Diversity, classification and evolution of CRISPR-Cas systems

Eugene V Koonin[1], Kira S Makarova[1] and Feng Zhang[2,3,4]

The bacterial and archaeal CRISPR-Cas systems of adaptive immunity show remarkable diversity of protein composition, effector complex structure, genome locus architecture and mechanisms of adaptation, pre-CRISPR (cr)RNA processing and interference. The CRISPR-Cas systems belong to two classes, with multi-subunit effector complexes in Class 1 and single-protein effector modules in Class 2. Concerted genomic and experimental efforts on comprehensive characterization of Class 2 CRISPR-Cas systems led to the identification of two new types and several subtypes. The newly characterized type VI systems are the first among the CRISPR-Cas variants to exclusively target RNA. Unexpectedly, in some of the class 2 systems, the effector protein is additionally responsible for the pre-crRNA processing. Comparative analysis of the effector complexes indicates that Class 2 systems evolved from mobile genetic elements on multiple, independent occasions.

Addresses
[1] National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD 20894, USA
[2] Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[3] McGovern Institute for Brain Research at MIT, Cambridge, MA 02139, USA
[4] Departments of Brain and Cognitive Science and Biological Engineering, Cambridge, MA 02139, USA

Corresponding author: Koonin, Eugene V (koonin@ncbi.nlm.nih.gov)

Introduction

Thanks to the unprecedented utility of the engineered Cas9 endonucleases harnessed as genome editing tools, comparative genomics, structures as well as biochemical activities and biological functions of CRISPR-Cas systems and individual Cas proteins have become a subject of intensive study [1–4]. The CRISPR-Cas is a bona fide adaptive (acquired) immunity system with immune memory (Lamarckian-type inheritance) that is stored in the form of spacer sequences derived from foreign genomes and inserted into CRISPR arrays. CRISPR-Cas is a programmable form of immunity that can adapt to target any sequence and therefore is not under pressure to evolve an immense diversity of specificities as is the case for restriction-modification enzymes, the most abundant form of innate immunity in prokaryotes [5]. Nevertheless, like any defense system, CRISPR-Cas is engaged in an incessant arms race with viruses, which results in rapid evolution of *cas* genes [6] and notable diversity of the gene repertoires and architectures of the CRISPR-*cas* loci translating into diversification of the actual defense mechanisms [7,8••]. More specifically, the diversification of the CRISPR-Cas systems is likely to be partly driven by their competitive coevolution with virus-encoded dedicated anti-CRISPR proteins [9•,10•,11•,12].

Despite this extensive diversification, comprehensive comparative analysis has revealed major unifying themes in the evolution of CRISPR-Cas. These common trends include multiple contributions of mobile genetic elements to the emergence of the CRISPR-Cas immunity and its distinct variants; extensive duplication of *cas* genes yielding functionally versatile effector complexes; and modular organization, with extensive recombination of the modules [4,8••,13]. The two principal modules of the CRISPR-Cas systems consist of the suites of genes encoding proteins involved in adaptation (spacer acquisition) and effector functions, that is, pre-crRNA processing, and target recognition and cleavage. The adaptation module is largely uniform across the diversity of the CRISPR-Cas systems and consists of the endonuclease Cas1 and the structural subunit Cas2 [14]. In contrast, the effector modules are highly variable between CRISPR-Cas types and subtypes. Various proteins involved in ancillary roles, such as regulation of the CRISPR response and other, still poorly characterized functions, can be assigned to a third, accessory module [4,8••,13].

In this brief review, we discuss the recent advances in the characterization of the CRISPR-Cas diversity, which result in rapid expansion of the CRISPR-Cas classification, the distinct structural and functional features of the newly discovered CRISPR-Cas variants, and the emerging evolutionary insights.

Exploration of CRISPR diversity and the evolving classification of CRISPR-Cas systems

The fast evolution and variability of the CRISPR-Cas systems makes their classification a daunting task. Given the absence of universal *cas* genes and the frequent

modular recombination, a single classification criterion is impractical. Therefore, a multipronged approach to CRISPR-Cas classification has been adopted that takes into account the signature *cas* genes that are specific for individual types and subtypes of CRISPR-Cas, sequence similarity between multiple shared Cas proteins, the phylogeny of Cas1 (the best conserved Cas protein),

the organization of the gene in the CRISPR-cas loci and the structure of the CRISPR themselves [8**,15]. The combined application of these criteria resulted in the currently adopted classification scheme that partitions the CRISPR-Cas systems into two distinct classes characterized by different principles of the effector module design (Figure 1).

Figure 1
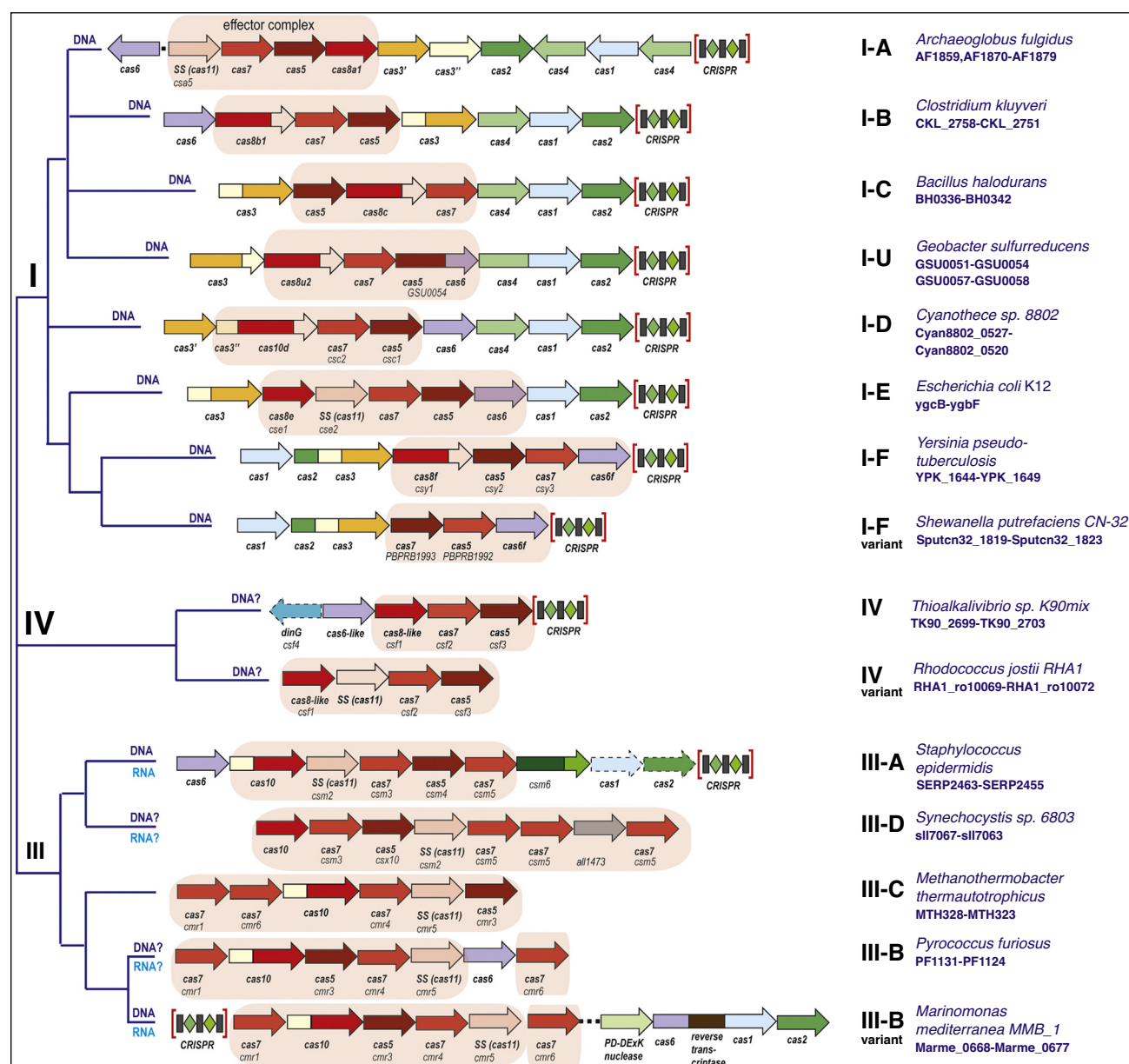
Updated classification of CRISPR-Cas systems.
(a) Class 1
(b) Class 2
The organization of the CRISPR-*cas* loci and domain architectures of the effector proteins as well as the (predicted) target (DNA or RNA, or both) are shown for each subtype. For subtype III-D, a locus with reverse transcriptase fused to *cas1* is included; other reverse transcriptase-containing variants, from subtypes III-A and III-D, are not shown. The genes for Class 1 effector subunits are shaded. SS, small subunit; TM, predicted transmembrane segment.

