRPPTDPINSMLSLGYTLLSQNVYSFIQAMGLHTHFGNLHVPRDNHPALVSDLVEQFRAQLVDSLIA
YLVNSNIFTIEDFTNPDERGGVYFQPHALKKFLKHWEEKLQSEITHPHTEHKVAFRRAIELQVREYIACLMGE
VEVYRPMIWKL
>74_03029  Cas_Cas1
MSTIYITEQDAMLKIQHHYLKVFHQQKQCISIRISNVSQIIIFGKIPLPREVIKTVLAHQIPVLYLTQDSEYI
GRLENTSQKPAKYLIHQRRRARDVEFNRATAESIIWAKLHNQNIFLQSWTRHYTSDATQRASNYLMLLMDNLA
LATSLDELRQYSEEADNVYYCAIGSLLNFYNGCPQTTAKRIRRFLNLGNQLLHQYIYTLVNTAGLHPDYAILH
YDTDHELPLVWDFMAEFHAPIVDDLVLNFARHLSNSHSNGNGKTQPRSLLQRFLQYWEAKLRTFVLHPYAGEV
SYRQCIDLQVREYLACLLGDIEYYRPLALKFHPNDPSFTNIAETHKPALTLVKK
>PCC7105_03680   Cas_Cas1
MGTVYIAQEDAFLGKVDERLSVKYQRDRLLDIPLIKINGIVVLGRATVSPAVVEELLDRNIPLSFLTRTGRYR
GRLEPEVTRNIFVRRAQWGASGETSISTDAVRGFVRGKLKNYRSLLQRKQRENPELTLGEAIEQLHRAIAALK
KTHHIDRLRGYEGAGSAAYFGCFGQLVRGEGFSFTHRNRRPPLDPINSLLSLGYALLRHDIQSALNIVGFDPY
LGYLHTERYGRPSLALDVMEEFRPLVVDVVALAVTNKRLLTPGDFETEPVSGAVSLTSDGLKVFLRAYEEKKQ
SKFKHPVMGCKCTYQEAFEIQGRLLAKFLMGETDKYPPLVLR
>PCC7105_02309   Cas_Cas4+Cas_Cas1
```

```
MNAQITVETLRVCALHALAYCPRLFYLEEVEELYTQNQSVFEGRRLHAELDRTEDGEWEDLFLTDETLGLRGR
VDALRTRDGHLIPYEHKKGRSHRDAQKQPQAWESDRLQILAYCCLLETVAQHPIREGRIRYHGDNALVKVPFD
ENGRAEVKAAVELARTLRESPERPPIAENERLCQNCSLAIVCLPEETRFFQDATHPPSRLFPKDDDRQIVHVL
EPGTRVGRTGEQLKITRRDRSVETLAVRQTSQVVLHSFSQISTQALHLCASQDIGVHWISGGGRYVGSFDGRR
GSVQRRCQQYAALSQPDMCLNLARRLVACRGRGQRQFLMRCQRNLNGTRDAISASVKQMKLLLKQVEEARSLA
QLLGLEGNLAALYFQGLARTISSKASPELQFDGRNRRPPKDRFNTLLGFGYGLLLKDVMNAILAVGLEPALGF
YHQPRTQAPPLALDLMEIFRVSLVDMTVLASVNRGQWDVTADFDPRGNGVWLSDRGRRKFVELYERRKGESWK
HPVLGYSLTYRRLLELEVRLLEKEWLGEGDGFARLVVR
>PCC7105_02125   Cas_Cas1
MTTLYLTEPGTVVGYGNESLTVKRGEEKRRCRLAELRLLVVLPGVQLTAQAIAALLDRGVETLFLRGDGRFRG
RLQGQFQTNPTLRLAQYRVVDTTFGMALAQMLVLGKIRNQRVLLQRRNREVQGRVAVLAEAVETMGTYIRQLQ
GLESPLERSQLMGIEGICARTYYQALAQVLPPRWNFSGRNRRPPRDPGNALLSWGYGVLLSRVFSACVRAGFD
PYLGFFHATEPYRPNLALDVMEEFRPVAIDRTVWGMLESDLLDPLDFEPSPDGEGVWLGTTAKKLFLAQLERR
FQTRFFYEPQNRQLTLSQILLEQARWLGRCCVELRSDYRAFTVR
>PCC7105_01177   Cas_Cas1
MSVLYITQPDAVVSKHYEAFIVALKQEDGSWEKKKIPAQTVEQIVLMGNPSITGEALTYALELGISVHYLSRF
GKYLGSALPGESRNGQLRLAQYAAHCDLPQKLEIVKAIVAGKIHNQYGLLYRRGCPDNPLQQRKKAVSSQETL
DAVRGMEGLAAREYFGLWSSILKSPWTFPGRNRRPPTDPANALLSFAYGLLRVQVVTAVHVTGLDPYIGYLHE
TTRGQPAMVLDLMEEFRPIVADSVVLSVLNMGALKPKDFQESLGSYRLSDGGRKTFLEAFDRKLTDEFKHPVF
GYRCSYRRAIELQARLLGRHLQEGVPYKPLALR
>PCC7414_04671   RVT_1+Cas_Cas1
MFTLEQITSAWLLVRAGTKAAGVDDISLDLFASAINDQLPILLRQLQQESYRASPAKGFYLSKSSGGKRLVGI
PTVRDRIVQRLLLEELYFPLEDTFLDCSYAYRPGRNIQQAVQHLYSYYQFQPKSIIKADIAEFFDSLCWALLL
TALEELSLEPIVVQLLEGQLKSGIVIAGKLIYPGQGVLQGGILSGALANLYLTAFDKKCLSHGINLVRYGDDF
AIACNSWLEANRTLDKVTTWLGELYLHLQPDKTQIFAPGDEFTFLGYRFAGGEVYAPPSPVVTREGEWVTNGS
GLPYFRPKQRPTKSVSRPPKACSIDKPIKFPTAPISHLWQDFMSTLYVTDQGAYLSVKNQQFQVFYQGELRIK
IPVSRVSNVVLFGCCNVSHGAASLALRRRIPIMYLSQKGRYFGRLQVEGDAKVEYLMRQVECCQNPEFTRRQA
EAIVTAKIHNSRILLMRLKRRKSSEFDDNLVKEAIKDLDTLMDKLPFAESMDALRGYEGRAATVYFQALGSLF
SSVFKFEKRTKRPPTDPVNSLLSLGYTLLSQNVFSFIQAIGLHGHFGNLHVPRDNHPALVSDLMEEFRAQLVD
SLVAYLINSNIFTEDDFTPPDEKGGVYLQPHALKKFLKHWEEKLQSELTHPNTGYKVSFRRCLELQVREYVAC
