Published in final edited form as:

Curr Opin Microbiol. 2017 August ; 38: 36–43. doi:10.1016/j.mib.2017.04.004.

Casposons: mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery

Mart Krupovic¹, Pierre Béguin¹, and Eugene V. Koonin²

¹Unité Biologie Moléculaire du Gène chez les Extrêmophiles, Institut Pasteur, 25 rue du Docteur Roux, 75015

²National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD 20894, USA

Abstract

A casposon, a member of a distinct superfamily of archaeal and bacterial self-synthesizing transposons that employ a recombinase (casposase) homologous to the Cas1 endonuclease, appears to have given rise to the adaptation module of CRISPR-Cas systems as well as the CRISPR repeats themselves. Comparison of the mechanistic features of the reactions catalyzed by casposases and the Cas1–Cas2 heterohexamer, the CRISPR integrase, reveals close similarity but also important differences that explain the requirement of Cas2 for integration of short DNA fragments, the CRISPR spacers.

Graphical Abstract

Introduction

Most of the archaea and many bacteria possess an adaptive (acquired) immunity system known as CRISPR (Clustered Regularly Interspaced Palindromic Repeats)-Cas (CRISPR-associated proteins). CRISPR-Cas systems specifically recognize and degrade, in a sequence-dependent manner, invading foreign DNA and/or RNA of viral or plasmid origin [1–5]. The system keeps memory of previously encountered mobile genetic elements (MGE)

Correspondence to: Mart Krupovic.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

by maintaining an archive of foreign DNA fragments within the CRISPR arrays, in which the fragments are integrated as unique spacers between identical, often palindromic, repeats of about 25–50 bp each [6,7]. The CRISPR array, including the spacers, is subsequently transcribed and processed into short RNA guides that consist of a spacer and part of an adjacent repeat [8,9]. The guide RNAs form a complex with a distinct set of Cas proteins which then recognizes and cleaves invading nucleic acids bearing a sequence that is identical or, in some cases, closely related to the spacer [10,11].

The CRISPR-Cas systems display remarkable diversity with respect to the protein components responsible for CRISPR RNA processing ('expression' stage) and target degradation ('interference' stage), and are currently partitioned into 2 classes, each encompassing several types and subtypes [3,12]. By contrast, the Cas1 and Cas2 proteins, which play the central roles in the 'adaptation' (immunization) stage, are nearly universally conserved across all CRISPR-Cas types and are considered to be the hallmark of the prokaryotic adaptive immunity [3,6]. Thus, the adaptation module appears to predate the evolution of the components involved in the expression and interference stages of the CRISPR-Cas response which, in all likelihood, joined the adaptation machinery on multiple independent occasions [12,13].

The mechanistic details of Cas1 and Cas2 action have been extensively studied both *in vivo* and *in vitro* by a combination of structural, biochemical and genetic approaches, resulting in a detailed picture of the CRISPR-Cas adaptation process. These findings are covered by several recent, extensive reviews [5–7]. By contrast, the origin of the adaptation machinery remained largely unresolved. Recently, however, a new superfamily of Cas1-encoding MGE, dubbed *Casposons*, has been discovered [14–16], suggesting a plausible scenario for the evolution of the CRISPR-Cas adaptation mechanism [17]. In the next sections, we briefly describe the currently observed casposon diversity and draw parallels between the reactions catalyzed by Cas1-like endonucleases encoded by casposons and CRISPR-Cas systems, respectively.

Casposons: a new superfamily of mobile genetic elements

Casposons have been discovered in the course of the analysis of genomic neighborhoods of *casI* genes that are not embedded within typical CRISPR-Cas loci. One such *casI* group is associated with genes encoding family B DNA polymerases (PolB) and several other conserved proteins. It was noticed that the loci encompassing *casI* and *polB* are surrounded by inverted repeats and further flanked by shorter direct repeats which, respectively [14], resemble terminal inverted repeats (TIR) and target site duplications (TSD) featured by various DNA transposons [18]. However, none of these putative novel transposons encoded typical transposases or integrases which are known to mediate integration of MGE into host chromosomes [19]. The only protein with activities predicted to be compatible with the integration reaction was a homolog of the Cas1 endonuclease that is conserved in CRISPR-Cas systems (hence the name Casposons for these novel MGEs). This prediction has been fully validated by subsequent *in vitro* experiments (see below). To distinguish between the *bona fide* Cas1 of CRISPR-Cas systems and Cas1 homologs from casposons, the latter enzyme is denoted 'casposase' [20,21].

