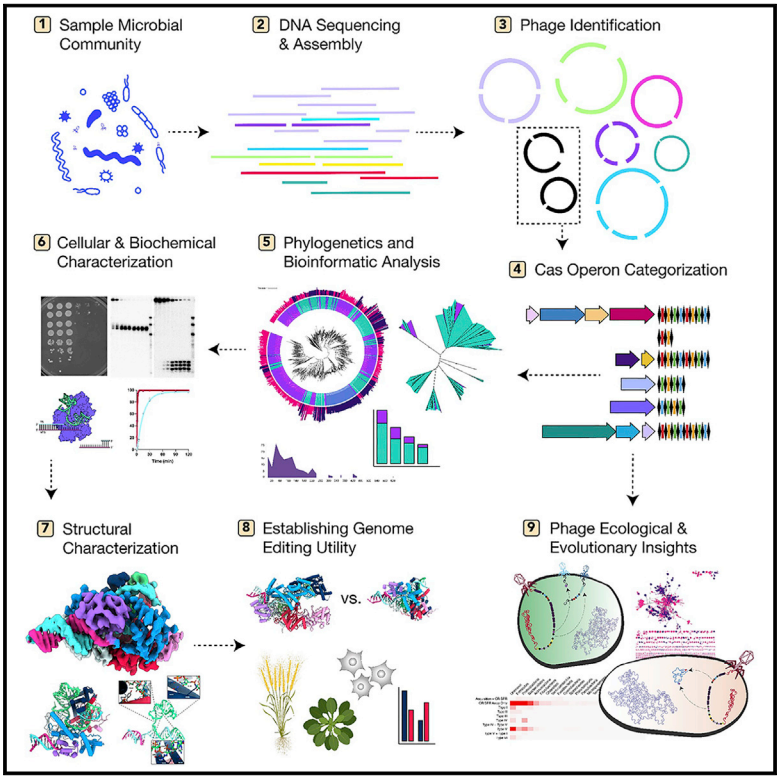
Cell

Article

Diverse virus-encoded CRISPR-Cas systems include streamlined genome editors

Graphical abstract



Highlights

- CRISPR pathways encoded in diverse bacteriophage are hypercompact anti-viral systems
- Phage-encoded CRISPR systems encompass all known CRISPR-Cas types
- The Casλ enzyme's compact structure is capable of plant and human cell genome editing
- These findings reveal a new source of CRISPR-Cas enzymes with value as genome editors

Authors

Basem Al-Shayeb, Petr Skopintsev, Katarzyna M. Soczek, ..., Steven E. Jacobsen, Jillian F. Banfield, Jennifer A. Doudna

Correspondence

jbanfield@berkeley.edu (J.F.B.), doudna@berkeley.edu (J.A.D.)

In brief

A tour-de-force metagenomics analysis of phage-encoded CRISPR-Cas enzymes provides a rich resource for new CRISPR-based tools.

Al-Shayeb et al., 2022, Cell *185*, 4574–4586
November 23, 2022 © 2022 The Authors. Published by Elsevier Inc.
https://doi.org/10.1016/j.cell.2022.10.020

CellPress

CellPress
OPEN ACCESS

Cell

Article

Diverse virus-encoded CRISPR-Cas systems include streamlined genome editors

Basem Al-Shayeb,[1,2,8,9,10,11,12,13] Petr Skopintsev,[2,3,4] Katarzyna M. Soczek,[2,3,4] Elizabeth C. Stahl,[2,3,4] Zheng Li,[5] Evan Groover,[1,2] Dylan Smock,[2] Amy R. Eggers,[2,3] Patrick Pausch,[2,3,14] Brady F. Cress,[2] Carolyn J. Huang,[2,3] Brian Staskawicz,[1,2] David F. Savage,[2,3,6] Steven E. Jacobsen,[5,7] Jillian F. Banfield,[2,8,9,10,*] and Jennifer A. Doudna[2,3,4,6,11,12,13,15,*]

[1]Department of Plant and Microbial Biology, University of California, Berkeley, CA, USA
[2]Innovative Genomics Institute, University of California, Berkeley, CA, USA
[3]Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA
[4]California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, CA, USA
[5]Department of Molecular, Cellular and Developmental Biology, University of California, Los Angeles, CA, USA
[6]Howard Hughes Medical Institute, University of California, Berkeley, CA, USA
[7]Howard Hughes Medical Institute, University of California, Los Angeles, CA, USA
[8]Department of Earth and Planetary Science, University of California, Berkeley, CA, USA
[9]Department of Environmental Science, Policy and Management, University of California, Berkeley, CA, USA
[10]University of Melbourne, Melbourne, Australia
[11]Department of Chemistry, University of California, Berkeley, CA, USA
[12]MBIB Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[13]Gladstone Institutes, University of California, San Francisco, CA, USA
[14]Present address: VU LSC-EMBL Partnership for Genome Editing Technologies, Life Sciences Center, Vilnius University, Vilnius, Lithuania
[15]Lead contact
*Correspondence: jbanfield@berkeley.edu (J.F.B.), doudna@berkeley.edu (J.A.D.)
https://doi.org/10.1016/j.cell.2022.10.020

SUMMARY

CRISPR-Cas systems are host-encoded pathways that protect microbes from viral infection using an adaptive RNA-guided mechanism. Using genome-resolved metagenomics, we find that CRISPR systems are also encoded in diverse bacteriophages, where they occur as divergent and hypercompact anti-viral systems. Bacteriophage-encoded CRISPR systems belong to all six known CRISPR-Cas types, though some lack crucial components, suggesting alternate functional roles or host complementation. We describe multiple new Cas9-like proteins and 44 families related to type V CRISPR-Cas systems, including the Casλ RNA-guided nuclease family. Among the most divergent of the new enzymes identified, Casλ recognizes double-stranded DNA using a uniquely structured CRISPR RNA (crRNA). The Casλ-RNA-DNA structure determined by cryoelectron microscopy reveals a compact bilobed architecture capable of inducing genome editing in mammalian, *Arabidopsis*, and hexaploid wheat cells. These findings reveal a new source of CRISPR-Cas enzymes in phages and highlight their value as genome editors in plant and human cells.

INTRODUCTION

CRISPR-Cas systems confer resistance in prokaryotes against invading extrachromosomal elements, including viruses and plasmids[1] (Figure 1A). To generate immunological memory, microbes capture fragments of foreign genetic elements and incorporate them into their genomic CRISPR array using the Cas1-Cas2 integrase.[1] Subsequent transcription of the array creates CRISPR RNAs (crRNAs) that bind to and direct CRISPR-associated (Cas) nucleases to target complementary nucleic acids. These systems comprise two classes, each with three different types, defined by the architectures of their nuclease effector modules involved in crRNA processing and DNA or RNA interference.[2]

Reports of type I-F CRISPR-Cas loci encoded in bacteriophages (phage) that infect *Vibrio cholera*,[3,4] or type V CRISPR-Cas loci in huge phage genomes reconstructed from microbial community DNA sequences,[5] hinted at a wider distribution of phage-encoded CRISPR systems that might function in novel ways and reveal important insights into prokaryotic biology. However, these were the only reported examples of phage-encoded CRISPR systems, prompting us to perform a comprehensive study of the abundance, distribution, and diversity of CRISPR-Cas systems encoded throughout the virosphere and to begin to explore the biochemical activity of novel systems.

Here, using metagenomic analysis of microbial samples isolated from soil, aquatic, human, and animal microbiomes, we report the widespread occurrence of diverse, compact

 © 2022 The Authors. Published by Elsevier Inc.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

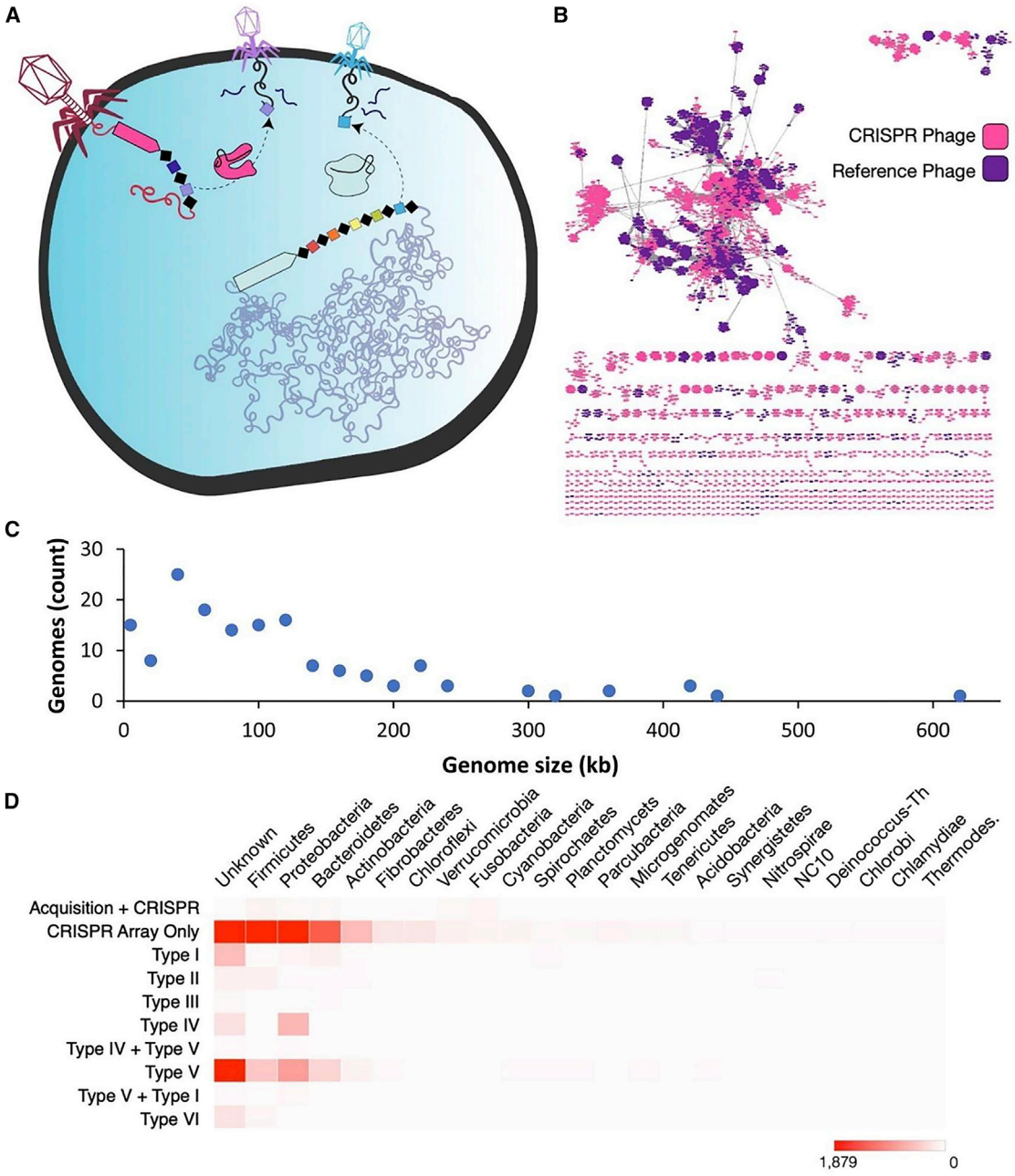


Figure 1. Diversity of CRISPR-encoding phages and the hosts they predate

(A) An illustration of the mechanism of CRISPR interference as an anti-viral system used by bacteria and bacteriophages.

(B) Protein-clustering network analysis based on the number of shared protein clusters between the CRISPR-encoding phages in this study and RefSeq phages. The plot is composed of viral clusters where each node represents a phage genome, and each edge is the hypergeometric similarity between genomes based on shared protein clusters.

(C) Genome size distribution of circularized CRISPR-containing phages from this study (n = 152).

(D) A heatmap showing the number of CRISPR phage genomes containing each CRISPR type with respect to major bacterial phyla which they infect. "Unknown" indicates CRISPR phages that could not be assigned to any of the known types. Phyla are organized in the plot from left to right based on CRISPR array abundance and are concordant with the number of bacterial genomes available for each phylum.

See also Figure S2.

CRISPR-Cas systems encoded in phage genomes, demonstrating an unexpected biological reservoir of anti-viral machinery within infectious agents. Phage-encoded CRISPR-Cas systems include members of all six CRISPR types (types I–VI) as defined by bacterially encoded examples. We found evidence for new or alternative modes of nucleic acid interference involving phage-encoded type I, III, IV, and VI systems. In addition, the phage and phage-like sequences result in a several-fold expansion of CRISPR-Cas9 and -Cas12 enzymes belonging to the type II and type V families that are widely deployed for genome editing applications. Casλ, one of the most divergent in sequence of the phage-encoded type V enzymes identified in this study, was found to have robust biochemical activity as an RNA-guided double-stranded DNA (dsDNA) cutter. Its cryoelectron microscopy (cryo-EM)-determined molecular structure explains its use of a natural single-guide RNA for DNA binding, and cell-based experiments demonstrated robust endogenous genome editing activity in plant and human cells. The compact architecture of Casλ and other phage-encoded CRISPR-Cas proteins holds significant promise for vector-based and direct delivery into cells for wide-ranging biotechnological applications.

RESULTS

A wide diversity of phages across many bacterial phyla encode divergent CRISPR-Cas systems

Using genome-resolved metagenomics, we analyzed over 660 giga base pairs (bp) of assembled genomic DNA from both environmental and animal-associated microbiomes to reveal a surprising diversity of over 6000 CRISPR-encoding phages (Figure 1B). Our analysis of publicly available phage genomes revealed that CRISPR-Cas systems occur in only 0.4% of phages, making them exceptionally rare compared to their abundance in prokaryotic genomes where they occur in 40% of bacteria and 85% of archaea. CRISPR-encoding phages, rather than being limited to specific phylogenetic clusters, are found within many diverse phage subtypes (Figure 1B). This finding is consistent with previous work that determined phage phylogeny from protein-clustering analysis.[5] At least two phages harboring CRISPR arrays were alternatively coded such that the TAG stop codon was recoded to glutamine. Although circularized CRISPR-encoding phages included huge phages such as a >620 kbp megaphage (Figure 1C), most had a genome size close to the average of 52 kbp.[5] Notably, however, relatively few phages encode complete CRISPR-Cas systems. Fewer than 10% of CRISPR-encoding phages were found to contain machinery for the acquisition of new spacer sequences into their CRISPR arrays, consistent with observations in huge phages.[5] Many phages encode CRISPR arrays, but few of these (~6%) include Cas effectors encoded nearby (Figure 1D). In such situations, phages may produce their own guide RNAs but hijack the Cas effectors provided by their hosts. Consistent with this possibility, ~1% of phages encode only the Cas1-Cas2 integrase used for the acquisition of new spacers, but no other Cas enzymes. In some cases, phage-encoded Cas1 contained a fusion to another protein such as reverse transcriptase, suggesting the possibility of the acquisition of RNA protospacers into the phage array. Notably, only 27 of the thousands of phage-encoded CRISPR-Cas loci identified in this study target RNA and can be classified as new homologs of previously described RNA-targeting systems. Thus, the vast majority of phage encoded CRISPR systems target DNA.

Phage-encoded CRISPR-Cas systems include all six known types but with phage-specific properties

We found that all six known types of CRISPR-Cas systems occur in phages, and relative to host-encoded systems, they have various unique properties associated with their existence within phage genomes. These include missing sequence integration or targeting machinery as mentioned above, modified type III and VI systems that mitigate the abortive infection mechanism, and spacers that target other mobile genetic elements.

For example, some of the rare phage-encoded type III systems are associated with CRISPR arrays targeting vital or highly abundant RNA transcripts of other mobile elements, such as phage tail proteins or transposases (Figure 2). In well-studied type III systems, the Cas10 protein converts ATP into a cyclic oligoadenylate (cOA) product, which allosterically activates an auxiliary Csm6 ribonuclease.[6] The activated Csm6 amplifies the immune response by degrading RNA transcripts indiscriminately, thereby destroying the invasive transcriptome or inducing host cell dormancy or death, aborting the phage infectious cycle.[6–9] Interestingly, in phage-encoded type III systems, the Cas10 subunit contains multiple mutations, hinting at an inability to produce cOA (Figure S1), and Csm6 or a related CARF-domain ribonuclease is absent, similarly to archaeal Borg elements.[10] Notably, the key residues for DNA cleavage in the Cas10 HD domain, and for RNA cleavage in Cas7, remain intact (Figure S1). Unless the cOA production and Csm6 ribonuclease functionalities are complemented by orthogonal type III systems from the host genome, this suggests that the type III phage systems may be capable of cleaving key RNA transcripts and genomic DNA of competing mobile elements to interfere with their infectious cycle without activating abortive infection in which cOA signaling triggers *trans*-cleavage of transcripts in the host cell.

In addition to type III systems, we found the first examples of phage-encoded type VI (Cas13) ribonucleases, most of which belong to the Cas13b and the relatively small Cas13d superfamilies (Figure 3). Analogous to the findings above with type III systems in abortive infection, the lack of signature Csx27 and Csx28 proteins, which are transmembrane factors that enhance abortive infection mechanisms,[11] may indicate the absence of an abortive infection pathway unless supplemented by the host.

