## REVIEW



# CRISPR in mobile genetic elements: counter-defense, inter-element competition and RNA-guided transposition



Eugene V. Koonin<sup>1\*</sup> and Kira S. Makarova<sup>1</sup>

## Abstract

CRISPR are adaptive immunity systems that protect bacteria and archaea from viruses and other mobile genetic elements (MGE) via an RNA-guided interference mechanism. However, in the course of the host-parasite co-evolution, CRISPR systems have been recruited by MGE themselves for counter-defense or other functions. Some bacteriophages encode fully functional CRISPR systems that target host defense systems, and many others recruited individual components of CRISPR systems, such as single repeat units that inhibit host CRISPR systems and CRISPR mini-arrays that target related viruses contributing to inter-virus competition. Many plasmids carry type IV or subtype V-M CRISPR systems that appear to be involved in inter-plasmid competition. Numerous Tn7-like and Mu-like transposons encode CRISPR-associated transposases (CASTs) in which interference-defective CRISPR systems of type I or type V mediate RNA-guided, site-specific transposition. The recruitment of CRISPR systems and their components by MGE is a manifestation of extensive gene shuttling between host immune systems and MGE, a major trend in the coevolution of MGE with their hosts.

Keywords CRISPR systems, CRISPR-associated transposases, Mobile genetic elements, Exaptation

## Background

CRISPR-Cas (CRISPR for short) are bacterial and archaeal adaptive immune systems that consist of an array of direct repeats separated by unique spacers and the adjacent CRISPR-associated (*cas*) genes [1-3]. Cas1 and Cas2 proteins, in some CRISPR subtypes assisted by additional proteins, constitute the CRISPR adaptation (spacer acquisition) module that inserts a fragment of foreign (in particular, viral) DNA between the proximal repeats in the array, yielding a spacer complementary to a segment of the target genome (protospacer), which is

\*Correspondence:

koonin@ncbi.nlm.nih.gov

accompanied by repeat duplication. Transcripts of the CRISPR array are processed into small, unit size CRISPR (cr) RNAs by a distinct complex of Cas proteins or an external enzyme and serve as guides for highly specific recognition of the target DNA or RNA that is typically followed by the target cleavage by a dedicated Cas nucle-ase. Characteristically of defense systems that evolve in a perennial arms race with mobile genetic elements (MGE), CRISPR show pronounced diversity, with two classes, 7 types, and numerous subtypes that differ with regard to the composition of Cas proteins and the organization of the CRISPR-*cas* genomic loci [4, 5]. In particular, each of the 7 CRISPR types encompasses a distinct effector nuclease.

In 2018, one of the authors of this review published a BMC Biology Editorial on open questions in CRISPR biology [6]. Altogether, 10 such unresolved problems have been formulated, dealing primarily with the distribution



This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Eugene V. Koonin

<sup>&</sup>lt;sup>1</sup> Computational Biology Branch, Division of Intramural Research, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

of CRISPR systems across the diversity of archaea and bacteria, the biological roles of different types of CRISPR, and their evolution. It appears that these questions were indeed quite challenging because 6 years after the publication, most of them remain as open as they were when first asked. However, on one of these problems, number 9 on the list, "What are the functions of CRISPR-Cas systems encoded by transposons and plasmids?", the progress has been impressive. Thus, here, we focus on the recruitment of CRISPR systems and their components by mobile genetic elements (MGE) including viruses and their functional repurposing (exaptation).

Almost all life forms harbor multiple, diverse MGE, some of which integrate into the host genome, whereas others replicate as free elements [7, 8]. The three main classes of MGEs are transposons, plasmids, and viruses which employ substantially different mechanisms of mobility. Remarkably, in prokaryotes, members of all these three types of MGE have on multiple occasions recruited CRISPR systems or their components that in some cases retain their full functionality but more commonly, lose some of their activities and are repurposed for distinct roles in MGE reproduction.

In this brief review article, we discuss the spread and diversity of CRISPR systems and their components in MGE, modifications associated with their recruitment, and the latest findings on their functions in MGE reproduction.

#### Main text

## CRISPR in viruses: counter-defense and inter-virus competition

The study of CRISPR recruitment by MGE started in the now classic 2013 study by Seed and colleagues on the type I-F CRISPR encoded by the Vibrio phage ICP1 [9]. This phage-encoded CRISPR system is fully functional, complete with a CRISPR array and adaptation module (Fig. 1A), and contains multiple spacers that target a *Vibrio* antiphage-defense module, phage-induced islandlike element (PLE), enabling ICP1 reproduction. Notably, this CRISPR system is only present in a subset of ICP1 strains where it occupies a specific antidefense locus that in other strains is occupied by the GIY-YIG family nuclease Odn which inhibits PLE via a completely different mechanism [10, 11]. The type I-F CRISPR system encoded by ICP1 is unrelated to the I-E system encoded by the host bacterium, so it remains unclear where, when, and how ICP1 captured the CRISPR locus.