LMGEVKIYRPMIWKP
>PCC7414_00001   Cas_Cas1
LKKPGFFKYLIHQRRRARDVEFNRATAESIIWAKLHNQNVFLQSWTRHYTSDAIQRASNYLMLLMDNLALATS
IDELRQYSEEADNVYYCAIASLLNFYNGCPQTTAKRIRRFLNLGNQLLHQYIYTLVNTVGLHPDYAILHHDAD
YELPLAWDFMAEFHPPIVDDLVLNFARHLPHSNGNGKSQPRTLLQRFLQHWEAKLRTFVLHPYAGEVSYRQCI
DLQVREYLACLLGDVEYYRPLALKFNPARSEFADVIAPQQVPLTLVKR
>PCC7414_00734   Cas_Cas1
MGTVYITQEDAFIGKVDERLNVKYEKKTILDVPLIKIDGLVVMGRASISPATIAELVNQKIPLTFLTNNGKYI
ARLEPEMSKNIFLRSAQWKARGESPQAIHASKGFVRGKLKNYRYSLLKSQRSYPDLDLNGGLNQLATAIANLD
QATSIDSIRGLEGAGSAAYFSCFNQLIRVDNFQFHTRNRRPPTDPVNSLLSLGYALLRHDVQGAINIVGFDPY
LGYLHTERYGRPSLALDLMEEFRPVIVDAVVLTTINRRILVPTDFITEPVSNAVSLTKEGLHKFLRLYQEKKQ
TKFKHPVMQKQYTYQESFELQARLLAKYLMGEIEKYPPLIMK
>NIES592_00223   Cas_Cas1
MSTIYITEQDAMLKIQHHYLKVFHQQKQCICIRISNVSQIIIFGNIPLPREVIKTVLAHQIPVLYLTQDSEYL
GRLENTSQQPAKYLIHQRRRARDVEFNRATAESIIWAKLHNQNVFLQSWTRHYTSDAIQRASNYLMLLMDNLA
LATSIDELRQYSEEADNVYYCAIASLLNFYNGCPQTTAKRIRRFLNLGNQLLHQYIYTLVNTVGLHPDYAILH
HDADYELPLVWDFMAEFHPPIVDDLVLNFAPHLPHSNGNGKSQPRTLLQRFLQHWEAKLRTFVLHPYAGEVSY
RQCIDLQVREYLACLLGDVEYYRPLALKFNPARSEFADVIAPQQVPLTLVKR
>NIES592_00222   RVT_1+Cas_Cas1
MFTLEQITSAWLLVRAGTKAAGVDDISLDLFASAINDQLPILLRQLQQESYRASPAKGFYLSKSSGGKRLVGI
PTVRDRIVQRLLLEELYFPLEDTFLDCSYAYRPGRNIQQAVQHLYSYYQFQPKWIIKADIAEFFDSLCWALLL
TALEELSLEPIVVQLLEGQLKSGIVIAGKLIYPGKGVLQGGILSGALANLYLTAFDKKCLSHGINLVRYGDDF
AIACNSWLEANRTLDKVTTWLGELYLHLQPDKTQIFAPGDEFTFLGYRFAGGEVYAPPPPVVTREGEWVTNGS
GLPYFRPKQRPTKSVSRPPKACSIDKPIKFPTAPISHLWQDFMSTLYVTDQGAYLSVKNQQFQVFYQGELRIK
IPVSRVSNVVLFGCCNVSHGAASLALRRRIPIMYLSQKGRYFGRLQVEGDAKVEYLMRQVECCQNPEFTRRQA
EAIVTAKIHNSRILLMRLKRRKSSEFDDNLVKEAIKDLDTLMDKLPFAESMDALRGYEGRAATVYFQALGSLF
SSVFKFEKRTKRPPTDPVNSLLSLGYTLLSQNVFSFIQAIGLHGHFGNLHVPRDNHPALVSDLMEEFRAQLVD
SLVAYLINSNIFTEDDFTPPDEKGGVYLQPHALKKFLKHWEEKLQSELTHPNTGYKVSFRRCLELQVREYVAC
LMGEVKIYRPMIWKP
```

```
>PCC7120_01599   Alr1568 Cas_Cas1
MGTVYVTQDDSFISKVDERLNVKFEKKVILDVPLIKVDGLVVMGRASISPAAIFELIDKKIPLTFLTNNGKYL
ASLEPEMGKNIFVRAAQWKAAGESEQAIHVTQGFIRGKLKNYRHTLLEAQRRYEVDLNSNISQLSNAIASIDK
ANSINTIRGLEGAGSAAYFGCFNQLIRVDNFSFHTRNRRPPIDPVNSLLSLGYSLLRHDVQGALNIVGFDPYL
GYLHTERYGRPSLALDLMEEFRPLIVDAVVLTAINRRMLSPKDFITEPISNAVSLTKEGLHIFLRLYQEKKQT
QFKHPVMQKKYTYQETFEIQARLLAKYLLGELDKYPPLVMR
>PCC7120_01493   Alr1468 RVT_1+Cas_Cas1
MFTIEHLNFAWLQVRAGSKSAGVDGISVDLFESMATEQLQNIAYQLKEETYTANPAKGFYIPKKNGTKRLIGI
HTVRDRIIQRLLLDELYFPLEDTFLDCSYAYRPGHSIQQAVQHLYGYYQYQPKWIIKADVADFFDNLSWALLL
TYLEELSLEPSLLQLLEQQLKSGIIIAGQYRNFGKGVLQGGILSGALANLYLTSFDRKCLSQGINLVRYGDDF
VIACNSWLEANRILDKITGWLGEVYLTLQPEKTQIFTPNDEFTFLGYRFAGGEVYAPPPPEPVRPGEWVINDS
GTPYFRLKPKPRKPVSHPPKACSIDKPINFPKASLSHYWQETMTTLYITDQGAYLSVKNQQFQVFHQGELRIK
VPVMRVSNVVLFGCCNVSHGAVSMALRRRIPIMYLSQKGRYFGRLQTEGDAKVEYLMLQVERCQNHEFTRKQA
EAIVRAKLHNSRALLLKLNRRHPSKIAATAISGIAELMEKLSLAESMDSLRGYEGKAATLYFQGLGSLFTGAF
VFEKRTKRPPTDPVNSLLSLGYTLLSQNVFSFVQVIGLHTHFGNLHVPRDNHPALVSDLMEEFRAQLVDSLVA
YLINSKIFTFEDFTPPDEKGGVYLQPHALKKFLKHWEEKLQSEVTHPHTGYQVSLRRCLELQVREYITSLTGE
