The CRISPR Journal

ORIGINAL ARTICLE

Functional and Phylogenetic Diversity of Cas10 Proteins

Tanner Wiegand,[1] Royce Wilkinson,[1] Andrew Santiago-Frangos,[1] Mackenzie Lynes,[2] Roland Hatzenpichler,[1,2] and Blake Wiedenheft[1,*]

Abstract

Cas10 proteins are large subunits of type III CRISPR RNA (crRNA)-guided surveillance complexes, many of which have nuclease and cyclase activities. Here, we use computational and phylogenetic methods to identify and analyze 2014 Cas10 sequences from genomic and metagenomic databases. Cas10 proteins cluster into five distinct clades that mirror previously established CRISPR-Cas subtypes. Most Cas10 proteins (85.0%) have conserved polymerase active-site motifs, while HD-nuclease domains are less well conserved (36.0%). We identify Cas10 variants that are split over multiple genes or genetically fused to nucleases activated by cyclic nucleotides (i.e., NucC) or components of toxin–antitoxin systems (i.e., AbiEii). To clarify the functional diversification of Cas10 proteins, we cloned, expressed, and purified five representatives from three phylogenetically distinct clades. None of the Cas10s are functional cyclases in isolation, and activity assays performed with polymerase domain active site mutants indicate that previously reported Cas10 DNA-polymerase activity may be a result of contamination. Collectively, this work helps clarify the phylogenetic and functional diversity of Cas10 proteins in type III CRISPR systems.

Introduction

Type III CRISPR-Cas systems are classified into six phylogenetically distinct subtypes (i.e., A-F), most of which rely on four to six Cas proteins that assemble around a processed CRISPR RNA (crRNA) to form a large multisubunit surveillance complex.[1–8] The protein subunits of these crRNA-guided surveillance complexes are called Csm (subtypes III-A and III-D), Cmr (subtypes III-B and III-C), or gRAMP (i.e., Cas7–11)—due to historical naming conventions.

Cas10 proteins (i.e., Csm1 or Cmr2) are considered signature proteins of type III CRISPR systems, although the recently described type III-E systems do not contain Cas10 proteins.[1] Beyond the type III systems, Cas10 proteins are part of the type I-D CRISPR surveillance complexes, although these homologs (i.e., Cas10d) do not produce cyclic oligoadenylates (cOAs), and are thought to be an evolutionary intermediate between the large subunits of type III systems (i.e., Cas10) and type I systems (i.e., Cas8).[1,5,9,10]

Type III Cas10 proteins are comprised of an N-terminal HD nuclease domain, two Palm domains that are separated by a zinc-finger motif (ZF), and a C-terminal domain (CT) that anchors Cas10 to other subunits in the respective Csm or Cmr complex (Fig. 1A).[3,4,6,11,12] Single-stranded DNase activity has been reported for the HD domains of some Cas10 proteins, which is thought to promote nicking of ssDNA formed during transcription.[13–15]

While the HD-mediated ssDNA nicking activity of Cas10 promotes mutagenesis and accelerates the evolution of antibiotic resistance in *Staphylococci*, HD nuclease activity is dispensable for immunity in some type III systems.[16,17] Indeed, numerous Cas10 proteins lack an N-terminal HD domain, although the abundance and phylogenetic distribution of these truncated proteins have not been systematically evaluated.[1,5]

In 2017, two groups independently showed that target RNA recognition by the Csm/Cmr complex activates Cas10 cyclase activity, which polymerizes ATP into cOA.[18,19] Several groups have recently taken advantage of these cOA products, reprogramming type III surveillance complexes for detection of specific RNA sequences *in vitro*.[20–24] Cas10-mediated cyclization of ATP relies on a di-glycine di-aspartate (GGDD) motif that is associated with one of the two Palm domains (i.e., Palm2).[3,18,19,25,26]

In addition to generating cOAs, which are secondary messengers that activate collateral nucleases and other

---

Departments of ¹Microbiology and Cell Biology and ²Chemistry and Biochemistry, Montana State University, Bozeman, USA.

*Address correspondence to: Blake Wiedenheft, Department of Microbiology and Cell Biology, Montana State University, Bozeman 59717, Montana, USA, E-mail: bwiedenheft@gmail.com
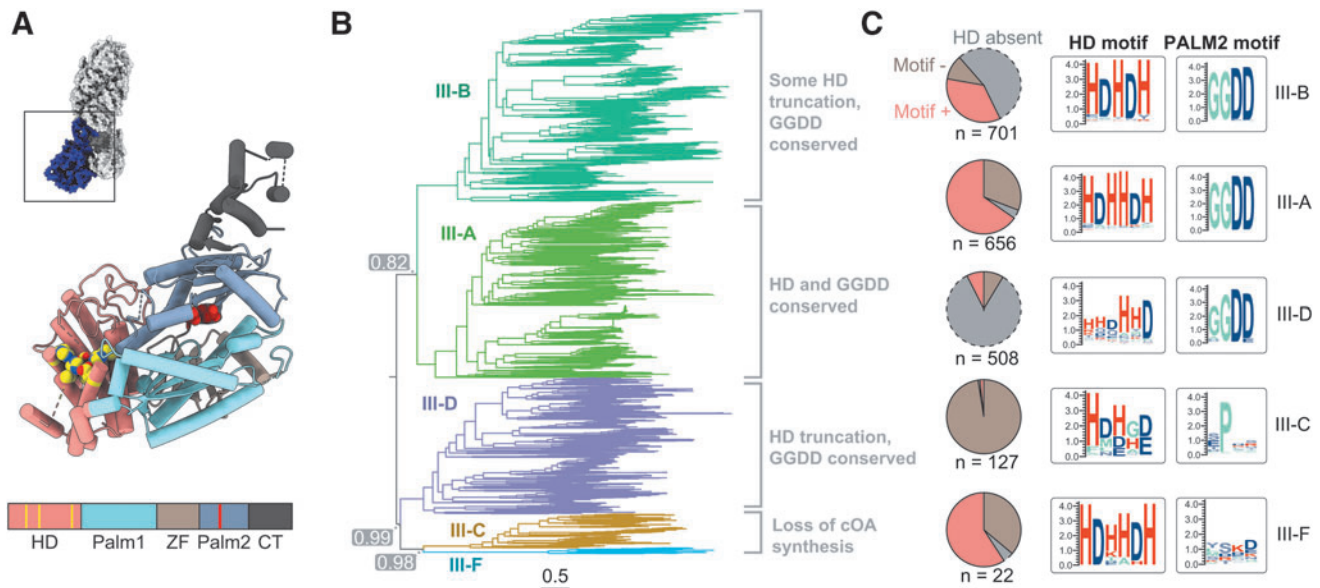
FIG. 1. Cas10 proteins have variable domain compositions, with subtype-specific patterns. **(A)** Cas10, shown in blue in upper left inset, is the large subunit of type III surveillance complexes (PDB: 6O1O). The structure of Cas10 from *Thermococcus onnurineus* NA1 (i.e., TonCas10) reveals five domains that are present in most structurally determined Cas10 proteins (PDB: 4UW2). HD and Palm2 active-site residues are mapped in yellow and red, respectively. **(B)** Phylogenetic reconstruction of 2014 Cas10 sequences recapitulates the recently updated type III subtype classifications.[1] Support values of critical nodes are highlighted in gray boxes, and scale bar indicates amino acid substitutions per site. **(C)** Pie charts depicting Cas10 sequences within each clade, where each section represents the number of sequences that are missing the HD (gray), contain mutated active-site residues (brown) or contain conserved HD-active-site residues (salmon). Active-site motif conservation of HD nuclease and Palm2 cyclase is shown for each Cas10 subtype on the right. Active-site residues were identified in one representative for each CRISPR subtype, which are shown in Supplementary Figure S1, and motif conservation within each Cas10 clade is quantified in Table S1. CT, carboxy-terminal domain; HD, histidine-aspartate nuclease domain; ZF, zinc-finger domain.