Miniature single-effector CRISPR-Cas systems are enriched in phage genomes

Class 2 CRISPR-Cas systems, including types II, V, and VI, generally employ single-subunit RNA-guided, nucleic acid-targeting interference enzymes. In addition to new Cas9 (a, b, c) and Cas12 (a, b, c, f, i) enzyme variants, we identified miniature CRISPR-associated nucleases in phages harboring both HNH and RuvC catalytic domains characteristic of Cas9. These

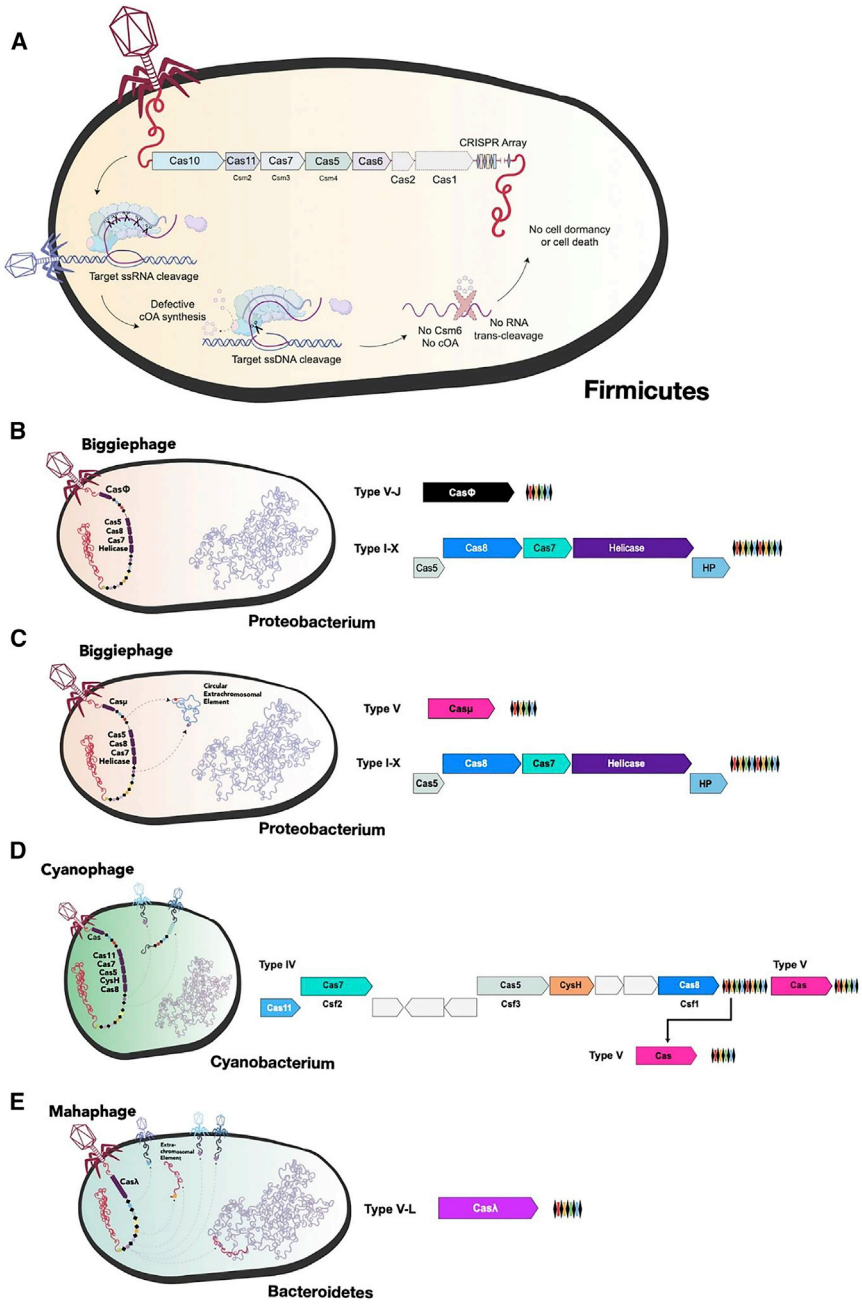


Figure 2. Diversity of phage-encoded CRISPR systems highlights anti-phage capability.

(A) Phage CRISPR spacers target other mobile genetic elements across bacterial phyla to abrogate superinfection via diverse mechanisms.

(B–E) Graphical illustrations of representative phage CRISPR loci harboring known and novel subtypes and their proposed mechanisms and functions as determined via spacer targeting and protein sequence analysis. Special consideration is given to phages carrying multiple loci.

See also Figure S2.

miniature nucleases constitute phylogenetically distinct clades denoted as types II-x, -y, and -z (Figure 3A). These systems lack the Cas1, Cas2, or Csn2 sequence acquisition machinery (Figures 1D and 3A) and have distinct domain organizations compared to previously studied Cas9 orthologs, with significant deletions across the proteins in comparison (Figure S2). The phylogenetic analyses indicated multiple evolutionary origins of the type II CRISPR-Cas systems in viruses, the evolutionary relatedness of virally and bacterially encoded CRISPR-Cas systems suggesting that those encoded by viruses were obtained from their host during prior infection.

Furthermore, we observed that bacteriophage genomes harbor an unusual enrichment of hypercompact type V effectors (Figures 1D and 3B) compared to abundance in bacteria,[2] including hundreds of variants comprising 44 protein families that are evolutionarily distant from previously reported and experimentally validated miniature type V CRISPR-Cas nucleases, including Cas12f and CasΦ[12,13] (Figure 3B). Evolutionary analysis suggests that distinct type V nuclease subtypes may have evolved multiple times from separate transposon-encoded TnpB families, which have recently been shown to be RNA-guided nucleases themselves,[14] and we observe that TnpB is also widely encoded on phages.

CRISPR arrays associated with the type V families contained spacer sequences targeting competing dsDNA-based extrachromosomal elements that are predicted to infect the same host (Figure 2). We found that in multiple related Biggiephages, miniature type V families including Casμ and CasΦ co-occurred with a type I system that we term type I-X, of which we only had one example previously,[5] bearing similarities to type I-C CRISPR systems but featuring a distinct helicase in place of the processive nuclease Cas3. Biggiephage genomes were recovered over a four-year time span, and remained identical save for their CRISPR arrays, which only exhibited minor differences (Figures S2C and S2D). Though we were unable to validate DNA cleavage by this system, it is possible that dsDNA binding silences the expression of target genes. In some cases, the arrays of the type I-X system target the same circular extrachromosomal element, albeit with distinct spacers, as the array associated with co-occurring type V systems. One such cryptic element harbored restriction enzymes and retron-based anti-phage defense systems that could limit Biggiephage infectivity, underscoring the dynamic nature of the evolutionary arms race between mobile elements in competition for host resources.

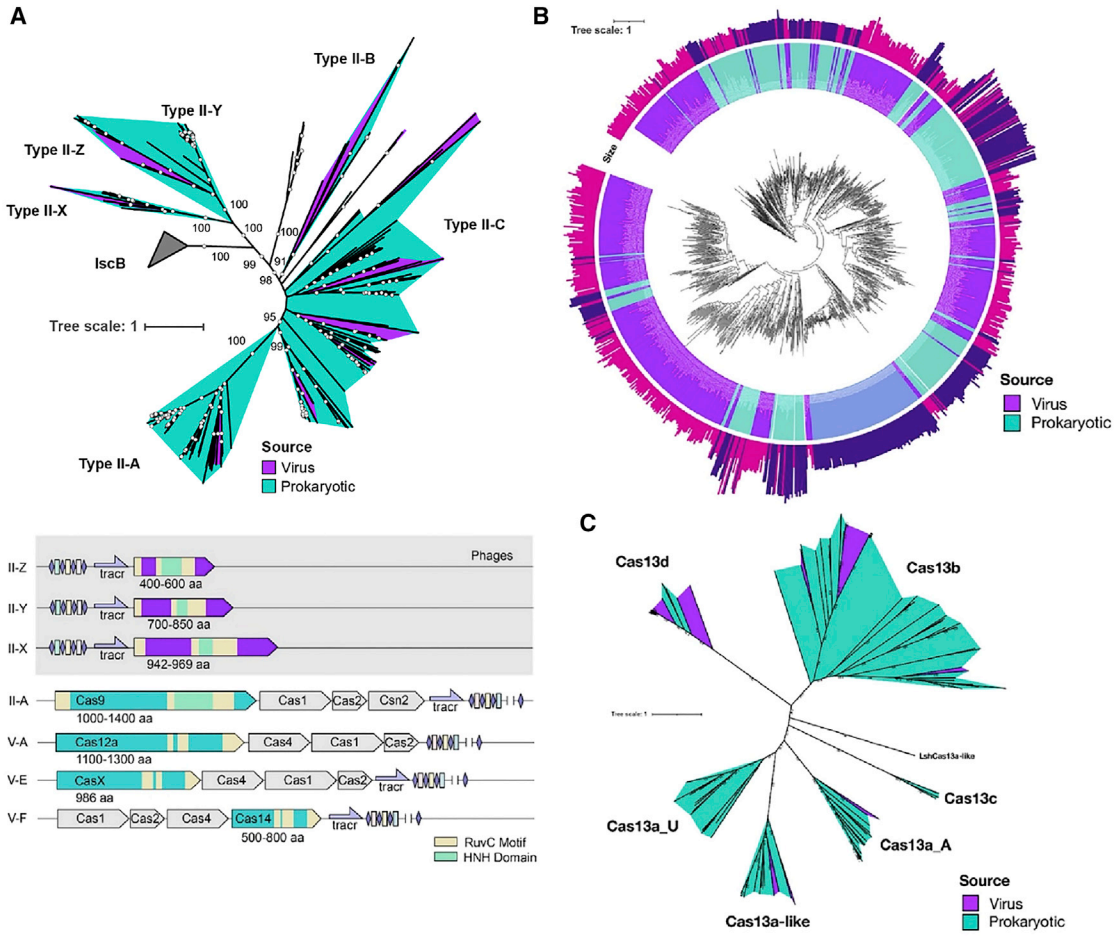


Figure 3. Diversity of Class 2 CRISPR-Cas systems on phage and phage-like genomes

(A) Maximum likelihood phylogenetic tree of phage encoded and bacterially encoded type II nucleases and respective predicted ancestral IscB nucleases. Bootstrap and approximate likelihood-ratio test values ≥ 90 are denoted on the branches, and the bootstrap support percentages at branch points are shown in numbers. Bottom illustration of genomic CRISPR-Cas loci of type II and representative type V systems previously employed in genome editing applications.

(B) Maximum likelihood phylogenetic tree of phage-like (purple) and previously reported (teal) bacterially encoded type V nuclease clades and respective predicted ancestral TnpB nucleases. Outer ring denotes protein sizes with purple indicating previously reported or publicly available sequences and pink denoting systems from this study.

(C) Maximum likelihood phylogenetic tree of phage and previously reported bacterially encoded type VI nucleases.

See also Figure S1.

We also found the first type IV systems encoded in lytic phage genomes. Type IV systems are predominantly found on plasmids, where their mechanisms of action are poorly understood and they sometimes lack a CRISPR array.[15] We report a type IV subtype that lacks the DinG hallmark gene and encodes in its place a CysH-like protein bearing limited similarity to non-CRISPR associated CysH phosphoadenosine 5′-phosphosulfate reductases. Remarkably, the CRISPR array associated with this type IV-F system and a neighboring type V targets the type V Cas gene encoded in a competing cyanophage (Figure 2).

Casλ is a divergent phage-specific CRISPR-Cas enzyme with a unique guide RNA

A distinctive phage-encoded enzyme family, Casλ, exists within huge bacteriophages that are evolutionarily linked to the recently reported Mahaphage clade.[5] Named using Greek nomenclature to indicate its phage origin, this family of 55 compact systems exhibited such sequence divergence that it had negligible sequence identity (<5%) to, and clustered separately from, type V and type II enzymes (Figure S3A). In addition, Casλ sequences have low similarity to these enzymes (<10%) but are phylogenetically closest to Cas14J[5] (Figure 3B). The protein is not encoded along with any other Cas proteins, and the RuvC nuclease was not immediately identifiable from the sequence. Difficulty in aligning this system to reported enzymes via remote homology (Figure S6) further suggested that a direct evolutionary relationship with known Cas superfamilies was questionable. CRISPR arrays associated with Casλ contain spacer sequences complementary to dsDNA-based extrachromosomal elements predicted to infect the same Bacteroidetes host (Figure 2). These observations implied that

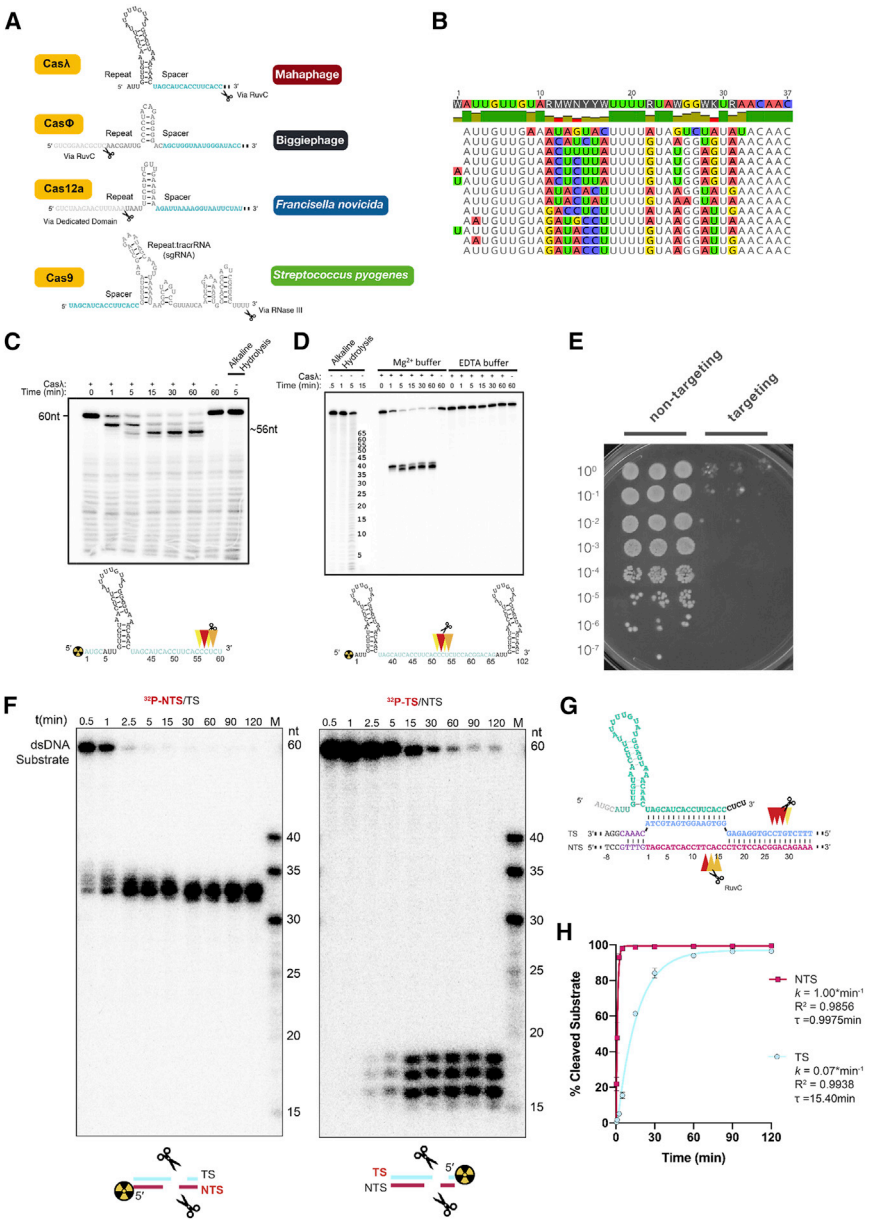


Figure 4. Casλ processes its own crRNA and cleaves dsDNA

(A) Casλ1 from huge Mahaphages displays a unique crRNA hairpin compared to known Cas12 enzymes, and is reminiscent of stem-loop 1 of the engineered SpyCas9 single gRNA (sgRNA).

(B) Casλ repeats uniquely display highly conserved nucleotide sequences at the 5′, 3′, and center of the RNA.

(C) 5′ radiolabeling of crRNAs indicates that Casλ1 uniquely processes its own crRNA in the spacer region (or 3′ end). OH-ladder enables the pre-crRNA processing sites (red triangles) to be derived.

(D) Processing of the repeat-spacer-repeat pre-crRNA substrate occurs similarly to (C) in the spacer region and does not occur in the absence of Mg^{2+}, indicating a role for the RuvC in the processing mechanism.

(E) Casλ with targeting or non-targeting guides validates its capacity to cleave DNA flanking experimentally determined PAMs in *E. coli* at different dilutions.

(F) Cleavage assay targeting dsDNA for mapping of the cleavage structure.

(G) Scheme illustrating the DNA cleavage pattern.

(H) Efficiency and kinetics of DNA cleavage of NTS and TS (n = 3 each, mean ± SD).

See also Figures S3 and S4.

Casλ may be targeting dsDNA in native contexts of the host similarly to Cas9 or Cas12 systems.

In any CRISPR-Cas system, processing of CRISPR array transcripts, consisting of repeats and spacer sequences acquired from previously encountered mobile genetic elements (MGEs),[16] is essential to generating mature crRNAs that guide Cas proteins[17] to destroy foreign viruses. Analogous to the distinct nature of the protein, the Casλ crRNA is predicted to form an elongated hairpin secondary structure not previously observed in guide RNAs associated with Cas12 (Figure 4A). Despite their divergent nucleotide sequences, crRNAs retain a similar predicted hairpin structure across the protein family (Figure S3B). Furthermore, Casλ crRNAs contain conserved sequences at their 5′ and-3′ ends and in the center of the RNA (Figure 4B). The overall sequence divergence of the protein, its putative RuvC domain, and the encoded crRNA prompted us to further analyze this protein family.

RuvC-mediated crRNA processing in the spacer region and dsDNA cutting by Casλ

The lack of a detectable tracrRNA encoded within the genomic locus begged the question of how this aberrant RNA, akin to a naturally occurring crRNA-tracrRNA hybrid (Figure 4A), may be processed by the CRISPR-Cas system or host factors to produce mature crRNA. Using radiolabeled precursor crRNAs (pre-crRNAs) as substrates, we first tested whether purified Casλ protein catalyzes RNA cleavage. Surprisingly, analytical denaturing gel electrophoresis showed that pre-crRNAs are cut by Casλ in the spacer region as opposed to the 5′ end of the RNA, where cutting has been observed in nearly all self-processing single-effector systems analyzed previously (Figures 4C, 4D, and S3D), with the exception of a type V-C system.[18,19] The Casλ-induced pre-crRNA processing yields a crRNA spacer sequence that is complementary to DNA target sites 14–17 nucleotides (nt) in length.