VELYRPMIWKL
>PCC7120_00385   Alr0381 Cas_Cas1
MTTLYLTEPGTILRYRNESLIIMKQEKSHNCRLAEITLIVVLPGVQLTDVVISQLLDRGIETIFLRQDGQFRG
RLQGHFATNMTIRLAQYRTVETTFGMALAQKLVIGKVRNQRVLLQRRNRATNGQISELTEAIDLISVYASQLN
NTTTPLNRNELMGVEGICARTYYQALKHWFPTQWNFNGRNRRPPLDPINALLSWGYGVLLARVFSACVQAGLD
PYLGFFHAIEPYRPNLVLDLMEEFRPVVVDQAVISLIQSDLLTQEDFQPSPDGVGIWLGTMGKKLLLGELERR
FRTSVLYPPQNRQLSLNQILLEQARSLGRCLLQSKLDYEAFVIK
>PCC8927_03718   Cas_Cas1
MTTLYLTEPGIVVRYHNQSFLIKQQDKTHHCRLVELTLVVILPGVQFTDEAIAQLLDRGIETLFLKRDGQFRG
RLQGQFQTNPTIRLGQYRMINTTFGMALAQLLVMAKIRNQRVLLQQRYRFSKGRIGELGEAIDAIAAYVLQLK
SVSTPLNRSELMGIEGICARTYYQGLRYYFPQRWKFTGRNRRPPRDPINALLSWGYGVLLARVFAACVQAGLD
PYLGFFHATEPYRPNLVLDLMEEFRPVVVDQVVISLIQGNLLQPEDFQPSPDGEGVWLGLAAKKLLLGQLERR
LREPLLYRPQNRKLSLSQIILEQARSLARCLVEMRLDYEGFVLK
>PCC8927_02461   Cas_Cas1
VRVSSLASIDRLETVIMLRALRLHNLQVNVFEVNTMGTVYITVDDSFIGKTDERLTVKADKKNVLDVPLIKVD
GIVILGRSTVSPAVVQELLERHIPLTFLTDTGRYLGRLEPEMTKNIFVRKAQWEAAGESEKALHLVRGFVRGK
LKNYRVLLLRQGRKYSNLQLDDNVTRLDNAIAPISYTNNINSLRGLEGAGSAAYFGVFDQLIRGDGFSFQTRN
RRPPTDPVNSLLSFGYSLLCHDVQSAVNLVGFDPYLGYLHVQHYGRPSLALDLMEEFRPLIVDSMVLSAINLK
QLTPDDFTTEPISNAVFLSNEGRRTFLKLYEQKKQSKFKHPVLGKQCTYQEAFELQARLLAKYLMGDIEKYPP
LVLG
>HeronIslandJ_05340    Cas_Cas1
MGTVYISTDDAFIGKTDERLKVRAKKETLLDVPLIKVDGVVVLGRATVSPAAMIELLERKIPLSFMTGTGRFL
GRLEPELNKNIFVRSAQWQAAGETPRSVHVVQGFIRGKLKNYRAMLLRAQREGITNLQSGITRLEQAIKPIDQ
TSIIDALRGLEGAGSAAYFGFFGRLLRNEEFTFKTRNRRPPTDPVNALLSFGYALLLHDIQGAINIVGFDPYL
GYLHTQRYGRPSLALDLMEEFRPLVVDAMVLSAINRKALTLDDFSDEPLSHAVSLTDEGRRKFLRLYEQKKQS
KFKHPVMGRQCTYQEAFEIQARLLAKYLMAETDQYPPLILK
>HeronIslandJ_06914    Cas_Cas1
VFGNCQLTTQVISVCLEQRITVVYLTQMGDYKGHLWSAEAAGMQAVMAQFERQQDPVFCLETARAIVTGKVQN
SRQLLLRLNRKRKIDIVAEIIQGLNNDLEAVAIANEINGLRGHEGAAAARYFKALGQLIVNPGFGFKGRHRRP
PIDPTNSLLSFGYTLLHNYILSLILAEGLTPYLGNLHGSERKQTFLAFDLIEEFRSVIVDALVIRLINRKSIS
PTDFTWPNQQGGVYLQGTARKLFLQRFEDRLSQPVLHEDVVDMVSYRRVIQLQVRRYKQAVLNNTSYGVYRRP
N
>NIES39_03409    Cas_Cas1
MITKRTVIIQQSKGSTTCRLAEVDLVVILPGVQLTDSTIAQLLDRGIETIFLRQDGRFRGRLQGHFHTNPTVR
LAQYRALDTTFSMAIAQKLILGKVRNQRYLLQRRNRETQGKITQLTEAIDTIASYLPRLKQTETPLDRNELMG
VEGVCARIYFQGLTHWFPPEWKFTGRNRRPPKDPVNALLSWGYGVLLARVFAATVQAGFDPYLGFFHANAPFR
PNLVLDLMEEFRPVVVDRVAIALINSGLIDPNDFEPAPGEEEGIWLGQTAKKLYLGQLERQLQDSILYPPQHR
KLKLSQILLEQGRSLGRCLVETRLDYEAFIFR
>NIES39_02745    Cas_Cas1
MTAIYLTEPQSYLEADLRDFAIVYNDRIQSRHSIDTTQQIIAFEGCTLGRSAIAAIYRYRIRLSFIGDNGHIT
AQITPQETPFYTVLQQERNSDRNFRIAFIHSLLPGRWYNATKVLQSLGYSGIGGCLEQQLEVLPGLNSWLALQ
TWQVKVFQRYQKALRSVLVSVKDFEPTLELAYALLSGELYVLLRSSGCCAEVGTLHLHCQNHLPLPCDLMEPF
```

```
RPVVEAWVFKKAPKFSDRGGLSQSYRATFIQNWESLMATQIFHAYAGQVSLRSALAWQVEEYVRAITETLEYR
PFLLKS
>NIES39_02743    RVT_1+Cas_Cas1
VRDRIVQRLLLRSLYPILEETFQDCSFAYRPGVGVKHAIERVAEVYSSQTWTIKADISQFFDSLCRTLLLSQL
EELSVDQTVVRYIKGQLEAGIVVGGMPILSGRGVLQGGILSGALANLYLSEFDRRCLDAGAYLTRYGDDFVIV
ARSLLEATRFLNLIEDWLSDIYLTLQPEKTHIFAPGEEFVFLGYGFQAGEIITPEYKQPRRNHNKKAVSVASR
PPFACSIVTPTRKSLSTRNFLDSWREGMTTLYVTDQGAYVKVKHQQFQVLLGNDLKVSIPVNVVDYIILFGCC