Figure 1 (*Continued*).

Current Opinion in Microbiology
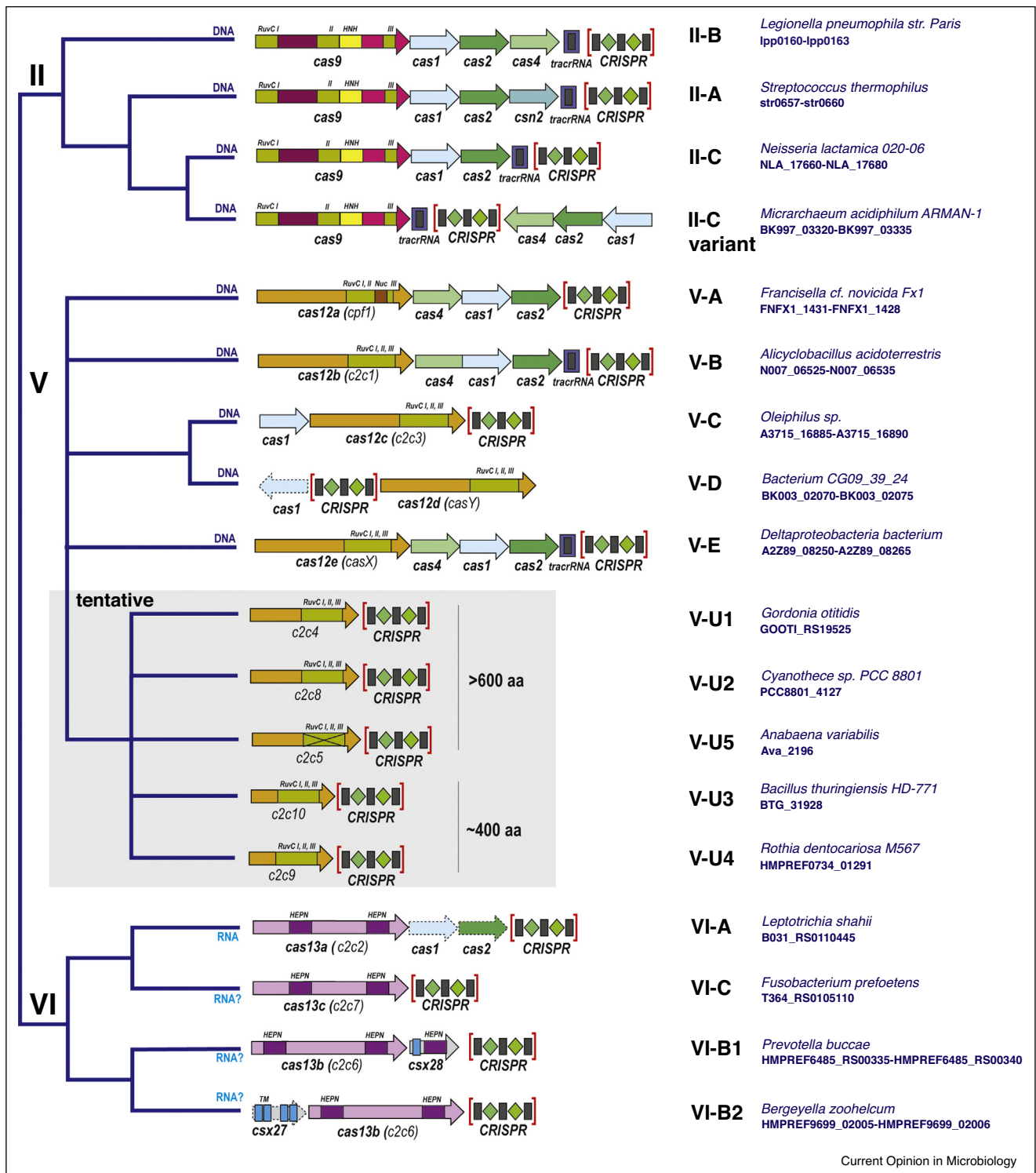
The Class 1 systems include the most common and diversified type I, type III that is represented in numerous archaea but is less frequent in bacteria, as well as the rare type IV that includes rudimentary CRISPR-*cas* loci lacking the adaptation module (Figure 1a). The effector complexes of type I and type III CRISPR-Cas display elaborate architectures, with a backbone consisting of paralogous RAMPs (Repeat-Associated Mysterious Proteins), such as Cas7 and Cas5, containing the RRM (RNA Recognition Motif) fold and additional 'large' and 'small'