The fact that Casλ can process its own pre-crRNA obviates the need for Ribonuclease III or other host factors required for the function of most known Cas9 and Cas12 family members. Although some CRISPR-Cas proteins process pre-crRNAs using

an internal active site distinct from the RuvC domain[20] or by recruiting Ribonuclease III to cleave a pre-crRNA:tracrRNA duplex,[21] we wondered whether phage-encoded Casλ, like phage-encoded CasΦ,[13] processes pre-crRNA using its RuvC active site. We thus tested Mg^{2+} dependence and showed that Casλ is indeed reliant on the presence of Mg^{2+} and thus, by extension, the RuvC active site for crRNA maturation (Figure 4D).

CRISPR–Cas systems target DNA sequences following or preceding a 2–5 bp Protospacer Adjacent Motif (PAM) for self-versus-non-self discrimination.[22] We determined the sequence requirements for DNA targeting by Casλ using a plasmid depletion assay in which targeting a library of putative PAM sites revealed sequence specificity. This assay demonstrated the ability of crRNA-guided Casλ to cleave dsDNA, without requirement for additional RNA components, and a putative TTR PAM sequence specificity (Figure S4A). Casλ with host genome-targeting guides showed a reduction in colony-forming units (as a proxy for cell viability) of multiple orders of magnitude, in comparison to negative control of Casλ with a non-targeting guide (Figure 4E).

In vitro incubation of purified Casλ with crRNAs, along with a linear dsDNA target substrate, generated cleavage products with surprisingly pronounced staggered 5′-overhangs of 11–16 nt (Figures 4F and 4G). Type V CRISPR-Cas enzymes such as Cas12a have also been observed to generate staggered overhangs, albeit smaller at only 5 nt.[23] Furthermore, the non-target strand (NTS) was cleaved faster than the target strand (TS) within the RuvC active site over a 2 h time period (Figure 4H).

Casλ ribonucleoproteins induce genome editing in endogenous genes in human and plant cells

The development of single-effector CRISPR-Cas systems for editing eukaryotic cells has revolutionized genome engineering.[24] However, the large sizes of Cas9 and Cas12a enzymes can inhibit delivery into many cell types, for which hypercompact genome editors with favorable kinetics imply great promise as an alternative. We conducted a head-to-head comparison of insertion and deletion efficiencies using Casλ and Cas12a ribonucleoproteins (RNPs) with identical guide RNA spacers targeting sequences recognizing Vascular Endothelial Growth Factor A (VEGFA) and Empty Spiracles Homeobox1 (EMX1) genes in HEK293T cells. Despite their miniature size, Casλ RNPs generated promising genome-editing outcomes compared to Cas12a, and in at least one case exceeded Cas12a insertion-deletion (indel) percentages (Figure 5A). Extending these experiments to *Arabidopsis thaliana*, we confirmed that Casλ exhibited editing efficiencies of up to 18% at the endogenous *PDS3* gene (Figure 5B), notably higher than what we observed previously using CasΦ.[13] The efficiency of editing was highly dependent on temperature, with no editing occurring at 23°C, an intermediate level of editing occurring at 28°C, and the highest level of editing occurring at 32°C. Furthermore, we were able to achieve editing in the endogenous disease resistance gene *Snn5* in hexaploid wheat protoplasts, where a genome ~5x larger than the human genome poses a scanning challenge to achieve successful editing at the target site (Figure 5C). Next-generation sequencing for both human and plant cells revealed indel profiles with large deletions (Figures 5D and S5C), consistent with the staggered cuts observed *in vitro* at the PAM distal region.

Casλ protein structure explains interference mechanism

To explore the mechanism by which Casλ achieves RNA-guided DNA recognition, we generated a purified Casλ-crRNA-dsDNA ternary complex suitable for analysis by cryo-EM. CryoEM maps of this complex revealed a bilobed architecture analogous to Cas9 and Cas12 enzymes,[25] despite the divergence in both sequence and size between Casλ and these much larger enzymes (Figures 6A, 6B, S5, S6, and S7). The 3 Å resolution structure revealed the shape and domain organization of Casλ and the unique structure of the guide RNA (Figures 6A–6C, S6, and S7). Notably, the RuvC domain of Casλ is split into four parts across the C-terminal half of the protein, likely hindering reliable alignment and clustering with reported Cas12 systems (Figure 6D). The REC I and REC II domains are also segmented in the protein sequence, with the PAM-interacting domain wedged within REC I as opposed to the N terminus of the protein as seen in CasΦ,[26] but similar to Cas12i.[19] In contrast to CasΦ, Casλ contains a Target Strand Loading (TSL) domain that likely functions to load the single-stranded DNA (ssDNA) substrate, in a position analogous to the "Nuc" domain that was incorrectly hypothesized in other type V CRISPR-Cas enzymes to be a second nuclease domain responsible for DNA cleavage.[27] Casλ also exhibits a distinct structure in the REC I domain compared to CasΦ (Figure S6D).

The crRNA assumes an unexpected shape that blankets the protein, with a divergent recognition lobe in Casλ that binds to distinct sequences and structural features of the guide RNA (Figures 6C, and S6A–S6C). Specifically, we observed possible interactions between primarily polar or charged residues within the REC II domain in Casλ with the conserved motifs of the crRNA hairpin (Figures 4B and S6C). These residues are conserved across the protein family and likely interact either directly with the RNA nucleobases (Q452, N510), or with the RNA phosphate backbone to stabilize the guide (S451, K496, E444, N445, K503, Y619) (Figures 6C and S6C). We did not observe protein contacts to the unpaired nucleobases A9, A30, A31 motif in the middle of the guide RNA stem-loop (Figure 3A), which is further supported by the lack of sequence conservation. However, interestingly, the non-complementarity is conserved, which is likely important for the hairpin kink geometry.

CRISPR-Cas proteins initiate the unwinding of target dsDNA following PAM recognition. In Casλ, this recognition is achieved via interactions with the oligonucleotide-binding domain (OBD), REC I, and a five α-helical bundle referred to as the PAM-interacting domain (PID). Residues within the three domains interact with the sugar-phosphate backbone of the target DNA (Figure S6B) and, in some cases such as residue N102, interact directly with the nucleobases. The interaction between N102 and nucleobase G(-1) may explain the preference for purines in this position as opposed to pyrimidines, since a pyrimidine substitution would result in a base that is too distant from the interacting asparagine (Figure S6B). In examining the aftermath of *cis*-cleavage of DNA, we found that Casλ had a very low level of ssDNA or ssRNA cleavage in *trans* upon DNA recognition in *cis* (Figure S4B). Incubation of the Casλ protein with non-cognate guides from other orthologs within the protein family replicated the ssDNA *trans* cleavage effect despite differences

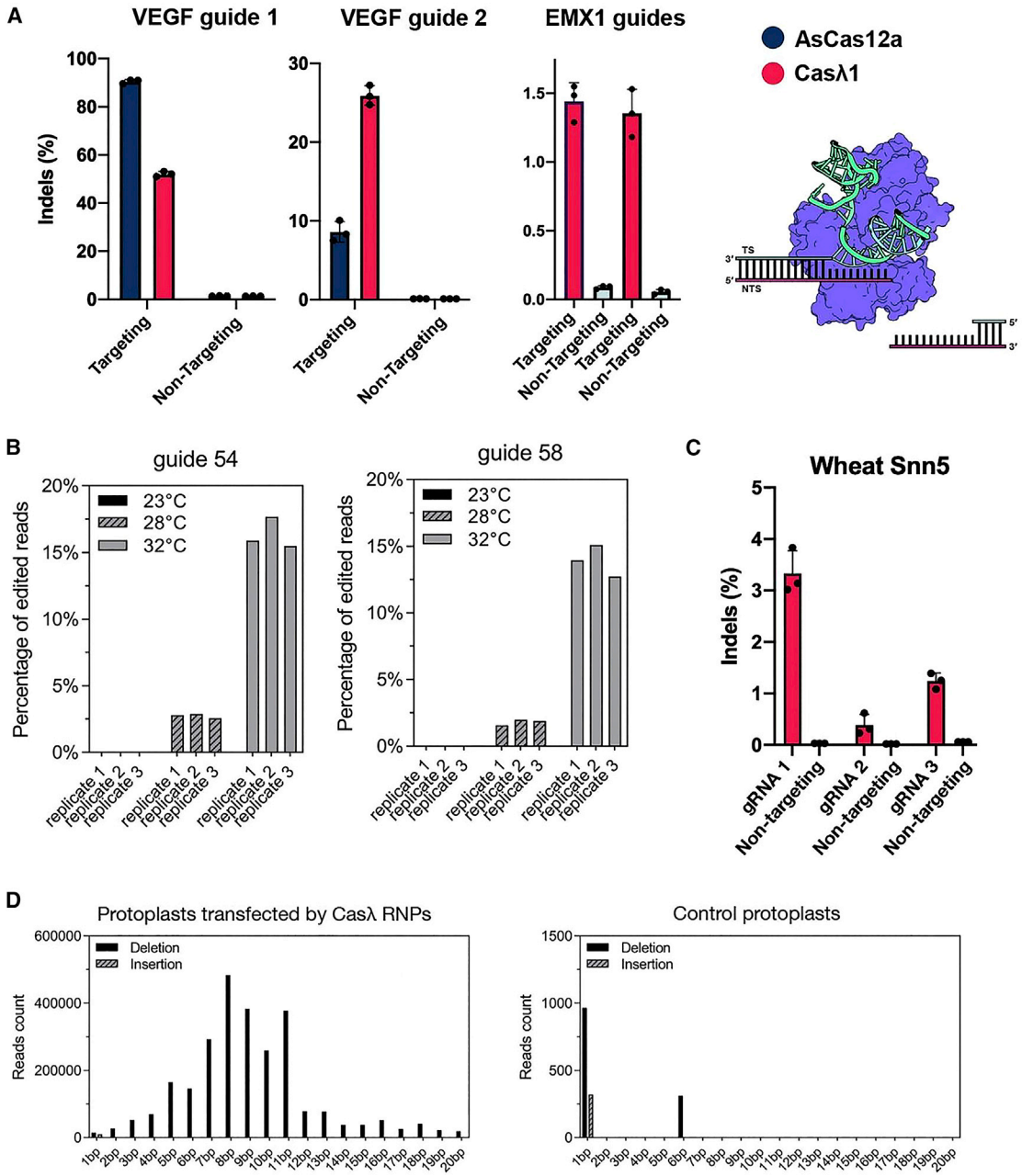


Figure 5. Casλ RNPs are functional for editing endogenous genes in human, *Arabidopsis*, and wheat cells with large deletion profiles

(A) Indel efficiency using Casλ and Cas12a RNPs with identical spacers targeting VEGF, and Casλ RNPs targeting EMX1 genes in HEK293T cells, and a schematic of the *in vitro* model of DNA cleavage outcomes following DNA cleavage by Casλ.

(B and C) Indel efficiencies in *Arabidopsis thaliana* protoplasts show significantly higher levels of editing than previously achieved by CasΦ for the same *PDS3* gene, and (C) in wheat protoplasts targeting the disease resistance gene *Snn5*.

(D) Indel profiles generated by Casλ RNP administration show primarily large deletions, and little change without Casλ.

in their sequence (Figure 6E), suggesting that guides within the Casλ family may be interchangeable, unlike Cas9. These findings warrant further investigation to examine the determinants of guide RNA binding by Casλ. Single mismatches across the ssDNA target revealed that the seed region of the target DNA (1–5) and the region extending from bases 7–13 are required to match the spacer sequence of the guide RNA for efficient cleavage (Figure 6F). Investigation of positions that possibly interact with the DNA in these regions (Figure S6D) or the corresponding RNA revealed conserved residues in REC, OBD, PID, and RuvC domains that may account for the complex's intolerance to target mismatches, and, therefore, the possibility of relatively

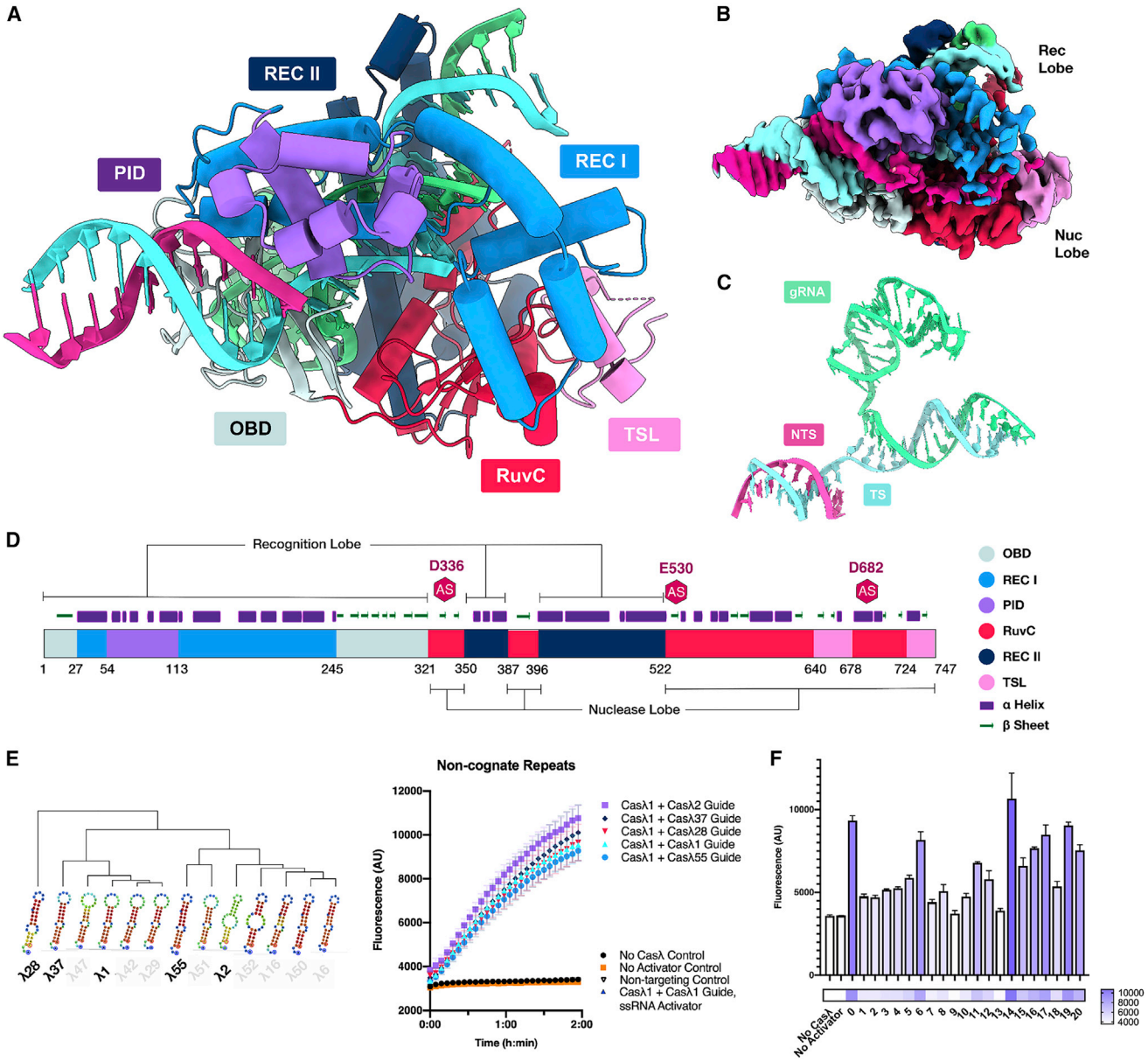


Figure 6. Structure of Casλ-gRNA-DNA complex

(A) Schematic representation of the Casλ–gRNA–DNA complex. Disordered linkers are shown as dotted lines. Insets for protein-DNA interactions are shown in Figure S5.

(B) Cryo-EM maps of the Casλ-guide-RNA-DNA complex. The target strand is shown in cyan and the non-target strand is shown in magenta.

(C) Cryo-EM-based model of guide RNA-target DNA complex.

(D) Schematic of the domain organization and secondary structure of Casλ.

(E) Hierarchical clustering dendrogram of different repeats with their predicted secondary structures. Casλ can still cleave ssDNA in *trans* with guide RNAs consisting of non-cognate repeats that are divergent at the sequence level.

(F) Fluorescence output using oligonucleotide activators with mismatches at each respective position along the target DNA. "0" indicates no mismatches (control). Insets relating to protein-DNA interactions are shown in Figure S6.

See also Figures S5, S6, and S7.

high fidelity in the context of genome editing. Overall, more domains within Casλ, such as the RECI, RECII, OBD, RuvC, and TSL domains, exhibit segmentation and rearrangement compared to known type V systems.

DISCUSSION

CRISPR-Cas systems occur rarely in bacteriophage genomes, but large-scale analysis of all known and novel phages

uncovered thousands of viruses encoding an unexpected diversity, distribution, and potential function of RNA-guided proteins across the biosphere. CRISPR-encoding bacteriophages are predicted to infect a remarkable diversity of bacterial phyla, and the spacer sequences associated with their CRISPR arrays show complementarity to dsDNA-based viruses and extrachromosomal elements predicted to infect the same hosts. These observations underscore the general strategy of phages to protect their hosts against superinfection by competing elements. We find that phage genomes harbor CRISPR-Cas systems corresponding to all six known types, and we also note differences observed bioinformatically or biochemically between the viral CRISPR-Cas systems and their bacterial counterparts. These differences hint that phages have co-opted CRISPR anti-viral systems with modifications that tune RNA-guided pathways to their needs.