NLSHGAIGLALRRRIPILFLSDQGRYFGRLQTDGMTRVDYLSRQVHCAEDETFVLRQAKVIVAGKLHNCRILL
RRLNRDRQISQVIEAIEELGVWQEKIAEVELLESLLGYEGFGTRIYFQALRALVQPPFTFEHRTRRPPTDPVN
SLLSLGYTLLHQNIHSLILAVGLHPHYGNLHVPRSNHPALVSDLIEEFRAPVVDSLVIYLVNSGIFTPEDFTP
SDERGGVYIYSDALKKYLKHWHDKLSLKTTHPHTGYKVSYYRCLELQVWEYISCLMGEREVYRPMLTKM
>CR12_00314 Cas_Cas1
MTILFLLEQDTLVRQIDERLEIFKHDKRLTDVPLCKLESVVVYGRVILTIPALKILNQRSIPITYLSEEGRTV
ATLLPEPNPNAILRSKQYQASFDTHKTLAIAREIIRGKLFNQHTILARFSRQNERSEKVISALKSLKACQKSI
DQTTSLNELRGYEGQGASSYFGVLGELLTGTPWSFSHRTRRPPTDPVNALLGFGYALLYGDCRAALHTVGLDP
YQGFLHGERYGRANLALDLMEEFRPIFVDGLVLQLLRNNSLEKESFINYPGGAVHLNEQGMKKFIQSYEQRKN
TTFQHSVFEIQTDYRRALILQARLLAKHLNQEIEKYLPLKVR
>CR12_00004 Cas_Cas1
MQTLYVSEQNCYICLQKEILVVKRGEVTQAQVQLPLLEQILIFGSSQVTTQVVRACLLRDIPIAYLSRMGYCY
GRILPISRGYRQLSRYQQQLSAMEKLVTAREIIKSKIKNSRVLLQRQKKKTASANLERVLESLNYLAEQTASA
DSCERLMGIEGAGAAQYFSAFGECLSNENFIFLGRSRRPPGNQVNAMLSFGYQVLWNHLLALIEIQGLDPYYG
CLHQGHDGHPALASDLIEEFRSPLIDSLVMWLINRNIFDVGADFEFIKGGCYLNNSGRKKFLQAFLQRMTEEV
QTEDGTKQPKWDLLTQQVRAFKQFVYNPAHHYRPYRIH
>NMC1_01837 Cas_Cas1
MTTLYVIQPDAVLSKDYEAFQVALRQEDGSWKKQKVAAQTVNEVILMGNPQVTGDAFVYALELGMPVHYLSSF
GKYLGSALPKSSRNGQLRLAQYEVYHDSVRRLELVKAIVSAKIHNQYNVLYRHNEKDNPLKERKNLVKTQQTL
DQVRGVEGLAAREYFACWQRMVGKQWTFNGRNRRPPTDPVNSLLSFAYALLQGQVMAGVHIAGLDPYIGYLHE
VHHGQPAMVLDLMEEFRALIADNLVLSLLHKHEIKPEDFHESLGAYRLKDNARKTFLQAFDKKMNDEFKHPVF
DYRCTYRRAIELQARLLSRHLQEGIPYKPLSLR
>NMC1_03058 Cas_Cas1
MGTLYVTQNDSFIGKVDERIHVKFEKKILQDIPFIHLKDVVVLGRATVSPAVVFELMTRHIPLSFIDERGHYL
GRLEPEMTGNIFIRKAQWQAAGESPQSIHAVQGFVRGKLKNYRQMLMRRQRESNDVDLSKYITRIEQAIAPID
STHNINSLRGLEGAGSAAYFGCFNQLIRNPEFSFTKRVRRPATDPVNSLLSFGYSLLRHDVQGAVNIVGFDPY
LGYLHYDRYNRPSLPLDLMEEFRPLVVDTVVLSILNQQLLTPADFISEPLSGAVSLTPEGRKIFLTQYERKKQ
SEFKHPVMGRKCTYRQAFEWQARLLAKYLMGETEKYPPLVLK
>UAMMAO_03900    Cas_Cas1
MDNLSLATSFDQLWEYSEEADGVYYNAVTSLLKFHHPYLYFQRGRIGGLFNLGYQLLHQHIYTLLDTSGLHPH
CAILHRNVHHELPLAWDFTAEFRTAIVDDLVLHFAKKFPHTNGNGNGNSNGKSQPRTIVQHFLQHWEAKLKTF
VLHPQAGTVSYRQCMDFQVREYLGYVLSDVEYYRPLALKFHPLDSKLASITKPQKASLTVVK
>UAMMAO_00873    Cas_Cas1
MSILYVTQPDAVLSKAAEAFKVALKQEDGGWEKKSVPAQTVEQVVLMGNPQITADALVYALELGMPVHYLSGF
GKYLGSALPDYFLLITPYSSMKPPRGIETNISRDVLQYVPTQLSNQ
>UAMMAO_01631    Cas_Cas1
VKISAIDTFANMQTLYVSEQNCYVYLKQETLLVKQGETICAEVQLPLLEQIFIFGHSQLTTQVIQACLWRDIS
IAYLSSMGYCYGRILPISKGYRQLSRYQQQLSPVDKLITARAIVRGKLKNSRVLLRRQKKRRESEILERVLQT
LDYLADQAVQADTWEKLIGFEGAGAAQYFSAFSECLTNPSFVFSGRSRRPPGNPVNAMLSFGYQVLWNHLLAL
IEMQGLDPYHACLHQGNDGHAALASDLMEEFRAPLVDSLVMWLINRNIVDAESDFEFKNGGCYLNNFGRKKFL
RGFLQRMTEEIQTNPGEQQPKWDLLTQQVRAFKQFVYNPSHHYQPYRID
>NIES81_02515    Cas_Cas1
MLSFGYQILWNHLLTLIELQGLDPYYACLHQGTERHAALTSDLIEEFRAPIIDSLVLWLVNTKIMNIDNDFEY
HDSGCYLNNSGRPKYIKYFLQRLEEEVQNSEGEKQPRWDLMTQQIKAFKDFVYQPSKLYKPYQIR
>NIES81_01397    Cas_Cas1
MGTLYIVQDDAFIGKVDERINVKYEKKILQDVPFIHLEGVVVLGRATISPAVVDELMQRHIPLSFINKIGHYL
GRLEPEMTGNIFIRKAQWQAAGETPQSIHLVQGFVRGKLKNYRQTLMRYQREFNDIDLSKSIPRIEQVIAAIN
STESINSLRGLEGSGSAAYFGCFDNLIRNSQFSFTKRVRRPPTDPVNSLLSFGYSLLRHDLQSAVNIVGFDPY