The Casposons vary in size from ~8 kb to ~20 kb and have been identified in the genomes of many archaea and some bacteria. Based on comparative genomics, phylogenetic analysis of casposases and taxonomic distribution, casposons can be classified into 4 families (Figures 1 and 2) [14,16,22]. Family 1 elements are exclusive to members of the archaeal phylum *Thaumarchaeota* and encode PolBs of the protein-primed variety, closely related to the corresponding proteins from archaeal viruses. By contrast, casposons from the 3 other families encode divergent PolBs that are not closely similar to either protein-primed or RNA-primed PolBs [23]. Due to the fact that casposons encode the major protein required for their replication (i.e., DNA polymerase) they are included in the class of self-synthesizing (or self-replicating) MGEs, which also includes eukaryotic virus-like transposons of the Polinton/Maverick superfamily [16,24–26]. The latter encode protein-primed PolBs, RVE (retrovirus-like) family integrases and a set of proteins required for virion formation. Accordingly, it has been predicted that polintons actually form virus particles, under conditions that remain to be identified, and possibly, should be rebranded as polintoviruses [26,27]. Although virions so far have not been detected for polintons, related *bona fide* viruses, such as virophages, are known [26,28,29]. By contrast, casposons do not encode viral structural proteins or proteins involved in virion morphogenesis, and thus appear to be typical transposons [14,16].

Family 2 casposons are widespread in members of the archaeal order Euryarchaeota, particularly in methanogens, whereas Family 3 is specific to Bacteria and is represented in actinobacteria and proteobacteria. Family 4 is currently the least populated and is thus far restricted to certain species of the euryarchaeon *Methanosarcina mazei*. Notably, some *M. mazei* strains contain elements of both family 2 and family 4. Besides casposase and PolB, casposons of families 2, 3 and 4 share genes for an HNH nuclease and one or two helix-turn-helix (HTH) proteins. In addition, casposons encode a variable set of proteins, including different helicases and nucleases (Figure 1). Most notably, some casposons carry genes for Cas4-like nucleases, another protein that is present in many CRISPR-Cas systems, albeit not as widely as Cas1, and has been implicated in the adaptation process.

Casposons are integrated into diverse genomic targets. Thaumarchaeal casposons (family 1) target the 3'-distal region of the gene encoding translation elongation factor aEF-2, whereas all bacterial casposons (family 3) and family 4 casposons are inserted within intergenic regions. In contrast, family 2 casposons are integrated either into intergenic regions or into the 3'-distal region of diverse tRNA genes. Unlike typical transposons, casposons are present in only few (typically, one) copies per host genome. Nevertheless, evidence of casposon mobility *in vivo* was obtained through comparative genomic analysis of a large collection of *M. mazei* strains, which revealed striking variation of the casposon insertion sites among 62 strains of this archaeon and insertion of closely related casposons into different sites, indicative of multiple, recent gains and losses [22]. Interestingly, casposons compete with other types of MGE for integration sites and are occasionally inserted into other MGE that might provide vehicles for the casposon horizontal transfer [22].

Phylogenetic analysis has shown that casposases do not cluster with any particular group of CRISPR-associated Cas1 proteins, which is compatible with a basal position of casposons in the evolutionary tree of the Cas1 family [14,17]. These observations prompted an

evolutionary scenario under which a casposon was the ancestor of the adaptation module of CRISPR-Cas systems, having contributed *cas1* and possibly additional genes, such as *cas4*; the Cas2 protein also might have been encoded by this ancestral casposon although so far *cas2* genes have not been identified in casposons. Furthermore, the TIR of that ancestral element could have given rise to the CRISPR repeats [17]. This scenario is remarkably parallel to the proposed route for the origins of the vertebrate adaptive immunity systems and the chromatin diminution system in the ciliate macronucleus although the transposons involved are unrelated in each case [17,30]. Collectively, these findings on the role of MGE in defense systems imply that various MGEs are regularly recruited as “weapons” for cellular defense in the evolutionary arms race between pathogens and hosts [31]. Below we compare the activities and mechanistic details of DNA integration by CRISPR-Cas1 and the casposase and revise the scenario for the origin of the CRISPR adaptation module.