At least two aspects of phage-encoded CRISPR systems differ notably from cellular systems, highlighting the versatility of these pathways and also the potential for phage-mediated functional evolution. First, some RNA-targeting type III and type VI systems recognize abundant or essential transcripts of competing phages, and type III systems retain the catalytic residues required to cleave ssDNA and enable nicking of the target DNA, but lack components used for non-specific transcript cleavage following RNA target recognition. The absence of these components, which trigger abortive infection by analogous host-encoded systems, suggests that some phages prefer to avoid self-destruction of transcripts or induction of a dormant state in the host, both of which may be disadvantageous to the phage life cycle. Consistent with this idea, the minimal *trans*-cutting of ssDNA and RNA observed for Casλ implies limited ability to target single-stranded replication intermediates of MGEs (Figures S4 and S8). A second important difference between cellular and phage-encoded CRISPR systems is the absence of a processive nuclease, such as Cas3, in some phage type I systems. Together with the presence of CysH, which may be recruited as a putative effector in type IV systems, these observations suggest alternate outcomes to nucleic acid-targeting by these phage-encoded pathways. In particular, the lack of a Cas3 nuclease in type I systems targeting plasmid-like elements suggests a gene silencing mechanism that precludes DNA cutting. We speculate that such phage-based type I systems could assist the activity of the co-occurring CasΦ systems that are found in the same genome. Because the targeted plasmid-like elements harbor restriction enzymes and retron-based anti-phage defense systems that could limit the infectivity of the CRISPR-encoding phage, coordinated activities of orthogonal CRISPR systems could assist competition between mobile elements.

This study demonstrates that phage genomes are a natural reservoir of miniature single-effector CRISPR-Cas systems, including DNA targeting type II and type V enzymes belonging to the Cas9 and Cas12 superfamilies. We use Greek nomenclature to indicate the phage origins of Casμ, CasΩ, and Casλ, extending the naming convention established by phage-encoded CasΦ. In contrast to the prevalence of multi-subunit type I and type III CRISPR systems in prokaryotic genomes,[2] the notable abundance of miniature Cas12-family enzymes in phages may reflect the size restriction of many phage genomes. Because phages evolve quickly, they serve as important sources of new, divergent, or hypercompact CRISPR systems. Some of these, such as Casλ, bear sufficient sequence-level divergence to cluster separately from Cas12 and Cas9 systems and obscure a direct evolutionary relationship with known Cas superfamilies. Nonetheless, Casλ's structure, domain composition, and biochemical mechanism are similar to other type V enzymes.[25] This finding implies that within phage genomes, distinct type V nucleases may have evolved multiple times from ancestral transposon-encoded TnpB families, which also function as RNA-guided nucleases.[14] Despite being from different clades of phages and having divergent sequences and domain organizations, we observe a convergent evolution of Cas12-like architecture in the Casλ and CasΦ protein structures. In addition, both can process their own pre-crRNA and rely on the same RuvC active site used for DNA cleavage for this activity. This extreme compression of enzymatic activities within one active site has not been observed for bacterially encoded CRISPR-Cas proteins. Nonetheless, the phage-encoded enzymes diverge functionally from one another in other ways, including guide RNA structure and maturation process. This may reflect the interplay between rapid phage evolution, which generates diversity, and selective pressure to maintain CRISPR compatibility in a variety of host environments over time, which favors pathway conservation. In both cases, phages that encode their own Cas variants that do not rely on host factors may eliminate the possibility that the ongoing evolution of essential host proteins or cofactors will result in incompatibility with phage-encoded anti-viral systems.

The molecular structure of the Casλ-crRNA-dsDNA complex reported in this study illustrates possible convergent evolution of RNA-guided effectors, despite extreme sequence divergence and distinct ancestral protein origins. The domain architecture of Casλ exhibits more segmentation and likely structural rearrangements than have been seen in other Cas12-family enzymes, with multiple functional domains split at the sequence level into separate segments that assemble during protein folding. This unique domain organization may explain the difficulty in accurately aligning Casλ to previously reported enzymes, despite overall structural similarity. Remarkably, however, this segmented domain composition does not compromise genome editing activity as observed for Casλ-based editing of human, *Arabidopsis*, and wheat cells. The finding that Casλ can induce efficient genome editing of endogenous genes in these diverse cell types, in some cases exceeding the efficacy of Cas12a-mediated genome editing, shows that there isn't necessarily a tradeoff between Cas effector size and function. This result, together with the compact size of phage-encoded CRISPR-Cas proteins that is advantageous for vector-based cellular delivery (Figure S7), shows that nature's phage reservoir is an important future source of enzymes useful for genome editing in heterologous cell types.

Overall, the discovery of thousands of viruses encoding CRISPR systems representing all six CRISPR-Cas types highlights the sparsity but broad diversity of RNA-guided systems within viruses. Genome-resolved metagenomics and bioinformatics-enabled phylogenetic insights enabled us to analyze these systems from uncultivated viruses and infer their mechanisms of action within their biological contexts. We investigated

hundreds of novel hypercompact and divergent CRISPR-Cas systems, with particular focus on the unique Casλ family. We report the utility of Casλ as a valuable tool for genome editing in plant and human cells. In addition, our data show how the structural compaction of this protein family preserves robust biochemical and cell-based functionality essential to both natural activities and biotechnological applications. Finally, our structural understanding of Casλ proteins provides a starting point for the future design of variants with expanded genome editing capabilities that combine the advantages of a small protein with the versatility of a robust RNA-guided DNA recognition machine.

Limitations of the study

Some mechanistic aspects of Casλ remain to be defined, including the possible interchangeability of different variant guide RNAs to support catalytic activity. In addition, details of target binding, including the seed region within the target binding site as well as target recognition accuracy, remain to be established. Potential interactions between phage-encoded and host-encoded CRISPR systems have not been explored and will be fascinating to understand. Future research is needed to identify the functions of phage-encoded CRISPR systems that may extend beyond anti-phage defense and to further test the applications of Casλ as a genome editing tool in plants, other organisms, and human cells.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mammalian models
  - Plant models
- METHOD DETAILS
  - Phylogenetic analysis
  - crRNA sequence analysis
  - PAM depletion analysis
  - PAM depletion sequencing analysis
  - Programmable DNA targeting
  - Protein purification
  - Pre-crRNA processing assays
  - *In vitro* cleavage assays - Radiolabeled nucleic acids
  - Fluorophore quencher and DNA mismatch tolerance assay
  - I-X PAM binding assay
  - Mammalian genome editing
  - Plant genome editing
  - Ternary complex reconstitution for cryo-EM
  - Electron microscopy grid preparation and data collection
  - Single-particle cryo-EM data processing and 3D volume reconstruction
  - Model building and refinement
  - Data deposition and figure preparation
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2022.10.020.

ACKNOWLEDGMENTS

The authors are indebted to Marena Trinidad for her careful collation of final manuscript files and data analysis, and to Marena Trinidad and Lin-Xing Chen for providing computational biology support. The authors thank Jamie Cate for assistance with figure preparation and critical reading of the manuscript. We also thank members of the Doudna, Banfield, Staskawicz, Savage, and Jacobsen labs for many helpful discussions. Financial support was provided by grants from the National Institutes of Health and the National Science Foundation. J.A.D., S.E.J., and D.S. are Howard Hughes Medical Institute investigators. P.P. was supported by the NIH Somatic Cell Genome Editing consortium (NIH U01AI142817-02). P.S. was supported by the Swiss National Science Foundation Mobility fellowship (P2EZP3_195621)

AUTHOR CONTRIBUTIONS

B.A.-S. conceived of and designed the study with guidance from J.F.B. and J.A.D. B.A.-S. and J.F.B. designed and implemented the computational searches, host matching, phylogenetic analyses, and classification of phages and CRISPR-Cas systems. B.A.-S. performed PAM and cell viability assays, radiolabeling, and DNA cleavage biochemical work and analyzed the data, with guidance from P.P. and B.F.C. C.J.H. performed the crRNA processing assay. B.A.-S. and D.S. performed protein purifications and *trans*-cleavage assays. B.A.-S., P.S., and K.M.S. prepared cryoEM grids and collected cryoEM dataset. B.A.-S. generated the initial protein model. P.S. and B.A.-S. processed cryoEM data with guidance from K.M.S. B.A.-S. designed guide RNAs and purified protein for eukaryotic genome editing experiments that were performed by Z.L., E.C.S., E.G., and S.E.J. with oversight from S.E.J., D.F.S. and B.S. B.A.-S. and A.R.E. performed phage-type I-X analyses. J.A.D. and J.F.B. provided advisory support and funding. J.A.D. and B.A.-S. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

The Regents of the University of California have filed patents related to this work on which the authors are inventors. J.A.D. is a co-founder of Caribou Biosciences, Editas Medicine, Scribe Therapeutics, Intellia Therapeutics, and Mammoth Biosciences. J.A.D. is a scientific advisory board member of Vertex, Caribou Biosciences, Intellia Therapeutics, Scribe Therapeutics, Mammoth Biosciences, Algen Biotechnologies, Felix Biosciences, The Column Group, and Inari Agriculture. J.A.D. is Chief Science Advisor to Sixth Street, a Director at Johnson & Johnson, and Altos and Tempus, and has research projects sponsored by Apple Tree Partners and Roche. J.F.B. is a co-founder of Metagenomi. S.E.J. is a scientific co-founder of Inari Agriculture, and a science advisory board member to Invaio Sciences, Senda Biosciences, FL70, and Sixth Street.

Received: June 9, 2022
Revised: September 10, 2022
Accepted: October 18, 2022
Published: November 23, 2022

REFERENCES

1. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR Provides

Acquired Resistance Against Viruses in Prokaryotes. Science *315*, 1709–1712. https://doi.org/10.1126/science.1138140.

2. Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., et al. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat. Rev. Microbiol. *18*, 67–83. https://doi.org/10.1038/s41579-019-0299-x.
3. O'Hara, B.J., Barth, Z.K., McKitterick, A.C., and Seed, K.D. (2017). A highly specific phage defense system is a conserved feature of the Vibrio cholerae mobilome. PLoS Genet. *13*, e1006838. https://doi.org/10.1371/journal.pgen.1006838.
4. Seed, K.D., Lazinski, D.W., Calderwood, S.B., and Camilli, A. (2013). A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. Nature *494*, 489–491. https://doi.org/10.1038/nature11927.
5. Al-Shayeb, B., Sachdeva, R., Chen, L.-X., Ward, F., Munk, P., Devoto, A., Castelle, C.J., Olm, M.R., Bouma-Gregson, K., Amano, Y., et al. (2020). Clades of huge phages from across Earth's ecosystems. Nature *578*, 425–431. https://doi.org/10.1038/s41586-020-2007-4.
6. Niewoehner, O., Garcia-Doval, C., Rostøl, J.T., Berk, C., Schwede, F., Bigler, L., Hall, J., Marraffini, L.A., and Jinek, M. (2017). Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. Nature *548*, 543–548. https://doi.org/10.1038/nature23467.
7. Jiang, W., Samai, P., and Marraffini, L. (2016). Degradation of Phage Transcripts by CRISPR-Associated RNases Enables Type III CRISPR-Cas Immunity. Cell *164*, 710–721. https://doi.org/10.1016/j.cell.2015.12.053.
8. Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G., and Siksnys, V. (2017). A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. Science *357*, 605–609. https://doi.org/10.1126/science.aao0100.
9. Koonin, E.V., Makarova, K.S., and Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. Curr. Opin. Microbiol. *37*, 67–78. https://doi.org/10.1016/j.mib.2017.05.008.
10. Al-Shayeb, B., Schoelmerich, M.C., West-Roberts, J., Valentin-Alvarado, L.E., Sachdeva, R., Mullen, S., Crits-Christoph, A., Wilkins, M.J., Williams, K.H., Doudna, J.A., et al. (2021). Borgs Are Giant Extrachromosomal Elements with the Potential to Augment Methane Oxidation. Microbiology. https://doi.org/10.1101/2021.07.10.451761.
11. VanderWal, A.R., Park, J.-U., Polevoda, B., Kellogg, E.H., and O'Connell, M.R. (2021). CRISPR-Csx28 Forms a Cas13b-Activated Membrane Pore Required for Robust CRISPR-Cas Adaptive Immunity. Biochemistry. https://doi.org/10.1101/2021.11.02.466367.
12. Karvelis, T., Bigelyte, G., Young, J.K., Hou, Z., Zedaveinyte, R., Budre, K., Paulraj, S., Djukanovic, V., Gasior, S., Silanskas, A., et al. (2020). PAM recognition by miniature CRISPR–Cas12f nucleases triggers programmable double-stranded DNA target cleavage. Nucleic Acids Res. *48*, 5016–5023. https://doi.org/10.1093/nar/gkaa208.
13. Pausch, P., Al-Shayeb, B., Bisom-Rapp, E., Tsuchida, C.A., Li, Z., Cress, B.F., Knott, G.J., Jacobsen, S.E., Banfield, J.F., and Doudna, J.A. (2020). CRISPR-CasΦ from huge phages is a hypercompact genome editor. Science *369*, 333–337. https://doi.org/10.1126/science.abb1400.
14. Karvelis, T., Druteika, G., Bigelyte, G., Budre, K., Zedaveinyte, R., Silanskas, A., Kazlauskas, D., Venclovas, Č., and Siksnys, V. (2021). Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. Nature *599*, 692–696. https://doi.org/10.1038/s41586-021-04058-1.
15. Pinilla-Redondo, R., Mayo-Muñoz, D., Russel, J., Garrett, R.A., Randau, L., Sørensen, S.J., and Shah, S.A. (2020). Type IV CRISPR–Cas systems are highly diverse and involved in competition between plasmids. Nucleic Acids Res. *48*, 2000–2012. https://doi.org/10.1093/nar/gkz1197.
16. McGinn, J., and Marraffini, L.A. (2019). Molecular mechanisms of CRISPR–Cas spacer acquisition. Nat. Rev. Microbiol. *17*, 7–12. https://doi.org/10.1038/s41579-018-0071-7.
17. Hille, F., Richter, H., Wong, S.P., Bratovič, M., Ressel, S., and Charpentier, E. (2018). The Biology of CRISPR-Cas: Backward and Forward. Cell *172*, 1239–1259. https://doi.org/10.1016/j.cell.2017.11.032.
18. Harrington, L.B., Ma, E., Chen, J.S., Witte, I.P., Gertz, D., Paez-Espino, D., Al-Shayeb, B., Kyrpides, N.C., Burstein, D., Banfield, J.F., and Doudna, J.A. (2020). A scoutRNA Is Required for Some Type V CRISPR-Cas Systems. Mol. Cell *79*, 416–424.e5. https://doi.org/10.1016/j.molcel.2020.06.022.
19. Huang, X., Sun, W., Cheng, Z., Chen, M., Li, X., Wang, J., Sheng, G., Gong, W., and Wang, Y. (2020). Structural basis for two metal-ion catalysis of DNA cleavage by Cas12i2. Nat. Commun. *11*, 5241. https://doi.org/10.1038/s41467-020-19072-6.
20. Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A., and Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. Nature *532*, 517–521. https://doi.org/10.1038/nature17945.
21. Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602–607. https://doi.org/10.1038/nature09886.
22. Westra, E.R., Semenova, E., Datsenko, K.A., Jackson, R.N., Wiedenheft, B., Severinov, K., and Brouns, S.J.J. (2013). Type I-E CRISPR-Cas Systems Discriminate Target from Non-Target DNA through Base Pairing-Independent PAM Recognition. PLoS Genet. *9*, e1003742. https://doi.org/10.1371/journal.pgen.1003742.
23. Stella, S., Mesa, P., Thomsen, J., Paul, B., Alcón, P., Jensen, S.B., Saligram, B., Moses, M.E., Hatzakis, N.S., and Montoya, G. (2018). Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity. Cell *175*, 1856–1871.e21. https://doi.org/10.1016/j.cell.2018.10.045.
24. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science *337*, 816–821. https://doi.org/10.1126/science.1225829.
25. Wang, J.Y., Pausch, P., and Doudna, J.A. (2022). Structural biology of CRISPR–Cas immunity and genome editing enzymes. Nat. Rev. Microbiol. *20*, 641–656. https://doi.org/10.1038/s41579-022-00739-4.
26. Pausch, P., Soczek, K.M., Herbst, D.A., Tsuchida, C.A., Al-Shayeb, B., Banfield, J.F., Nogales, E., and Doudna, J.A. (2021). DNA interference states of the hypercompact CRISPR–CasΦ effector. Nat. Struct. Mol. Biol. *28*, 652–661. https://doi.org/10.1038/s41594-021-00632-3.
27. Liu, J.-J., Orlova, N., Oakes, B.L., Ma, E., Spinner, H.B., Baney, K.L.M., Chuck, J., Tan, D., Knott, G.J., Harrington, L.B., et al. (2019). CasX enzymes comprise a distinct family of RNA-guided genome editors. Nature *566*, 218–223. https://doi.org/10.1038/s41586-019-0908-x.
28. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.
29. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX : Structure visualization for researchers, educators, and developers. Protein Sci. *30*, 70–82. https://doi.org/10.1002/pro.3943.
30. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of *Coot*. Acta Crystallogr D Biol Crystallogr *66*, 486–501. https://doi.org/10.1107/S0907444910007493.
31. Clement, K., Rees, H., Canver, M.C., Gehrke, J.M., Farouni, R., Hsu, J.Y., Cole, M.A., Liu, D.R., Joung, J.K., Bauer, D.E., and Pinello, L. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat. Biotechnol. *37*, 224–226. https://doi.org/10.1038/s41587-019-0032-3.

32. Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods *14*, 290–296. https://doi.org/10.1038/nmeth.4169.

33. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol. Biol. Evol. *32*, 268–274. https://doi.org/10.1093/molbev/msu300.