immune effectors, some Cas10 proteins may have DNA-dependent DNA-polymerase (DdDp) activity.[12,27] Degeneration of GGDD motifs has been reported for some type III-associated Cas10 proteins (i.e., some III-C and III-F), but like HD motifs, it remains unclear how widespread GGDD motifs are in Cas10 proteins.[1,5,28]

Here, we evaluate the phylogenetic diversity of Cas10 proteins and test the functional diversity of a subset of these proteins. We identify 2014 Cas10 sequences in genomic and metagenomic databases, and show that Cas10 sequences form two main lineages, one that includes Cas10 proteins that form the III-A and III-B clades, while the other includes Cas10 proteins that form the III-C, III-D, and III-F clades. Although Cas10 proteins from the same CRISPR subtype form monophyletic clades, some clades have greater diversity in the domains responsible for catalytic activities.

HD nuclease active-site motifs are most strongly conserved in Cas10 proteins from III-A and -F systems,

whereas Palm2 GGDD motifs are only strongly conserved in Cas10 proteins from III-A, -B, and -D loci. Further, Cas10 domain composition is more diverse than previously reported. We find that Cas10 subunits are sometimes split across multiple genes or fused to innate immune effectors (e.g., nucleases activated by cyclic nucleotides [NucC] from cyclic oligonucleotide-based anti-phage signaling systems [CBASS]) and toxin–antitoxin systems (e.g., AbiEii from AbiE).

To evaluate the functional diversity of Cas10 proteins, we clone and purify five representatives from three clades (i.e., III-A, -B, and -D) and perform activity assays. None of the isolated Cas10 proteins are functional cyclases,[18] but isolated Cas10 proteins have been previously reported to have DdDp activity.[27] We repeat the primer extension assays and detected weak, but reproducible DdDp activity for each of the purified proteins. However, mutations to Cas10 GGDD motifs do not abolish polymerase activity,[27] demonstrating that the polymerase

activity is not dependent on the Cas10 Palm2 domain, suggesting that this activity is from low-level contamination of a copurifying polymerase.

Methods

Identification of Cas10 proteins

Prodigal was used to translate all open-reading frames (ORFs) in 58,864 genomes downloaded from the National Center for Biotechnology Information (NCBI) database, 9687 genomes from the Joint Genomics Institute (JGI), and 21,210,363 metagenomic scaffolds from samples isolated from hot springs in Yellowstone National Park or enrichment cultures from a salt marsh.[29–31] Profile hidden Markov models (HMMs) of Cas10 from CasFinder were used to query the database of predicted ORFs.[32]

To identify III-F Cas10 proteins, we queried the database with an additional profile HMM that was constructed from 11 III-F Cas10 sequences identified by Makarova et al.[1] HMM queries were conducted in HMMER with an E-value cutoff of 0.01.[33] Search results were further filtered in R to exclude matches that did not meet a sequence E-value threshold of 1e-05. This search resulted in the identification of 5277 Cas10-like sequences. Redundant sequences were filtered using CD-HIT, resulting in a set of 3859 unique Cas10-like proteins.[34]

Phylogenetic and domain composition analysis

Three thousand eight hundred fifty-nine unique Cas10-like sequences were aligned in MAFFT using the "auto" setting.[35] This multiple sequence alignment (MSA) revealed numerous Cas10-like sequences that were truncated by the ends of incomplete metagenomic scaffolds. Sequences <500 residues in length were removed. This threshold was chosen to keep Cas10 sequences without an HD domain and was based on the sizes of structurally determined Cas10s (i.e., Cas10 lengths = 758 to 871 amino acids; lengths without HD domains = 519 to 669 amino acids).[4,11,36,37] Size filtering resulted in a set of 2014 Cas10 protein sequences (amino acid sequences and accession numbers in Supplementary Data S1).

This dataset was then separated into individual FASTA files for each type III subtype, based on which HMM had initially produced the best result, and Cas10 sequences within each subset were aligned in MAFFT using the LINSI setting. The resulting Cas10 MSAs were merged in MAFFT, and this MSA was used to construct a phylogenetic tree in FastTree with -wag and -gamma model settings.[38] The phylogenetic tree was rooted with the midpoint.root() function in R using the

phytools package,[39] and the resulting tree was visualized using the ggTree package from BioConductor.[40]

To determine if ancillary domains were present in Cas10-like sequences, we used the PFAM database and profile HMMs of Cas proteins extracted from the Cas-Finder program to annotate each Cas10-like sequence.[32,41] Results were manually analyzed, and only novel Cas10 gene fusions with E-values less than 1e-10 were reported.

Analysis of HD and Palm2 catalytic motifs

To assess the HD nuclease and Palm2 cyclase active sites of Cas10 proteins, we selected one representative from each CRISPR-Cas subtype (Supplementary Fig. S1 and Supplementary Data S1). Experimentally determined structures of III-A and III-B Cas10 representatives were available from the PDB (6KC0, 4W8Y; respectively), and representative III-C, III-D, and III-F Cas10 structures were predicted using an Alphafold Collab server or a local version of the program from DeepMind Technologies (v2.2.0).[42]

These structural models were used to guide analyses of each subtype-specific Cas10 MSA, and columns corresponding to active-site residues were extracted from each MSA using the BioStrings package in R.[43] A local version of the WebLogo application was used to plot the conservation of HD and Palm2 active-site motifs for each Cas10-associated CRISPR subtype (Fig. 1C),[44] and ChimeraX was used to visualize each protein.[45]

Cas10 cloning and site-directed mutation of active-site residues in Palm2 domains

Five Cas10 homologs (TonCsm1: WP_012571853.1, SthCsm1: WP_014621547.1, TthCsm1: WP_011229152.1, SsoCsm1: WP_231918241.1, and PfuCmr2: WP_011012269.1) were cloned into pRSFDuet-1 expression vectors. Cas10 genes were codon optimized for expression in *Escherichia Coli*, and include N-terminal 6x-histidine, 2x-strep-affinity, and SUMO solubilization tags.

Using the same approach, an additional vector was also constructed containing TonCsm1 (with an N-terminal six histidine tag) and TonCsm4, as previously described by Jia et al.[3] To test how mutations to GGDD motifs in Palm2 domains impact Cas10 polymerase activity, Q5 mutagenesis was used to mutate the two aspartate residues in these motifs to alanine residues (i.e., GGAA) (primers in Supplementary Table S2).