ISP1 currently remains the only CRISPR-encoding phage for which the antidefense activity of the CRISPR system has been studied experimentally. Phage genome surveys show that CRISPR systems are rarely encoded by phages in the typical genome size range (below 200 kb) [12]. By contrast, analysis of the large (> 200 kb) genomes of jumbo phages assembled from metagenomic sequences led to the identification of numerous CRISPR systems of different types some of which appear to be fully functional whereas others lack either the adaptation module or the nuclease required for interference, or both (Fig. 1A) [13].

A broader variety of phages and archaeal viruses encompass minimal CRISPR components, namely, CRISPR mini-arrays (that is, arrays including only one or two spacers) and single repeat units (Fig. 1B, C) [12]. The mini-arrays contain repeats identical to those in the CRISPR systems of the respective hosts, and the majority of the spacers target viruses related to the given mini-array containing virus. This targeting specificity immediately prompted the hypothesis that the viral mini-arrays exploited the host CRISPR system to block replication of related viruses in coinfected host cells. This mechanism was validated in competition experiments with two related archaeal viruses one of which carried a CRISPR mini-array with a spacer against the other virus [14] (Fig. 1B).

The single CRISPR repeat units present in numerous viral genomes have been proposed to be transcribed into small RNAs mimicking CRISPR RNAs and acting as dominant-negative inhibitors of Cas proteins (RNA anti-CRISPR, Racr), forming non-productive effector

(See figure on next page.)

**Fig. 1** CRISPR systems and their components encoded in viral genomes. **A** Organization of the CRISPR loci carried by viruses. Genes are shown by block arrows, roughly to scale. CRISPR repeats are shown by rectangles, and spacers are shown by diamonds; spacers with detectable matches are shown with colors, and spacers without matches are shown in grey. *Cas* genes are denoted by their systematic names, with some legacy names shown in grey. CEE, circular extrachromosomal element; HD, nuclease of HD family (named for its catalytic dyad); HP; hypothetical protein; PLE, phage-like element; TerS (terminase small subunit), TerL (terminase large subunit) and phage minor tail protein U, essential phage proteins; tracrRNA, transactivating CRISPR RNA. For the phage-encoded Cas9c protein, inactivation of the three catalytic motifs of the RuvC-like nuclease is indicated. **B** Competition between two archaeal viruses mediated by viral CRISPR mini-arrays; Spacers with matching protospacers in the respective viral genomes are shown in black and maroon, and spacers without matches are shown in grey. SPV1, SPV2, Saccharolobus portogloboviruses 1 and 2, two closely related, non-lytic archaeal viruses [14]. The other designations are as in **A**. **C** Inhibition of a host CRISPR systems by virus-encoded CRISPR systems and CRISPR. SRU, single repeat unit. Other designations are as in **A**. **D** General schematic of the functions of virus-encoded CRISPR systems and CRISPR components



complexes (Fig. 1C). Subsequently, it has been shown that a prophage-encoded Racr strongly inhibits the cognate type I-F CRISPR-Cas system by interacting specifically with Cas6f and Cas7f proteins, resulting in the formation of an aberrant Cas subcomplex [15]. In addition, Racrs have been shown to inhibit CRISPR systems of almost all other CRISPR types.

Altogether, virus-encoded CRISPR systems have been shown to contribute to all types of biological conflict in which viruses are involved including counter-defense, in particular, inactivation of defensive extrachromosomal elements, and intervirus competition (Fig. 1D).

Many viruses of bacteria and archaea encode diverse anti-CRISPR proteins (Acrs) [16, 17]. It would seem plausible that at least some of the Acrs would be Cas proteins or individual domains of these recruited by viruses and acting as dominant negative inhibitors of host CRISPR systems. Some complex viruses of eukaryotes, such as poxviruses, indeed have evolved multiple counterdefense proteins via this route [18]. Apparently, however, among the hundreds identified Acrs [19], very few were derived from Cas proteins. One of such rare cases includes homologs of the Cas4 nuclease encoded by some archaeal viruses one of which has been shown to inhibit CRISPR adaptation in model experiments [20]. Another example is AcrIII-1, an inhibitor of type III CRISPR systems which is a homolog of RING nucleases that cleaves cyclic oligoA, the second messenger synthesized by Cas10 protein, mitigating programmed cell death induction by these systems [21, 22]. Another notable case is the phage-encoded Cas9 protein with an inactivated RuvClike nuclease domain that likely suppresses the host CRISPR immunity by sequestering crRNAs via the formation of stable complexes via a phage-encoded transactivating CRISPR (tracr) RNA [5] (Fig. 1A). Finally, AcrIF3 appears to be a structural mimic of the domain of Cas8f protein (the large subunit of the Cascade complex) that recruits the Cas3 nuclease, and thus acts as a dominant-negative inhibitor of CRISPR interference [23]. These exceptions notwithstanding, it remains an enigma why segments of CRISPR RNAs (Racr) were broadly recruited by viruses as CRISPR inhibitors whereas Cas proteins apparently were not.