34. Kaur, S., Gomez-Blanco, J., Khalifa, A.A.Z., Adinarayanan, S., Sanchez-Garcia, R., Wrapp, D., McLellan, J.S., Bui, K.H., and Vargas, J. (2021). Local computational methods to improve the interpretability and analysis of cryo-EM maps. Nat. Commun. *12*, 1240. https://doi.org/10.1038/s41467-021-21509-5.

35. Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. *30*, 772–780. https://doi.org/10.1093/molbev/mst010.

36. Bland, C., Ramsey, T.L., Sabree, F., Lowe, M., Brown, K., Kyrpides, N.C., and Hugenholtz, P. (2007). CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. BMC Bioinf. *8*, 209. https://doi.org/10.1186/1471-2105-8-209.

37. Afonine, P.V., Klaholz, B.P., Moriarty, N.W., Poon, B.K., Sobolev, O.V., Terwilliger, T.C., Adams, P.D., and Urzhumtsev, A. (2018). New tools for the analysis and validation of cryo-EM maps and atomic models. Acta Crystallogr D Struct Biol *74*, 814–840. https://doi.org/10.1107/S2059798318009324.

38. Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A.J., et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. Acta Crystallogr D Struct Biol *75*, 861–877. https://doi.org/10.1107/S2059798319011471.

39. Mastronarde, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. J. Struct. Biol. *152*, 36–51. https://doi.org/10.1016/j.jsb.2005.07.007.

40. Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A.J., and Berger, B. (2019). Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. Nat. Methods *16*, 1153–1160. https://doi.org/10.1038/s41592-019-0575-8.

41. Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms Mol Biol *6*, 26. https://doi.org/10.1186/1748-7188-6-26.

42. Burstein, D., Harrington, L.B., Strutt, S.C., Probst, A.J., Anantharaman, K., Thomas, B.C., Doudna, J.A., and Banfield, J.F. (2017). New CRISPR–Cas systems from uncultivated microbes. Nature *542*, 237–241. https://doi.org/10.1038/nature21059.

43. Harrington, L.B., Burstein, D., Chen, J.S., Paez-Espino, D., Ma, E., Witte, I.P., Cofsky, J.C., Kyrpides, N.C., Banfield, J.F., and Doudna, J.A. (2018). Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. Science *362*, 839–842. https://doi.org/10.1126/science.aav4294.

44. Shmakov, S., Abudayyeh, O., Makarova, K., Wolf, Y., Gootenberg, J., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K., et al. (2015). Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Mol. Cell *60*, 385–397. https://doi.org/10.1016/j.molcel.2015.10.008.

45. Yan, W.X., Hunnewell, P., Alfonse, L.E., Carte, J.M., Keston-Smith, E., Sothiselvam, S., Garrity, A.J., Chong, S., Makarova, K.S., Koonin, E.V., et al. (2019). Functionally diverse type V CRISPR-Cas systems. Science *363*, 88–91. https://doi.org/10.1126/science.aav7271.

46. Shan, Q., Wang, Y., Li, J., and Gao, C. (2014). Genome editing in rice and wheat using the CRISPR/Cas system. Nat. Protoc. *9*, 2395–2410. https://doi.org/10.1038/nprot.2014.157.

47. Murray, M.G., and Thompson, W.F. (1980). Rapid isolation of high molecular weight plant DNA. Nucl Acids Res *8*, 4321–4326. https://doi.org/10.1093/nar/8.19.4321.

48. Punjani, A., Zhang, H., and Fleet, D.J. (2020). Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. Nat. Methods *17*, 1214–1221. https://doi.org/10.1038/s41592-020-00990-8.

49. Asarnow, D., Palovcak, E., and Cheng, Y. (2019). asarnow/pyem. UCSF pyem v0.5. https://doi.org/10.5281/ZENODO.3576630.

50. Zhang, H., Li, Z., Xiao, R., and Chang, L. (2020). Mechanisms for target recognition and cleavage by the Cas12i RNA-guided endonuclease. Nat. Struct. Mol. Biol. *27*, 1069–1076. https://doi.org/10.1038/s41594-020-0499-0.

51. Yamano, T., Zetsche, B., Ishitani, R., Zhang, F., Nishimasu, H., and Nureki, O. (2017). Structural Basis for the Canonical and Non-canonical PAM Recognition by CRISPR-Cpf1. Mol. Cell *67*, 633–645.e3. https://doi.org/10.1016/j.molcel.2017.06.035.

52. Wu, D., Guan, X., Zhu, Y., Ren, K., and Huang, Z. (2017). Structural basis of stringent PAM recognition by CRISPR-C2c1 in complex with sgRNA. Cell Res. *27*, 705–708. https://doi.org/10.1038/cr.2017.46.

53. Takeda, S.N., Nakagawa, R., Okazaki, S., Hirano, H., Kobayashi, K., Kusakizako, T., Nishizawa, T., Yamashita, K., Nishimasu, H., and Nureki, O. (2021). Structure of the miniature type V-F CRISPR-Cas effector enzyme. Mol. Cell *81*, 558–570.e3. https://doi.org/10.1016/j.molcel.2020.11.035.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
E. coli BL21(DE3)	New England Biolabs	C2527H
E. coli BL21(DE3)-Star	QB3-Macrolab, UC Berkeley	N/A
E. coli 5-alpha Competent cells	New England Biolabs	C2987H
Chemicals, peptides, and recombinant proteins		
384 well flat bottom black polystyrene assay plate	Corning	3820
2xYT media	Sigma-Aldrich	Y2377
Acrylamide/Bis 19:1, 40% solution	Fisher	187,328
Agarose	Omnipur	2125
Alkaline Hydrolysis Buffer	Invitrogen	AM2283
Alt-R Cas12a (Cpf1) Electroporation Enhancer	IDT	1,076,301
ATP, [g-32P]- 3000 Ci/mmol	Perkin Elmer	BLU002A001MC
DNase Alert	IDT	11-02-01-04
Dulbecco's Modified Eagle Medium (DMEM)	Corning	10-017-CV
EDTA	Fisher Chemical	S311500
fetal bovine serum (FBS)	VWR	1500–500
Glycerol	Fisher bioreagents	BP229
HEPES	Sigma	PHG0001
Imidazole	Sigma-Aldrich	IX0005
isopropyl-β-D-thiogalactopyranoside (IPTG)	Goldbio	I-902
Luria-Bertani Broth (LB)	USBiological Life Sciences	L1505
Magnesium chloride	Sigma-Aldrich	M2670
penicillin-streptomycin	Gibco	10,378,016
Potassium chloride	Sigma-Aldrich	P9541
Q5 High-Fidelity DNA Polymerase	NEB	M0491
Quick Extract	Lucigen Corporation	QE09050
Recovery Medium	Lucigen Corporation	NC9996840
RNase Alert	IDT	11-02-01-02
SF nucleofection buffer	Lonza	S50-642
SOC media	NEB	B9020
Sodium chloride	JT Baker	4058
T4 PNK	NEB	M0236S
Terrific Broth (TB)	VWR	97,063–420
Tris (2-carboxyethyl) phosphine hydrochloride (TCEP)	Sigma-Aldrich	C4706
UREA	Sigma	u5378
Critical commercial assays		
DNA Clean and Concentrator kit	Zymo Research	D4033
HiSpeed Plasmid Maxi Kit	QIAGEN	12,663
QIAprep Spin Miniprep kit	QIAGEN	Cat#27106
Deposited data		
Cryo-EM density map of the Casλ-crRNA-dsDNA ternary complex	This paper	EMDB: EMD-27320
Coordinates of the Casλ-crRNA-dsDNA ternary complex	This paper	PDB: 8DC2

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Cell lines		
HEK293T	University of California Berkeley Cell Culture Facility	N/A
Oligonucleotides		
rBAS80 ssRNA oligo AUUGUUGUAACUCUUAUUUU GUAUGGAGUAAACAACUAGCAUCACCUUCACC	This paper	N/A
dBAS608 ssDNA oligo CTTGACCGTTTGATCGTAG TGGAAGTGGGAGATAGTAATGTTAATG	This paper	N/A
dBAS609 ssDNA oligo CATTAACATTACTAAGAGGG TGAAGGTGATGCTACAAACGGTCAAG	This paper	N/A
Recombinant DNA		
Cas12L1 Non-Targeting array and Cas12L1 ORF	This paper	pBAS12
pBAS12 with GFP targeting spacer	This paper	pBAS18
L1_orf L2 guide 19 targeting GFP	This paper	pBAS40
L1_orf L1 guide non-targeting GFP	This paper	pBAS41
L1_orf L1 guide non-targeting GFP	This paper	pBAS42
L1_orf L1 guide non-targeting GFP	This paper	pBAS43
L1_orf L1 guide 19 targeting GFP	This paper	pBAS44
Software and algorithms		
AlphaFold2	Jumper et al.[28]	https://github.com/deepmind/alphafold
ChimeraX version 1.2.5	Pettersen et al.[29]	https://www.cgl.ucsf.edu/chimerax/download.html
Coot version 0.9.4.1	Emsley et al.[30]	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
CRISPResso2	Clement et al.[31]	https://github.com/pinellolab/CRISPResso2
cryoSPARC version 3.2.0	Punjani et al.[32]	https://cryosparc.com/download
ImageQuant TL 8.1	Cytiva Life Sciences	https://www.cytivalifesciences.com/en/us/shop/molecular-biology/nucleic-acid-electrophoresis–blotting–and-detection/molecular-imaging-for-nucleic-acids/imagequant-tl-10-2-analysis-software-p-28619
IQTREE version 1.6.6	Nguyenet al.[33]	https://github.com/Cibiv/IQ-TREE/releases/tag/v1.6.6
LocSpiral	Kaur et al.[34]	https://github.com/1aviervargas/LocSpiral-LocBSharpen-LocBFactor-LocOccupancy
MAFFT version 7	Katoh and Standley[35]	https://mafft.cbrc.jp/alignment/software/
MinCED	Bland et al.[36]	https://github.com/ctSkennerton/minced
Mtriage	Afonine et al.[37]	https://phenix-online.org/documentation/reference/mtriage.html
Phenix version 1.19.2–4158	Liebschner et al.[38]	https://phenix-online.org/download/
SerialEM version 3.8.7	Mastronarde[39]	https://bio3d.colorado.edu/SerialEM/download.html
Topaz	Bepler et al.[40]	https://github.com/tbepler/topaz#installation
ViennaRNA	Lorenz et al.[41]	https://github.com/ViennaRNA/ViennaRNA
Other		
1.2/1.3 300 mesh UltrAuFoil gold grids	Mitigen	Q350AR13A
10 kDa MWCO concentrator	Amicon Ultra, Merck	UFC9010
4D-nucleofector	Lonza	N/A
Amersham Typhoon phosphorimager	GE Healthcare	N/A
Cytation 5 plate reader	BioTek	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
HiLoad 16/600 Superdex 200pg column	GE Healthcare	28,990,944
MicroPulser Electroporator	Bio-Rad	N/A
NanoDrop 8000 Spectrophotometer	Thermo Scientific	N/A
Ni-NTA Superflow Cartridge	QIAGEN	30,761
SH800 Cell Sorter	Sony Biotechnology	N/A

RESOURCE AVAILABILITY

Lead contact

Requests for additional information and reagents should be directed to the lead contact, Jennifer Doudna (doudna@berkeley.edu).

Materials availability

Reagents generated in this study are available, upon reasonable request, from the lead contact with a Materials Transfer Agreement.

Data and code availability

- Structural data were deposited in EMDB and PDB with accession numbers reported in the Key Resources Table.
- This paper does not report original code. Analyses were performed as previously described in Al-Shayeb et al. 2020.
- Any additional information required to reanalyze the data reported in this work is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mammalian models

Mammalian gene-editing experiments were performed in HEK293T cells obtained from University of California Berkeley Cell Culture Facility. HEK293T cells were female in origin and grown in DMEM media (Corning) containing 10% fetal bovine serum (VWR) and 100U/mL of penicillin-streptomycin (Gibco) at 37°C with 5% CO_2.

Plant models

PDS3 gene-editing was tested in *A. thaliana* protoplasts isolated from the leaves of 4-week-old plants. Following RNP screening experiments, protoplasts were incubated in W5 solution (4 mM MES pH 5.7, 0.5 M mannitol, 20 mM KCl) at RT for 12 h, then moved to 37°C for 2.5 h, followed by a final incubation at room temperature for 48 h.

Additional experiments in wheat (*Triticum aestivum*. L cv. Fielder) were performed using protoplasts extracted from leaves of 2-week-old plantlets. Edited cells were incubated in darkness with W5 solution at 30°C for 24 h.

METHOD DETAILS

Phylogenetic analysis

For analysis of publicly available phage genomes, we analyzed Genbank-recorded phages, complete RefSeq-recorded phages, IMG-VR-recorded phages. Cas protein sequences and representatives from the TnpB superfamily were collected from literature.[2,5,42–45] The resulting set was clustered at 90% amino acid identity to reduce redundancy. A new alignment of Casλ with the resulting sequence set was generated using MAFFT with 1000 iterations and filtered to remove columns composed of gaps in 95% of sequences. The phylogenetic tree was inferred using IQTREE v1.6.6 using automatic model selection and 1000 bootstraps.

crRNA sequence analysis

CRISPR-RNA (crRNA) repeats from Phage-encoded CRISPR loci were identified using MinCED (github.com/ctSkennerton/minced). The repeats were compared by generating pairwise similarity scores using the Needleman-Wunsch algorithm. A heatmap was built using the similarity score matrix and hierarchical clustering produced dendrograms that were overlaid onto the heatmap to delineate different clusters of repeats. The RNA structures were predicted with ViennaRNA.[41]

PAM depletion analysis

PAM depletion assays were performed with plasmids containing the casλ protein coding sequence as derived from metagenomics and a mini CRISPR targeting guide (pBAS18), or with plasmids that contained only the *casλ* gene and a non-targeting guide (pBAS12). Assays were performed as three individual biological replicates. Plasmids containing *casλ* and mini CRISPRs were transformed into *E. coli* BL21(DE3) (NEB). Subsequently, electrocompetent cells were prepared by ice-cold H_20 and 10% glycerol washing. A plasmid

library was constructed with 8 randomized nucleotides upstream of the (5′) end of the target sequence. Competent cells were transformed in triplicate by electroporation with 200 ng library plasmids (0.1 mm electroporation cuvettes (Bio-Rad) on a Micropulser electroporator (Bio-Rad)). After a 2 h recovery period, cells were plated on selective media and colony forming units were determined to ensure appropriate coverage of all possible combinations of the randomized 5′ PAM region. Strains were grown at 25°C for 48 h on media containing appropriate antibiotics (either 100 μg/mL carbenicillin and 34 μg/mL chloramphenicol, or 100 μg/mL carbenicillin and 50 μg/mL kanamycin) and 0.05 mM isopropyl-β-D-thiogalactopyranoside (IPTG), or 200 nM anhydrotetracycline (aTc), depending on the vector to ensure propagation of plasmids and Casλ effector production. Subsequently, propagated plasmids were isolated using a QIAprep Spin Miniprep Kit (Qiagen).

PAM depletion sequencing analysis

Amplicon sequencing of the targeted plasmid was used to identify PAM motifs that are preferentially depleted. Sequencing reads were mapped to the respective plasmids and PAM randomized regions were extracted. The abundance of each possible 8 nucleotide combination was counted from the aligned reads and normalized to the total reads for each sample. Enriched PAMs were computed by calculating the log ratio compared to the abundance in the control plasmids, and were used to produce sequence logos.

Programmable DNA targeting

A flp recombination assay was performed in *E. coli* to eliminate the Kanamycin resistance cassette from *E. coli* strains that contain GFP and RFP expression cassettes integrated into the genome. Individual colonies of the *E. coli*ΔKan were picked to inoculate three 5 mL (LB) starter cultures to prepare electrocompetent cells the following day. 100 mL (LB) main cultures were inoculated from the starter cultures and grown vigorously shaking at 37°C to an OD_{600} of 0.6–0.7 before preparation of electrocompetent cells by repeated ice-cold H_20 and 10% glycerol washes. Cells were resuspended in 10% glycerol and 50 μL aliquots were flash frozen in liquid nitrogen and stored at −80°C. Casλ vectors were generated containing codon optimized *casλ1* gene and a guide comprised of its cognate repeat element and selections of spacers targeting the GFP DNA within the resulting *E. coli*ΔKan strain (pBAS41, pBAS42, pBAS43, pBAS44) were subcloned from pBAS12. Casλ vectors containing Casλ1 and a guide composed of a non-cognate repeat unit from *casλ2* and a GFP-targeting spacer (TAGCATCACCTTCACCCTCTCCACGGACAG) guide were also subcloned to form pBAS40. The Casλ vectors and Casλ vectors with a non-targeting guide control plasmid were transformed into 25 μL of electrocompetent cells with 100 ng of plasmid via electroporation in 0.1 mm electroporation cuvettes (Bio-Rad) on a Micropulser electroporator (Bio-Rad), cells were recovered in 1 mL recovery medium (Lucigen) shaking at 37°C for 1 h 10-fold dilution series were then prepared and 3.5 μL of the respective dilutions were spot-plated on LB-Agar containing the appropriate antibiotics and IPTG inducer. Plates were incubated overnight at 37°C and colonies were counted the following day to determine the transformation efficiency. To assess the transformation efficiency, the mean and standard deviations were calculated from the cell forming units per ng transformed plasmids for the electroporation triplicates. The experiment showed marked reduction of GFP *E. coli* using Casλ vectors with their cognate guides (pBAS44) in comparison to the non-targeting control, indicating a dsDNA break at the target region. The growth of primarily RFP-positive/GFP-negative colonies under blue light further supports the ability to confer targeted programmable genome editing to result in strains lacking GFP production. Growth inhibition using Casλ vectors with guides from a separate Casλ ortholog (pBAS40), with colonies observed expressing primarily RFP and no GFP, also indicate that Casλ orthologs may function using guides from related CRISPR-Cas systems to confer editing in cells, with a precise ablation of GFP production. This can be further expanded to HEK293T mammalian cells with integrated GFP, which indicate activity in mammalian cells. The sickly phenotype of *E. coli* colonies that have grown in both cases even in undiluted samples is also indicative of possible *trans*-cleavage of nucleic acids (RNA or DNA), which can be used for diagnostic purposes by providing a sample containing the target nucleic acid with the Casλ RNP and a ssDNA fluorophore-quencher (ssDNA-FQ) reporter or RNA fluorophore-quencher (ssRNA-FQ) reporter molecule, generating a strong fluorescence signal in the presence of the target nucleic acid compared to a markedly lower fluorescence signal in its absence.