Cas10 protein purification

Protein expression was induced in 1-L cultures of *E. coli* BL21-DE3 cells when they reached an OD_{600} reading of 0.4–0.5 through the addition of 0.2 mM IPTG. Protein

expression continued overnight at 16°C before cells were collected through centrifugation at 3000 g (10 min at 4°C). Cells were lysed in 20 mM Tris-HCl, pH 7.5, 500 mM KCl, 10% glycerol, and 1 mM DTT through sonication, centrifuged at 10,000 g (25 min at 4°C), and soluble fractions of each protein were bound to Streptactin resin (IBA Lifesciences).

Contaminants were removed with 5 column volumes (CVs) of Wash Buffer 1 (20 mM Tris-HCl, pH 8.0, 500 mM KCl, 10% glycerol, 1 mM DTT), 20 CV Wash Buffer 2 (Wash Buffer 1 with 5 mM ATP and 20 mM MgCl$_2$), and 5 CV Wash Buffer 2 without the ATP and MgCl$_2$ before the bound proteins were eluted in 20 mM Tris-HCl, pH 7.5, 250 mM NaCl, 5% glycerol, 1 mM TCEP, 2.5 mM desthiobiotin. Proteins were concentrated (10k MWCO concentrator) before size exclusion chromatography.

SEC was performed using a Superdex 200 column equilibrated in 20 mM Tris-HCl, pH 7.5, 200 mM KCl, 10% glycerol, and 1 mM DTT (SEC profiles in Supplementary Fig. S4). This protocol yielded high-quality purifications of all five wildtype Cas10 proteins, TonCas10-Csm4, and four of the Palm2 Cas10 mutants, but we were unable to purify a Palm2 (GGAA) mutant for SsoCas10 (i.e., SsoCas10$_{mut}$) in quantities sufficient for subsequent biochemical activity assays.

Cas10 cyclase activity assays

To assess the cyclase (i.e., cOA-synthase) activities of five Cas10 proteins, we assembled reactions comprised of 1 μM Cas10, 25 μM rATP, 0.08 μM ^{32}P-α-rATP, and a reaction buffer (20 mM Tris-HCl pH 7.8, 250 mM monopotassium-glutamate, 10 mM ammonium sulfate, 5 mM magnesium sulfate, 1 mM TCEP). Reactions were incubated at either 37°C (SthCsm1) or 60°C (TthCsm1, SsoCsm1, TonCsm1, PfuCmr2) for 1 h. The TthCsm and Csm from *Streptococcus thermophilus* (SthCsm) complexes with crRNAs targeting the N gene of SARS-CoV-2 served as positive controls.[21]

TthCsm and SthCsm reactions were prepared as described above, with the addition of 10^{12} copies of *in vitro*-transcribed target RNA containing a 76-nucleotide fragment of the SARS-CoV-2 N gene sequence (GGG AACUGAUUACAAAC**AUUGGCCGCAAAUUGCAC AAUUUGCCCCCAGCGCUUCAGCGU**UCUUCGG AAUGUCGCGC; TthCsm target in bold, SthCsm target underlined). Nucleic acid products were phenol-chloroform extracted from each reaction and were diluted 2:1 with 100 mM sodium acetate at pH 5.2. One microliter of extracted nucleic acids was spotted ∼1.5 cm above the bottom edge of a TLC (thin-layer chromatography) Silica gel plate (Sigma-Aldrich).

The TLC plate was placed in a 2 L beaker filled with a small volume of mobile phase (0.2 M ammonium bicarbonate pH 9.3, 70% ethanol and 30% water), the beaker covered with aluminum foil, and the mobile phase allowed to wick up the TLC plate for 2 h at room temperature. The TLC plate was then dried and exposed to a phosphor screen, which was imaged with a Typhoon 5 phosphorimager (GE Healthcare).

Cyclase activity was also assessed using a previously described calcein-based fluorescence assay.[21] Cas10 proteins were incubated in a calcein-containing buffer (20 mM Tris-HCl pH 8.8, 100 mM KCl, 10 mM ammonium sulfate, 6 mM magnesium sulfate, 0.5 mM MnCl$_2$, 1 mM TCEP, 25 μM calcein) with 1 mM dNTP, rNTP, or rATP alone. Reactions were held at 4°C for 150 s in a QuantStudio RT-qPCR thermocycler (ThermoFisher), before being heated to 37°C and held at this temperature for 1 h. Fluorescence was measured every 10 s using a FAM-detection setting. Fluorescence data were quantified and analyzed in R.

Cas10 polymerase activity assays

To evaluate the DdDp activity of 5 wild-type Cas10 and 4 Palm2 Cas10 mutant (i.e., GGAA) proteins, we first annealed a 20-nucleotide-long 5′-^{32}P-labeled DNA primer (5′-ACTGCTAATAACGAGCGTTG-3′) to a 40-nucleotide DNA template (5′-TATCATTGGCTCCTTC AATCCAACGCTCGTTATTAGCAGT-3′) at a ratio of 1:3 (primer:template). Reactions consisting of 1 μM Cas10, 1 nM DNA-primer DNA-template, and a reaction buffer (20 mM Tris-HCl pH = 7.0, 100 mM NaCl, 2.5 mM MgCl$_2$, 10% (v/v) glycerol, 2 mM DTT, 250 μM dNTP, 0.05 mg/mL BSA) were incubated at 37°C for 1 h.

The reaction was stopped by adding one volume of 2×RNA loading buffer (95% formamide, 0.02% SDS, 0.02% bromophenol blue, 1 mM EDTA). Samples were heat denatured for 5 min at 95°C, and then cooled on ice before separation using a 19% PAGE gel containing 7M urea in 1X TBE. Gels were exposed to a phosphor screen overnight at −20°C, before the screen was imaged on a Typhoon 5 scanner.

Results

Phylogenetic diversity and proposed evolutionary history of Cas10 proteins

To evaluate the phylogenetic diversity of Cas10 proteins, we used profile HMMs to identify Cas10 sequences in 68,551 genomes from the NCBI and the JGI databases. In addition, we queried 21,210,363 metagenome-scaffolds from hot spring sediment and microbial samples from Yellowstone National Park and salt marsh sediment enrichments.[30,31]

Collectively, these queries identified 2014 unique Cas10 sequences, which were aligned and used to construct a phylogenetic tree (Fig. 1B). This analysis reveals five phylogenetically distinct clades that recapitulate the previously established CRISPR-Cas subtype designations.[1,5] Cas10 sequences from type III-A and III-B systems form two clades on one arm of a bifurcated tree, while Cas10s associated with III-C, III-F, and III-D systems form clades on the other arm. While little is known about the antiviral roles of type III-C and III-F systems, the shared ancestry of their Cas10 proteins suggests that these systems may have mechanistic similarities.

Differences in Cas10 HD and Palm active sites

The majority of type III Cas10 proteins that have been biochemically or biologically investigated to date contain an HD nuclease and catalytically active Palm2 domains.[2–4,6,11,18,46] To assess the prevalence of these features, we evaluated HD and Palm active sites in MSAs of Cas10 proteins from each type III subtype (Fig. 1C and Supplementary Fig. S1).