Figure 2
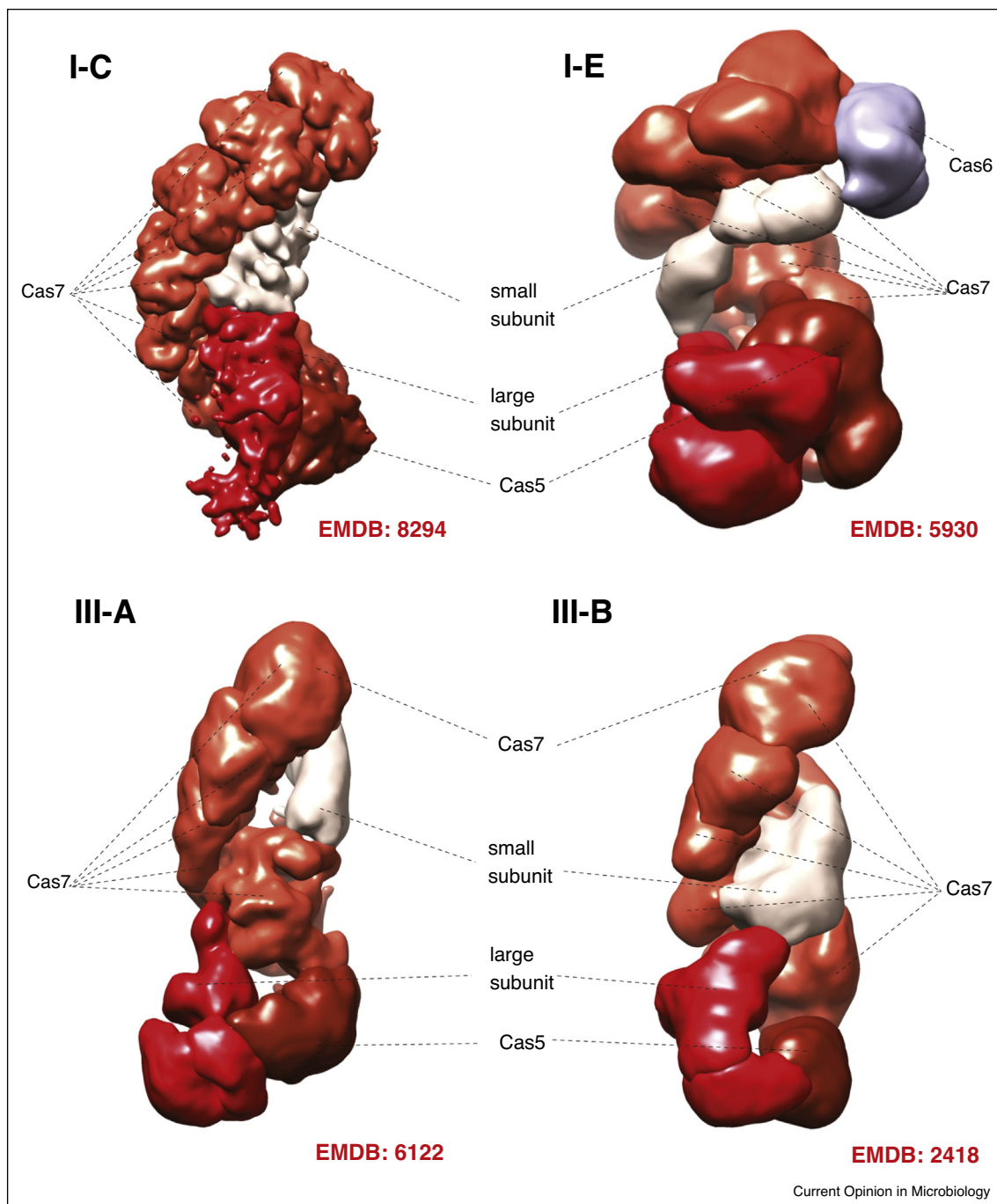
Cryo-electron microscopic structures of type I and type III effector complexes.
The Electron Microscopy Data Bank (EMDB) codes are indicated for each structure. The structures are from Refs. [21] (I-C), [20] (I-E), [23] (III-A), [22] (III-B).

subunits [16–23] (Figure 2). These effector complexes contain one Cas5 subunit and several Cas7 subunits. The complex accommodates the guide RNA that consists of the spacer and a portion of a repeat. The Cas5 subunit binds the 5′ handle of the crRNA and interacts with the large subunit (Cas8 in type I and Cas10 in type III). The small subunit is typically present in several copies and interacts with the crRNA backbone bound to Cas7. Notably, the length of the bound spacer correlates with the number of Cas7 subunits in the backbone of the complex

[24,25[•],26[•]]. Although the sequences of the protein sub-units in type I and type III effector complexes show little sequence similarity to each other, the presence of the homologous RAMPs in both types of complexes and the overall structural similarity revealed by cryo-electron microscopy (Figure 2) leave little doubt in the common origin of these complexes [8[••],15,19]. An additional RAMP, Cas6, is loosely associated with the effector complex and typically functions as the repeat-specific RNase in the pre-crRNA processing [27,28[•]].

The effector modules of Class 2 consist of a single, large, multidomain protein each, such that the respective CRISPR-cas loci have a much simpler and more uniform organization than those of Class 1 (Figure 1). At the time of the construction of the current formal CRISPR-Cas classification [8[••]], Class 2 included 3 subtypes of the thoroughly characterized type II, with the effector endo-nuclease Cas9, the widely used genome editing enzyme, and type V, with the predicted effector protein Cpf1. Type V was included into the classification scheme for the first time and remained experimentally uncharacter-ized. The subsequent demonstration that Cpf1 was indeed an active RNA-guided endonuclease that, unlike Cas9, did not require the additional trans-activating CRISPR (tracr)RNA for target cleavage [29[••]], raised interest in comprehensive characterization of the diver-sity of the Class 2 CRISPR-Cas systems by genome and metagenome mining.

To this end, a computational pipeline was constructed that (i) identified CRISPR-cas loci by using either the highly conserved cas1 gene or the CRISPR array as an anchor, (ii) discarded those loci that could be classified into one of the known subtypes using the previously developed automatic approach, and (iii) selected those that encoded a large protein, a putative Class 2 effector, for detailed, case by case analysis [30[••]]. Running this pipeline against all available bacterial and archaeal gen-omes along with the long metagenomics contigs resulted in the identification of 3 new subtypes of type V, which all contain a RuvC-like endonuclease domain, and 3 subtypes

of type VI, each with two HEPN domains predicted to possess RNase activity (Figure 1b) [31[••]]. The interference activities of the effectors of type V-B (Cas12b) [30[••]] type VI-A (Cas13a) [32] and type VI-B (Cas13b) [33[••]] effectors have been validated experimentally.

The design of the computational pipeline for Class 2 ensured effectively complete detection of all variants existing in the searched databases, up to the sensitivity of the anchor (cas1 gene or CRISPR array) detection. Indeed, all CRISPR-Cas loci that did not belong to the already identified subtypes but contained putative Class 2 effectors were analyzed in detail and classified accord-ingly [31[••]]. Obviously, however, new genomes, espe-cially those representing microbial groups not sampled previously, have the potential to yield new variants. Two such variants have been discovered in a recent search of genomes from uncultivated bacteria and can be classified as new type V subtypes [34[••]]. The predicted effectors of these new Class 2 variants contain the RuvC-like domain but no other detectable nuclease domains and have been originally denoted CasX and CasY but, according to the Cas protein nomenclature, should become, respectively, Cas12e and Cas12d, the latter related to the previously identified Cas12c (Figure 1).

Type II and type V effectors: functional convergence vs structural diversity

Comparative-genomic discovery of the new Class 2 CRISPR-Cas systems has been rapidly followed by structural and experimental characterization of their pre-dicted effectors (Table 1). The unifying feature of all type II and type V subtypes is the presence of a RuvC-like endonuclease domain that adopts the RNase H fold (Figures 1 and 3). However, the other parts of the Cas9 and Cas12 (the new designation for the type V effectors) proteins from different subtypes show no similarity to one another at the sequence or structure levels. Moreover, the similarity between the RuvC-like domain sequences of different subtypes is low so that they can be recognized as homologs only through the comparison of the conserved catalytic motifs and structures (Figure 3) [30[••],31[••]]. The

Table 1

Properties of experimentally characterized Class 2 CRISPR-Cas systems

Type/subtype/effector	Nuclease domains	Target	Cut structure	tracrRNA requirement	PAM/PFS[a]
II/Cas9	RuvC, HNH	dsDNA	blunt	Yes	3′ GC-rich PAM
V-A/Cas12a(Cpf1)	RuvC, NUC(?); uncharacterized domain for pre-crRNA processing	dsDNA	Staggered, 5′-overhangs	No	5′ AT-rich PAM
V-B/Cas12b(C2c1)	RuvC	dsDNA	Staggered, 5′-overhangs	Yes	5′ AT-rich PAM
VI-A/Cas13a(C2c2)	2xHEPN; uncharacterized domain for pre-crRNA processing	ssRNA	Guide-dependent RNA cuts + collateral RNA cleavage	No	3′PFS: non-G
VI-B/Cas13b	2xHEPN	ssRNA	Guide-dependent RNA cuts + collateral RNA cleavage	No	5′PFS: non-C; 3′PFS: NAN/NNA

[a] PAM, protospacer adjacent motif; PFS, protospacer-flanking sequence.