Cas1 family of endonucleases

The Cas1 family of endonucleases, which includes the casposases and CRISPR-Cas1 proteins, was unknown prior to the onset of intensive research on CRISPR-Cas systems but in the course of the early comparative analysis of the CRISPR-*cas* loci, it has been predicted that Cas1 proteins comprise a novel nuclease family [32]. Indeed, these proteins display a novel fold and are structurally unrelated to other characterized enzymes (Figure 2) [33]. CRISPR-Cas1 and casposase from a family 2 casposon of *Aciduliprofundum boonei*, the only casposase that has been biochemically studied thus far [20,21], are both dimers in solution [20,33–35]. Crystal structures have been solved for a number of CRISPR-Cas1 proteins. The fold consists of two structurally distinct domains, the N-terminal β -strand domain and the C-terminal α -helical domain, which are connected by a flexible linker [33–35]. The dimeric structure of Cas1 resembles a butterfly (Figure 2c). Although there are no high-resolution structures available for casposases, the relatively high sequence similarity between Cas1 and casposases and the conservation of the active site strongly suggest that the overall structure of the enzyme is the same in both cases.

Similar to the majority of CRISPR-Cas1, casposases from casposon families 1, 4 and some members of family 3 do not contain any additional domains (Figure 2b). However, the casposases of family 2 encompass a characteristic C-terminal HTH domains predicted to bind the casposon TIRs, whereas certain members of family 3 contain terminal extensions including Zn-ribbon motifs [14]. Notably, CRISPR-Cas1 is also occasionally fused to other proteins involved in the adaptation process, such as Cas2, Cas4 or a reverse transcriptase (Figure 2b) [3,36,37].

The active site of the Cas1 family endonucleases includes a conserved histidine (H208 in Cas1 from *Escherichia coli*) and three acidic residues (E141, D218 and D221), which are essential for the endonucleolytic activity [33,34,38]. Similar to active sites of many endonucleases, the histidine and two acidic residues (E141 and D221) coordinate a divalent metal cation. For both casposases and Cas1, the catalytic activities are much more pronounced in the presence of Mn^{2+} , compared to Mg^{2+} [20,33,39]. In characterized single-metal-ion nucleases, the metal ion serves to coordinate and subsequently destabilize the scissile phosphate for nucleophilic attack [40]. Cas1 enzymes from different organisms

display divalent metal ion-dependent nucleolytic activity against single-stranded (ss) RNA and ssDNA as well as various double-stranded DNA substrates [33–35,41], indicating that in the absence of a protospacer, Cas1 acts as a non-specific nuclease whereby a water molecule acts as a nucleophile. In presence of the protospacer, however, Cas1 catalyzes site-specific transesterification reaction. By contrast, the *A. boonei* casposase appears to be less promiscuous and cleaves the target DNA only in the presence of a linear DNA substrate corresponding to the TIR of the cognate casposon, which provides a 3' OH group as a nucleophile [20,21]. This observation implies that, unlike with the majority of CRISPR-Cas1, a water molecule cannot be substituted for the nucleophilic 3' OH group provided by the TIR (or protospacer) in the case of the transesterification reaction catalyzed by the casposase.

Despite strong indications of recent casposon mobility obtained by comparative genomics [22] and experimental characterization of the integration reaction, the mechanism of casposon excision from the host chromosome and the involvement of the casposase in this process remain unresolved. Attempts to detect cleavage of DNA substrates corresponding to the contiguous TSD-TIR region so far have not been successful [20]. It cannot be ruled out that the casposase acts exclusively as an integrase, whereas new linear copies of the casposon are synthesized *de novo* from the integrated element by the casposon-encoded PolB.