This analysis revealed that 58.6% of Cas10 proteins contain an N-terminal HD domain, and that HD active-site motifs are strongly conserved in III-A (HDHHDH motif) and III-F (HDHHDH motif) Cas10 proteins (Fig. 1C and Supplementary Table S1). Type III-B, -C, and -D Cas10 proteins have weaker conservation in their HD active sites (HDHDH, HDHHD, and HHDHHD motifs, respectively) or lack an N-terminal HD domain altogether, suggesting that many Cas10 proteins from these clades lack HD nuclease activity. The predicted loss of HD nuclease activity in III-D Cas10s supports previous bioinformatic observations.[47]

A di-glycine, di-aspartate (i.e., GGDD) motif is necessary for cOA synthesis in Cas10 Palm2 domains (Supplementary Fig. S1).[3,18,46] Evaluation of subtype-specific alignments of Cas10 sequences revealed strong conservation of GGDD motifs in Cas10s from III-A, III-B, and III-D (Fig. 1C and Supplementary Table S1), supporting reports that cOA signaling is important for immunity in these type III systems. Cas10s from III-C and III-F systems lack this conserved GGDD motif, in agreement with previous reports that III-F loci lack cOA-activated (e.g., CARF) effectors (Fig. 1C).[1,5,9,28,47]

We hypothesized that some of these sequences have a GGDD motif in Palm1 to compensate for their catalytically inactive Palm2 domains, but alignments of these sequences do not reveal a compensatory GGDD motif in Palm1. Interestingly, III-C and III-F Cas10s share a common ancestor (Fig. 1B), suggesting that cOA signaling was lost before the split and diversification of these CRISPR system subtypes.

Cas10 fusions and split Cas10s

To determine if Cas10 proteins contain any additional domains, we queried the Cas10 sequence library using profile HMMs of known Cas proteins and the PFAM database.[32,41] This analysis revealed genetic fusions of Cas10 to NucC, a widespread cA_3-activated effector of type III CRISPR and CBASS.[48] Four sequences were identified in *Bacillus* genomes with a ~ 300 residue NucC domain fused to the N-terminus of a type III-D Cas10 (i.e., NucC-Cas10) protein that lacks an HD domain (Fig. 2A and Supplementary Data S1). In each case, these NucC-Cas10 fusions are flanked by a stand-alone *nucC* gene.

While it is unclear if NucC-Cas10 and standalone NucC proteins in these systems interact, the NucC proteins from *E. coli* and *Pseudomonas aeruginosa* assemble into functional homotrimeric nucleases that are activated by cA_3.[48] Structural predictions indicate that NucC-Cas10 could substitute for one NucC subunit in higher order assemblies (Supplementary Fig. S2), suggesting that these systems physically link the cOA-generator (Cas10) to the cOA-activated effector (NucC), which may accelerate the chemistry necessary for immune defense.

In addition to NucC-Cas10 fusions, we found eight examples of a type IV toxin from a toxin–antitoxin system (i.e., AbiEii) fused to the N-terminus of Cas10 (Fig. 2B and Supplementary Fig. S3). AbiEii proteins are nucleotidyltransferases that bind specific nucleotides (e.g., GTP in *Streptococcus agalactiae*).[49,50] The genetic association of *abiEii* genes with type III CRISPRs has been previously reported,[49,51] but its fusion to *cas10* has not. AbiEii-Cas10 fusions are encoded in *cas* gene operons, where they are flanked by their cognate antitoxins (i.e., *abiEi*) (Fig. 2B).

AbiEii-mediated toxicity is thought to occur through polymerization of nucleotides to acceptor stems of uncharged tRNAs,[50,52] and genetic fusion of this protein to Cas10 implies that this toxin may have been co-opted for an immune response that inhibits protein synthesis.

Initial efforts to produce a phylogenetic reconstruction of Cas10 proteins resulted in trees with long branch lengths and poor support values. Upon closer inspection, we noticed that many of these long branches corresponded to truncated Cas10 sequences. While many of these truncations result from incomplete metagenomic scaffolds, *cas10* in the III-A system of
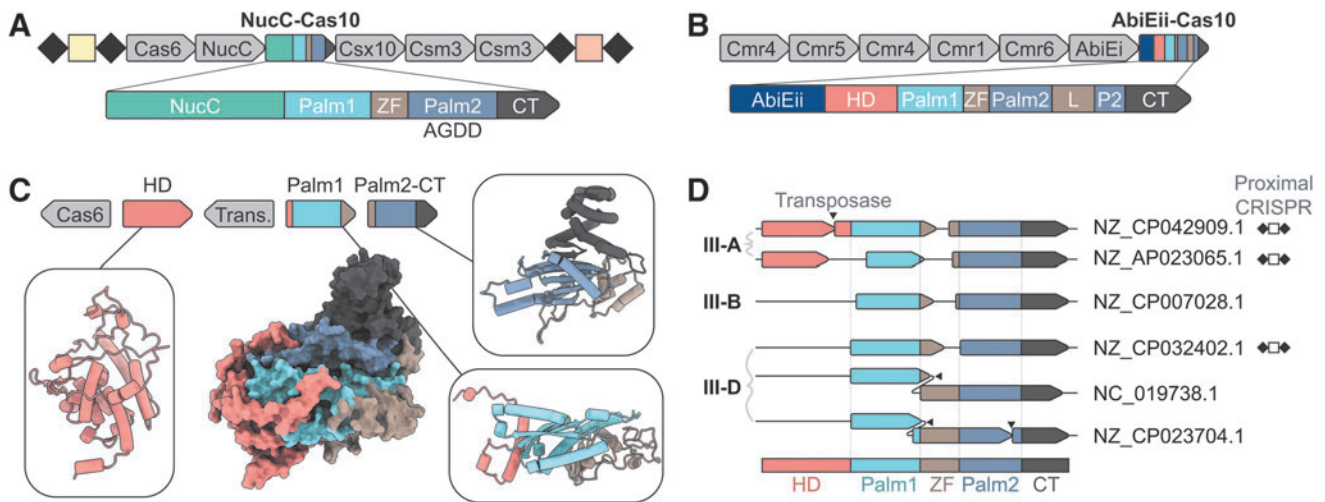
FIG. 2. Some Cas10 proteins are fused to additional domains or split over multiple genes. **(A)** Schematic of a CRISPR-Cas locus (JGI genome: 2654587738; JGI locus tag: Ga0100770_105251) illustrating an example of NucC fused to N-terminus of Cas10. **(B)** Schematic representation of CRISPR-Cas loci (NCBI genome: NC_022792.1; NCBI gene: WP_028843985.1) encoding a genetic fusion linking AbiEii to N-terminus of Cas10. **(C)** Cas10 from *Thermosulfurimonas marina* is split over three genes, with Cas10 domains defining the boundaries of each gene. An IS256 transposase ("Trans.") interrupts the first two partial Cas10 genes. AlphaFold predictions for each of the separate domains are shown as ribbons, and a multimeric structural prediction of all three proteins is shown as a surface representation (NCBI genome: NZ_CP042909.1). **(D)** Cas10 is sometimes split over multiple genes in type III-A, III-B, and III-D systems. Genomic accessions of each sequence and coverage of their domains are shown, with proximal CRISPR loci indicated to the right (repeats shown as black diamonds and white squares are spacers). Black triangles represent transposases of insertion elements. JGI, Joint Genomics Institute; NCBI, National Center for Biotechnology Information; NucC, nucleases activated by cyclic nucleotides.