Figure 3
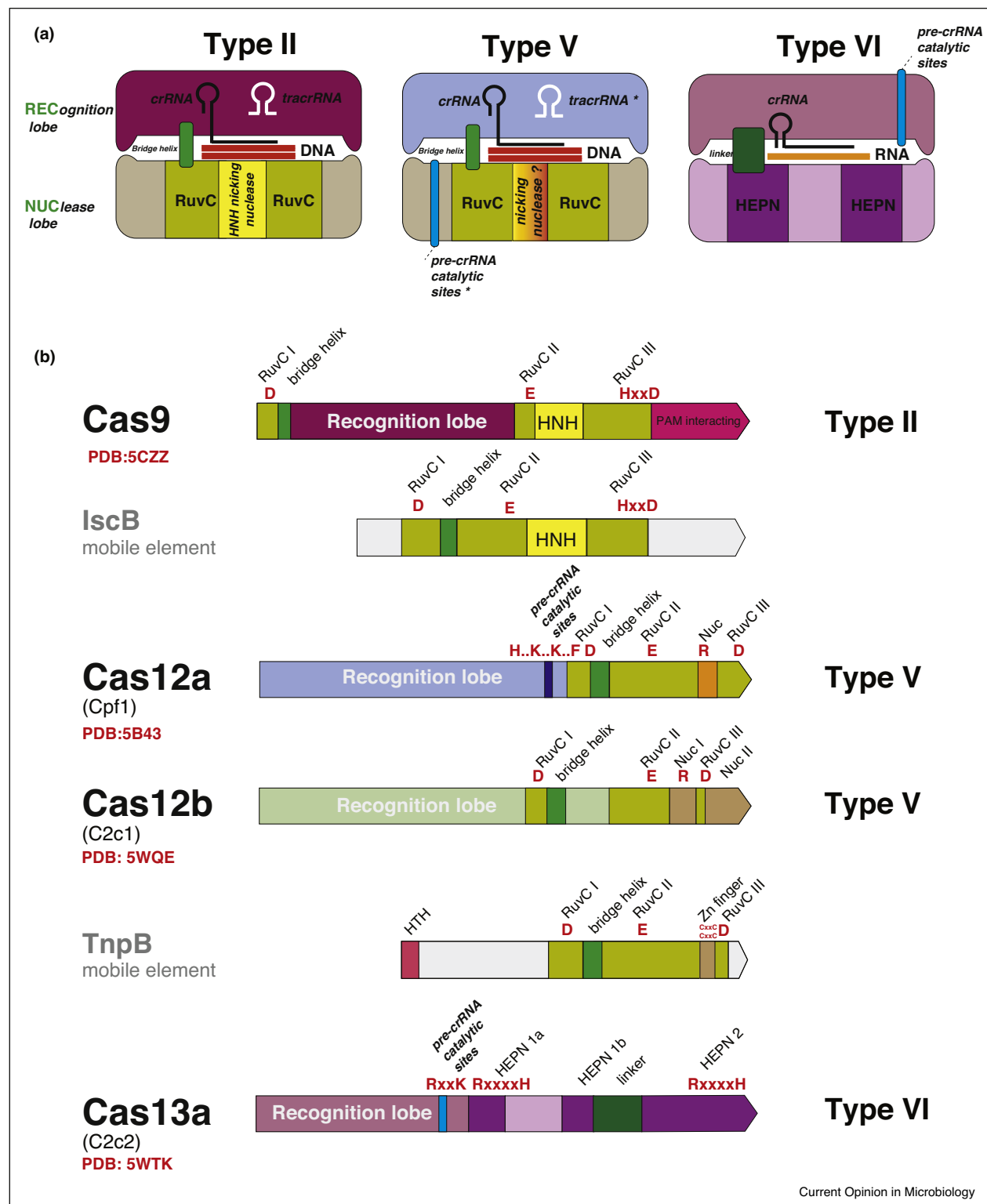
The Class 2 effector proteins.
(a) Schematic representation of the complexes of effector proteins with the target DNA or RNA, guide RNA and (for type II) tracrRNA
(b) Domain architectures of the effector proteins and their transposon-encoded evolutionary ancestors
IscB and TnpB are the inferred ancestors of the type II (Cas9) and type V (Cas12) effectors, respectively. The catalytic residues of the effector nuclease domain and, for Cas12a and Cas13a, the residues shown to be required for pre-crRNA processing are indicated in red. The Protein Data

structures of several Cas9 proteins [35–38] as well as structures of Cas12a (Cpf1) from three bacteria [39••,40••,41••,42••] and two structures of Cas12b (C2c1) [43••,44••] complexed with the guide RNA and target DNA have been solved. Comparison of these structures reveals a remarkable interplay of convergence and diversity. All three effectors are of closely similar size and share similar bilobed, 'jaw-like' shapes (Figure 3) but the structures could not be superimposed apart from the RuvC-like domains [45]. The nuclease activity of the RuvC-like domains of Cas9 is complemented by that of the unrelated HNH family nuclease that is inserted into the RuvC-like domain (Figure 3) and cleaves the target (complementary to the guide) strand of the substrate DNA [46,47]. Structural analysis of the CRISPR-Cas9 R-loop containing complex with the target DNA has shown that Cas9 induces the structural distortion of the DNA helix that is required for R-loop formation, which then causes a conformation change in Cas9, such that the RuvC and HNH catalytic sites are positioned in proximity of the displaced and target strands, respectively [48•].

The Cas12a protein lacks the HNH domain but contains a distinct domain with a unique fold inserted in a similar but not identical position within the RuvC-like domain (Figure 3) [45]. Mutation of the catalytic residues of the RuvC-like domain in Cas12a of the bacterium *Acidaminococcus* sp. completely prevented the cleavage of both strands, whereas mutation of some of the conserved residues in the insert domain specifically and completely inhibited cleavage of the target strand [40••]. These findings and the spatial proximity of this domain to the target strand suggest that the insert domain of Cas12a (denoted the Nuc domain) was responsible for the target strand cleavage [40••].

The RuvC-like domain of Cas12b (C2c1) also contains an insert domain with a unique fold that has been denoted Nuc by analogy with Cas12a and because some mutation in this domain substantially reduced the target strand cleavage efficiency [43••]. However, it has been shown that, in the case of Cas12b, both target and non-target DNA strands are accommodated and cleaved by the catalytic site of the RuvC-like domain, which undergoes a major conformational change triggered by the initial, non-target strand cleavage [44••]. In a recent analysis of the structure and activity of Cas12a from the bacterium *Francisella novicida*, mutations in the Nuc domain only partially inactivated the target strand cleavage, leading to the suggestion that, analogously to Cas12b, Cas12a cleaved both strands of the target DNA via the RuvC-like domain, whereas the Nuc domain only contributed to the guide and target binding [42••]. The details of the target recognition and cleavage by different type V effectors and the contributions of different domains remain to be thoroughly investigated. Nevertheless, the available data clearly indicate that, in type II and type V effectors of different subtypes, similar functional solutions have evolved via distinct routes.

Type VI: dedicated RNA-targeting CRISPR-Cas systems with toxin properties

In all type VI effectors (Cas13), the only recognizable features are two HEPN domains that, to an even greater extent than the RuvC-like domains in Cas9 and Cas12, show extreme sequence divergence such that their identification required extensive manual comparison [30••,33••]. These observations are in line with the previous analyses showing that the HEPN superfamily consists of extremely diverged RNases that are involved mostly in various defense-related functions in both prokaryotes and eukaryotes [49]. In particular, the HEPN RNase is the toxin domain of numerous bacterial and archaeal toxin-antitoxin modules. The discovery of the putative effectors containing HEPN domains prompted the prediction that type VI includes dedicated RNA-targeting CRISPR-Cas systems [30••] (RNA targeting has been demonstrated for certain type III systems but these target DNA as well [50,51,52••]). This prediction has been validated for both type VI-A and type VI-B effectors (Cas13a and Cas13b, respectively) that have been characterized as programmable, RNA-guided RNase that efficiently protect *Escherichia coli* from the RNA bacteriophage MS2 infection [32••,33••]. Analysis of the Cas13a structure confirmed the presence of the two HEPN domains and revealed yet another bilobed structure that generally resembles those of the type II and type V effectors (Figure 3) [53••].