Integrase activities of the casposase and Cas1

Efficient protospacer integration by Cas1 requires the presence of the Cas2 protein [39]. Two dimers of Cas1 are connected by a dimer of Cas2 within a heterohexameric complex [38]. Cas1 is the enzymatically active component required for protospacer integration, whereas the nuclease activity of Cas2 is not necessary, because Cas1–Cas2 complex containing Cas2 active site mutant is as active as the wild-type both *in vivo* and *in vitro* [38,39]. These results indicate that Cas2 plays a structural, scaffolding role during the integration process. Nevertheless, the conservation of the Cas2 nuclease active site is intriguing and the stage of the immune response in which Cas2 activity might be involved remains to be determined. Notably, protospacer integration can be achieved even in the absence of Cas2, albeit at low efficiency [39], whereas the reverse, “disintegration”, reaction is Cas2-independent [42]. In the case of casposons, the casposase is the only protein that is required and sufficient for integration *in vitro* [20,21].

Although initially it has been concluded that *A. boonei* casposase catalyzes casposon integration into random sites [20], subsequent study has demonstrated a strong sequence preference of the enzyme in the presence of a proper target site [21]. Importantly, the sequence features of the corresponding target sites during casposon and protospacer integration display remarkable similarity. In both cases, the functional target site consists of two components: (i) a sequence that gets duplicated upon integration of the incoming DNA duplex (the TSD segment and a CRISPR unit, respectively) and (ii) an upstream region which further determines the exact location of the integration (the Leader sequence located upstream of the CRISPR array and the 18-bp segment encoding the T^ΨC loop of tRNA-Pro in *A. boonei*) (Figure 3). Preferential integration of protospacers at the leader-repeat 1

border is an important property of the CRISPR-Cas systems, which ensures a sequential historical account of prior encounters between the host and MGEs [6,43]. Similarly, the inherent features of the target site can be predicted to allow multiple casposon integrations in tandem, each new casposon being inserted upstream of the preceding one. Such tandem casposon arrays, with individual casposons being separated from each other by a TSD segment have been indeed identified in the genomes of certain archaea (e.g., *Methanosarcina* sp. 2.H.A.1B.4 and 2.H.T.1A.3) [22]. Notably, integration by the *A. boonei* casposase does not depend on the DNA conformation and can proceed with relaxed, linear DNA [20,21], whereas the type I-E CRISPR Cas1–Cas2 complex of *E. coli* requires supercoiled substrates, unless bending of the target DNA is effected by the host IHF protein [44,45]. Although IHF is restricted to a subset of bacteria, it has been suggested that other organisms employ alternative DNA-binding proteins to provide the necessary deformation of the CRISPR Leader for Cas1–Cas2 recognition of the target site [44]; this prediction remains to be validated experimentally.

Despite differences in the composition of the integrase complex, protospacer and casposon integration reactions catalyzed by Cas1–Cas2 and casposases, respectively, are closely similar (Figure 3). In both systems, the integration reaction appears to proceed via a two-step mechanism, whereby the first nucleophilic attack occurs at the leader-repeat or the tRNA-TSD segment border by the 3'-OH end of the spacer or the casposon, respectively [21,44]. After the formation of the half-site intermediate, the second nucleophilic attack occurs on the opposite strand at the repeat-spacer border (for CRISPR spacer) or at the junction between the TSD segment and the flanking DNA, distal to the tRNA gene (for casposon). Notably, in the case of the *A. boonei* casposon, the half-site intermediate can be formed on either side of the target site [21]. A site for the second nucleophilic attack appears to be selected by a molecular ruler, i.e., the distance between the active sites in the Cas1 or casposase dimers rather than by the nucleotide sequence [21,38,39]. Consistent with this possibility is the observation that ectopic casposon integrations occurring outside of the canonical TSD result in 15-bp duplications [20,21]. The resulting single-stranded DNA gaps are repaired by still uncharacterized mechanisms, resulting in the duplication of the repeat and the target segment. Thus, the insertion of casposons appears to be deeply similar to the insertion of CRISPR spacers, further emphasizing the apparent direct evolutionary relationship between the two systems.