Thermosulfurimonas marina (strain SU872) is split into three genes (Fig. 2C). Comparison of these three *cas10* sequences with structurally determined Cas10 proteins indicates that each III-A *T. marina* gene starts and ends near the boundary of an expected Cas10 domain. One gene contains an HD domain, another Palm1, and the third contains the Palm2 and C-terminal domains (Fig. 2C).

When these three protein sequences were submitted to AlphaFold for structural prediction using a multimer modeling option, the three partial Cas10 proteins assembled to form a heterotrimer that superimposes on Ton-Cas10 (PDB: 4UW2) with a root-mean-squared deviation (RMSD) of 1.23Å over 131 α-carbon atoms. No major steric clashes are predicted, consistent with their assembly into a functional Cas10 protein.

Cas10 proteins with domains split over two or three neighboring genes were also identified in type III-B and III-D systems (Fig. 2D). Some of these *cas10* genes are interrupted by insertion elements (e.g., IS256 transposase splits Cas10 in the III-A system of *T. marina*; NZ_CP042909.1), while the stop and start codons of other *cas10* genes are separated by short intergenic regions.

CRISPR loci are associated with some of these split *cas10* genes (e.g., *cas10* from the III-A system of *Nitratiruptor* sp. YY09–18; NZ_AP023065.1) (Fig. 2D), suggesting that these *cas* gene cassettes are associated with active defense systems.

Cas10-mediated cOA production requires the type III surveillance complex

Cas10 and their associated Csm or Cmr complexes from different organisms have been reported to produce different cOAs. For example, SthCsm mainly produces cA_6, while Csm from *Thermus thermophilus* (TtCsm) produces greater quantities of cA_4.[18,19] These cOA-production profiles are expected to impact the efficacy of type III CRISPR-based diagnostics, which generally utilize nuclease effectors that are activated by a single species of cOA.[20–24] To determine if Cas10 polymerases from each clade have characteristic cOA-production profiles, we picked five representatives from three distinct clades and tested the cOA-synthesis activity of these proteins.

The five Cas10 homologs we purified all have GGDD motifs in their Palm2 domains, have been previously

studied by other groups (i.e., TonCas10, TthCas10, SthCas10, SsoCas10, and PfuCas10) (Figs. 3A, B and Supplementary Fig. S4), and are reported to have single-stranded DNase activity consistent with the presence of HD motifs reported in Figure 1 (Figs. 1B, C and Supplementary Fig. S1).[3,4,6,11,13,15,18,19,37,53–55] To measure the efficiency of cOA synthesis, we incubated each Cas10 protein with ^{32}P-labeled rATP for 1 h at 37°C.

Products of this reaction were then resolved using TLC. Fully assembled TthCsm and SthCsm complexes were incubated with complementary RNAs and ^{32}P-labeled rATP as positive controls. Consistent with previously published work, fully assembled TthCsm and SthCsm produce cOA molecules upon incubation with a complementary RNA target (Fig. 3C).[18–21] In contrast, none of the Cas10 proteins—nor a heterodimer of TonCas10 and TonCsm4—produced detectable levels of cOA (Fig. 3C). These data support a previous report that additional Csm/Cmr components are needed to license Cas10 cyclase activity.[18]

The requirement of fully assembled Csm/Cmr complexes for cyclase activity was further supported by results from a calcein-based fluorescence assay, which showed no detectable cyclase activity for any species of ribonucleotide (i.e., rNTP) or deoxyribonucleotide (i.e., dNTP) incubated with only Cas10 proteins (Supplementary Figs. S5–S7).[21] These results agree with recent experimental findings in the type III system from *Lactococcus lactis* (i.e., LlaCsm),

showing cOA synthesis is dependent on changes in Cas10 dynamics that are triggered by the surveillance complex binding its target.[56]

Although we were unable to detect cOA synthesis for any of the five Cas10 proteins we tested, a recent report by Zhang et al (2021)[27] describes a template-dependent DNA-polymerase activity for two of the Cas10 homologs we purified (i.e., TonCas10 and PfuCas10).[27] To better understand the structural basis for Cas10-mediated DdDp activity, we superimposed structures of Cas10 from *Thermococcus onnurineus* (TonCas10) onto the Palm domains of Y-family DNA polymerases.

Palm domains are common structural elements of diverse polymerases, including reverse transcriptases, RNA-dependent RNA polymerases, A-, B-, and Y-family DNA polymerases, nucleotidyltransferases, nucleotidylcyclases, and Cas10 proteins.[26] A recent review of polymerase structures indicates that the Palm2 domain from PfuCas10 is more similar to Palm domains from Y-family DNA polymerases than to Palm2 domains of other polymerases.[25] When the Palm2 domain of TonCas10—which contains the catalytic GGDD motif—is superimposed on the single Palm domain of Dpo4 (a Y-family DNA polymerase) from *Sulfolobus solfataricus*, the TonCas10 Palm1 domain clashes with the double-stranded DNA substrate (Fig. 3D).[57]

Similar clashing is also observed when Palm2 is superimposed on the Palm domains of other Y-family DNA polymerases. These results suggest that Cas10 would