In the case of type VI-B, a remarkable feature, so far not observed for other CRISPR-Cas systems, has been discovered: the activity of Cas13b proteins is either up- or down-regulated by accessory small Cas proteins encoded by the respective type VI-B variants [33••] (Figure 1b).

Strikingly, once activated by the cognate target RNA, Cas13a and Cas13b become promiscuous RNases that cleave RNA in a non-sequence-specific manner [32••,33••]. Furthermore, expression of the Cas13 proteins jointly with the guide and target RNAs leads to cellular toxicity, suggestive of induction of dormancy or programmed cell death in bacteria [32••]. This

(Figure 3 Legend Continued) Bank (PDB) codes are included for proteins with solved structures. HTH, helix-turn-helix DNA-binding domain. The tracrRNA, the pre-crRNA processing catalytic sites and the nicking, target strand-cleaving nuclease are denoted by asterisks to indicate that they are each present only in subsets of the type II and type V effectors (see text for details). The catalytic amino acid residues of the target-cleaving nucleases are shown in red, and those of the pre-crRNA processing nuclease are shown in blue. RuvCI, II and III stand for the three conserved motifs of the RuvC-like nuclease domain that contain the catalytic residues. In the motif signature, "x" stands for amino acid, and "." indicated that the catalytic residues are separated by a small, variable number of non-conserved residues.

phenomenon seems to be compatible with the previously proposed hypothesis on coupling between immunity and dormancy induction and/or programmed cell death (PCD) in prokaryotes [54,55]. Given that RNA viruses have a narrow host range and make up but a tiny part of the prokaryotic virosphere, it seems most likely that type VI systems target DNA viruses by cleaving viral transcripts and then inducing dormancy or PCD in response to intense expression of viral genes.

Pre-crRNA processing: multitasking Class 2 effectors

Most of the experimentally studied CRISPR arrays are expressed as a single, long transcript, known as pre-crRNA and are subsequently processed into mature crRNAs via mechanisms that differ between types and subtypes [27]. Pre-crRNA processing in type II CRISPR-Cas systems has been characterized in detail. In this case, pre-crRNA cleavage is catalyzed by RNase III, a ubiquitous bacterial enzyme that is encoded outside the CRISPR-cas locus, and requires tracrRNA, a distinct RNA species that is partially complementary to the CRISPR [27]. An alternative mechanism in some type II systems involves independent transcription of individual crRNAs [56]. Neither type V-A nor type VI systems require tracrRNA, and it has been shown that the corresponding effectors, Cas12a and Cas13a, are themselves fully responsible for pre-crRNA processing [53••,57•• ,58••,59••]. The catalytic sites involved in crRNA maturation have been at least partially identified and shown to be distinct from those involved in target cleavage but could not be associated with any distinct domains [53••,57••,58••] (Figure 3). The structures of Cas12a and Cas13a show no similarity, apart from the bilobed shape (Figure 3), implying that the pre-crRNA processing activities have evolved independently in types V and VI. Type V-B systems require tracrRNA for crRNA maturation and target cleavage [30••,43••,44••], but the mechanism of processing has not been established. It is thus a question of major interest whether pre-crRNA processing in type V-B involves RNase III or is catalyzed by Cas12b itself, or involves some other mechanism.

Origin and evolution of CRISPR-Cas: the pivotal contributions of mobile genetic elements

Over the last few years, comparative genomic analysis has shed light on the origins and evolution of CRISPR-Cas systems. Examination of the genomic context of *cas1* homologs that are not associated with CRISPR-*cas* loci led to the discovery of a novel superfamily of self-synthesizing transposons, the Casposons, named after the Cas1 homolog they encode and have been predicted to employ as integrase (recombinase) [60,61]. The integrase activity of the Casposon-encoded Cas1 (renamed casposase) has been validated experimentally [62], and similar target site specificities of Casposon integration and CRISPR spacer

incorporation have been demonstrated [63]. Although the currently identified Casposons do not encode Cas2, some encode Cas4 and additional nucleases [61]. It seems likely that the entire adaptation module and perhaps even additional Cas proteins were contributed to the emerging CRISPR-Cas system by the ancestral Casposon [64]. Furthermore, the prototype CRISPR repeats and the leader sequence could have originated from a duplicated target site of the casposon [65].

The ancestry of the effector module is far less clear. Given the wide spread existence of Class 1 CRISPR-Cas systems among archaea and bacteria, contrasted by the much lower abundance of the Class 2 systems, the multisubunit effector complexes of Class 1 are the likely ancestral form [30••]. Given that the core subunits of Class 1 effector complexes contain diverged version of the RRM, it appears likely that the effector module evolved by serial duplication of an ancestral RRM protein, with subsequent diversification driven by the host-parasite arms race[13]. The ultimate ancestor of the core Cas proteins could have been a RRM domain-containing nuclease, such as Cas10, that gave rise to the extant multitude of active and inactivated paralogs. The ancestral effector complex, before the recruitment of the Casposon, could function as an RNA-guided innate immunity system [64], perhaps similarly to the extant Argonaute-centered immunity [66].

The provenance of the Class 2 effector module is much clearer. Surprisingly, the sequences of the RuvC-like domains of Cas9 and Cas12 are more similar to the sequences of transposon-encoded TnpB proteins than they are to each other. The tnpB genes are highly abundant in bacterial and archaeal genomes and belong to autonomous but more frequently non-autonomous transposons. The role of TnpB in the transposon life cycle remains unclear given that this protein is not required for transposition [67] but the conservation of the RuvC-like endonuclease catalytic sites in most TnpB sequences indicates that these proteins are active nucleases. Remarkably, Cas9 and Cas12a, 12b and 12c showed the highest similarity to different groups of TnpB proteins, suggesting independent origins for the effectors of different types and subtypes of Class 2. The ancestry of Cas9 could be readily traced to a distinct family of transposons (hence denoted ISC, after Insertion Sequences Cas9-related) thanks to the shared domain architectures of the IscB and Cas9 proteins (Figure 3), in which an HNH endonuclease domain is inserted into the RuvC-like domain [68,69•]. The specific ancestry of Cas12a-c remained murky caused by low sequence conservation [30••].

The line of descent of type V effectors from TnpB was clarified with the identification of subtype V-U (after Uncharacterized) loci using CRISPR arrays as the anchor for the computational Class 2 discovery pipeline [31••].

Unlike the other Class 2 effectors, which are giant proteins with limited similarity to TnpB within the RuvC-like domains, type V-U loci encode TnpB homologs that are either the same size as transposon-encoded TnpB or slightly larger (Figure 1), and show highly significant similarity to the latter. The high similarity of the putative V-U effectors to TnpB enabled reliable phylogenetic analysis, which supported independent origin of Class 2 effectors from different TnpB subfamilies. The functionality of the type V-U systems remains to be demonstrated but at least some of them are likely to be active because the respective CRISPR arrays contain spacers homologous to phage genome sequences. The current evolutionary scenario for type II and type V effectors includes multiple insertions of non-autonomous TnpB-encoding transposons next to CRISPR arrays, followed by convergent 'domestication' of TnpB involving acquisition of additional domains [31[••]]. These domains are unrelated in different subtypes, but in each case, enable the effector proteins to accommodate the crRNA guide complexed with the target DNA (Figure 3).