Concluding remarks

We have previously hypothesized that CRISPR repeats have evolved from the casposon TIRs [17]. However, biochemical and structural work on CRISPR-Cas1 and the casposase outlined above, prompts us to revise this aspect of the original scenario for the evolution of the CRISPR-Cas adaptation module. In particular, given the similar modes of target selection during casposon and protospacer integration, it appears that, following the immobilization of the ancestral casposon via inactivation of the TIRs, the CRISPR repeats evolved directly from the preexisting casposon target site, rather than from the casposon TIRs (Figure 4). Indeed, it has been shown that Cas1, in the absence of Cas2, displays intrinsic sequence specificity towards the nucleotides flanking the integration site at the leader-repeat 1 boundary of the CRISPR locus [42]. Similarly, casposase alone is sufficient

to direct specific integration of casposons into the target site [21]. In addition, the new experimental data strongly suggests that the Leader sequence also evolved directly from the target site employed by the ancestral casposon, further extending the contribution of casposons to the evolution of the CRISPR-Cas adaptation module.

Another key step that likely precipitated the evolution of the CRISPR-Cas adaptation machinery from casposons, was recruitment of Cas2 for the formation of the pseudo-symmetrical heterohexameric complex. In the case of casposons, the two ends of the linear casposon DNA destined for integration are identical and can be both specifically recognized by the casposase. Potentially, each monomer within the casposase homodimer coordinates one terminus of the casposon DNA, with the rest of the casposon sequence forming a loop. This way, the two termini are poised for integration at a fixed distance, in accordance with the molecular ruler model outlined above. By contrast, the two ends of a protospacer are not identical. The protospacer end containing the protospacer adjacent motif (PAM)-complementary sequence is recognized in a base-specific fashion by one of the four Cas1 subunits, whereas the accommodation of the other protospacer terminus is likely to be dictated by the quaternary structure of the Cas1–Cas2 complex [46,47]. Furthermore, the persistence length of dsDNA is 35–50 nm (100–150 bp) [48,49], which precludes bending of the short (30–70 bp [50]) protospacers such that both ends of the molecule would be held by the homodimer. Thus, recruitment of Cas2 appears to be crucial for transforming the casposase into Cas1, i.e. an enzyme that specifically mediates integration of short DNA fragments. Cas2 is homologous to mRNA interferases found in toxin-antitoxin modules [32,51], which could have already been present in the casposon that gave rise to CRISPR-Cas adaptation machinery (Figure 4) [17]. However, a casposon carrying such a toxin-antitoxin-like module remains to be discovered. Alternatively, the Cas2-containing toxin-antitoxin system could have been associated with the ‘solo-Effector’ that gave rise to the Effector module. Notably, C2c3 (type V-C) [13] and CRISPR-CasY [52] loci that have been recently discovered in the genomes of uncultured microbes encode Cas1 but not Cas2. It will be interesting to test if in these systems Cas1 is sufficient for protospacer integration or whether Cas2 is recruited from a different CRISPR-Cas system located elsewhere in the respective genomes.

Acknowledgments

EVK was supported by intramural funds of the US Department of Health and Human Services (to the National Library of Medicine).

References

1. Barrangou R, Marraffini LA. CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Mol Cell*. 2014; 54:234–244. [PubMed: 24766887]
2. Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*. 2014; 346:1258096. [PubMed: 25430774]
- 3**. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJ, Charpentier E, Haft DH, et al. An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol*. 2015; 13:722–736. This article describes computational approaches for identification of CRISPR-cas loci and an updated classification of CRISPR-Cas systems. The two

classes are introduced for the first time, and a comprehensive census of CRISPR-Cas types and subtypes in archaeal and bacterial genomes is presented. [PubMed: 26411297]