Protein purification

Casλ overexpression vectors containing a His-Tag were transformed into chemically competent *E. coli* BL21(DE3)-Star (QB3-Macrolab, UC Berkeley) and incubated overnight at 37°C on LB-Kan agar plates (50 μg/mL Kanamycin). Single colonies were picked to inoculate 50 mL (LB, Kanamycin 50 μg/mL) starter cultures which were incubated at 37°C shaking vigorously overnight. The following day, 1.5 L TB-Kan media (50 μg/mL Kanamycin) were inoculated with 40 mL starter culture and grown at 37°C to an OD_{600} of 0.6, cooled down on ice, and gene expression was subsequently induced with 0.5 mM IPTG followed by incubation overnight at 16°C. The cells were harvested by centrifugation and resuspended in wash buffer (50 mM HEPES-Na pH 7.5 RT, 500 mM NaCl, 20 mM Imidazole, 5% glycerol and 0.5 mM TCEP), and then subsequently lysed by sonication. The soluble fraction was loaded on a 5 mL Ni-NTA Superflow Cartridge (Qiagen) which had been pre-equilibrated in the same wash buffer. Bound proteins were washed with 20 column volumes (CV) wash buffer and subsequently eluted in 5 CV elution buffer (50 mM HEPES-Na pH 7.5 RT, 500 mM NaCl, 500 mM imidazole, 5% glycerol, and 0.5 mM TCEP). The eluted proteins were concentrated to 1 mL before injection into a HiLoad 16/600 Superdex 200pg column (GE Healthcare) pre-equilibrated in size-exclusion chromatography buffer (20 mM HEPES-Na pH 7.5 RT, 500 mM NaCl, 5% glycerol, and 0.5 mM TCEP). Peak fractions were concentrated to 1 mL and concentrations were determined using a NanoDrop 8000 Spectrophotometer (Thermo Scientific). Proteins were purified at a constant temperature of

4°C and concentrated proteins were kept on ice to prevent aggregation, snap-frozen in liquid nitrogen, and stored at −80°C. SDS-PAGE gel electrophoresis of Casλ at varying stages of protein purification showed a protein size in line with computationally predicted values of 85 kDa.

Pre-crRNA processing assays

The reactions were carried out in RNA cleavage buffer containing 20 mM Tris-Cl (pH 7.5 at 37°C), 150 mM KCl, 5 mM $MgCl_2$, 1 mM TCEP, and 5% (v/v) glycerol. Pre-crRNA substrates were 5′-radiolabeled with T4 PNK (NEB) in the presence of gamma ^{32}P-ATP. In a typical pre-crRNA processing reaction, the concentrations of Casλ and ^{32}P-labeled pre-crRNA substrates were 100 and 3 nM, respectively. Reactions were incubated at 37°C, and an aliquot of each reaction was quenched with 2x Quench Buffer (94% (v/v) formamide, 30 mM EDTA, 400 μg/mL heparin, 0.2% SDS, and 0.025% (w/v) bromophenol blue) at 0, 1, 5, 15, 30, and 60 min. RNA hydrolysis ladders were prepared by incubating RNA probes in 1X RNA Alkaline Hydrolysis Buffer (Invitrogen) at 95°C before the addition of 2x Quench Buffer. Quenched reactions were incubated at 95°C for 3 min, and products were then resolved by denaturing PAGE (10% or 20% acrylamide:bis-acrylamide 19:1, 7 M urea, 1X TBE). Gels were dried (3 h, 80°C) on a Model 583 Gel Dryer (Bio-Rad) and exposed to a phosphor screen. Phosphor screens were imaged on an Amersham Typhoon phosphorimager (GE Healthcare). For assays in an EDTA-containing buffer, 25 mM EDTA was substituted for 5 mM $MgCl_2$.

In vitro cleavage assays - Radiolabeled nucleic acids

crRNA oligonucleotides were synthesized by IDT and dissolved in DEPC-treated ddH_20 to a concentration of 0.5 mM. Subsequently, the crRNA was heated to 65°C for 3 min and allowed to cool down to room temperature. Casλ RNP complexes were reconstituted at a concentration of 10 μM by incubation of 10 μM Casλ and 12 μM crRNA for 10 min at RT in 2x cleavage buffer (20 mM Hepes-Na pH 7.5, 300 mM KCl, 10 mM $MgCl_2$, 20% glycerol, 1 mM TCEP). RNPs were aliquoted to a volume of 10 μL, flash-frozen in liquid nitrogen, and stored at −80°C. RNP aliquots were thawed on ice before experimental use. Substrates were 5′-end-labelled using T4-PNK (NEB) in the presence of ^{32}P-γ-ATP. Oligonucleotide-duplex targets were generated by combining ^{32}P-labelled and unlabelled complementary oligonucleotides in a 1:1.5 M ratio. Oligos were hybridized to a DNA-duplex concentration of 50 nM in hybridization buffer (10 mM Hepes-Na pH 7.5 RT, 150 mM NaCl), by heating for 5 min to 95°C and a slow cool down to RT in a heating block. Cleavage reactions were initiated by combining 200 nM RNP with 2 nM substrate in CB buffer and subsequently incubated at 37°C. Reactions were stopped by the addition of two volumes of formamide loading buffer (96% formamide, 100 μg/mL bromophenol blue, 50 μg/mL xylene cyanol, 10 mM EDTA, 50 μg/mL heparin), heated to 95°C for 5 min, and cooled down on ice before separation on a 12.5% denaturing urea-PAGE. Gels were dried for 4 h at 80°C before phosphor-imaging visualization using an Amersham Typhoon scanner, v2.0.0.6 firmware version 208 (GE Healthcare). Bands were quantified using ImageQuant TL 8.1 (Cytiva) and the cleaved fraction was calculated as the product intensity sum divided by the combined substrate and product intensity sum. Curves were fitted to a One-Phase-Decay model to derive the rate of cleavage.

Fluorophore quencher and DNA mismatch tolerance assay

DNA oligo activators were ordered from IDT to contain mismatches at each respective position, (A- > C, T- > G, C- > A, G- > T). Casλ RNPs were prepared as described above. Reactions were started by combining 100 nM RNP (100 nM Casλ, 120 nM crRNA), 100 nM DNase Alert (IDT) FQ probe, with and without activator ssDNA and with the addition of a non-targeting guide or activator control in cleavage buffer in a 384 well flat bottom black polystyrene assay plate (#3820, Corning). Three replicates for each reaction were monitored (λex: 530 nm; λex: 590 nm) in a Cytation 5 plate reader (BioTek, software Gen v3.04) at 37°C every 1.5 min for the activator titration experiment. For the FQ-mismatch-assay, 2 nM activator oligonucleotides were used in singlicates. The data were background-subtracted using the mean values of the measurements taken for three no-activator controls at the respective time point.

I-X PAM binding assay

The PAM binding assay was conducted using NEB 5-alpha Competent *E. coli* cells. Plasmids containing the type I-X system included a targeting or non-targeting guide downstream of T7 promoters. PAM library plasmids contained sfGFP under the control of an araBAD promoter. downstream of the promoter was a six-nucleotide variable region of potential PAM sequences, resulting in loss of sfGFP fluorescence for a successful PAM binding event. All cultures used 2xYT media and were supplemented with kanamycin and ampicillin as needed for plasmid maintenance. Cell densities were maintained at greater than 100x library coverage throughout the assay.

Transformations of type I-X systems with guide and library plasmids were conducted consecutively. Type I-X systems with guides were transformed into NEB 5-alpha Competent *E. coli* cells following manufacturer's instructions. Individual colonies were incubated at 37°C overnight at 250 RPM. Non-transformed cultures were included for library only and no plasmid controls. Cells were back diluted 1000x and cultured to ABS_{600} ~0.6, pelleted, washed 3 times with sterile water, and resuspended in 10% glycerol to make them electrocompetent. Type I-X systems with guide, and non-transformed cultures were electroporated at 1800V with 100 ng of PAM library stock and recovered for 1 h in SOC media. Recovered cells were plated with appropriate antibiotics and incubated overnight. Plates were scraped, resuspended, and incubated at 37°C 250 RPM for 3 h. 25% glycerol stocks were stored at −80°C.

In preparation for fluorescence-activated cell sorting, stocks were back-diluted 100x and cultured to ABS_{600} ~1 OD. To induce proteins and guides, cultures were back diluted 1000x, supplemented with 0.5 mM IPTG and 1% arabinose. Strains were cultured overnight at 30°C until targeting and non-targeting strains reached an ABS_{600} ~1 OD. Prior to sorting on the Sony SH800 Cell Sorter, cells were pelleted and resuspended in PBS to 0.6 OD. The non-targeting strain was used to set forward scatter (Gate 1) and singlets (Gate 2) gates and to detect any decrease in fluorescence in the targeting strain. For the targeting strain, at least 270,000 events were sorted for the lowest ~0.13% of the fluorescent cell population (Gate 3) and 500,000 events for the next ~0.13 to 1.23% lowest fluorescent cells (Gate 4). Sorted cells were grown overnight on plates containing appropriate antibiotics along with non-targeting, library only, and no plasmid controls.

Targeting, non-targeting, and library only strains were individually prepared for next generation sequencing by first purifying plasmid DNA using a Qiagen HiSpeed Plasmid Maxi Kit. Plates were gently scraped and resuspended in ~50 mL 2xYT prior to pelleting. Concentrations were determined with a Nanodrop. In conjunction with the original naive PAM library stock control, PAM sequences were amplified using primers containing the 5′ stub sequence GCTCTTCCGATCT. Samples were submitted to the Innovative Genomics Core for completion of library preparation and iSeq sequencing at greater than 100x library coverage.

Mammalian genome editing

RNPs were formed in the SF nucleofection buffer (Lonza) with 100pmol protein & 120pmol crRNA at 10μM concentration for 10′ at RT. 78 pmol (1μL) of IDT Cas12a electroporation enhancer was then added. HEK293T cells (University of California Berkeley Cell Culture Facility) were added in a 10μL SF nucleofection buffer at 200,000 cells per nucleofection. 21μL reactions were loaded into cuvettes and electroporated with pulse code DS-150 in a 4D-nucleofector (Lonza). Nucleofections were performed in triplicate for each guide RNA tested. Cells were grown in duplicate in DMEM media (Corning) containing 10% fetal bovine serum (VWR) and 100U/mL of penicillin-streptomycin (Gibco) from each nucleofection in 24-well plates at 37°C with 5% CO_2. gDNA was collected after 72 h in Quick Extract (Lucigen) by heating at 65°C for 20 min followed by 95°C for 20 min. PCR1 was performed followed by bead clean-up to remove primers and submitted for PCR2, bead clean-up, and iSeq (Illumina) at the IGI Center for Translational Genomics. Approximately 20,000 reads per sample (2 x 150bp) were analyzed for genome editing using CRISPResso2 (https://crispresso.pinellolab.partners.org/login).

Plant genome editing

Guides were designed to target the PDS3 gene in *A. thaliana* protoplasts, incubated with protein as described for *in vitro* assays, and 26μL of 4μM RNP was transfected onto Arabidopsis protoplasts as previously described.[13]

Wheat (*T. aestivum*. L cv. Fielder) seedlings were grown in darkness on wet filter paper, wherein every third day seedlings were exposed to 6 h low light (~100 μE m-2 s-1). Protoplasts from 2-week-old plantlets were isolated from leaf tissue as previously described.[46] Leaves were cut into 0.5mm strips perpendicular to the leaf midrib in 0.6 M mannitol solution, then placed in freshly prepared enzyme solution (20 mM MES pH 5.7, 0.6 M mannitol, 10 mM KCl, 1.5% cellulase R10, 0.75% macerozyme R10). Leaf strips in solution were vacuum infiltrated for 30 min in darkness and then incubated for 6 h shaking at 70 rpm. After the incubation, the enzyme/protoplast solution was diluted with equal volume of W5 solution (2 mM MES pH 5.7, 154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl) and filtered through 40 μm cell strainers. Protoplasts were spun down at 80g for 3 min, then resuspended in 15mL W5 solution and left to aggregate in ice for 60 min. Supernatant was removed and protoplasts were resuspended in MMG solution (4mM MES pH 5.7, 0.4 M mannitol, 15mM $MgCl_2$) at $2.5x10^5$ cells/mL. Casλ RNP complexes were reconstituted with 6 μM Casλ protein, purified as described, and 10 μM guideRNA assembled in RNP reconstitution buffer (20mM Hepes-Na pH 7.5, 300 mM KCl, 10mM $MgCl_2$, 20% glycerol, 1mM TCEP) and incubated for 20 min at 37°C. 25 μL of 6 μM assembled RNP were added to a 1.5 mL tube, then mixed with 200 μL protoplasts. After flicking to mix, 220 μL PEG-CaCl2 solution (40% PEG 4000, 0.2 M mannitol, 100 mM $CaCl_2$) was added to tubes, and samples were mixed thoroughly by slowly inverting the Cryo-EM tube until streaks of PEG were no longer visible. Samples were incubated for 15 min in darkness at RT, then 880 μL of W5 solution was added and mixed by inverting. Protoplasts were harvested by centrifugation at 80*g* for 3 min, resuspended in 1.2 mL W5 solution (4 mM MES pH 5.7, 0.5 M mannitol, 20 mM KCl), and plated into 12-well plates. Plate edges were sealed with parafilm and cells were incubated for 24 h in darkness at 30°C. At the end of the incubation period, protoplasts were collected in 1.5 mL tubes, pelleted at 8000 g, and flash frozen. gDNA was extracted from protoplasts using 2X CTAB[47] and suspended in 50 μL DEPC-treated H_2O. 150–200 bp amplicons for amplicon sequencing were obtained using 30 cycles of Q5 High Fidelity DNA Polymerase, cleaned using the Zymo DNA Clean and Concentrator kit, then subsequently amplified using 15 cycles of PrimeSTAR GXL DNA Polymerase to add the requisite indices for sequencing. Amplicons were sequenced via paired-end 150 bp amplicon sequencing using an Illumina iSeq 100.

Ternary complex reconstitution for cryo-EM

Casλ was produced as described above. crRNA (rBAS80) was ordered as a synthetic RNA oligonucleotide from IDT and dissolved in DEPC-treated ddH_2O to a concentration of 0.5mM. Subsequently, the crRNA was heated to 65°C for 3 min and cooled down to RT to allow for hairpin formation. DNA oligonucleotides (dBAS608, dBAS609) were designed to contain a non complementary protospacer segment to produce 'bubbled' substrates and facilitate rapid R-loop formation during ternary complex reconstitution. Oligonucleotides were ordered from and synthesized by IDT. DNA oligonucleotides were combined in a 1:1.2 M ratio (target strand: non-target strand) and annealed to form a DNA duplex in hybridization buffer (10mM Hepes-Na pH 7.5 RT, 150mM NaCl) by heating for 5 min at 95°C and a subsequent slow cool down in a thermocycler.

Prior to reconstitution, thawed Casλ protein was incubated with crRNA in 1:1.1 ratio for 10 min at room temperature, and the DNA duplex was added. The ternary complex was reconstituted with the final Casλ: crRNA: TS: NTS strands stoichiometry of 1 : 1.1: 1.2 : 1.4, for another 10 min at RT, and further injected into a Superdex 200 prep grade 10/300 column (GE Healthcare) pre-equilibrated in low salt buffer (10mM Hepes-Na pH 7.5, 150mM NaCl) at 4°C to separate complexes from excess nucleic acids. Peak fractions were pooled and concentrated down to ~20 μM with a centrifugal filter device (Millipore 10 kDa Mw cutoff), as measured by absorbance at 260 nm with a NanoDrop 8000 Spectrophotometer (Thermo Scientific), and kept on ice before plunge-freezing.

Electron microscopy grid preparation and data collection

The resulting sample was frozen using FEI Vitrobot Mark IV, cooled to 8 °C at 100% humidity. 1.2/1.3 300 mesh UltrAuFoil gold grids (Electron Microscopy Sciences #Q350AR13A), were glow discharged at 15 mA for 25 s using PELCO easyGLOW. Total volume of 4 μL sample was applied to the grid and immediately blotted for 5 s with a blot force of 8 units. Micrographs were collected on a Talos Arctica operated at 200 kV and ×36,000 magnification (1.115 A pixel size), in the super-resolution setting of K3 Direct Electron Detector. Cryo-EM data was collected using SerialEM v.3.8.7 software. Images were obtained in a series of exposures generated by the microscope stage and beam shifts.