#### CRISPR in plasmids: inter-plasmid competition and more

Type IV CRISPR systems that appear to be derived descendants of type III systems are found primarily if not exclusively in plasmids, conjugative integrating elements, and some phages and prophages. Unlike typical Class 1 CRISPR systems, most of the type IV loci lack known or predicted effector nucleases, and in many cases, contain no adaptation module (Fig. 2A). Instead, subtype IV-A systems include a DinG helicase whereas subtype IV-B systems typically encompass a CysH-like protein, an inactivated derivative of phospho-adenylyl-sulphate reductase [24].

The CRISPR arrays in most type IV loci are strongly enriched with spacers targeting conjugative plasmids, suggesting a role in inter-plasmid competition [25–27]. Subsequently, it has been directly shown that, notwithstanding the lack of nuclease activity, plasmid-encoded subtype IV-A CRISPR can eliminate other plasmids from bacterial cells by crRNA-guided transcriptional repression of essential plasmid genes [28] (Fig. 2B). These findings were complemented by the demonstration that this plasmid-encoded CRISPR system can use the adaptation machinery of a host I-E CRISPR to incorporate spacers into its CRISPR array, compensating for the lack of an adaptation module in the type IV system [28]. Notably, one of the spacers in a type IV-A system from a Pseudomonas oleovorans megaplasmid targets host *pilN* gene, an essential component of type IV pili that are involved in DNA uptake, downregulating transcription of this gene [29]. In this case, the type IV system seems to contribute to plasmid competition indirectly, by preventing the uptake of other plasmids by a bacterium carrying the given one. Transcription inhibition might not be the only mechanism of interference by type IV-A CRISPR because they can block reproduction of target plasmids even via spacers matching non-coding regions [29, 30].

Although the most common type IV systems lack nucleases, in subtype IV-C, an HD nuclease domain is fused to the large subunit of the effector complex (Fig. 2). The presence of this fusion and the size of the large subunit, which is close to that of Cas10 and is in contrast to much smaller corresponding proteins in other type IV subtypes, suggests that IV-C systems are the evolutionary intermediate on the path from type III to type IV [27]. Furthermore, a recent largescale search for novel CRISPR systems in metagenomes led to the identification of a rare subtype IV-A variant that contains an HNH nuclease domain fused to DinG (Fig. 2). The dsDNA target cleaving, spacer-dependent activity of this nuclease was demonstrated experimentally and was shown to depend on the helicase activity of DinG which moves along the target molecule unwinding dsDNA and allowing the HNH nuclease to shred the target [5]. This interference mechanism mimics the mechanism of type I CRISPR systems in which Cas3 helicase containing a fused HD nuclease domain unwinds and shreds the target. However, DinG is only distantly related to Cas3 indicating that the targetshredding helicase-nuclease fusion proteins evolved independently in type I and type IV CRISPR systems. Notably, some variants of subtype IV-B encode an even



**Fig. 2** CRISPR systems carried by plasmids and their role in inter-plasmid competition. **A** Organization of the type IV and subtype V-M CRISPR loci carried by plasmids. Tra, Mob, PilN, and PelA are all essential plasmids genes; tr. regulator, transcription regulator. Block arrows with dashed outlines denote genes that are present in the majority but not all of the respective loci. Cas12m is an inactivated type V CRISPR nuclease. The other designations are as in Fig. 1. **B** Inter-plasmid competition mediated by a plasmid-encoded type IV CRISPR system via inhibition of transcription of essential plasmids genes. MobL, RecQ, TraE, and TraL are all essential plasmids proteins; P, promoter

more distantly related helicase, a RecD homolog [27], emphasizing multiple convergent acquisitions of helicases by different CRISPR systems.

Subtype IV-B systems lack both the adaptation module and a CRISPR array which, together with the CysH association, suggest a distinct function in plasmids and integrating conjugative elements that might not involve target recognition and remains to be characterized.

In addition to plasmids, albeit less commonly, type IV CRISPR systems are encoded in some prophages,

suggestive of roles in inter-phage or phage-plasmid competition [24, 27].

A systematic survey of CRISPR representation in plasmids showed that, apart from the plasmid-specific type IV, some plasmids encode CRISPR systems of all other types except for type VI, and subtypes III-B and V-F are specifically enriched in plasmids [31]. Particularly notable is subtype V-M that is strongly enriched in plasmids and also encoded by some prophages. The effector of this CRISPR subtype, Cas12m, contains an inactivated

RuvC-like nuclease domain and has been shown to inhibit plasmid replication by inhibition of transcription of essential plasmid genes, without cleaving the target [32]. Similarly to the case of viruses (see above), CRISPR systems are most common in large (above 200 kb) compared to smaller plasmids. Predominant plasmid-targeting specificity was observed for spacers from most plasmid-encoded CRISPR systems suggesting that not only type IV, for which inter-plasmid competition appears to be the primary function, but other CRISPR systems recruited by plasmids contribute to this type of conflict. However, a substantial minority of spacers in plasmid-carried CRISPR loci targeted viruses indicating that, in some cases, these CRISPR systems protect the plasmid-carrying host from viruses, thus ensuring propagation of the plasmid.