---

FIG. 3. Cas10 proteins do not have cyclase or polymerase activity in isolation. **(A)** Five Cas10 proteins with intact GGDD motifs were chosen from III-A, III-B, or III-D systems. These proteins were purified, and their cyclase and polymerase activities were tested *in vitro*. **(B)** SDS-PAGE gels of purified Cas10 proteins. Four wild-type proteins (e.g., PfuCas10) and their corresponding Palm2 (GGAA) mutants (e.g., PfuCas10$_{mut}$) are shown in the top gel. SsoCas10, for which we were unable to purify a Palm2 mutant, is shown on the bottom left, and TonCas10-Csm4 is shown on the bottom right. SEC profiles are shown in Supplementary Figure S4. **(C)** 1 μM Cas10, Cas10-Csm4, or fully assembled surveillance complex (i.e., TthCsm or SthCsm complex) were incubated with a mixture of 25 μM rATP and 0.08 μM ^{32}P-α-rATP for 1 h. Reactions with TthCsm and SthCsm complexes additionally contained RNA transcripts complementary in sequence to the respective crRNAs of each surveillance complex. Products from these reactions were resolved by TLC, and reveal an absence of cOA products for reactions with each of the five Cas10 proteins and for Cas10-Csm4. By contrast, cOA products are visible when TthCsm or SthCsm complex is incubated with a crRNA-complementary target. Unlabeled bands correspond to salt front on the TLC plate. **(D)** Left: to model Cas10 binding of a DNA primer–template substrate, a structure of TonCsm1 (PDB: 4UW2) was superimposed on a Y-family DNA-polymerase structure (Dpo4; PDB: 2AGQ) by aligning TonCas10 Palm2 to Dpo4 Palm. Conserved Cas10 cyclase and Dpo4 polymerase active-site motifs are shown in red, with acidic residues underlined. Right: steric clashes between Dpo4 DNA and Cas10 Palm1 atoms are shown as red lines. **(E)** Wild-type Cas10 proteins (e.g., PfuCas10) or Palm2 (GGAA) Cas10 mutants (e.g., PfuCas10$_{mut}$) were incubated with dNTPs and a 20nt ^{32}P-labeled primer annealed to 40nt template. Products from these reactions were resolved on a UREA-PAGE gel. Primer extension assays with SsoCas10, for which we were unable to purify a Palm2 mutant, are shown in Supplementary Figure S8. cOA, cyclic oligoadenylate; crRNA, CRISPR RNA; SDS-PAGE, sodium dodecyl-sulfate polyacrylamide gel electrophoresis; SthCsm, Csm from *Streptococcus thermophilus*; TLC, thin-layer chromatography.
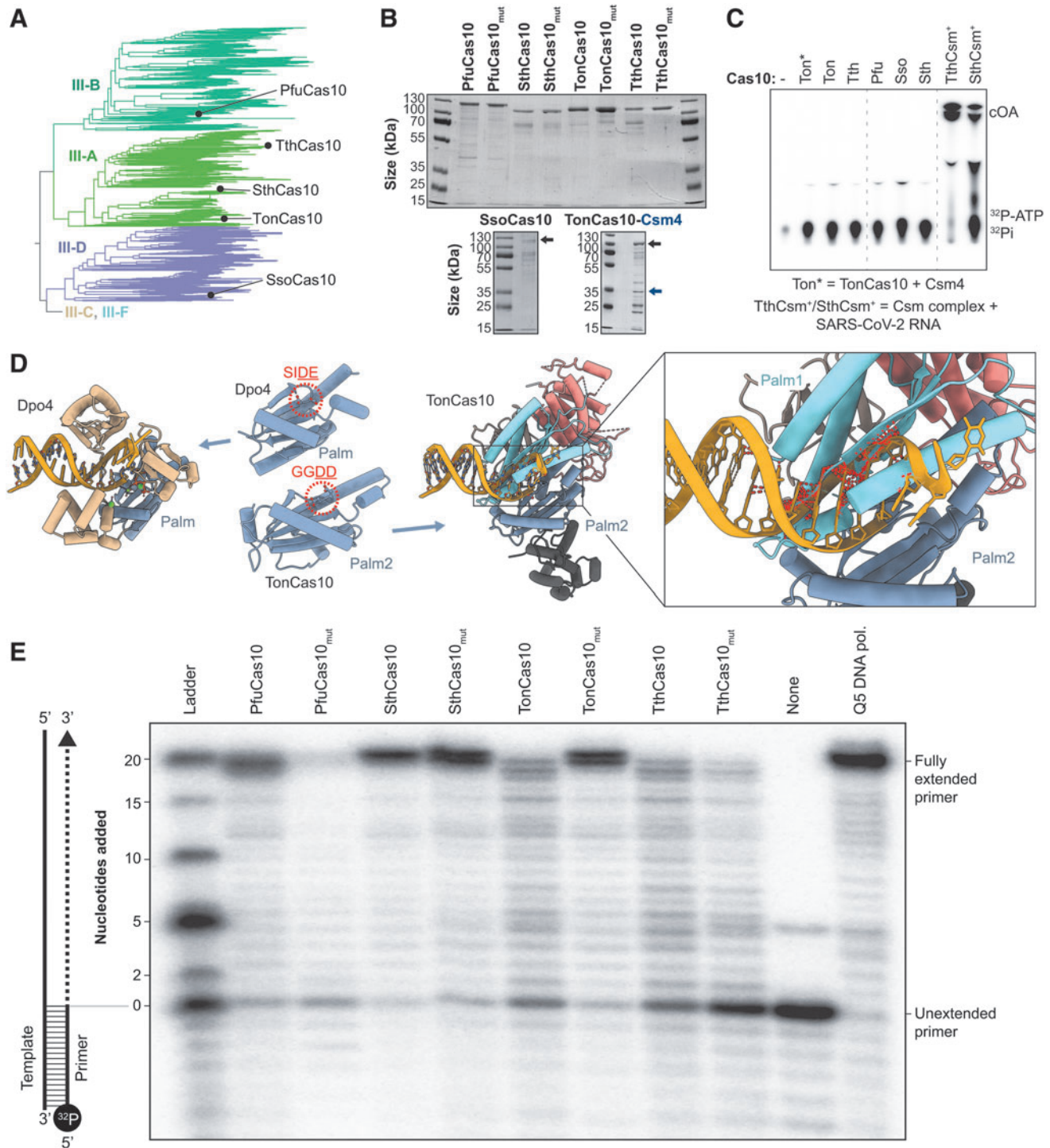
require significant conformational rearrangements to accommodate double-stranded DNA in its Palm2 active site. To determine how Cas10 might accommodate a DNA primer and DNA template, we attempted to copurify a Cas10-DNA complex for structural determination, but could not purify a stable complex.

To verify the DNA-polymerase activity of Cas10 proteins *in vitro*, we incubated each of the purified Cas10 proteins (Fig. 3A, B) with deoxyribonucleotides (dNTPs) and a 20-nucleotide ^{32}P-labeled primer annealed to a 40-nucleotide template (Fig. 3E and S8). Products corresponding to extended primers were visible after 1 h of

incubation with each of the purified proteins, consistent with previous reports of Cas10-mediated DdDp activity.[27]

A conserved motif corresponding to the Cas10 GGDD motif is required for DNA polymerization in Y-family DNA polymerases (Fig. 3D), and mutations to either of the acidic residues in the Y-family DNA-polymerase motif inhibit or eliminate DNA primer extension.[58–60] Thus, we expect that mutation of either aspartate residue in Cas10 GGDD motifs should abolish polymerase activity. However, mutations to the catalytic motif in the Cas10 Palm2 active site (i.e., GGDD to GGAA; Pfu-Cas10$_{mut}$, SthCas10$_{mut}$, TthCas10$_{mut}$, and TonCas10$_{mut}$) do not eliminate DNA-polymerase activity in Cas10 primer extension assays (Fig. 3E).

These data are consistent with previous work reporting Cas10 DNA-polymerase activity, in which mutations to aspartate residues in the GGDD motif did not abolish polymerase activity.[27] Collectively, these results suggest that the polymerase activity is from low-level contamination of DNA polymerase from *E. coli* that is not removed during Cas10 purification.[27]

Discussion

Cas10s are large subunits of Csm/Cmr complexes and, with the exception of type III-E, are the signature proteins of type III CRISPR systems. In 2011, Makarova et al. reported that these proteins contain HD and Palm domains that are characteristic of polymerases,[61,62] but the function of these domains remained obscure until 2017 when these proteins were shown to generate cOAs that function as critical signaling components of type III immune systems.[18,19] Here, we set out to evaluate the phylogenetic and functional diversity of Cas10 proteins.

Our work reveals that HD nuclease domains are absent from Cas10s in 54.2% of III-B and 84.1% of III-D CRISPR systems (Fig. 1C and Supplementary Table S1). These results suggest that most III-B and III-D Cas10s do not cleave ssDNA during CRISPR interference, and rely instead on cOA-regulated effectors (e.g., Csm6, Csx1). Accordingly, GGDD motifs—which catalyze cOA synthesis—are strongly conserved in Palm2 domains of III-A, III-B, and III-D Cas10 proteins (Fig. 1C).