4. Marraffini LA. CRISPR-Cas immunity in prokaryotes. *Nature*. 2015; 526:55–61. [PubMed: 26432244]
5. 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. [PubMed: 27493190]
6. Amitai G, Sorek R. CRISPR-Cas adaptation: insights into the mechanism of action. *Nat Rev Microbiol*. 2016; 14:67–76. [PubMed: 26751509]
7. Sternberg SH, Richter H, Charpentier E, Qimron U. Adaptation in CRISPR-Cas Systems. *Mol Cell*. 2016; 61:797–808. [PubMed: 26949040]
8. Hochstrasser ML, Doudna JA. Cutting it close: CRISPR-associated endoribonuclease structure and function. *Trends Biochem Sci*. 2015; 40:58–66. [PubMed: 25468820]
9. 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. [PubMed: 25994611]
10. Plagens A, Richter H, Charpentier E, Randau L. DNA and RNA interference mechanisms by CRISPR-Cas surveillance complexes. *FEMS Microbiol Rev*. 2015; 39:442–463. [PubMed: 25934119]
11. Tsui TK, Li H. Structure Principles of CRISPR-Cas Surveillance and Effector Complexes. *Annu Rev Biophys*. 2015; 44:229–255. [PubMed: 26048003]
- 12*. Shmakov S, Smargon A, Scott D, Cox D, Pyzocha N, Yan W, Abudayyeh OO, Gootenberg JS, Makarova KS, Wolf YI, et al. Diversity and evolution of class 2 CRISPR-Cas systems. *Nat Rev Microbiol*. 2017 This paper describes the latest developments in the discovery of novel variants of Class 2 CRISPR-Cas systems and reveals multiple contributions of mobile elements to the evolution of CRISPR-Cas.
13. Shmakov S, Abudayyeh OO, Makarova KS, Wolf YI, Gootenberg JS, Semenova E, Minakhin L, Joung J, Konermann S, Severinov K, et al. Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol Cell*. 2015; 60:385–397. [PubMed: 26593719]
14. 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 Biology*. 2014; 12:36. [PubMed: 24884953]
15. Hickman AB, Dyda F. CRISPR-Cas immunity and mobile DNA: a new superfamily of DNA transposons encoding a Cas1 endonuclease. *Mob DNA*. 2014; 5:23. [PubMed: 25180049]
16. 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. [PubMed: 26836982]
- 17**. Koonin EV, Krupovic M. Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nat Rev Genet*. 2015; 16:184–192. This work outlines the scenario of evolution of CRISPR-Cas from a casposon and an ancestral innate immunity locus, and draws parallels between the evolutionary trajectories of adaptive immunity in prokaryotes and animals. [PubMed: 25488578]
18. Hickman AB, Dyda F. Mechanisms of DNA Transposition. *Microbiol Spectr*. 2015; 3 MDNA3-0034-2014.
19. Craig, NL.Chandler, M.Gellert, M.Lambowitz, AM.Rice, PA., Sandmeyer, SB., editors. *Mobile DNA III*. 3. Washington, DC: ASM Press; 2015.
- 20**. 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. The first demonstration of the integrase activity of the casposase in vitro. [PubMed: 26573596]
- 21**. Béguin 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. This work investigates the mechanism of integration by the casposase and demonstrates the close similarity between the casposase target site and the leader-repeat junction in CRISPR-Cas. [PubMed: 27655632]