#### CASTs: recruiting CRISPR for RNA-guided transposition

Comparative genomic analysis of the diversity of CRISPR systems led to the unexpected discovery of distinct type I and type V CRISPR variants embedded in several groups of Tn7-like transposons (hereafter, CAST, CRISPRassociated transposase) [33, 34]. Tn7-like transposons encompass diverse cargo genes, in addition to the multiple transposase subunits, and CASTs are among the most common types of cargo [35]. Most of the CASTs lack both the adaptation module and the effector nuclease, that is, they encompass the minimal complement of cas genes required for crRNA processing and target recognition, while losing the capacity for target cleavage (Fig. 3A). Initially, two independent cases of type I-B CRISPR capture and a single case of type I-F capture by Tn7-like transposons were discovered by phylogenetic analysis of both the Cas7 protein, the most highly conserved component of type I CRISPR systems, and TnsA and TnsD (TniQ), two subunits of the Tn7 transposase [33]. The type I-F CAST appears to have been an ancient acquisition because it is conserved in a large branch of Tn7-like transposons, whereas I-B CASTs are shared by small groups of transposons and thus were apparently acquired more recently. Additional cases of CRISPR capture by Tn7-like transposons were discovered by subsequent mining of genomic and metagenomic databases including subtypes I-C, I-E, I-D, IV-A, and V-K (Fig. 3A) [12, 36, 37]. Most of the type I CASTs have lost the Cas3 helicase-nuclease, and with it, the target cleavage capacity while retaining the subunits of the effector complex (known as CASCADE in type I CRISPR) involved in crRNA binding and target recognition. In contrast, in Cas12k, the single-protein effector of type V-K CRISPR, the interference capacity is abolished due to the replacement of individual catalytic residues in the RuvC-like nuclease domain. These distinct modes of inactivation point to strong selection driving the CASTs to lose the target-cleaving nuclease activity. Remarkably, in the case of CAST I-D, one variant lost Cas3 and contains a mutationally inactivated target-cleaving HD nuclease domain in Cas10d, whereas the other variant is a complete, functional CRISPR system, containing both the adaptation module and the effector nuclease [37, 38] (Fig. 3A). Clearly, the functional version of CAST I-D represents an early stage in CAST evolution [37, 38]. Deep mining of metagenomes for new CRISPR systems led to the discovery of I-A and V-F CASTs in Mu transposons, a family of phage-like transposons distantly related to Tn7, further supporting the notion that CRISPR systems have been recruited by cargo-carrying transposons on multiple, independent occasions [5] (Fig. 3A).

The discovery of the CASTs prompted the straightforward hypothesis that the minimal CRISPR systems recruited by Tn7-like transposons enabled RNA-guided transposition by targeting the transposons to unique genomic sites defined by the spacer sequences [33]. This hypothesis was promptly validated experimentally by demonstrating that CAST I-D, CAST I-F, and CAST V-K each formed complexes with the respective Tn7 transposases and, when programmed to recognize a unique site in the target DNA via an engineered spacer, mediated transposon insertion at a fixed distance of 30 to 50 bp away from the protospacer (Fig. 3B) [34, 39].

The structures of CAST I-F [40] and CAST V-K [41– 44] transpososomes have been solved to atomic resolution by cryo-EM, revealing the details of the interactions between the Cas proteins and transposase subunit as well as target recognition. In CAST I-F, the Cascade complex is connected with the transposase via the dimer of

(See figure on next page.)

Fig. 3 The CASTs: recruitment of CRISPR systems by Tn7-like and Mu-like transposons for RNA-guided transposition. A Organization

of CAST-encoding Tn7-like and Mu-like transposons. Tns and tni genes encoded transposase subunits. Att, transposons attachment sites;

HTH, helix-turn-helix DNA-binding protein; LE, RE, transposon left end and right end, respectively; DR, direct repeat. The designations are as in Fig. 1, but the genes are shown not to scale. Crossed block arrows denote inactivated enzymes. **B** Schematic depiction of the structures and functionalities of two distinct CASTs. PAM, protospacer-adjacent motifs. **C** The two alternative routes of CAST insertion. Genetic organization of the I-B CAST is shown. Underneath, on the left, is a schematic depiction of homing to the attachment site via TniQ, and on the right, a schematic depiction of RNA-guided transposition via TnsD root (on the right). At the bottom, are schematics of the new Tn7-carrying loci emerging as the result of the transposon insertion. LE, RE, transposon Left End and Right END, respectively



Fig. 3 (See legend on previous page.)