LGYLHFDRYNRPSLALDLMEEFRPLVVDAVVLSILNKQLLTPADFVTEPLSGAVSLTPEGRKTFLTLYEKKKI
SEFKHPVMGRKCSYREAFELQARLLAKYLMGTIDKYPPLVLK
>NIES81_00215    Cas_Cas4+Cas_Cas1
```

```
MPTINLDLETLTKPALETLRVSSLHALAYCPRLFYLEEVEELYTQNAAVFAGRRLHTELEKQEDEEWEELFLE
SEELGLRGRLDALRTRDGQTIPYEHKRGRCHRDENKQPQAWESDKLQILAYSYLLESALGITVKEGRIRYHAD
NVLVHVPVDDSGRETVRQAIQQARKLRQSTHRPPVTDNERLCTRCSLAPVCLPEEARLAHDREWQPIRLFPED
DERLVIHVLEPGTLVGRTGEQIKITRRNQPVETIPALQVGQLVLHSFSQISTQALHFCADQEIGVHFISGGGR
YLGSFDNRNGSIQRRIRQYAALSNPDNCLELARKLVLCRGQGQRKFLMRGTRGKAEVPEKLEKAIAQMKAVLK
QVPQVQSLESLLGLEGNIAALYFSALPCLISKDVAQELHFNGRNRRPPKDRFNALLSFGYALVLKDVMNAILS
VGLEPALGFYHQPRSQAAPLALDLLEIFRVPLVDMTVISSINRGQWNPGEDFDIRGQQVWLSELGRRKFVELY
ERRKQESWKHPATGYSLTYRRLFELEVRLLEKEWTGEGGLFGHLILR
>NIES81_04594    Cas_Cas1
MQTLYVSEQNCYVCLQKETLLVKQGETIYAEVQLPLLEQILIFGKSQLTTQVIRACLWRDIPIAYLSRMGYCY
GRILPISRGYRQLSRYQQQLSPVEKLITARAMIKGKLKNSRVLLRRQRKKKESELLERVLQSLDYLADQVAQA
DNLERLMGFEGAGAAQYFSAFSECLTNPDFVFSGRSRRPPGNPVNAMLSFGYQVLWNHLLALIEIQGLDPYYA
CLHQAHDGHAALASDLIEEFRAPVVDSLVMWLINRKIMDAGADFEFKNGGCYLNDGGRKKFLRAFLQRMNEEI
QTDDIKQPKWDLLTQQVKAFKQFVYNPSHHYQPYRIN
>NIES81_04295    Cas_Cas1
MPTLYIIQPDAVLSKDYEAFQVALKQEDGSWKKQKVAAQTVDEVILMGNPQVTGDAFVYALELGMPVHYLSTF
GKYLGSALPKSSRNGQLRLAQYAVYHDPVRRLELVKAIVSAKIHNQYNVLYRHNEKDNLLKERKILIKTQQNI
DQVRGVEGLAAREYFACWQRMIGKQWTFNGRNRRPPTDPVNSLLSFAYALLQGQVMAGVHVAGLDPYIGYLHE
VHHGQPAMVLDLMEEFRALIADNLVLSLLHKHEIKPEDFNESLGAYRLKDNAKKTFLQAFDKKMNDEFKHPLF
DYRCTYRRAIELQARLLSRHLQEGIPYKPLSLR
>MBR210_00417    Cas_Cas4+Cas_Cas1
MQALELSRATEVQISTIRVSALHALAYCPRLFYLEEVEELYTQDAAVFAGRRLHTQLEKQEDEEWEELFLESE
ELGLRGRLDALRTRDGQTIPYEHKRGRAHRDENNQPQAWESDRLQILAYAYLLESAHGITVKEGRIRYHADNV
IVHVPLDEAGRLKVKEAIQQARILRQSTYRPPVIDNERMCARCSLAPVCLPEEARLAHNREWQPIRLFPEDDE
RQVIHILEPGTAVGRTGEQIKITRRNQPLETVPARQVGQIVLHSFSQISTQALHFCAQQGIGVHFISGGGRYL
GSFDTRQGSIQSRIRQYGALSNPDACLELARKLVICRGQGQRKFLMRGARGKTKTDKLEKAIAQMKAVLKQVP
QAKSLESLLGYEGNLAALYFAALPDLISPEVPKELHFAGRNRRPPLDRFNAMLSFGYALLLKDVMNAILTVGL
EPALGFYHQPRSQAAPLALDLLEIFRVPLVDMTVMAAVNRGQWDAKADFEIRGKQVWLSELGRRKFVELYESR
KQESWKHPVTGYSLTYRRLFELEVRLLEKEWSGEGGLFGQLILR
>JSC1_01058 Cas_Cas4+Cas_Cas1
MNSSELPDSGQETLRVSALHALAYCPRLFYLEEVEELYTQDEAVFAGRRLHVEIEKEEGDVYESLMLESEALG
LRGKVDALRTRSGQTIPYEHKQGRCRRDEQNQPQAWDSDRLQILAYACLIEAALGIMVTEGRIRYHADNVLVK
VPLDEQGRQQVRQAIEQARQLRRSTERPPVTSNERLCVRCSLAPVCLPEETRFAQAEATQPMRLFPEDDDREI
VHVLEPGACVGRSGDQIKISRKDEPTITLPARQVGQIVLHSYSQISTQALYFCAEQQIGVHFISGGGRYLGSL
NTRQGSIQRRIRQYRALSDSEMSLKLARQLVLCRGQSQRRFLMRGQRNLNQAGEPLQQTIDQMERLLRLVPQA
ESLERLLGLEGNLASLYFAALPCLIGDNVPAELRFIRRSRRPPQDRFNALLGFGYALLLKDVVNSILVVGLEP
SLGFYHQPRTQAPPLGLDLMEIFRVPLVDMTVMGAVNRGQWDIWEDFEVRGHRVWLSDSGRRKFINLYERRKE
ETWKHPILKYSLTYRRLIELEVRLLEKEWMGEGNLFGHLVIR
>JSC1_06626 Cas_Cas1
MSILYITQPDAVLSKSHEAFTVALKQEDGRWQKRAVPAQTVEQVVLMGNPQVTGDALVYALELGMPVHYLSSF
GTYLGSALPGASRNGQLRLAQYALYLDGTRRLDLVKAIIAAKIHNQYTVLYRHGQTENPLKLRKKQVEQQTTL
DKARGIEGMAAREYFALWQAMLGDSWSFTGRNRRPPTDPVNSLLSFAYSLLRGQVTTAVHLAGLDPYIGYLHE