Moreover, a recent analysis of cOA-binding domains (i.e., CARF and SAVED) shows that these domains are genetically associated with III-A, III-B, and III-D CRISPR systems.[51] By contrast, type III-C and III-F systems are unlikely to employ cOA-activated effectors as their Palm2 domains do not contain GGDD motifs (Fig. 1C), and compensatory motifs are also absent in Palm1. Cas10s from these subtypes form clades with a shared ancestor (Fig. 1B), suggesting that cOA signaling was lost before the emergence of III-C and III-F CRISPR systems.

In addition to the loss of HD domains and catalytic Palm2 motifs in many Cas10 proteins, we identified Cas10 sequences that are split across multiple genes and Cas10s that are fused to additional domains. The existence of partial Cas10 proteins corresponding to Palm2 domains might explain the origin of minimal CRISPR polymerases (mCpols) (Fig. 2C, D).[63] While the function of mCpols remains unclear, some mCpols are fused to cOA-sensing (i.e., CARF) domains and HEPN RNase or transmembrane domains. These fusion proteins are thought to constitute a self-contained signal-detector-effector system, and inspired the initial hypothesis that type III CRISPR systems adopted cOA-signaling networks for immune purposes.[19,28,51]

Genetic fusions are not unique to mCpols, and in the course of analyzing the domain compositions of Cas10 sequences, we identified fusions of Cas10 to a toxin (AbiEii-Cas10) and an innate immune effector (NucC-Cas10) (Fig. 2A, B). Additional work is needed to elucidate the roles played by AbiEii-Cas10 and NucC-Cas10 fusions in CRISPR immunity, but it is conceivable that these proteins are involved in triggering senescence or programmed cell death.[48,49]

Type III CRISPR systems have recently been developed as programmable diagnostics for detection of RNA.[20–24] These diagnostics utilize cOA-activated nucleases that cleave fluorescent reporters, but many of these effectors are only activated by specific species of cOA (e.g., Csm6 from *Enteroccocus italicus* is specifically activated by cA$_6$).[64] In an effort to understand how phylogeny impacts Cas10 cyclase activity, we screened five Cas10 proteins from three clades with the aim of revealing their cOA-production profiles (Fig. 3A, B).

Results from these experiments concur with a previous report that Cas10 proteins alone are incapable of synthesizing cOA (Fig. 3C).[18] These results may not be surprising since cyclase activity must be tightly regulated by the recognition of invading nucleic acids. Promiscuous cOA production would activate cOA-regulated effectors, inducing cellular dormancy or cell death in the absence of bacteriophage infection.

To better understand the mechanism of DdDp activity for Cas10, we repeated a previously reported primer extension assay (Fig. 3E and Supplementary Fig. S8).[27] Products corresponding to extended primers were visible for all five Cas10 proteins tested, consistent with a previous report of Cas10-mediated DNA-polymerase activity.[27] However, mutations (i.e., GGAA) to the aspartate residues in Palm2 GGDD motifs do not eliminate this activity (Fig. 3E).[18,19] While we cannot exclude the possibility that Cas10 proteins have DNA-polymerase activity that is catalyzed by residues outside the GGDD motif, this scenario would represent a departure from the known

mechanisms of structurally similar Y-family polymerases that rely on acidic residues at structurally similar positions (Fig. 3D).[25,58–60]

Four lines of evidence suggest that the DNA-polymerase activity observed here may not be a function of Cas10: structural superpositions of Cas10 and Y-family DNA polymerases reveal steric clashes that would prevent primer–template loading in the absence of dramatic Cas10 conformational rearrangements (Fig. 3D), attempts to purify Cas10 bound to DNA substrates failed, mutations in the Palm2 domain do not eliminate DNA-pol activity, and the well-established cyclase activity of Cas10 relies on ribonucleotides rather than deoxyribonucleotides.[18,19,27] Overall, these data suggest that low levels of a contaminating polymerase, rather than Cas10, may be responsible for the polymerase activity observed in primer extension assays.

Collectively, the work presented here helps clarify the phylogenetic and functional diversity of Cas10 proteins, illustrating that nuclease and cyclase domain conservation is confined to specific clades. Our work also suggests that additional work is needed to understand how III-C and III-F systems provide immunity without cOA production, and how exaptation of toxin or innate immune components—through genetic fusion to *cas10* (i.e., NucC-Cas10 and AbiEii-Cas10)—impacts type III defense.

Authors' Contributions

B.W., T.W., A.S.F., and R.W. designed the research. R.H. and M.L. performed metagenomic sequencing. T.W., A.S.F., and R.W. performed biochemical activity assays. T.W. conducted bioinformatic analyses. B.W., T.W., A.S.F., and R.W. wrote the article with input from all authors.

Author Disclosure Statement

B.W. is the founder of SurGene, LLC and VIRIS Detection Systems Inc. and is an inventor on patent applications related to CRISPR-Cas systems and applications thereof. A.S.-F. is a co-inventor on a patent for a CRISPR-based diagnostic.

Funding Information

Supplementary Material

Supplementary Data S1
Supplementary Table S1
Supplementary Table S2
Supplementary Figure S1
Supplementary Figure S2
Supplementary Figure S3
Supplementary Figure S4
Supplementary Figure S5
Supplementary Figure S6
Supplementary Figure S7
Supplementary Figure S8