the TniQ subunit, with one TniQ monomer interacting with Cas6 and the other one interacting with one of the 6 Cas7 subunits of the Cascade (Fig. 3B). In this system, the Cascade-TniQ complex scans numerous sites in the target DNA, but recruits TnsB and TnsC, the two transposase subunits, only to a few sites with sufficient complementarity to the spacer, suggesting a proofreading step [45]. In the CAST V-K transpososome, Cas12k protein, although structurally unrelated to Cascade, also interacts with TniQ which in turn recruits TnsC. Unexpectedly, the CAST V-K transpososome was found to include an additional subunit interacting with Cas12k, ribosomal protein S15 (Fig. 3B). In contrast to the other experimentally characterized CASTs, CAST I-D can mediate RNA-guided transposition even in the absence of TniQ, apparently, via direct interaction with TnsC, a core transposase subunit [37].

The role of the CASTs in the transposon life cycle remains incompletely understood. Tn7-like transposons propagate via two types of transposition, inserting either into variable sites in MGE (plasmids) or into conserved homing sites adjacent to tRNA genes or glmS genes in bacterial genomes (Fig. 3C) [46]. The very short CRISPR arrays associated with the CASTs contain some spacers targeting MGE, such as conjugative plasmids [37]. However, although the CASTs efficiently drive spacer-guided transposon insertion in model experiments and likely insert into MGE via the spacers from the array, this is not what happens during homing given that integrated copies of CASTs are not adjacent to protospacers. Instead, both CAST I-F and CAST V-K home to sites near tRNA or glmS genes through the so-called delocalized crRNA which is encoded near the CAST but outside of the array and contains sequences that only partially match the repeats in the array along with spacer-like sequences targeting short sequences adjacent to the homing sites [47, 48]. Furthermore, CAST variants containing only delocalized crRNA but no array have been discovered [37], suggesting that the respective transposons have lost the ability to integrate into MGE or integrate via a distinct mechanism independent of a guide RNA.

Whereas CAST I-F and CAST V-K apparently insert both into MGE and into conserved homing sites in bacterial genomes via the RNA-guided mechanism, CAST I-B employs two distinct mechanisms for these two types of transposition [48]. In this case, homing does not involve the Cascade or any guide RNA and is mediated by TnsD via a protein-only mechanism (Fig. 3C).

The intricacy and diversity of the interactions between the transposition machinery and Cas proteins and molecular mechanisms of the CASTs remain to be explored in detail. In particular, it is unclear how the CASTs acquire spacers and what is the origin of the delocalized crRNA. An unexpected functionality of the CASTs is suggested by the recent demonstration that CAST I-B, CAST I-F, and CAST V-K all can bind crR-NAs from the host chromosomal CRISPR systems via a sequence-independent interaction between Cas6 and the repeat in the crRNA [49]. The spacer of the bound crRNA can then be used by the CAST for guided transposition. This could be an additional mechanism for integration of the CASTs into MGE given that spacers from the host arrays mostly target viruses and plasmids.

Last but not least, the potential of the CASTs for precise insertion of large pieces of DNA into bacterial and eukaryotic genomes is obvious. In spite of the technical difficulties due to the necessity of simultaneous expression of multiple proteins, the first successful experiments in this direction in bacterial [50–52] and human [53] models have been reported.

### Conclusions

CRISPR systems or their components have been recruited and repurposed by all major types of MGE in prokaryotes, viruses, plasmids, and transposons. There are three major types of functionality of CRISPR in MGE: (1) counter-defense, (2) inter-MGE competition, (3) RNA-guided transposition. The repeated cooption (exaptation) of RNA-guided target recognition machinery that originally evolved as an adaptive anti-MGE immunity mechanism to function in the propagation of MGEs themselves is a striking manifestation of the "guns for hire" principle, a major trend in the coevolution of MGE with their hosts whereby MGE extensively exchange components with host immune systems [54]. Indeed, recruitment of CRISPR by MGE discussed here is only one side of this relationship. At least two types of CRISPR systems, II and V, evolved from RNA-guided nucleases encoded by transposons of the IS200/605 family and involved in distinct modes of transposon propagation [55–59]. Furthermore, a distinct variety of prokaryotic transposons, the casposons, apparently gave rise to the CRISPR adaptation module [60]. Undoubtedly, further studies will reveal new connections between CRISPR and MGE and yield insights into the ecology of MGE-encoded CRISPR systems that currently remain poorly understood.

#### Abbreviations

CAST	CRISPR associated transposase
CRISPR	Clustered regularly interspaced palindromic repeats
crRNA	CRISPR RNA
MGE	Mobile genetic element
tracrRNA	TRans-activating crRNA

#### Acknowledgements

Not applicable.

#### Authors' contribution

E.V.K. and K.S.M. wrote the manuscript.

#### Funding

Open access funding provided by the National Institutes of Health The authors are supported through the Intramural Research Program of the National Library of Medicine, National Institutes of Health.