As with the type V effectors, the exact ancestry of the type VI, HEPN-containing effector proteins is difficult to pinpoint, but HEPN-domain containing Cas proteins of Class 1 CRISPR-Cas systems, such as Csm6 and Csx1, that target viral transcripts appear plausible candidates [31[••],70[••]]. Ultimately, there is little doubt that these Cas proteins evolved from HEPN-containing RNase toxins. Thus, the evolutionary history of Class 2 systems is largely a story of the second, after the Casposons, major contribution of mobile elements to the evolution of the CRISPR-Cas adaptive immunity. Of all, apparently independent cases of recruitment of nucleases from mobile elements giving rise to Class 2 systems, the capture of IscB that yielded type II was by far the most successful one: taken together, all other variants in Class 2 area about an order of magnitude less abundant among bacteria than type II [31[••]]. Whether the type II systems provide some distinct advantage to bacteria or are for some reason more prone to horizontal gene transfer than other Class 2 systems, is an intriguing question for future work on CRISPR biology.

Conclusions
The recent progress in the study of the diversity of CRISPR-Cas systems has led to the discovery of not only novel effector protein and locus architectures but also fundamentally new molecular mechanisms. These include the exclusive RNA-targeting by type VI systems that appear to directly couple CRISPR immunity with dormancy induction or PCD; pre-crRNA processing by the effector proteins of types V-A and VI-A; differential regulation type VI-B effectors by small Cas proteins; and also the use of RNA for adaptation via CRISPR-associated reverse transcriptases by some type III systems [71[••]]. There is little doubt that more functional diversity

remains to be discovered, in particular, with regard to the mechanisms of regulation of CRISPR-Cas activity and coupling of immunity with dormancy induction. During the short time interval of only about 2 years, focused study of Class 2 systems has revealed a broad spectrum of effector nuclease activities that offer ample choice of functionalities for genome editing and regulation tools (Table 1). Furthermore, the parallel discoveries of Casposons and the type V-U loci suggests the possibility of experimental investigation of different stages in CRISPR-Cas evolution that can shed light on the transition from mobile genetic elements to an adaptive immunity system.

Author declaration
We confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He/she is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Sorek R, Lawrence CM, Wiedenheft B: **CRISPR-mediated adaptive immune systems in bacteria and archaea**. *Annu. Rev. Biochem*. 2013, **82**:237-266.

2. Wright AV, Nunez JK, Doudna JA: **Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering**. *Cell* 2016, **164**:29-44.

3. Komor AC, Badran AH, Liu DR: **CRISPR-based technologies for the manipulation of eukaryotic genomes**. *Cell* 2017, **168**:20-36 http://dx.doi.org/10.1016/j.cell.2016.10.044.

4. Mohanraju P, Makarova KS, Zetsche B, Zhang F, Koonin EV, van der Oost J: **Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems**. *Science* 2016, **353** aad5147.

5. Pingoud A, Wilson GG, Wende W: **Type II restriction endonucleases—a historical perspective and more**. *Nucleic Acids Res.* 2014, **42**:7489-7527.

6. Takeuchi N, Wolf YI, Makarova KS, Koonin EV: **Nature and intensity of selection pressure on CRISPR-associated genes**. *J. Bacteriol.* 2012, **194**:216-1225.

7. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J *et al.*: **Evolution and classification of the CRISPR-Cas systems**. *Nat. Rev. Microbiol.* 2011, **9**:467-477.

8. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA,
●● Saunders SJ, Barrangou R, Brouns SJ, Charpentier E, Haft DH, Horvath P *et al.*: **An updated evolutionary classification of CRISPR-Cas systems**. *Nat. Rev. Microbiol.* 2015, **13**:722-736.
This paper presents the methods for identification and comparative analysis of CRISPR-Cas systems and the resulting classification. The two classes of CRISPR-Cas systems as well as type V are introduced for the first time.

9. Bondy-Denomy J, Pawluk A, Maxwell KL, Davidson AR:
● **Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system**. *Nature* 2013, **493**:429-432.
This work demonstrates that different virus-encoded anti-CRISPR proteins interfere with the activity of CRISPR-Cas systems by binding different Cas proteins.

10. Bondy-Denomy J, Garcia B, Strum S, Du M, Rollins MF, Hidalgo-
● Reyes Y, Wiedenheft B, Maxwell KL, Davidson AR: **Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins**. *Nature* 2015, **526**:136-139.

11. Pawluk A, Amrani N, Zhang Y, Garcia B, Hidalgo-Reyes Y, Lee J,
● Edraki A, Shah M, Sontheimer EJ, Maxwell KL, Davidson AR: **Naturally occurring off-switches for CRISPR-Cas9**. *Cell* 2016, **167**:1829-1838 e1829.
This paper expands the repertoire of virus-encoded anti-CRISPR proteins to include, for the first time, inhibitors of Class 2 systems.

12. Pawluk A, Staals RH, Taylor C, Watson BN, Saha S, Fineran PC, Maxwell KL, Davidson AR: **Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species**. *Nat. Microbiol.* 2016, **1**:16085.

13. Makarova KS, Wolf YI, Koonin EV: **The basic building blocks and evolution of CRISPR-cas systems**. *Biochem. Soc. Trans.* 2013, **41**:1392-1400.

14. Amitai G, Sorek R: **CRISPR-Cas adaptation: insights into the mechanism of action**. *Nat. Rev. Microbiol.* 2016, **14**:67-76.

15. Makarova KS, Aravind L, Wolf YI, Koonin EV: **Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems**. *Biol. Direct.* 2011, **6**.

16. Zhao H, Sheng G, Wang J, Wang M, Bunkoczi G, Gong W, Wei Z, Wang Y: **Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli***. *Nature* 2014, **515**:147-150.

17. van der Oost J, Westra ER, Jackson RN, Wiedenheft B: **Unravelling the structural and mechanistic basis of CRISPR-Cas systems**. *Nat. Rev. Microbiol.* 2014, **12**:479-492.

18. Jackson RN, Golden SM, van Erp PB, Carter J, Westra ER, Brouns SJ, van der Oost J, Terwilliger TC, Read RJ, Wiedenheft B: **Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli***. *Science* 2014, **345**:1473-1479.

19. Jackson RN, Wiedenheft B: **A conserved structural chassis for mounting versatile CRISPR RNA-guided immune responses**. *Mol. Cell.* 2015, **58**:722-728.

20. Hochstrasser ML, Taylor DW, Bhat P, Guegler CK, Sternberg SH, Nogales E, Doudna JA: **CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference**. *Proc. Natl. Acad. Sci. U. S. A.* 2014, **111**:6618-6623.

21. Hochstrasser ML, Taylor DW, Kornfeld JE, Nogales E, Doudna JA: **DNA Targeting by a minimal CRISPR RNA-guided cascade**. *Mol. Cell.* 2016, **63**:840-851.

22. Staals RH, Agari Y, Maki-Yonekura S, Zhu Y, Taylor DW, van Duijn E, Barendregt A, Vlot M, Koehorst JJ, Sakamoto K, Masuda A *et al.*: **Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of *Thermus thermophilus***. *Mol. Cell.* 2013, **52**:135-145.

23. Staals RH, Zhu Y, Taylor DW, Kornfeld JE, Sharma K, Barendregt A, Koehorst JJ, Vlot M, Neupane N, Varossieau K, Sakamoto K *et al.*: **RNA targeting by the type III-A CRISPR-Cas Csm complex of *Thermus thermophilus***. *Mol. Cell.* 2014, **56**:518-530.