Single-particle cryo-EM data processing and 3D volume reconstruction

In total, 2795 movies were collected with a defocus range of −0.8 to −2.2 μm and a 20° tilt. Data processing was further performed in cryoSPARC v3.2.0[40]. Movies were corrected for beam-induced motion using patch motion, and CTF parameters were calculated using patch CTF. Two rounds of Topaz training[40] were applied to the data to enrich the amounts of Casλ ternary complex particles picked as follows. In the first round, as a result of initial curation, a subset of 562 micrographs with seemingly best ice quality and CTF fit were selected. Further, 3931 particles were manually picked and submitted to Topaz particle training. The resulting Topaz model was used to pick particles from the micrographs, and a total of 153,537 particles were extracted with bin factor 2**,** and applied to 2D classification. Following the selection of the best classes, 113,638 particles were used for *ab initio* reconstruction with three classes. The 55,587 particles constituting the best class in terms of resolution and resemblance to an RNP were subject to non-uniform map refinement,[48] and an initial complex map was obtained. In the second round, the latter particles were used to train a new Topaz model. Following the second round of curation, a total of 1931 micrographs were selected, and the new Topaz model was applied to pick and extract the particles. In total, 884,595 particles were subject to a round of 2D classification. After excluding a minor subset of classes, a total of 874,119 particles were selected and submitted to *ab initio* reconstruction with three classes. Three resulting maps and all particles were applied to a round of heterogeneous refinement. Particles constituting the best class in terms of resolution were subject to the remove duplicates procedure, and further to non-uniform map refinement. As a result, a 2.99 Å map reconstructed from a total of 369,389 particles was obtained. Half-maps from this refinement were used to generate the final LocSpiral map with improved weaker density regions.[34] This map was further used for model building.

Model building and refinement

The initial model of the Casλ protein was obtained with the AlphaFold2 program.[28] The predicted model was split into two parts (eventually constituting REC and Nuc lobes), and each was docked independently into the map with the fitmap tool in UCSF ChimeraX v1.2.5[38]. The dsDNA and crRNA models were built *de novo*. The combined ternary complex model was refined using the real-space refinement and rigid body fit tools in Coot v0.9.4.1[11]. Finally, the model was subject to a round of real_space_refine tool in Phenix v1.19.2–4158,[38] using secondary structure, Ramachandran, and rotamer restraints.

Data deposition and figure preparation

Cryo-EM maps and model coordinates were deposited to the EMDB (code EMD-27320) and PDB (code 8DC2). The structure figures were generated in UCSF ChimeraX v1.2.5[38]. Cryo-EM map σ levels were calculated as: map level/root-mean-square deviation from zero. The orientation distribution plots were either obtained from CryoSPARC or generated using pyem csparc2star.py and star2bild.py programs.[49] Map versus model Fourier shell correlation (FSC) graphs were calculated in Mtriage, as implemented in Phenix.[38] Gold standard FSC plot was generated in cryoSPARC.

QUANTIFICATION AND STATISTICAL ANALYSIS

Computational analyses and figures were prepared as described in Method Details. Where applicable, the SEM was graphed to depict variability across replicates.

Supplemental figures

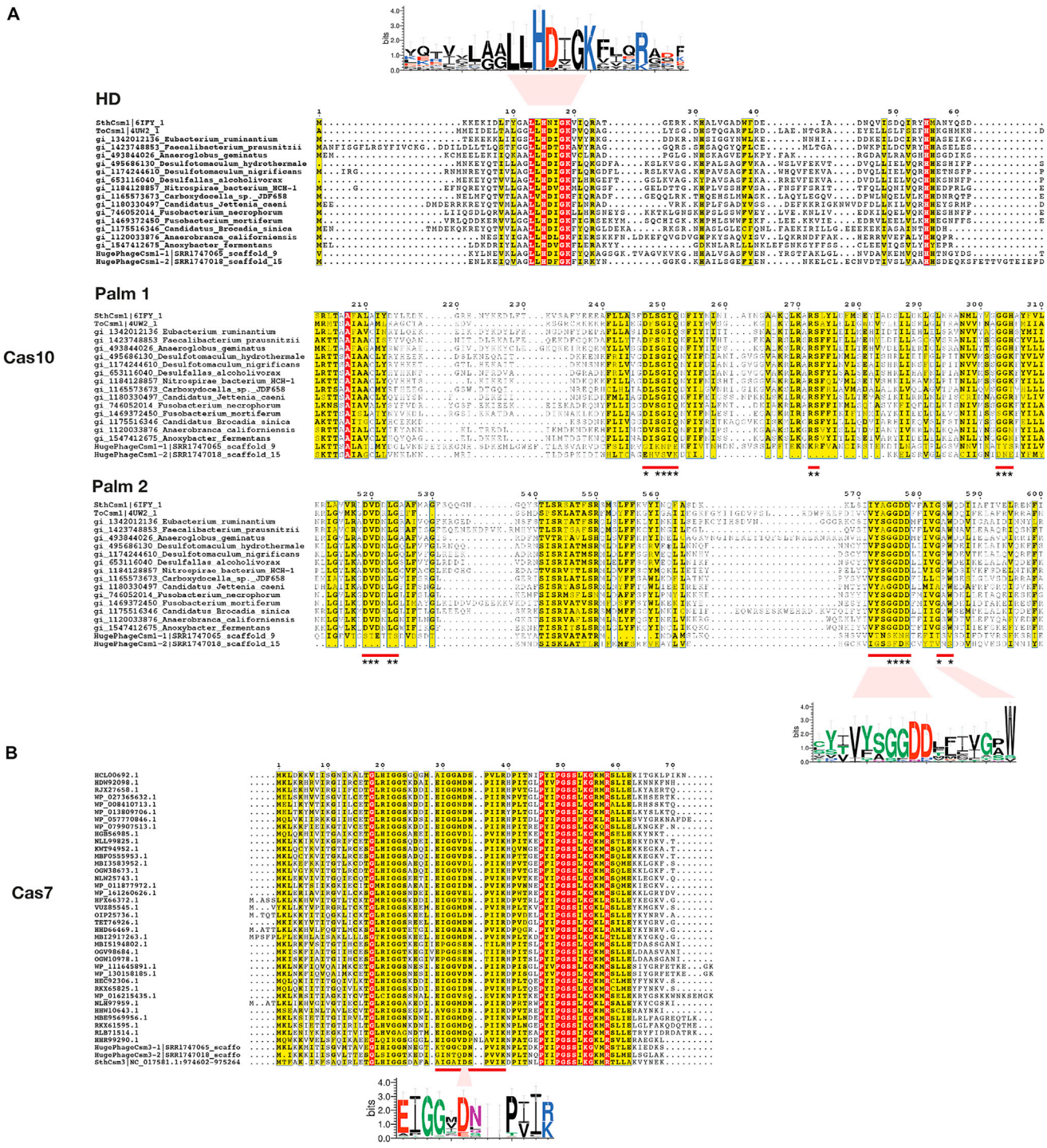


Figure S1. Sequence similarity of phage-encoded CRISPR-Cas systems, related to Figure 3

(A) Alignment of Cas10 effectors from Huge Phages with those sourced from bacterial genomes. Phage type III Cas10's are predicted to cleave DNA via Cas10 HD nuclease, but lack the residues required for the Palm domain to generate cyclic oligonucleotide signaling molecules.

(B) Alignment of Cas7 proteins from phages with those sourced from bacterial genomes. Phage type III Cas7's have conserved motifs that are predicted to cleave RNA.

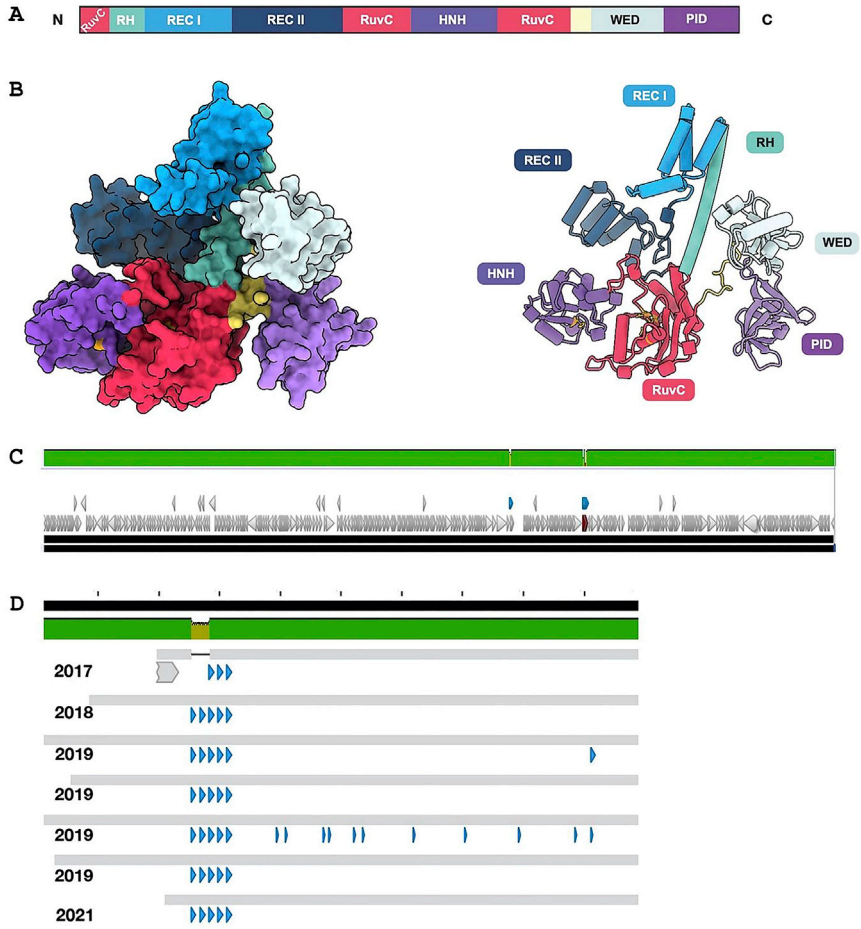


Figure S2. Structure of phage-encoded Cas9-like systems and comparison of type I-X CRISPR arrays, related to Figures 1 and 2

(A) Predicted domain organization for hypercompact phage-encoded Cas9-like systems.

(B) Predicted models for Cas9-like phage-encoded systems.

(C) comparison of type I-X and CasΦ-encoded Biggiephages recovered across a four-year time frame using Mauve, with the CRISPR repeat locations denoted in blue. Identical sequences at the nucleotide level are shown in green, with differences shown in brown or red.

(D) Phage type I-X CRISPR arrays from metagenomes sampled from the same site over the span of four years show remarkably stable arrays.

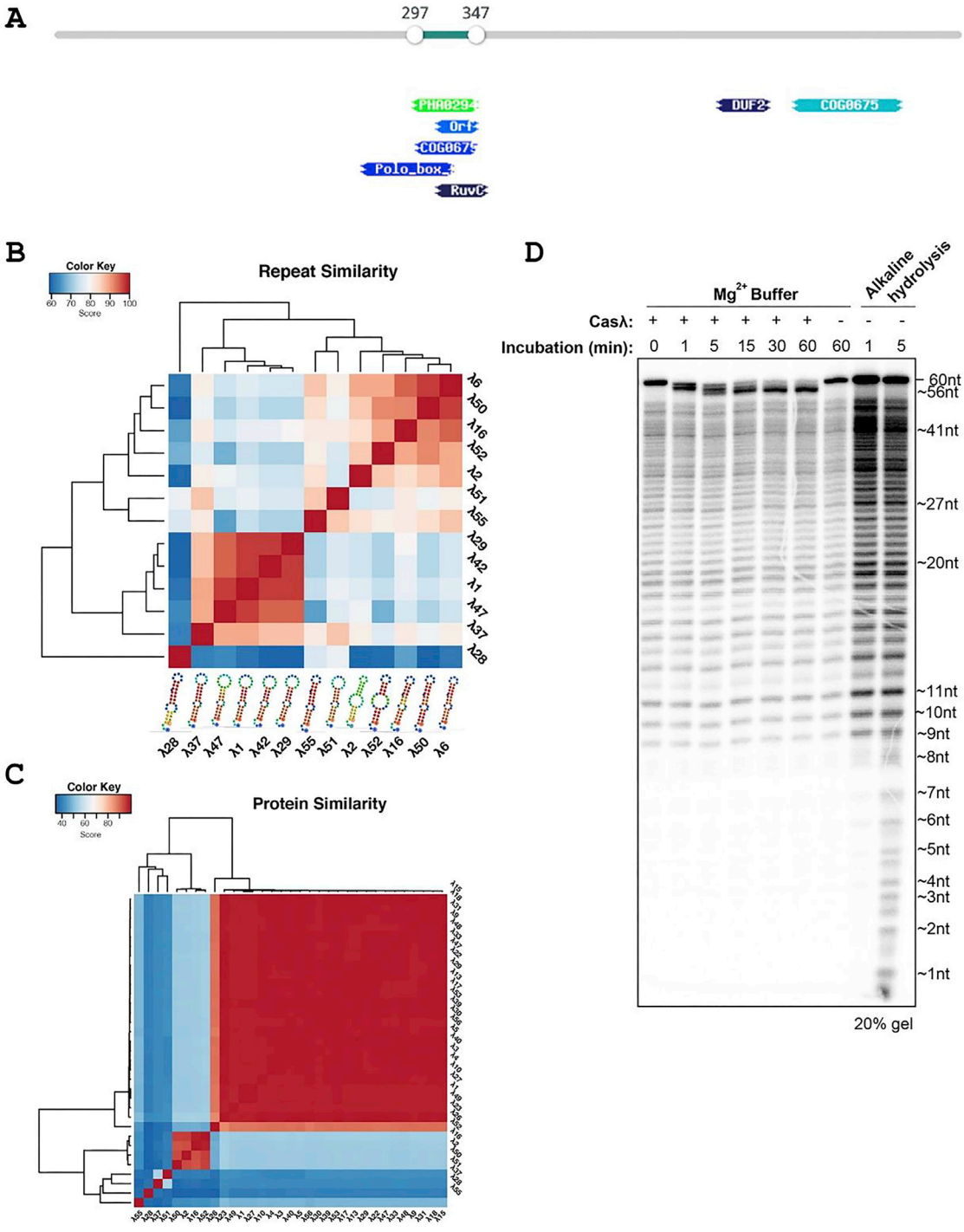


Figure S3. Divergent properties of Casλ, related to Figure 4

(A) Casλ remote homolog searches across public databases led to poor hits and no similarity to known CRISPR-Cas proteins, where only poor hits (green-black) were observed in one RuvC motif.

(B) Comparison of crRNA repeat similarity across orthologs.

(C) Comparison of protein similarity across orthologs.

(D) A time-series experiment incubating Casλ with 5' radiolabeled crRNAs with the product run on a 20% Urea PAGE gel supports the finding that Casλ uniquely processes its own single crRNA in the spacer region (or 3' end).

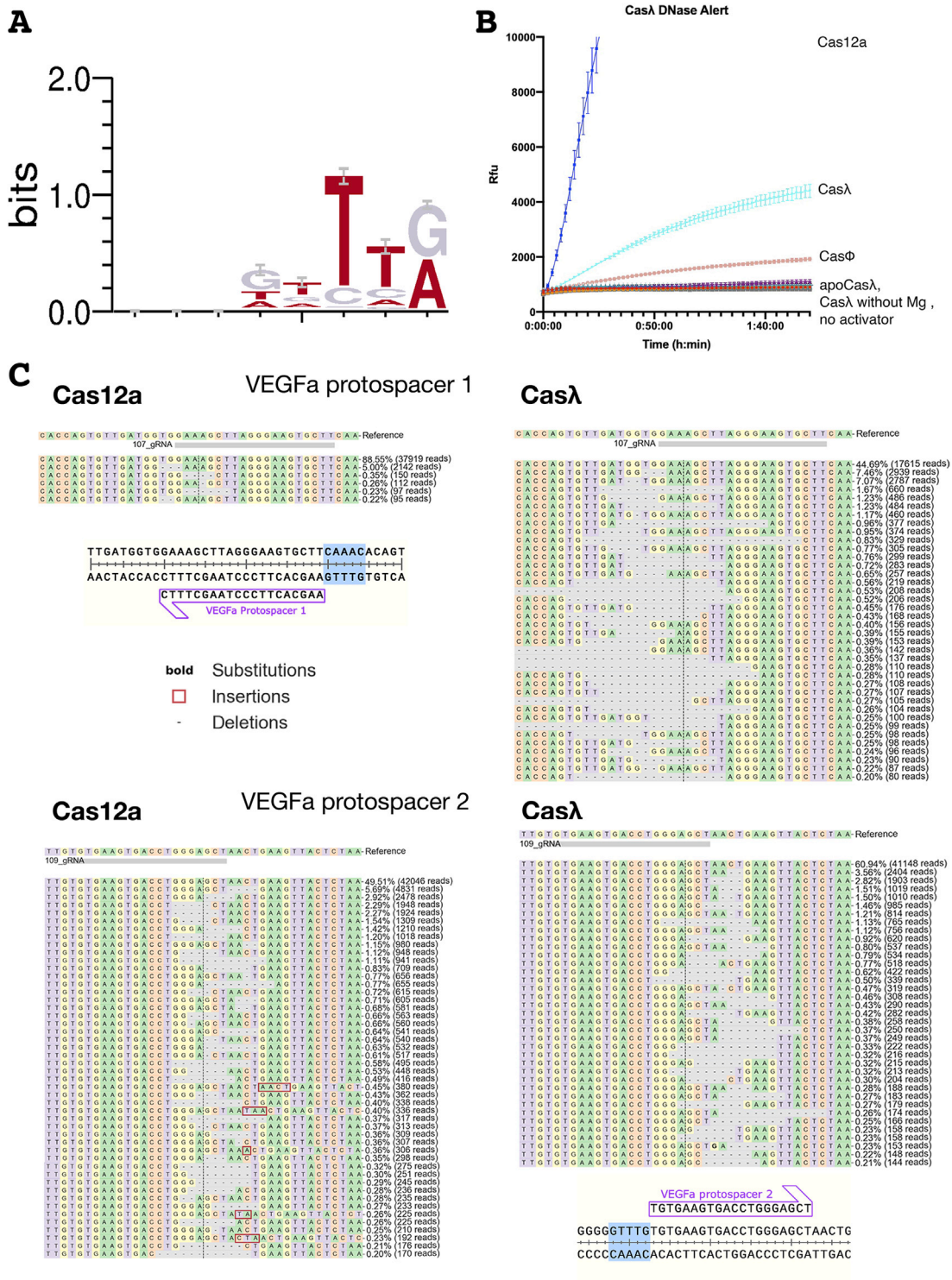


Figure S4. Casλ PAM specificity and comparison with other Cas otholog *trans*-cleavage and indel profiles, related to Figure 4

(A) The most depleted 5' PAMs resulting from the PAM depletion assay, indicating DNA recognition and cleavage preferences for Casλ1. (B) DNAse alert *trans*-cleavage assay with the same molarities of Cas12a, Casλ, and CasΦ targeting the same ssDNA activator. (C) Casλ indel profile in HEK293T cells compared to AsCas12a. Guide 107 targets the antisense strand, while guide 109 targets the sense strand of VEGFa.