#### Data availability

No datasets were generated or analysed during the current study.

## Declarations

**Ethics approval and consent to participate** Not applicable.

#### Consent for publication

Not applicable.

#### **Competing interests**

The authors declare no competing interests.

Received: 16 September 2024 Accepted: 3 December 2024 Published online: 18 December 2024

#### References

- 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(7):479–92.
- Marraffini LA. CRISPR-Cas immunity in prokaryotes. Nature. 2015;526(7571):55–61.
- 3. Barrangou R, Horvath P. A decade of discovery: CRISPR functions and applications. Nat Microbiol. 2017;2:17092.
- Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, Charpentier E, Cheng D, Haft DH, Horvath P, et al. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol. 2020;18(2):67–83.
- AltaeTran H, Kannan S, Suberski AJ, Mears KS, Demircioglu FE, Moeller L, Kocalar S, Oshiro R, Makarova KS, Macrae RK, et al. Uncovering the functional diversity of rare Cas systems with deep terascale clustering. Science. 2023;382(6673):eadi1910.
- 6. Koonin EV. Open questions: CRISPR biology. BMC Biol. 2018;16(1):95.
- Koonin EV, Dolja VV. A virocentric perspective on the evolution of life. Curr Opin Virol. 2013;3(5):546–57.
- Koonin EV. Viruses and mobile elements as drivers of evolutionary transitions. Philos Trans R Soc Lond B Biol Sci. 2016;371(1701):20150442.
- Seed KD, Lazinski DW, Calderwood SB, Camilli A. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. Nature. 2013;494(7438):489–91.
- Barth ZK, Nguyen MH, Seed KD. A chimeric nuclease substitutes a phage CRISPR-Cas system to provide sequence-specific immunity against subviral parasites. Elife. 2021;10:e68339.
- Boyd CM, Angermeyer A, Hays SG, Barth ZK, Patel KM, Seed KD. Bacteriophage ICP1: A Persistent Predator of *Vibrio cholerae*. Annu Rev Virol. 2021;8:285–304.
- Faure G, Shmakov SA, Yan WX, Cheng DR, Scott DA, Peters JE, Makarova KS, Koonin EV. CRISPR-Cas in mobile genetic elements: counter-defence and beyond. Nat Rev Microbiol. 2019;17(8):513–25.
- Al-Shayeb B, Skopintsev P, Soczek KM, Stahl EC, Li Z, Groover E, Smock D, Eggers AR, Pausch P, Cress BF, et al. Diverse virus-encoded CRISPR-Cas systems include streamlined genome editors. Cell. 2022;185(24):4574–4586. e4516.
- Medvedeva S, Liu Y, Koonin EV, Severinov K, Prangishvili D, Krupovic M. Virus-borne mini-CRISPR arrays are involved in interviral conflicts. Nat Commun. 2019;10(1):5204.
- Camara-Wilpert S, Mayo-Munoz D, Russel J, Fagerlund RD, Madsen JS, Fineran PC, Sorensen SJ, Pinilla-Redondo R. Bacteriophages

suppress CRISPR-Cas immunity using RNA-based anti-CRISPRs. Nature. 2023;623(7987):601–7.