24. Hatoum-Aslan A, Samai P, Maniv I, Jiang W, Marraffini LA: **A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs**. *J. Biol. Chem.* 2013, **288**:27888-27897.

25. Kuznedelov K, Mekler V, Lemak S, Tokmina-Lukaszewska M,
● Datsenko KA, Jain I, Savitskaya E, Mallon J, Shmakov S, Bothner B, Bailey S *et al.*: **Altered stoichiometry *Escherichia coli* Cascade complexes with shortened CRISPR RNA spacers are capable of interference and primed adaptation**. *Nucleic Acids Res.* 2016, **44**:10849-10861.
Together with Ref. [26●], this paper shows that the Cas7 subunits of the Class 1 effector complexes function as a ruler that determines the length of the accommodated guide RNA.

26. Luo ML, Jackson RN, Denny SR, Tokmina-Lukaszewska M,
● Maksimchuk KR, Lin W, Bothner B, Wiedenheft B, Beisel CL: **The CRISPR RNA-guided surveillance complex in *Escherichia coli* accommodates extended RNA spacers**. *Nucleic Acids Res.* 2016, **44**:7385-7394.
Together with Ref. [25●], this paper shows that the Cas7 subunits of the Class 1 effector complexes function as a ruler that determines the length of the accommodated guide RNA.

27. Charpentier E, Richter H, van der Oost J, White MF: **Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity**. *FEMS Microbiol. Rev.* 2015, **39**:428-441.

28. Niewoehner O, Jinek M: **Structural basis for the
● endoribonuclease activity of the type III-A CRISPR-associated protein Csm6**. *RNA* 2016, **22**:318-329.
Demonstration of the involvement of a HEPN domain in the RNase activity of the accessory Cas protein Csm6.

29. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM,
●● Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, Koonin EV *et al.*: **Cpf1 is a single RNA-guided endonuclease of a Class 2 CRISPR-Cas system**. *Cell* 2015, **163**:759-771.
The first experimental characterization of a single-protein CRISPR effector endonuclease distinct from Cas9. Unexpectedly, this work shows that Cpf1 (now Cas12a) does not require tracrRNA and functions as a single guide RNA-dependent endonuclease. In addition, Cpf1 introduces a staggered cut into the target DNA, as opposed to the blunt cut made by Cas9. These features make Cpf1 an attractive alternative genome editing tool.

30. Shmakov S, Abudayyeh OO, Makarova KS, Wolf YI,
●● Gootenberg JS, Semenova E, Minakhin L, Joung J, Konermann S, Severinov K, Zhang F *et al.*: **Discovery and functional characterization of diverse Class 2 CRISPR-Cas systems**. *Mol. Cell.* 2015, **60**:385-397.
This paper presents the computational pipeline for the discovery of novel Class 2 CRISPR-Cas systems in genomes and metagenomes using cas1 genes as the anchor. Subtypes V-B, V-C and VI-A are described. The crRNA expression and processing are demonstrated for V-B and VI-A, and the interference activity is validated for V-B.

31. Shmakov S, Smargon A, Scott D, Cox D, Pyzocha N, Yan W,
●● Abudayyeh OO, Gootenberg JS, Makarova KS, Wolf YI, Severinov K *et al.*: **Diversity and evolution of class 2 CRISPR-Cas systems**. *Nat. Rev. Microbiol.* 2017.
This paper describes a comprehensive search for Class 2 CRISPR-Cas systems using either cas1 or CRISPR array as anchors. The discovery of subtype V-U allows a reconstruction of the evolutionary scenario for type V systems.

32. Abudayyeh OO, Gootenberg JS, Konermann S, Joung J,
●● Slaymaker IM, Cox DB, Shmakov S, Makarova KS, Semenova E,

Minakhin L, Severinov K *et al.*: **C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector**. *Science* 2016, **353** aaf5573.

The first experimental characterization of a dedicated RNA-cleaving CRISPR-Cas syste, type VI-A. It is shown that guide-dependent RNA cleavage stimulates non-specific collateral RNA cleavage by C2c2 (Cas13a) protein. A hypothesis is proposed that the collateral activity induces dormancy or cell death in infected bacteria, defending the population via an altruistic mechanism.

33. Smargon AA, Cox DB, Pyzocha NK, Zheng K, Slaymaker IM,
•• Gootenberg JS, Abudayyeh OA, Essletzbichler P, Shmakov S, Makarova KS, Koonin et al EV: **Cab is a type VI-B CRISPR-associated RNA-guided RNase differentially regulated by accessory proteins Csx27 and Csx 28**. *Mol. Cell.* 2017, **65**:618-630.

34. Burstein D, Harrington LB, Strutt SC, Probst AJ, Anantharaman K,
•• Thomas BC, Doudna JA, Banfield JF: **New CRISPR-Cas systems from uncultivated microbes**. *Nature* 2017, **542**:237-241.

Discovery of novel Class 2 variants in genomes of uncultivated prokaryotes obtained by genome-resolved metagenomics. The first finding of type II in archaea.

35. Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, Kaplan M *et al.*: **Structures of Cas9 endonucleases reveal RNA-mediated conformational activation**. *Science* 2014, **343**:1247997.

36. Nishimasu H, Cong L, Yan WX, Ran FA, Zetsche B, Li Y, Kurabayashi A, Ishitani R, Zhang F, Nureki O: **Crystal Structure of *Staphylococcus aureus* Cas9**. *Cell* 2015, **162**:1113-1126.

37. Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O: **Crystal structure of Cas9 in complex with guide RNA and target DNA**. *Cell* 2014, **156**:935-949.

38. Hirano H, Gootenberg JS, Horii T, Abudayyeh OO, Kimura M, Hsu PD, Nakane T, Ishitani R, Hatada I, Zhang F, Nishimasu H *et al.*: **Structure and Engineering of *Francisella novicida* Cas9**. *Cell* 2016, **164**:950-961.

39. Dong D, Ren K, Qiu X, Zheng J, Guo M, Guan X, Liu H, Li N,
•• Zhang B, Yang D, Ma C *et al.*: **The crystal structure of Cpf1 in complex with CRISPR RNA**. *Nature* 2016, **532**:522-526.

Along with Refs. [40–42], this work reports the structure of Cas12a (Cpf1) and reveals the details of the interaction between Cas12a and the guide crRNA.

40. Yamano T, Nishimasu H, Zetsche B, Hirano H, Slaymaker IM, Li Y,
•• Fedorova I, Nakane T, Makarova KS, Koonin EV, Ishitani R *et al.*: **Crystal structure of Cpf1 in complex with guide RNA and target DNA**. *Cell* 2016, **165**:949-962.

This paper reports the structure of Cas12a complexed with both the crRNA and the target DNA and presents evidence that this effector protein employs a unique nuclease domain (Nuc) for the target strand cleavage.

41. 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**:901-913.

This paper describes the detailed analysis of the crRNA-mediated target DNA recognition by Cas12a and highlights the differences between Cas12a and Cas9.

42. Swarts DC, van der Oost J, Jinek M: **Structural basis for guide**
•• **RNA processing and seed-dependent DNA targeting by CRISPR-Cas12a**. *Mol. Cell.* 2017, **66**:221-233 e224.

This paper describes structural requirements for pre-crRNA processing by Cas12a and presents indications that both the target and the non-target strands of the substrate DNA are cleaved by the RuvC-like nuclease domain, whereas the Nuc domain is required only for crRNA and target binding.