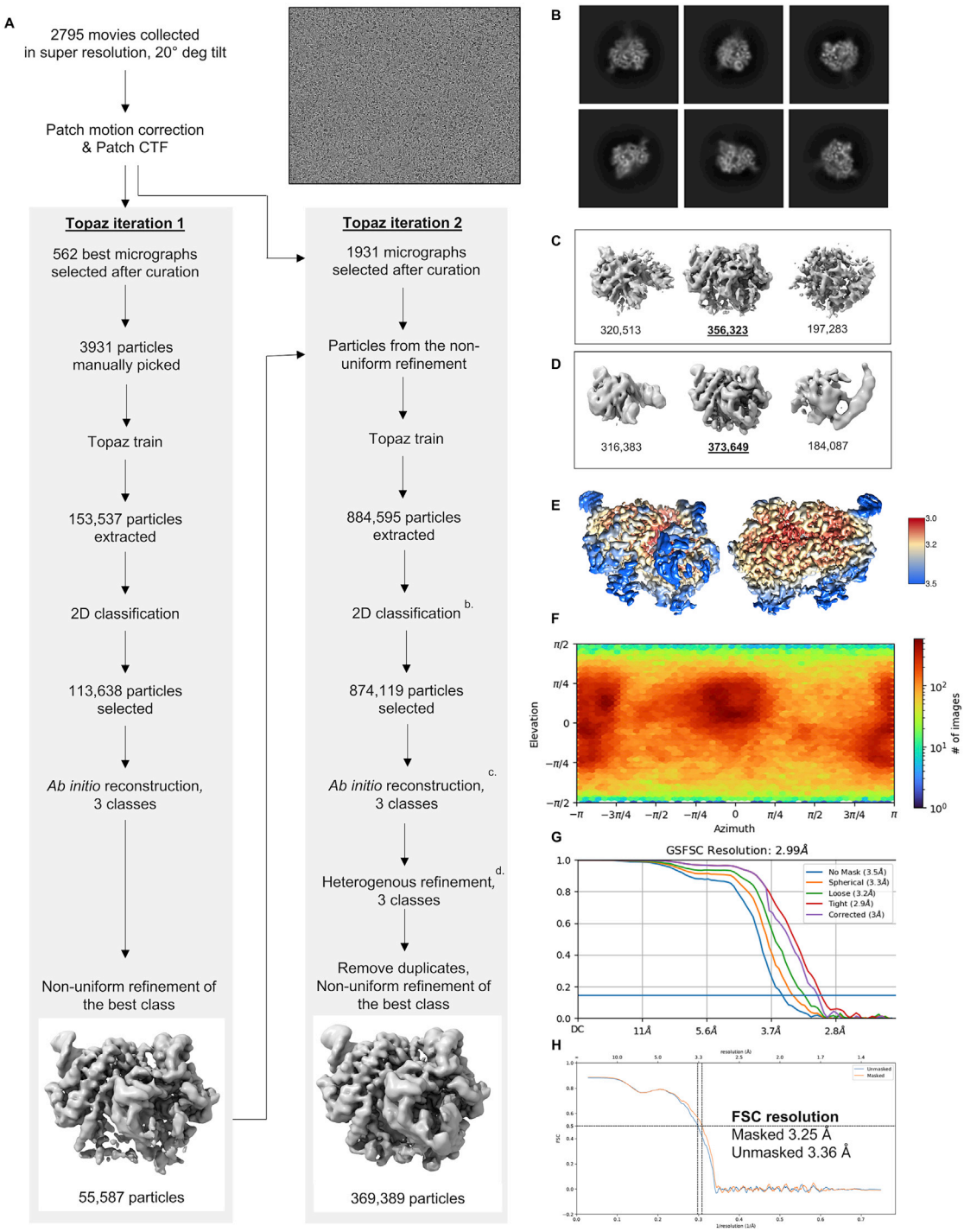


Figure S5. Cryo-EM workflow, related to Figure 6

(A) Map generation pipeline in cryoSPARC.

(B–D) Representative 2D class averages of the final set of particles, (C) the corresponding 3D maps resulting from *ab initio* reconstruction, and further (D) from heterogeneous refinement.

(E) Local resolution map as calculated in cryoSPARC v3.3.

(F) Orientation distribution of the final set of refined particles.

(G and H) gold standard, and (H) map versus model FSC curves of the model refined to the LocSpiral map and plotted with the final cryoSPARC sharp experimental map.

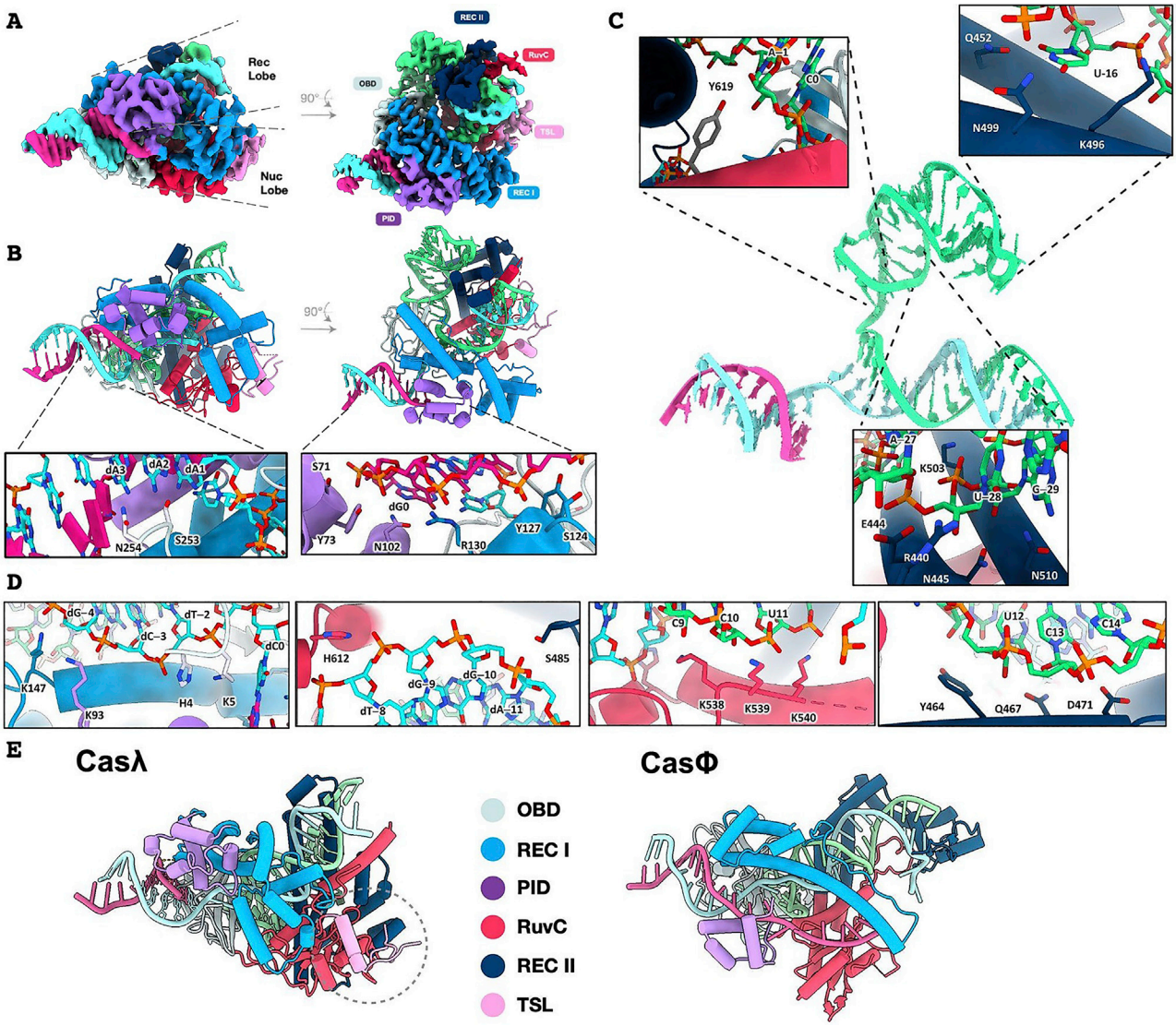


Figure S6. Structure of Casλ ternary complex, related to Figure 6

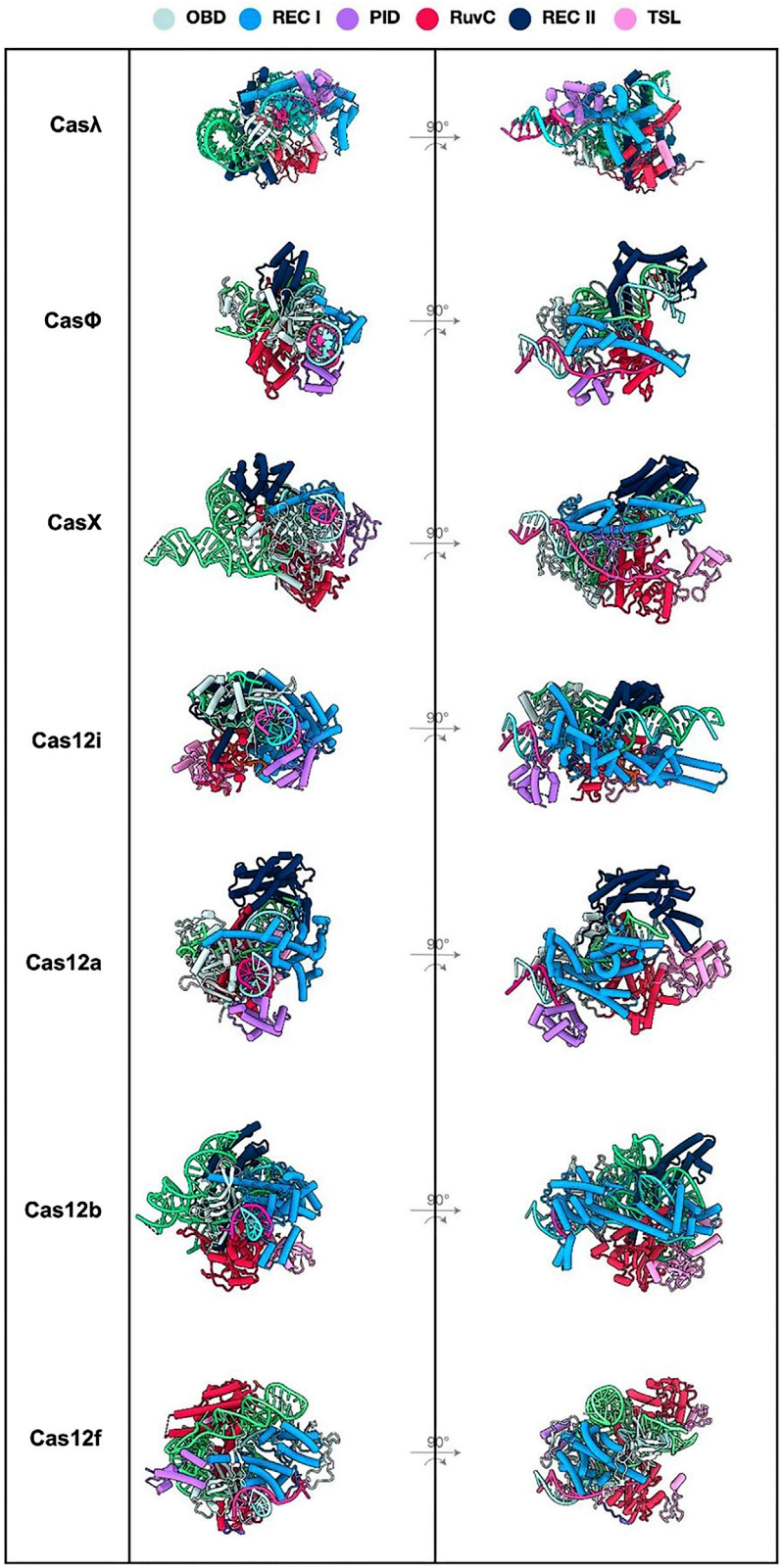
(A) Cryo-EM maps of the Casλ-guide RNA- DNA complex in two 90°-rotated orientations.

(B) Cartoon representation of the Casλ–gRNA–DNA complex. Insets highlight residues N102, S253, N254 predicted to be responsible for PAM recognition. Hydrogen bonds are shown as dashed lines.

(C) Model of guide RNA–target DNA complex, with insets highlighting residues conserved across the protein family that are predicted to be interacting with the RNA.

(D) Close-up views of the residues predicted to be responsible for recognition of the seed and low mismatch tolerance regions observed in (Figure 6F).

(E) Direct comparison of Casλ and CasΦ (PDB-ID: 7LYS) with a dashed bubble highlighting the Casλ TSL domain. Differences in RecI (Blue) can also be observed between the two proteins.



(legend on next page)

Figure S7. Structural comparison of Cas12 orthologs, related to Figure 6

Structural comparison of all DNA-targeting Cas12's in order of increasing RNP size: CasΦ (7LYS[26]), CasX (6NY3[27]), Cas12i (6W5C[50]), Cas12a (5XUS[51]), Cas12b (5WTI[52]), Cas12f (7C7L[53]).

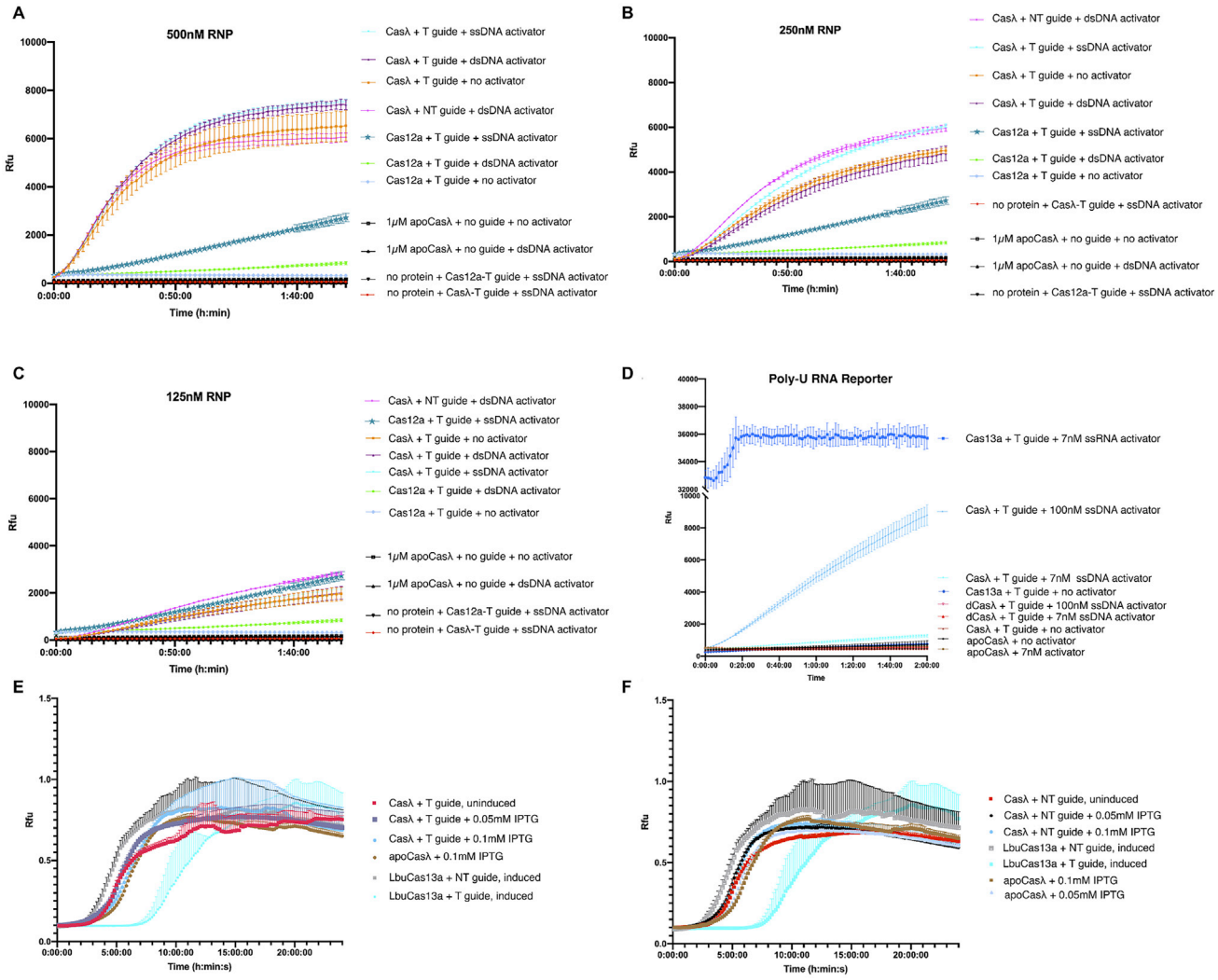


Figure S8. *Trans*-cleavage assay, related to Figure S4

(A–F)*Trans*-cleavage assays conducting with RNase Alert reporter substrate at decreasing RNP concentrations (A–C) for binary and ternary complexes of Casλ, and with (D) PolyU RNA reporter substrates, and testing cell viability assays with cells expressing Casλ in conjunction with (E) targeting and (F) non-targeting guides.