- Maxwell KL. The Anti-CRISPR Story: A Battle for Survival. Mol Cell. 2017;68(1):8–14.
- Mayo-Munoz D, Pinilla-Redondo R, Camara-Wilpert S, Birkholz N, Fineran PC. Inhibitors of bacterial immune systems: discovery, mechanisms and applications. Nat Rev Genet. 2024;25(4):237–54.
- Mutz P, Resch W, Faure G, Senkevich TG, Koonin EV, Moss B. Exaptation of Inactivated Host Enzymes for Structural Roles in Orthopoxviruses and Novel Folds of Virus Proteins Revealed by Protein Structure Modeling. mBio. 2023;14(2):e0040823.
- Sahakyan H, Makarova KS, Koonin EV. Search for Origins of Anti-CRISPR Proteins by Structure Comparison. CRISPR J. 2023;6(3):222–31.
- Zhang Z, Pan S, Liu T, Li Y, Peng N. Cas4 Nucleases Can Effect Specific Integration of CRISPR Spacers. J Bacteriol. 2019;201(12):e00747–18.
- Athukoralage JS, McMahon SA, Zhang C, Gruschow S, Graham S, Krupovic M, Whitaker RJ, Gloster TM, White MF. An anti-CRISPR viral ring nuclease subverts type III CRISPR immunity. Nature. 2020;577(7791):572–5.
- Athukoralage JS, Graham S, Rouillon C, Gruschow S, Czekster CM, White MF. The dynamic interplay of host and viral enzymes in type III CRISPRmediated cyclic nucleotide signalling. Elife. 2020;9:e55852.
- Rollins MF, Chowdhury S, Carter J, Golden SM, Miettinen HM, Santiago-Frangos A, Faith D, Lawrence CM, Lander GC, Wiedenheft B. Structure Reveals a Mechanism of CRISPR-RNA-Guided Nuclease Recruitment and Anti-CRISPR Viral Mimicry. Mol Cell. 2019;74(1):132–142.e135.
- 24. Taylor HN, Laderman E, Armbrust M, Hallmark T, Keiser D, Bondy-Denomy J, Jackson RN. Positioning Diverse Type IV Structures and Functions Within Class 1 CRISPR-Cas Systems. Front Microbiol. 2021;12:671522.
- Pinilla-Redondo R, Mayo-Munoz D, Russel J, Garrett RA, Randau L, Sorensen SJ, Shah SA. Type IV CRISPR-Cas systems are highly diverse and involved in competition between plasmids. Nucleic Acids Res. 2020;48(4):2000–12.
- Newire E, Aydin A, Juma S, Enne VI, Roberts AP. Identification of a Type IV-A CRISPR-Cas System Located Exclusively on IncHI1B/IncFIB Plasmids in Enterobacteriaceae. Front Microbiol. 1937;2020:11.
- Moya-Beltrán A, Makarova KS, Acuña LG, Wolf YI, Covarrubias PC, Shmakov SA, Silva C, Tolstoy I, Johnson DB, Koonin EV, et al. Evolution of Type IV CRISPR-Cas systems: insights from CRISPR loci in Integrative Conjugative Elements of Acidithiobacillia. CRISPR J. 2021;4(5):656–72.
- Benz F, Camara-Wilpert S, Russel J, Wandera KG, Cepaite R, Ares-Arroyo M, Gomes-Filho JV, Englert F, Kuehn JA, Gloor S, et al. Type IV-A3 CRISPR-Cas systems drive inter-plasmid conflicts by acquiring spacers in trans. Cell Host Microbe. 2024;32(6):875–886.e879.
- Guo X, Sanchez-Londono M, Gomes-Filho JV, Hernandez-Tamayo R, Rust S, Immelmann LM, Schafer P, Wiegel J, Graumann PL, Randau L. Characterization of the self-targeting Type IV CRISPR interference system in Pseudomonas oleovorans. Nat Microbiol. 2022;7(11):1870–8.
- Crowley VM, Catching A, Taylor HN, Borges AL, Metcalf J, Bondy-Denomy J, Jackson RN. A Type IV-A CRISPR-Cas System in Pseudomonas aeruginosa Mediates RNA-Guided Plasmid Interference In Vivo. CRISPR J. 2019;2(6):434–40.
- Pinilla-Redondo R, Russel J, Mayo-Munoz D, Shah SA, Garrett RA, Nesme J, Madsen JS, Fineran PC, Sorensen SJ. CRISPR-Cas systems are widespread accessory elements across bacterial and archaeal plasmids. Nucleic Acids Res. 2022;50(8):4315–28.
- Wu WY, Mohanraju P, Liao C, Adiego-Perez B, Creutzburg SCA, Makarova KS, Keessen K, Lindeboom TA, Khan TS, Prinsen S, et al. The miniature CRISPR-Cas12m effector binds DNA to block transcription. Mol Cell. 2022;82(23):4487–4502.e4487.
- Peters JE, Makarova KS, Shmakov S, Koonin EV. Recruitment of CRISPR-Cas systems by Tn7-like transposons. Proc Natl Acad Sci U S A. 2017;114(35):E7358–66.
- Strecker J, Ladha A, Gardner Z, Schmid-Burgk JL, Makarova KS, Koonin EV, Zhang F. RNA-guided DNA insertion with CRISPR-associated transposases. Science. 2019;365(6448):48–53.
- Benler S, Faure G, Altae-Tran H, Shmakov S, Zheng F, Koonin E. Cargo Genes of Tn7-Like Transposons Comprise an Enormous Diversity of Defense Systems, Mobile Genetic Elements, and Antibiotic Resistance Genes. mBio. 2021;12(6):e0293821.