43. Liu L, Chen P, Wang M, Li X, Wang J, Yin M, Wang Y: **C2c1-sgRNA**
•• **complex structure reveals RNA-guided DNA cleavage mechanism**. *Mol. Cell.* 2017, **65**:310-322.

Along with Ref. [44••], this paper presents the details of the interaction between C2c1 (Cas12b), guide RNA, tracrRNA and the target DNA, and demonstrates that the RuvC-like nuclease cleaves both strands of the target.

44. Yang H, Gao P, Rajashankar KR, Patel DJ: **PAM-dependent**
•• **target DNA recognition and cleavage by C2c1 CRISPR-Cas endonuclease**. *Cell* 2016, **167**:1814-1828 e1812.

45. Lewis KM, Ke A: **Building the Class 2 CRISPR-Cas Arsenal**. *Mol. Cell.* 2017, **65**:377-379.

46. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E: **A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity**. *Science* 2012, **337**:816-821.

47. Gasiunas G, Barrangou R, Horvath P, Siksnys V: **Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria**. *Proc. Natl. Acad. Sci. U. S. A.* 2012, **109**:E2579-E2586.

48. Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, Thompson AJ,
• Nogales E, Doudna JA: **Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage**. *Science* 2016, **351**:867-871.

This paper details the structural changes that are induced in Cas9 by the R-loop formation and position the two nuclease domains for the specific cleavage of the target and non-target strands.

49. Anantharaman V, Makarova KS, Burroughs AM, Koonin EV, Aravind L: **Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing**. *Biol. Direct.* 2013, **8**:15.

50. Hale CR, Cocozaki A, Li H, Terns RM, Terns MP: **Target RNA capture and cleavage by the Cmr type III-B CRISPR-Cas effector complex**. *Genes. Dev.* 2014, **28**:2432-2443.

51. Hale CR, Majumdar S, Elmore J, Pfister N, Compton M, Olson S, Resch AM, Glover CV, 3rd, Graveley BR, Terns RM, Terns MP: **Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs**. *Mol. Cell.* 2012, **45**:292-302.

52. Samai P, Pyenson N, Jiang W, Goldberg GW, Hatoum-Aslan A,
•• Marraffini LA: **Co-transcriptional DNA and RNA cleavage during Type III CRISPR-Cas immunity**. *Cell* 2015, **161**:1164-1174.

This paper demonstrates the requirement of transcript cleavage by the Csm3 protein for the co-transcriptional cleavage of the target DNA by the Cas10 protein in type III systems.

53. Liu L, Li X, Wang J, Wang M, Chen P, Yin M, Li J, Sheng G, Wang Y:
•• **Two distant catalytic sites are responsible for C2c2 RNase activities**. *Cell* 2017, **168**:121-134.

This paper reports the structure of Cas13a (C2c2) complexed with the guide and target RNAs, validates the role of the two HEPN domains in the target cleavage and shows that pre-crRNA processing depends on a distinct catalytic site outside the HEPN domains.

54. Koonin EV, Zhang F: **Coupling immunity and programmed cell suicide in prokaryotes: life-or-death choices**. *Bioessays* 2017, **39**:1-9.

55. Makarova KS, Anantharaman V, Aravind L, Koonin EV: **Live virus-free or die: coupling of antivirus immunity and programmed suicide or dormancy in prokaryotes**. *Biol. Direct.* 2012, **7**:40.

56. Zhang Y, Heidrich N, Ampattu BJ, Gunderson CW, Seifert HS, Schoen C, Vogel J, Sontheimer EJ: **Processing-independent CRISPR RNAs limit natural transformation in *Neisseria meningitidis***. *Mol. Cell.* 2013, **50**:488-503.

57. Fonfara I, Richter H, Bratovic M, Le Rhun A, Charpentier E: **The**
•• **CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA**. *Nature* 2016, **532**:517-521.

Along with Ref 59**, this work shows that Cpf1 (Cas12a) is responsible not only for target cleavage but also for the pre-crRNA processing via a catalytic site that is located outside the RuvC-like and NUC domains.

58. East-Seletsky A, O'Connell MR, Knight SC, Burstein D, Cate JH,
•• Tjian R, Doudna JA: **Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection**. *Nature* 2016, **538**:270-273.

This paper shows that, analogously to Cas12a, Cas13a (C2c2) catalyzes the pre-crRNA processing, in addition to the target RNA cleavage.

59. Zetsche B, Heidenreich M, Mohanraju P, Fedorova I, Kneppers J,
•• DeGennaro EM, Winblad N, Choudhury SR, Abudayyeh OO, Gootenberg JS, Wu WY *et al.*: **Multiplex gene editing by**

CRISPR-Cpf1 using a single crRNA array. *Nat. Biotechnol.* 2017, **35**:31-34.

60. Krupovic M, Koonin EV: **Self-synthesizing transposons: unexpected key players in the evolution of viruses and defense systems**. *Curr. Opin. Microbiol.* 2016, **31**:25-33.

61. Krupovic M, Makarova KS, Forterre P, Prangishvili D, Koonin EV: **Casposons: a new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-Cas immunity**. *BMC Biol.* 2014, **12**:36.

62. Hickman AB, Dyda F: **The casposon-encoded Cas1 protein from Aciduliprofundum boonei is a DNA integrase that generates target site duplications**. *Nucleic Acids Res.* 2015, **43**:10576-10587.

63. Beguin P, Charpin N, Koonin EV, Forterre P, Krupovic M: **Casposon integration shows strong target site preference and recapitulates protospacer integration by CRISPR-Cas systems**. *Nucleic Acids Res.* 2016, **44**:10367-10376.

64. Koonin EV, Krupovic M: **Evolution of adaptive immunity from transposable elements combined with innate immune systems**. *Nat. Rev. Genet.* 2015, **16**:184-192.

65. Krupovic M, Beguin P, Koonin EV: **Casposons: the mobile elements that gave rise to the adaptation module of CRISPR-Cas systems**. *Curr. Opin. Microbiol.* 2017, **38**:36-43.

66. Swarts DC, Makarova K, Wang Y, Nakanishi K, Ketting RF, Koonin EV, Patel DJ, van der Oost J: **The evolutionary journey of Argonaute proteins**. *Nat. Struct. Mol. Biol.* 2014, **21**:743-753.

67. Pasternak C, Dulermo R, Ton-Hoang B, Debuchy R, Siguier P, Coste G, Chandler M, Sommer S: **ISDra2 transposition in *Deinococcus radiodurans* is downregulated by TnpB**. *Mol. Microbiol.* 2013, **88**:443-455.

68. Chylinski K, Makarova KS, Charpentier E, Koonin EV: **Classification and evolution of type II CRISPR-Cas systems**. *Nucleic Acids Res.* 2014, **42**:6091-6105.

69. Kapitonov VV, Makarova KS, Koonin EV: **ISC, a novel group of**
• **bacterial and archaeal DNA transposons that encode Cas9 homologs**. *J. Bacteriol.* 2015, **198**:797-807.
This paper describes the identification of the direct evolutionary ancestor of Cas9, the transposon-encoded IscB proteins that contain a HNH nuclease domain inserted into a RuvC-like nuclease domain.

70. Jiang W, Samai P, Marraffini LA: **Degradation of phage**
•• **transcripts by CRISPR-associated RNases enables type III CRISPR-Cas immunity**. *Cell* 2016, **164**:710-721.
This work demonstrates that virus transcript cleavage is essential for target cleavage by type III CRISPR-Cas systems in vivo.

71. 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-Ca fusion protein**. *Science* 2016, **351** aad4234.
The first report on spacer acquisition from RNA, via reverse transcription by a reverse transcriptase encoded in a type III CRISPR-cas locus.