- 22*. Krupovic M, Shmakov S, Makarova KS, Forterre P, Koonin EV. Recent Mobility of Casposons, Self-Synthesizing Transposons at the Origin of the CRISPR-Cas Immunity. *Genome Biol Evol.* 2016; 8:375–386. This work takes advantage of the availability of many genomes of *Methanosarcina mazei* strains with multiple casposon insertions to demonstrate recent mobility of casposons, thus showing that these are active MGE. [PubMed: 26764427]
23. Makarova KS, Krupovic M, Koonin EV. Evolution of replicative DNA polymerases in archaea and their contributions to the eukaryotic replication machinery. *Front Microbiol.* 2014; 5:354. [PubMed: 25101062]
24. Kapitonov VV, Jurka J. Self-synthesizing DNA transposons in eukaryotes. *Proc Natl Acad Sci U S A.* 2006; 103:4540–4545. [PubMed: 16537396]
25. Pritham EJ, Putliwala T, Feschotte C. Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene.* 2007; 390:3–17. [PubMed: 17034960]
26. Krupovic M, Koonin EV. Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. *Nat Rev Microbiol.* 2015; 13:105–115. [PubMed: 25534808]
27. Krupovic M, Bamford DH, Koonin EV. Conservation of major and minor jelly-roll capsid proteins in Polinton (Maverick) transposons suggests that they are bona fide viruses. *Biol Direct.* 2014; 9:6. [PubMed: 24773695]
28. Koonin EV, Dolja VV, Krupovic M. Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology.* 2015; 479–480:2–25.
29. Fischer MG, Suttle CA. A virophage at the origin of large DNA transposons. *Science.* 2011; 332:231–234. [PubMed: 21385722]
30. Nowacki M, Shetty K, Landweber LF. RNA-Mediated Epigenetic Programming of Genome Rearrangements. *Annu Rev Genomics Hum Genet.* 2011; 12:367–389. [PubMed: 21801022]
31. Koonin EV, Krupovic M. A movable defense. *Scientist.* 2015; 29:46–53.
32. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct.* 2006; 1:7. [PubMed: 16545108]
33. Wiedenheft B, Zhou K, Jinek M, Coyle SM, Ma W, Doudna JA. Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure.* 2009; 17:904–912. [PubMed: 19523907]
34. Babu M, Beloglazova N, Flick R, Graham C, Skarina T, Nocek B, Gagarinova A, Pogoutse O, Brown G, Binkowski A, et al. A dual function of the CRISPR-Cas system in bacterial antiviral immunity and DNA repair. *Mol Microbiol.* 2011; 79:484–502. [PubMed: 21219465]
35. Kim TY, Shin M, Huynh Thi Yen L, Kim JS. Crystal structure of Cas1 from *Archaeoglobus fulgidus* and characterization of its nucleolytic activity. *Biochem Biophys Res Commun.* 2013; 441:720–725. [PubMed: 24211577]
36. Plagens A, Tjaden B, Hagemann A, Randau L, Hensel R. Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon *Thermoproteus tenax*. *J Bacteriol.* 2012; 194:2491–2500. [PubMed: 22408157]
37. Silas S, Mohr G, Sidote DJ, Markham LM, Sanchez-Amat A, Bhaya D, Lambowitz AM, Fire AZ. Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science.* 2016; 351:aad4234. [PubMed: 26917774]
38. Nunez JK, Kranzusch PJ, Noeske J, Wright AV, Davies CW, Doudna JA. Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat Struct Mol Biol.* 2014; 21:528–534. [PubMed: 24793649]
- 39*. Nunez JK, Lee AS, Engelman A, Doudna JA. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature.* 2015; 519:193–198. A detailed characterization of the role of the Cas1–Cas2 complex and its two subunits in spacer acquisition. [PubMed: 25707795]
40. Yang W. An equivalent metal ion in one- and two-metal-ion catalysis. *Nat Struct Mol Biol.* 2008; 15:1228–1231. [PubMed: 18953336]

41. Wilkinson ME, Nakatani Y, Staals RH, Kieper SN, Opel-Reading HK, McKenzie RE, Fineran PC, Krause KL. Structural plasticity and in vivo activity of Cas1 from the type I-F CRISPR-Cas system. *Biochem J*. 2016; 473:1063–1072. [PubMed: 26929403]
- 42*. Rollie C, Schneider S, Brinkmann AS, Bolt EL, White MF. Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *Elife*. 2015; 4:e08716. This work shows that Cas2 is not required for the disintegration reaction catalyzed by Cas1 and suggests that the inherent sequence specificity of cas1 is the principal specificity determinant of spacer acquisition.
- 43**. Wright AV, Doudna JA. Protecting genome integrity during CRISPR immune adaptation. *Nat Struct Mol Biol*. 2016; 23:876–883. This work reveals the details of the recognition of the spacer insertion site at the leader-repeat junction by the Cas1–Cas2 complex and provides evidence in support of the model in which the palindromic ends of the repeat are recognized by distal Cas1 active sites. [PubMed: 27595346]
- 44*. Nunez JK, Bai L, Harrington LB, Hinder TL, Doudna JA. CRISPR Immunological Memory Requires a Host Factor for Specificity. *Mol Cell*. 2016; 62:824–833. This work demonstrates the requirement of *E. coli* integration host factor (IHF) for spacer integration into linear DNA by Cas1–Cas2. It is shown that IHF introduces a sharp bend into the target DNA molecule, allowing the Cas1–Cas2 integrase to catalyze the first integration reaction. [PubMed: 27211867]
45. Yoganand KN, Sivathanu R, Nimkar S, Anand B. Asymmetric positioning of Cas1–2 complex and Integration Host Factor induced DNA bending guide the unidirectional homing of protospacer in CRISPR-Cas type I-E system. *Nucleic Acids Res*. 2017; 45:367–381. [PubMed: 27899566]
- 46**. Wang J, Li J, Zhao H, Sheng G, Wang M, Yin M, Wang Y. Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell*. 2015; 163:840–853. This paper together with ref [47] describe X-ray structures of the tertiary complexes of Cas1–Cas2 heterohexamers with the bound protospacers. The protospacer in ref [46] contains the PAM-complementary sequence, revealing site-specific interactions between this motif and the active site of Cas1. [PubMed: 26478180]
47. Nunez JK, Harrington LB, Kranzusch PJ, Engelman AN, Doudna JA. Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature*. 2015; 527:535–538. [PubMed: 26503043]
48. Brinkers S, Dietrich HR, de Groote FH, Young IT, Rieger B. The persistence length of double stranded DNA determined using dark field tethered particle motion. *J Chem Phys*. 2009; 130:215105. [PubMed: 19508104]
49. Mitchell JS, Glowacki J, Grandchamp AE, Manning RS, Maddocks JH. Sequence-dependent persistence lengths of DNA. *J Chem Theory Comput*. 2016; doi: 10.1021/acs.jctc.6b00904
50. Sorek R, Kunin V, Hugenholtz P. CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol*. 2008; 6:181–186. [PubMed: 18157154]
51. Kwon AR, Kim JH, Park SJ, Lee KY, Min YH, Im H, Lee I, Lee BJ. Structural and biochemical characterization of HP0315 from *Helicobacter pylori* as a VapD protein with an endoribonuclease activity. *Nucleic Acids Res*. 2012; 40:4216–4228. [PubMed: 22241770]
52. Burstein D, Harrington LB, Strutt SC, Probst AJ, Anantharaman K, Thomas BC, Doudna JA, Banfield JF. New CRISPR-Cas systems from uncultivated microbes. *Nature*. 2016; doi: 10.1038/nature21059