VHHGQPALVLDLMEEFRPLVADNIVLAVLNNRELQPKDFTESLGAYRLSDDGRKRFLQAFERKMTSEFTHPVF
GYRCTYRRAIELQARLLGRHLQEEIPYKPLILR
>JSC1_06420 Cas_Cas1
MQTLYVSQQGCYVSLRQESLVIKQGQAVLEEVQLPLLEQVLIFGRSQMTTQAIRACLWREIPVVYLSRLGWCY
GRVLPISRGYRQLGRYQQELAFGERIVAARAIVLAKLKNSRVMLQRQQRRRGTARIEAAIEQLEHLSTQVNRA
VTLEVLMGLEGAGASSYFGAFGDCLVNPEFGFVMRSRRPPADPVNALLSFGYQMLWNHLLSLIELQGLDPHEA
CLHQGSERHAALASDLLEEFRAPIVDSLVLYLVNRSVVKPEDFEYRNEGCFLNELGRKKYLRAFVGRMEELLQ
TEEGMQPRWDLLMQQVKAFKRFVYSPAQGYQPYLIR
>VB61278_05635   Cas_Cas1
VWEVLPPSNEKQSISEVVSSLEPENEDIAFLKNVATLHQPFIAEIEIESPLQTEEDNSPFQKNIWHREMAAIY
LIEQGTNIYKDYQRFIIHVSEKPKLEVPIREVQQILVFGNIQLSTPVIQACLQEQIPVLFLSQSGQYHGHLWS
EESTNLDNQLIQIERRQDDYFQFQVSRAIVFGKLMNSKQLLLRFNRKRKVTEVEKAIFGISQDIDSLDLVDNL
DTLRGYEGIAAARYFPAFGQLITNSKFEFSLRNRQPPTDPVNSLLSFGYTLLFNNVLSFLVAEGLSPYLGNFH
YGEKQKPYLAFDLMEEFRSPIVDSLILKILNNGEFKPKDFDVVASTGGVYLSQSARKVFLKQFENRMNEEISH
PDLISPVTYRHAIQLQVRRYKRCLLSGVPYESFLRAG
>VB61278_04582   Cas_Cas1
```

```
MGTLYITQDDAFIGKIDERLHVKFDKKTILDVPLLKIDGVVVLGRATVSPAAINELLQRQIPLTFLTETGRYL
GRLEPEVTKNIFVRKAQWQAAGDTAQAIHVVQGFVRGKLKNYRSLLVRRQRECSGLDLSNSIARLDQAIAPID
TTHNIDSLRGLEGAGSAAYFGCFDQMIRNTGFTFTKRVRRPPTDPVNSLLSFGYSLLCHDVQSAVNIAGFDSY
LGYLHCDRYGRPSLALDLMEEFRPLVVDAVVLSALNKQFLKVEDFVTESLSGAVSLTNEPRKTFLRLYGQKKL
SEFKHPVLGRKCTYQEAFELQARLLAKYLMGEIEKYPPLVLK
>PCC6313_00038   Cas_Cas1
MRSLYVSQQGCTVTLTQETLIVATQGAVITTVPLPLLEQVLIFGQSQVTTQAIRACLSRDIPIVYLSRLGYCY
GRTLALQRGYRALARQQQALTESERLQVARAIAQAKLANCRTLLLRQQRRRSLDLDLAIQTLAYLADQVSTAT
SAAQIMGYEGAAAGSYFAAFGQCLQGSEFAFLARSRRPPRNPVNALLSFGYMVVWNHLLTLIELQGLDPYAGC
LHSGSDRHAALASDLIEEFRAPIVDSLVLQVVNGRMVTSEDFDYRDGGCFLGATGRRTFLRSLLQRMEERLSD
GEPRWDLLSRQVRQFKQFVYHPSQGYSPYRIR
>PCC6313_00021   RVT_1+Cas_Cas1
LFKGIAPEQLSHIQTQLNAATYGLSPVRGIHLRKENGGKRLIGIPTVRDRIVQRWLLEDLYPELEHQFADCSY
AYRPGRSIQQAVQHLYLHYQYQPAWVIKADITQFFDSLCWGLLMNSLEQMELTQNHLALLHQSLTVPVILRGQ
RLRRTQGVVQGSVLSGALANLYLSEFDARCLQQGMNLVRYGDDFVVVCHSAEEARIRLAQIQDGLTQLYLRLN
SEKTAIAPPDQEFTFLGHRFVNGKVYAPPPPSPLKLRDGYWVLSPSGVPRWRIRRPERLLPAPPKVCALRKQP
PPSVVANPNHLFSDLMTTLYITDQGAYLKAQQYQFQVFHRQELRCKIPINQVSHVVLFGCCNLSHGAVRLALA
RRIPVLYLSQRGRYFGRLQTDGQAKVNYLSRQVEWSQDEEQVRNTAETIIRAKLQNSRVLLQRLNRRRATDAA
QQAIAQLAIAMDNLPLAESLEALRGMEGNAAKLYFQGLGSLFKGDLSFSKRSKRPPTDPVNSLLSLGYTLLHQ
NLHSFIETLGLHPHFGNLHVPRDNHPALVSDLVEEFRALVVDSLVSYVVNKNIITAEDFTPADERGGVFLQPD
ALKRFLKHWEEKLHSEVTHPTTGHQVSLRRCFELQVREYIAALLGDVPAYRPMLWGK
>PCC6313_00010   Cas_Cas1
MRSLYVSQQGCTVTLTQETLIVATQGAVITTVPLPLLEQVLIFGQSQVTTQAIRACLSRDIPIVYLSRLGYCY
GRTLALQRGYRALARQQQALTESERLQVARAIAQAKLANCRTLLLRQQRRRSLDLDLAIQTLAYLADQVSTAT
SAAQIMGYEGAAAGSYFAAFGQCLQGSEFAFLARSRRPPRNPVNALLSFGYMVVWNHLLTLIELQGLDPYAGC
LHSGSDRHAALASDLIEEFRAPIVDSLVLQVVNGRMVTSEDFDYRDGGCFLGATGRRTFLRSLLQRMEERLSD
GEPRWDLLSRQVRQFKQFVYHPSQGYSPYRIR
>PCC6313_00829   Cas_Cas1
MTTLYLLQPDAILSKKHEAFAIALKQPDDTWLNRSVAAQTIEQIILMGHPTITGEALTTALDLGIPVHYLSSF
GKYLGSALPQQSRNGQLRLAQYGLYHDPIPRLHSVQAIVRAKIHNQALVLAKHATSADRLRHHKAQIKSQASI
DQVRGIEGIAARDYFAGWQAMLDPVWQFTGRNRRPPQDPVNALLSFAYSLLQVQVMAAVHLVGLDPYIGYLHT
