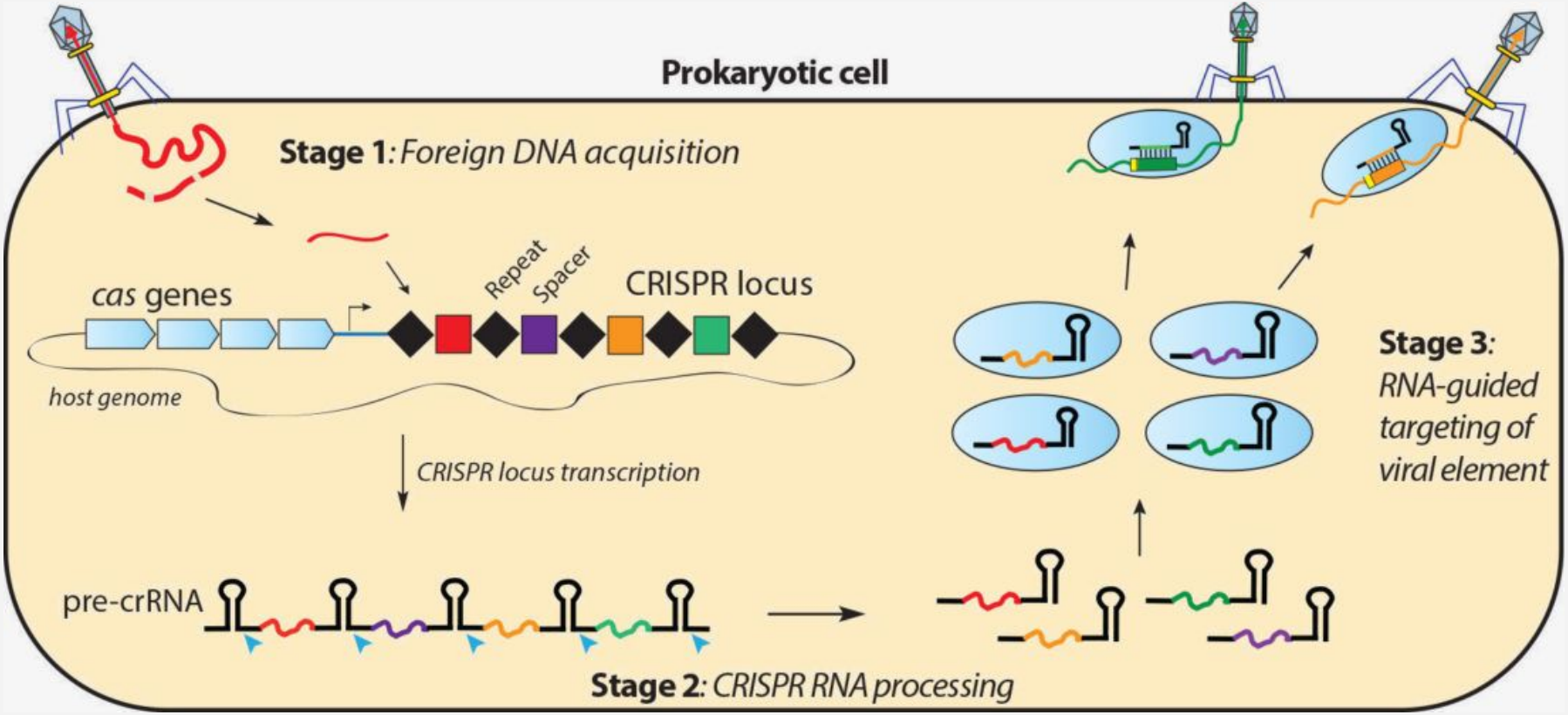
Diversity and evolution of class 2 CRISPR–Cas systems