Highlights

- Casposons are self-replicating mobile genetic elements found in bacteria and archaea
- Casposase is an active casposon integrase homologous to CRISPR-Cas1
- Casposons gave rise to the adaptation machinery of CRISPR-Cas
- CRISPR repeats and the Leader sequence have evolved from the casposon target site
- Recruitment of Cas2 was crucial for transforming the casposase into Cas1

Figure 1.

Genome organizations the casposons. Genome maps of representatives of the four casposon families. Family 1: NitSJ-C1 from *Nitrosopumilus* sp. SJ; Family 2: AciBoo-C1 from *Aciduliprofundum boonei* T469; Family 3: HenMar-C1 from *Henriciella marina* DSM 19595; Family 4: MetMaz1FA1A3-C1 from *Methanosarcina mazei* strain 1.F.A.1A.3. Abbreviations: TIR, terminal inverted repeat; Cas1, casposase; (p)DNAP, (protein-primed) family B DNA polymerase; HNH, HNH family endonuclease; HTH, helix-turn-helix proteins; MTase, methyltransferase; S1H and S2H, superfamily 1 and 2 helicases.

Figure 2.

Cas1 family endonucleases. (a) Phylogenetic relationship between Cas1 family endonucleases encoded by CRISPR-Cas systems (grey triangle) and casposons (blue triangles). (b) Diversity of domain organizations of Cas1/casposase endonucleases in different casposon families and CRISPR-Cas systems. Cas1/casposase endonuclease domains are in light blue, helix-turn-helix (HTH) domain in pink, Zn-ribbon (ZnR) domain in yellow, Cas4 endonuclease in red, Cas2 in dark blue, and reverse transcriptase in green. The representative proteins are labeled with their GenBank accession numbers. (c) X-ray structure of the CRISPR-Cas1 dimer from *Pseudomonas aeruginosa* (PDB id: 3GOD) [33]. The 4 active site residues are indicated using stick representation and are colored dark blue. The Mn²⁺ ions are shown as green spheres.

Figure 3. Comparison of models of CRISPR spacer (left) and casposon (right) integration. The figure has been adapted from Bguin et al [21].

Figure 4.

A revised scenario for the origin of the adaptation module of CRISPR-Cas systems as well as the CRISPR repeats (R) and the Leader sequence from a casposon. The ancestral casposon is postulated to encode Cas1, Cas2 and Cas4 although *cas2* genes so far have not been identified in casposons. Alternatively, Cas2 might originate from a toxin-antitoxin system associated with the solo-Effector module.