VHHGQPALVLDLMEEFRPLVSDRLVLSVINRQEIKPDDFTTTLGACRLADRARSRFLQAIEQKFNETFTHPVF
DYQCTYRRAIELQARLFARHLQEDIPYKPLQLR
>Ana39858_03619  Cas_Cas1
MGTVYVTQDDSFISKIDERLNVKFEKKVILDVPLIKIDGLVVMGRASISPAAIFELIDKKIPLTFLTNNGKYL
ASLEPEMGKNIFVRAAQWKAAGESDQAIHVTQGFVRGKLKNYRHSLLEAQRRYEVDLNSNISQLSHAITSIDK
ANSINTIRGLEGAGSAAYFGCFNQLIRVDNFSFHTRNRRPPIDPVNSLLSLGYSLLRHDIQGALNIVGFDPYL
GYLHTERYGRPSLALDLMEEFRPLIVDAVVLTAINRRMLSPKDFVTEPISNAVSLTKEGLHIFLRLYQEKKQT
QFKHPVMQKKYTYQETFEIQARLLAKYLMGELDKYPPLVMR
>CS951_04287     Cas_Cas1
MGTIYIIQEDAFIGKTDERLIVKVGKEKHIDVPLIKVEGIVILGRATVSPAVVAELLQRHIPITFLTKTGRYL
GRLEPEVTKNIFVRKSQWLAAGESPQAVHAVRGFVRGKLKNYRNTLLRRSREYSELDLQTKITQIEQAIAPIN
KTENINSLRGLEGAGSAAYFGCFNELINNSAFFFQTRNRRPPLDPVNSLLSLGYALLRNDVQSAVNIVGFDPY
LGYLHVERYGRPSLALDLMEEFRPLIVDSVVLSTLNKKLLNPEHFITEPLSHAVSLTPEGLKIFLQLYEHKKQ
SKFTHPVLKRQCTYQESFEIQARLLAKYLMEETNQYPPLVLR
>CS951_03879     Cas_Cas1
MSVLYVTQPDAILNKKYEAFIVSLKQEDGTWSKHSVAAQTVEQVVLMGNPQITGDALSYALELGMPVHYLSQY
GKYIGTTLPGYSRNGQLRLAQYATHCHEEKRLELVKIIVAGKIHNQSSVLYRYGQKDNPLKLRKKQVCEQKTL
DQVRGIEGIAAREYFACWLNILDEPWSFQGRHHRSPSDPMNVLMNFAYGLLRIQVTAAVHIAGLDPYIGYLHE
TTRGQPAMVLDLIEEFRPLVADSFLLSLLSHKELKLSDFSESLGAYRLKDAGRKIFLEAFERKLNSEFKHPVF
GYRCTYRRAIELQARLLARHLQENVVYKPLKIR
>PCC9631_01765   Cas_Cas1
MSILYITQPQAILGKKYEAFTVATQEESGKWNKQSVPAQTVEQVVLMGNPQVTGDALTYALELGMPIHYLTSF
GKYLGSALPGYSRNGALRLAQFELYQDTERRLELVKTIVSAKIQNQATVLYRQKIADSPLKDHKKLVFKKATL
DEVRGVEGIAAKDYFREWSAILNEPWFFPGRNRRPPTDPVNALLSFAYGLLRVQVTAAVHLVGLDPYIGYLHE
VSRGQPSMVLDLMEEFRPLVADSLVFSVLNKREIQPEDFTESLGAYRLKDEGRKRFLQAFERKMNDEFKHPVF
EYRCTYRRAIELQARLLGRYLQEDLVYKPLVLR
>MVCC14_00299    Cas_Cas1
```

```
MTILFLLEQDTLVRQTDERLEIFKYDKRLTDVPLCKLEAVVVFGKVILTIPVLKVLNERSIPITYLSEEGRTV
ATLLPEPNPNAILRAKQYQASFDQRKTLDIAREFVRGKLFNQHTVLARFSRQNERSERIGVALKSLKSCQKSL
DQTTSLNELRGYEGQGASSYFGVLGELLTDTPWEFSHRTRRPPTDPVNALLGFGYALLYGDCRAALHAVGLDP
YQGFLHGERYGRANLALDLMEEFRPIFVDGLVLQLLRNNGLEQDCFRNYPGGAVYLNEEGIKKFIQSYEQRKN
TTFQHPVFEIQTDYRRALILQSRLLAKHLNQEIDNYLPLKIR
>MVCC14_02328    Cas_Cas1
MQTLYISEQNCYVCLQKETLIVKQGDTVHVEVQLPLLEQILIFGRSQITTQVIRACLWRDIPIAYLSRMGYCY
GRILPISRGYRQLSRYQQELSPVDRLITARAIVKGKLKNSRILLRRQRKKRESETLERTLQSLDYLAEQAVQA
DTWERLMGFEGTGAAQYFSVFSECLTNPDFVFSGRSRRPPGNPVNAMLSFGYQVLWNHLLALIEVQGLDPYYA
CLHQANDGHAALASDLIEEFRAPLVDSLVMWLINRNIMDVKEDFQFKNGGCYLNDSGRKKFLRGFLQRMTEEI
QTNDGKQPKWDLLTQQVRNYKQFVYNPSHHYKPYQID
>WA113_01966     Cas_Cas1
MQTLYVSEQNCYVCLQKETLLVKQGDTVYIEVQLPLLEQILIFGQSQITTQVIRACLWRDIPIAYLSRMGYCY
GRILPISRGYRQLSRYQQQLSPIDRLITAREIVKGKLKNSRVLLRRQQKKRDLELIERVLDSLDYLATQAATA
DNLERLRGFEGAGAAQYFSAFGECLTNSDFVFSGRSRRPPGNPVNAMLSFGYQVLWNHLLALIEIQGLDPYYA
CLHQGNDGHAALASDLIEEFRAPLVDSLVMWLINRKIMDVKDDFEFKNGGCYLNNLGRKKFLRGFLQRMTEEI
KTDQEIKQPKWDLLTQQVKAFKQFVYNPSHHYKSYRID
>PCC9333_00876   Cas_Cas1
LRGLEGAGSAAYFGSFEKLIRAEGFTFEARRRRPPTDAINSLLSFGYSLLAHDIESAVNIVGFDPYLGYLHTE
RYGRPSLALDLMEEFRPLVVDAVVLSVVNKRSLVPSDFTAEPISGAVYLSEEGRRTFLRMYEQKKQTKFKHPV
LQKQCTYQEAFEIQARLLGKYLMDETELYPPLVLK