DOI: 10.1038/nrmicro.2016.184

CRISPR–Cas systems provide adaptive immunity in archaea and bacteria

Adaptation	**Expression and processing**	**Interference**
spacer acquisition	maturation of pre-crRNA	DNA target cleavage



http://doudnalab.org/wp-content/uploads/2018/02/crispr-pathway-1100x501.jpg

Classification of CRISPR-Cas systems

Class 1 CRISPR–Cas systems

- multisubunit effector complexes

  4–7 Cas protein subunits in an uneven stoichiometry

- most common in bacteria and archaea
- ~90% of CRISPR–cas loci

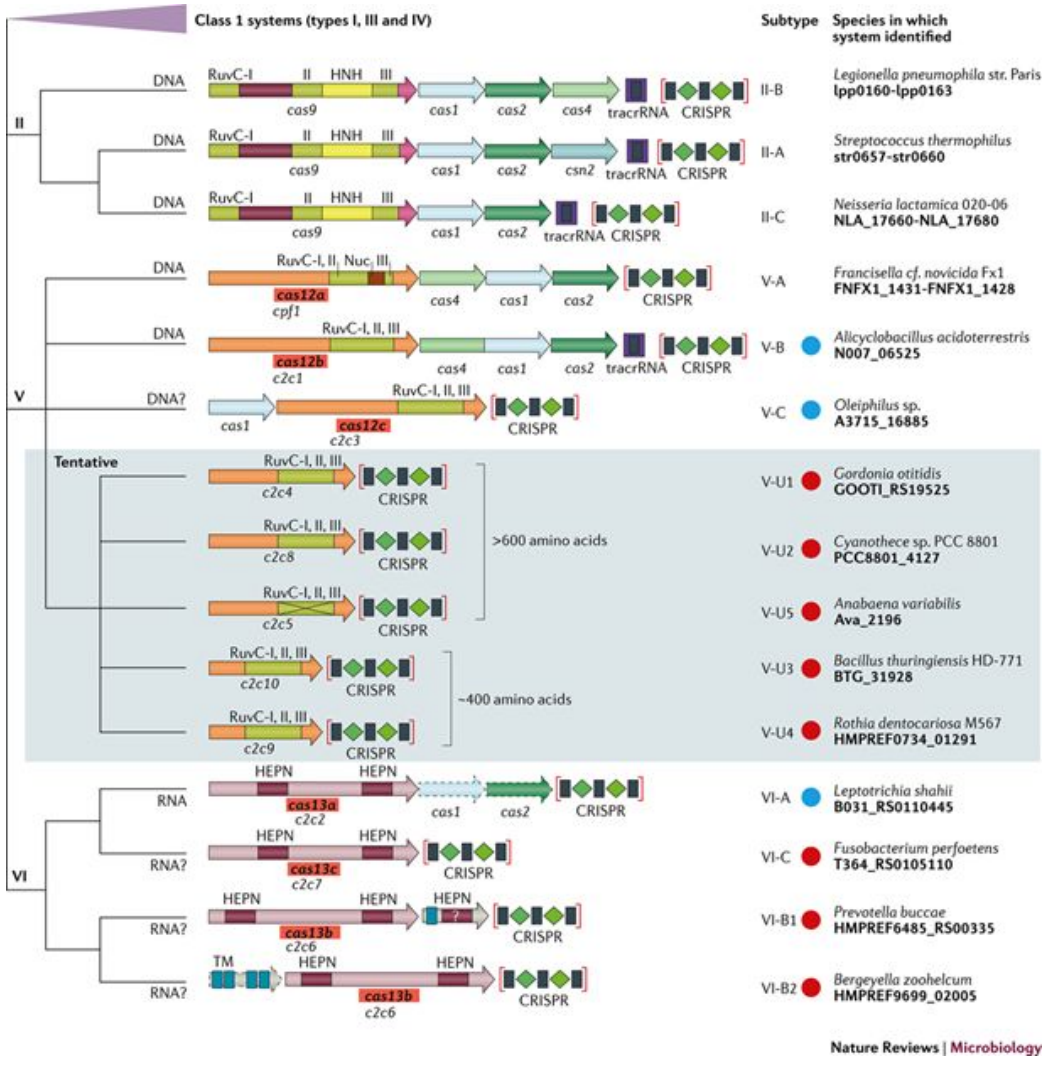
Class 2 CRISPR–Cas systems

- type II, type V or type VI effector protein

  single, multidomain protein

- almost exclusively in bacteria
- ~10% of CRISPR–cas loci

The updated classification scheme for class 2 CRISPR-Cas systems



Type II and type V

The distinctive feature **of type II and type V** CRISPR–Cas sequences is **the presence of a RuvC-like nuclease domain** in their multidomain effector proteins