- Rybarski JR, Hu K, Hill AM, Wilke CO, Finkelstein IJ. Metagenomic discovery of CRISPR-associated transposons. Proc Natl Acad Sci USA. 2021;118(49):e2112279118.
- Faure G, Saito M, Benler S, Peng I, Wolf YI, Strecker J, Altae-Tran H, Neumann E, Li D, Makarova KS, et al. Modularity and diversity of target selectors in Tn7 transposons. Mol Cell. 2023;83(12):2122–2136.e2110.
- Hsieh SC, Peters JE. Discovery and characterization of novel type I-D CRISPR-guided transposons identified among diverse Tn7-like elements in cyanobacteria. Nucleic Acids Res. 2023;51(2):765–82.
- Klompe SE, Vo PLH, Halpin-Healy TS, Sternberg SH. Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. Nature. 2019;571(7764):219–25.
- Halpin-Healy TS, Klompe SE, Sternberg SH, Fernandez IS. Structural basis of DNA targeting by a transposon-encoded CRISPR-Cas system. Nature. 2020;577(7789):271–4.
- Xiao R, Wang S, Han R, Li Z, Gabel C, Mukherjee IA, Chang L. Structural basis of target DNA recognition by CRISPR-Cas12k for RNA-guided DNA transposition. Mol Cell. 2021;81(21):4457–4466.e4455.
- Schmitz M, Querques I, Oberli S, Chanez C, Jinek M. Structural basis for the assembly of the type V CRISPR-associated transposon complex. Cell. 2022;185(26):4999–5010.e4917.
- Park JU, Tsai AW, Rizo AN, Truong VH, Wellner TX, Schargel RD, Kellogg EH. Structures of the holo CRISPR RNA-guided transposon integration complex. Nature. 2023;613(7945):775–82.
- Tenjo-Castano F, Sofos N, Stutzke LS, Temperini P, Fuglsang A, Pape T, Mesa P, Montoya G. Conformational landscape of the type V-K CRISPRassociated transposon integration assembly. Mol Cell. 2024;84(12):2353– 2367.e2355.
- Hoffmann FT, Kim M, Beh LY, Wang J, Vo PLH, Gelsinger DR, George JT, Acree C, Mohabir JT, Fernandez IS, et al. Selective TnsC recruitment enhances the fidelity of RNA-guided transposition. Nature. 2022;609(7926):384–93.
- 46. Peters JE. Tn7. Microbiol Spectr. 2014;2(5). https://doi.org/10.1128/microbiolspec.mdna3-0010-2014.
- Petassi MT, Hsieh SC, Peters JE. Guide RNA Categorization Enables Target Site Choice in Tn7-CRISPR-Cas Transposons. Cell. 2020;183(7):1757-1771. e1718.
- Saito M, Ladha A, Strecker J, Faure G, Neumann E, Altae-Tran H, Macrae RK, Zhang F. Dual modes of CRISPR-associated transposon homing. Cell. 2021;184(9):2441–2453.e2418.
- Hu K, Chou CW, Wilke CO, Finkelstein IJ. Distinct horizontal transfer mechanisms for type I and type V CRISPR-associated transposons. Nat Commun. 2024;15(1):6653.
- 50. Yap ZL, Rahman A, Hogan AM, Levin DB, Cardona ST. A CRISPR-Cas-associated transposon system for genome editing in Burkholderia cepacia complex species. Appl Environ Microbiol. 2024;90(7):e0069924.
- Arevalo S, Perez Rico D, Abarca D, Dijkhuizen LW, Sarasa-Buisan C, Lindblad P, Flores E, Nierzwicki-Bauer S, Schluepmann H. Genome Engineering by RNA-Guided Transposition for Anabaena sp. PCC 7120. ACS Synth Biol. 2024;13(3):901–12.
- Gelsinger DR, Vo PLH, Klompe SE, Ronda C, Wang HH, Sternberg SH. Bacterial genome engineering using CRISPR-associated transposases. Nat Protoc. 2024;19(3):752–90.
- Lampe GD, King RT, Halpin-Healy TS, Klompe SE, Hogan MI, Vo PLH, Tang S, Chavez A, Sternberg SH. Targeted DNA integration in human cells without double-strand breaks using CRISPR-associated transposases. Nat Biotechnol. 2024;42(1):87–98.
- Koonin EV, Makarova KS, Wolf YI, Krupovic M. Evolutionary entanglement of mobile genetic elements and host defence systems: guns for hire. Nat Rev Genet. 2020;21(2):119–31.
- 55. Koonin EV, Makarova KS. Origins and evolution of CRISPR-Cas systems. Philos Trans R Soc Lond B Biol Sci. 2019;374(1772):20180087.
- Altae-Tran H, Kannan S, Demircioglu FE, Oshiro R, Nety SP, McKay LJ, Dlakic M, Inskeep WP, Makarova KS, Macrae RK, et al. The widespread IS200/ IS605 transposon family encodes diverse programmable RNA-guided endonucleases. Science. 2021;374(6563):57–65.
- Karvelis T, Druteika G, Bigelyte G, Budre K, Zedaveinyte R, Silanskas A, Kazlauskas D, Venclovas C, Siksnys V. Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. Nature. 2021;599(7886):692–6.

- Meers C, Le HC, Pesari SR, Hoffmann FT, Walker MWG, Gezelle J, Tang S, Sternberg SH. Transposon-encoded nucleases use guide RNAs to promote their selfish spread. Nature. 2023;622(7984):863–71.
- Altae-Tran H, Shmakov SA, Makarova KS, Wolf YI, Kannan S, Zhang F, Koonin EV. Diversity, evolution, and classification of the RNA-guided nucleases TnpB and Cas12. Proc Natl Acad Sci USA. 2023;120(48):e2308224120 in press.
- Krupovic M, Beguin P, Koonin EV. Casposons: mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery. Curr Opin Microbiol. 2017;38:36–43.

## **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.