>PCC9333_00875   Cas_Cas1
MGTVYISQEDSFIGKTDERLQVKAEKKTLLDVPMIKIDGVVVMGRSTISPAALMELLNRKIPLTFLTRTGSYL
GRLEPPLTKNIFIRSAQWKTRENAAKVTHVVRGFVRGKLKNYRNSLLRAQREYESLNFQNAITRLDQLLRRSI
KLKVLIHCED
>PCC9333_00607   Cas_Cas1
MATIYLLEPGSTLYKEHLRFVIWVPDEQKIEVPIREVERILVFGNIQVTTQAINACLEQQILVLFLSTTGQYK
GHIWNWESNHLGNELIQIEKRQDPQFQFLVSQAIVRGKLANSRQLLMRLNRKRKVAEVSQAIAGIGTDISALD
TVNNLDSLRGYEGVGAARYFPAFGQLITNPEFSFSLRYRQPPTDPVNSLLSFGYTLLFNNVLSLIVAEGLSPY
FANLHYGEDKKPYLAFDLMEEFRSPIVDSLVLKLINSAVFKSRDFETVVSTGGVYLRKESRRIFLQHFEKRMN
EELSHPDLLSQVSYRHAIQLQIRRYKRSLLHSIAYEPFLRTA
>AWQC310F_01824  Cas_Cas1
MQTLYVSEQNCYVCLQKETLLLKQGDTVYIEVQLPLLEQILIFGQSQITTQVIRACLWRDIPIAYLSKMGYCY
GRILPISRGYRQLSRYQQQLSPIDRLITARELVKGKLKNSRVLLRRQQKKRESELIQRVLESLDYLAAQAATA
DNLERLMGFEGAGAAQYFSAFGECLTNSDFVFSGRSRRPPGNPVNAMLSFGYQVLWNHLLALIEIQGLDPYYA
CLHQGHDGHAALASDLIEEFRAPLVDSLVMWLINRKIMDVKADFEFKNGGCYLNNLGRKKFLRGFLQRMTEEI
KTDQEIKQPKWDLLTQQVKAFKQFVYNPSHHYQCYRID
>AWQC310F_02850  Cas_Cas1
MGTLYITQDDAFIGKVDERIHVKFEKKILQDIPFIHLEGVVILGRATISPAVVFELMERHIPLSFISNIGHYL
GRLQPEMTGNIFIRKAQWQAAGETPKSIHAVQGFVRGKLKNYRQTLMRYQREFSDLDLSKDILRIEQVITAIN
STENISSLRGLEGSGSAVYFGCFDRLIRNSQFSFTKRVRRPPTDPVNSLLSFGYSLLRHDLQSAVNIVGFDPY
LGYLHFDRYNRPSLALDLMEEFRPLVVDAVVLSILNKQLLTPADFVTEPLSGAVSLTPEGRKTFLTLYEKKKL
SEFKHPVMGRKCTYREAFELQARLLAKYLMGTTDKYPPLILK
>PCC9006_00139   Cas_Cas1
MRTLYLSRQGCFVSLQQEMLVVKQGDRTLEEVRLPLLDQVLVFGQSQLTTQAIRACLWRNIPIAYLSRMGYCY
GRVVAIEGGYRQLARYQRELTALERLLLAQALVKAKLLNSRVILQRQQRRRSSLMLETAIEQLAYLIHQISDT
DQIDRLMGLEGAAAAQYFQGFGDCLRHADFAFVSRSRRPPGNPVNALLSFGYQVLWNHLLTLVELQGLDPYEG
CLHQGSYRHAALVSDLIEPFRAPLVDSLVLQLINNGELDAQGDFEYRDGGCFLGERGRRKWLGAFVQRMEMPV
QSEEDAAQPSWDLLNRQVRAYKEFVYSPVQGYAPYRIR
```

Table S2. Desoxyoligonucleotide primers used in this work. The sequences are given in 5’-3’ direction. Bold and underlined nucleotides correspond to the T7 promoter.

Primer name	**Sequence (5’ to 3’)**
CR_1_fwd	CCCGCTGACCCGCGCGTCC
CR_1_rev	**TAATACGACTCACTATAGG**CGGCGGCACCGCGGCCTG
CR2f	TGATTTTCAACCACTTCGGGTTGAG
CR2r_T7	**TAATACGACTCACTATAGGG**AATTATCCCCCAAGACCAATCCCC
CR_3_fwd	TATCAGAAATTGTGTGTTTGGGTCTAACTTATTAAAAAG
CR_3_rev	**TAATACGACTCACTATAGGG**CTAGAATTCCAGGAATAAATTTTCTAGGAAGCGT
CR_4_fwd	CGGGCAATCTTAAAATACCCTGACCTGAA
CR_4_rev	**TAATACGACTCACTATAGGG**TAACTTCCATTAGCTTGGCAGGCATTAGC
CR6f	TCGCTTATACAACAAAGGGATCTTC
CR6r_T7	**TAATACGACTCACTATAGGG**TTGAGGCTAGATTATTTATTTTTGATGCTG
CR_67_fwd	CTTACTGATGCCACATCAATCGCTGTAGA
CR_67_rev	**TAATACGACTCACTATAGGG**AGCGATTAGATTGCTGGGGTTAGACATCA
CR_8_fwd	GGGATGGGGCAAGTGTGGGTGAA
CR_8_rev	**TAATACGACTCACTATAGGG**GTAGAGACTGGTTATTAGGTAAAATAGGTCTTTACTT
CR_9_fwd	TTTGAATATTCAGAACTTTATATTGTGCGCGAT
CR_9_rev	**TAATACGACTCACTATAGGG**GCAAGCTGATTTGGTAGAAGCTGTTAAT
CR_10_fwd	CCTCTCTATACAAGCGTTGTAGAAGTGAAC
CR_10_rev	**TAATACGACTCACTATAGGG**TCTTGCATACAAAGCTGCATTTCTAGATGACAA
CR_11_fwd	CCAAATACAGCAAGGCTTTGAGAGCCGAG

CR_11_rev	**TAATACGACTCACTATAGGG**AAGTGGTGGCAGCGCTGCTAGGATTC
5S rRNA	TAGCAGCGTTTCACCTCTGAGTTCGG