RuvC-I, RuvC-II and RuvC-III are the three distinct motifs that contribute to the nuclease catalytic centre

In **the type II** effector Cas9, the RuvC-like domain contains an inserted **HNH nuclease domain**

All of the **class V effectors** that were identified at this stage share **a similar, large size** (typically, 1,000–1,300 amino acid residues) and **a single common domain**, the RuvC-like endonuclease domain, although **the sequence similarity between** the effector proteins of different subtypes **is extremely low**

The search for homologues of the type II and type V effectors showed that the **RuvC-like nuclease domains are related to TnpB proteins**, an extremely abundant but poorly characterized family of nucleases that are encoded by many autonomous and even more numerous non- autonomous bacterial and archaeal transposons

The closest relatives and possible ancestors of Cas9 were identified on the basis of readily detectable sequence similarity and on the presence of the HNH insert in the RuvC-like nuclease domain of a distinct family of TnpB proteins that was denoted **IscB** (insertion sequences Cas9-like protein B). It is difficult to confidently trace a direct connection between type V effector proteins and a particular group of TnpB proteins, because type V effector proteins show less similarity to TnpB proteins than Cas9 shows to IscB proteins

Subtype V-U

The search for CRISPR–cas loci that lack the adaptation module yielded several additional variants of putative type V systems.

The putative effector proteins of these loci, which we have provisionally assigned to subtype V-U (*U - uncharacterized*) share two features that distinguish them from type II and type V effectors that are found at CRISPR–cas loci that contain Cas1:

1) First, these proteins are **much smaller than class 2 effectors that contain Cas1**, comprising between ~500 amino acids (only slightly larger than the typical size of TnpB) and ~700 amino acids (between the size of TnpB and the typical size of the bona fide class 2 effectors)
2) Second, these putative effectors show **a higher level of similarity to TnpB** proteins than the larger type I and type V effectors

The subtype V-U TnpB-like proteins are too small to adopt a bilobed structure of sufficient size to accommodate the crRNA–target DNA complex, as the typical class 2 effectors do, and, therefore, are **unlikely to function in that capacity without additional partners**

Furthermore, the subtype **V-U loci lack any additional cas genes**, which, together with the above structural considerations, calls for caution in predicting that they have fully fledged CRISPR activity.

Type VI

The signature of **type VI** systems is the presence of an effector protein that contains **two HEPN domains** (*higher eukaryotes and prokaryotes nucleotide-binding domains*)

HEPN domains are common in various defence systems, the experimentally characterized of which, such as the toxins of numerous prokaryotic toxin–antitoxin systems or eukaryotic RNase L, all have RNase activity

Therefore, the first putative type VI effector, denoted C2c2, was predicted to function as an **RNA-guided RNase**

In addition, a novel feature of C2c2 is that, **once primed with the cognate target RNA, the effector becomes a promiscuous RNase** that has a toxic, growth-inhibitory effect on bacteria

These findings demonstrate **a coupling between adaptive immunity and programmed cell death**

More recently, the C2c2 protein was shown to mediate not only interference but also the processing of pre-crRNA

The search for CRISPR–cas loci using the CRISPR seed identified two additional large putative effectors that contained two HEPN domains and which we assigned to **subtype VI-B** and **subtype VI-C**, respectively (accordingly, the **C2c2-encoding loci** became **subtype VI-A**). This classification of the type VI systems into separate subtypes is justified by **the extremely low sequence similarity between the three groups of effectors**, which is practically limited to the catalytic motif of the HEPN domain, **the different positions of the HEPN domains with the large protein sequences**, and **the additional features of the locus architecture in the case of subtype VI-B**

Moreover, given that RNA viruses only represent a minor part of the prokaryotic virome, **type VI systems might primarily elicit toxin activity in response to the active transcription of foreign DNA**

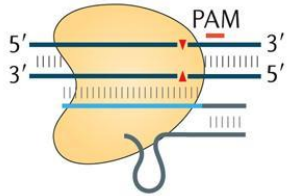
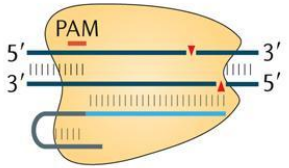
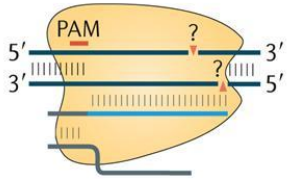
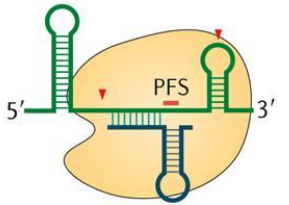
A comprehensive census of class 2 CRISPR–Cas systems in bacterial and archaeal genomes

	Subtype						
	II	V-A	V-B	V-U*	VI-A	VI-B	VI-C
Effector[a]	Cas9	Cas12a (Cpf1)	Cas12b (C2c1)	C2c4, C2c5; five distinct subgroups (V-U 1–5)	Cas13a (C2c2)	Cas13b (C2c6)	Cas13c (C2c7)
Number of loci in bacterial and archaeal genomes	• 3,822 in total • 2,109 II-A • 130 II-B • 1,573 II-C • 10 unassigned	70	18	92	30	94	6
Representation	Diverse bacteria	Diverse bacteria and two archaea	Diverse bacteria	Diverse bacteria	Diverse bacteria	Bacteroidetes	Fusobacteria and Clostridia
Other cas genes	85% *cas1* and *cas2*; 55% *csn2*; 3% *cas4*	70% *cas1* and *cas2*; 55% *cas4*	65% *cas1*, *cas2* and *cas4*	None	25% *cas1* and *cas2*	None	None
Percent of loci that contain CRISPR array	65%	68%	60%	~50%	73%	90%	83%

Applications in genome engineering: **Cas9**

Cas9	**dCas9**	*Challenges*
genome editing ➔ gene knockout ➔ precise editing	➔ transcriptional control ➔ epigenetic modulation ➔ imaging	➔ the potential for off-target effects ➔ delivery ➔ the difficulty of targeting RNA rather than DNA

Functional diversity of the experimentally characterized class 2 CRISPR-Cas systems

		Nuclease domains	tracrRNA	PAM	Substrate	Cleavage pattern
Type II Cas9		RuvC and HNH	Yes	3′, GC-rich	dsDNA	Blunt ends
Type V-A Cas12a (Cpf1)		RuvC and Nuc	No	5′, AT-rich	dsDNA	Staggered ends, 5′overhangs
Type V-B Cas12b (C2c1)		RuvC	Yes	5′, AT-rich	dsDNA	Staggered seven-nucleotide cut of target DNA
Type VI-A Cas13a (C2c2)		2 HEPN domains	No	5′, non-G PFS	ssRNA	Cleaves ssRNA near uracil and collateral activity

Applications in genome engineering: **Cas9 alternatives**

Cas12a (Cpf1, V-A)	**Cas13a** (C2c2, VI-A)	**dCas13a**
genome editing ➔ simpler, single RNA-guided and more specific enzymes than Cas9 ➔ alternative PAM (editing in AT-rich genomes, such as the genome of *Plasmodium falciparum*	new RNA-guided RNA-targeting technologies ➔ perturbation, modulation, modification and monitoring of specific RNA transcripts in cells the selective ablation of cell types based on expression profiles	new tools for RNA-biology ➔ sensing of different cellular states ➔ manipulation of translation ➔ tracking of RNA levels and localization in live cells