References

1. Makarova KS, Wolf YI, Iranzo J, et al. Evolutionary classification of CRISPR-Cas systems: A burst of class 2 and derived variants. Nat Rev Microbiol 2019;18(2):67–83; doi: 10.1038/s41579-019-0299-x
2. Huo Y, Li T, Wang N, et al. Cryo-EM structure of Type III-A CRISPR effector complex. Cell Res 2018;28:1195–1197; doi: 10.1038/s41422-018-0115-6
3. Jia N, Jones R, Sukenick G, et al. Second messenger cA4 formation within the composite Csm1 palm pocket of type III-A CRISPR-Cas Csm complex and its release path. Mol Cell 2019;75:933–943 e936; doi: 10.1016/j.molcel.2019.06.013
4. Liu TY, Liu JJ, Aditham AJ, et al. Target preference of Type III-A CRISPR-Cas complexes at the transcription bubble. Nat Commun 2019;10:3001; doi: 10.1038/s41467-019-10780-2
5. Makarova KS, Wolf YI, Alkhnbashi OS, et al. An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol 2015;13:722–736; doi: 10.1038/nrmicro3569
6. Sofos N, Feng M, Stella S, et al. Structures of the Cmr-beta complex reveal the regulation of the immunity mechanism of type III-B CRISPR-Cas. Mol Cell 2020;79:741–757 e747; doi: 10.1016/j.molcel.2020.07.008
7. Ozcan A, Krajeski R, Ioannidi E, et al. Programmable RNA targeting with the single-protein CRISPR effector Cas7–Cas11. Nature 2021;597:720–725; doi: 10.1038/s41586-021-03886-5
8. van Beljouw SPB, Haagsma AC, Rodriguez-Molina A, et al. The gRAMP CRISPR-Cas effector is an RNA endonuclease complexed with a caspase-like peptidase. Science 2021;373:1349–1353; doi: 10.1126/science.abk2718
9. Manav MC, Van LB, Lin J, et al. Structural basis for inhibition of an archaeal CRISPR-Cas type I-D large subunit by an anti-CRISPR protein. Nat Commun 2020;11:5993; doi: 10.1038/s41467-020-19847-x
10. Schwartz EA, McBride TM, Bravo JPK, et al. Structural rearrangements allow nucleic acid discrimination by type I-D Cascade. Nat Commun 2022;13:2829; doi: 10.1038/s41467-022-30402-8
11. Jung TY, An Y, Park KH, et al. Crystal structure of the Csm1 subunit of the Csm complex and its single-stranded DNA-specific nuclease activity. Structure 2015;23:782–790; doi: 10.1016/j.str.2015.01.021
12. Makarova KS, Aravind L, Grishin NV, et al. A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. Nucleic Acids Res 2002;30:482–496; doi: 10.1093/nar/30.2.482
13. Elmore JR, Sheppard NF, Ramia N, et al. Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system. Genes Dev 2016;30:447–459; doi: 10.1101/gad.272153.115
14. Estrella MA, Kuo FT, Bailey S. RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex. Genes Dev 2016;30:460–470; doi: 10.1101/gad.273722.115
15. Kazlauskiene M, Tamulaitis G, Kostiuk G, et al. Spatiotemporal control of type III-A CRISPR-Cas immunity: Coupling DNA degradation with the target RNA recognition. Mol Cell 2016;62:295–306; doi: 10.1016/j.molcel.2016.03.024
16. Mo CY, Mathai J, Rostol JT, et al. Type III-A CRISPR immunity promotes mutagenesis of Staphylococci. Nature 2021;592:611–615; doi: 10.1038/s41586-021-03440-3
17. Hatoum-Aslan A, Maniv I, Samai P, et al. Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system. J Bacteriol 2014;196:310–317; doi: 10.1128/JB.01130–13
18. Kazlauskiene M, Kostiuk G, Venclovas C, et al. A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. Science 2017;357:605–609; doi: 10.1126/science.aao0100
19. Niewoehner O, Garcia-Doval C, Rostol JT, et al. Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. Nature 2017;548:543–548; doi: 10.1038/nature23467

20. Nemudraia A, Nemudryi A, Buyukyoruk M, et al. Sequence-specific capture and concentration of viral RNA by type III CRISPR system enhances diagnostic. Nat Commun 2022;13:7762; doi: 10.1038/s41467-022-35445-5

21. Santiago-Frangos A, Hall LN, Nemudraia A, et al. Intrinsic signal amplification by type III CRISPR-Cas systems provides a sequence-specific SARS-CoV-2 diagnostic. Cell Rep Med 2021;2:100319; doi: 10.1016/j.xcrm.2021.100319

22. Steens JA, Zhu Y, Taylor DW, et al. SCOPE enables type III CRISPR-Cas diagnostics using flexible targeting and stringent CARF ribonuclease activation. Nat Commun 2021;12:5033; doi: 10.1038/s41467-021-25337-5

23. Gruschow S, Adamson CS, White MF. Specificity and sensitivity of an RNA targeting type III CRISPR complex coupled with a NucC endonuclease effector. Nucleic Acids Res 2021;49:13122–13134; doi: 10.1093/nar/gkab1190

24. Santiago-Frangos A, Nemudryi A, Nemudraia A, et al. CRISPR-Cas, Argonaute proteins and the emerging landscape of amplification-free diagnostics. Methods 2022;205:1–10; doi: 10.1016/j.ymeth.2022.06.002

25. Monttinen HA, Ravantti JJ, Poranen MM. Common structural core of three-dozen residues reveals intersuperfamily relationships. Mol Biol Evol 2016;33:1697–1710; doi: 10.1093/molbev/msw047

26. Yang W. An overview of Y-Family DNA polymerases and a case study of human DNA polymerase eta. Biochemistry 2014;53:2793–2803; doi: 10.1021/bi500019s

27. Zhang S, Li T, Huo Y, et al. Mycobacterium tuberculosis CRISPR/Cas system Csm1 holds clues to the evolutionary relationship between DNA polymerase and cyclase activity. Int J Biol Macromol 2021;170:140–149; doi: 10.1016/j.ijbiomac.2020.12.014

28. Kolesnik MV, Fedorova I, Karneyeva KA, et al. Type III CRISPR-Cas systems: Deciphering the most complex prokaryotic immune system. Biochemistry (Mosc) 2021;86:1301–1314; doi: 10.1134/S0006297921100114

29. Hyatt D, Chen GL, Locascio PF, et al. Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010;11:119; doi: 10.1186/1471-2105-11-119

30. Kohtz AJ, Jay ZJ, Lynes M, et al. Culexarchaeia, a novel archaeal class of anaerobic generalists inhabiting geothermal environments. bioRxiv 2022:2022.2004.2006.487207; doi: 10.1101/2022.04.06.487207

31. Lynes MM, Krukenberg V, Jay ZJ, et al. Diversity and function of methyl-coenzyme M reductase-encoding archaea in Yellowstone hot springs revealed by metagenomics and mesocosm experiments. bioRxiv 2022:2022.2008.2018.504445; doi: 10.1101/2022.08.18.504445

32. Couvin D, Bernheim A, Toffano-Nioche C, et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res 2018;46:W246–W251; doi: 10.1093/nar/gky425

33. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol 2011;7:e1002195; doi: 10.1371/journal.pcbi.1002195

34. Huang Y, Niu B, Gao Y, et al. CD-HIT Suite: A web server for clustering and comparing biological sequences. Bioinformatics 2010;26:680–682; doi: 10.1093/bioinformatics/btq003

35. Katoh K, Kuma K, Toh H, et al. MAFFT version 5: Improvement in accuracy of multiple sequence alignment. Nucleic Acids Res 2005;33:511–518; doi: 10.1093/nar/gki198

36. You L, Ma J, Wang J, et al. Structure studies of the CRISPR-Csm complex reveal mechanism of co-transcriptional interference. Cell 2019;176:239–253 e216 doi: 10.1016/j.cell.2018.10.052

37. Osawa T, Inanaga H, Numata T. Crystal structure of the Cmr2-Cmr3 subcomplex in the CRISPR-Cas RNA silencing effector complex. J Mol Biol 2013;425:3811–3823; doi: 10.1016/j.jmb.2013.03.042

38. Price MN, Dehal PS, Arkin AP. FastTree 2—Approximately maximum-likelihood trees for large alignments. PLoS One 2010;5:e9490; doi: 10.1371/journal.pone.0009490

39. Revell LJ. Phytools: An R package for phylogenetic comparative biology (and other things). Methods Ecol Evol 2012;3:217–223; doi: 10.1111/j.2041-210X.2011.00169.x

40. Yu G, Lam TT, Zhu H, et al. Two methods for mapping and visualizing associated data on phylogeny using ggtree. Mol Biol Evol 2018;35:3041–3043; doi: 10.1093/molbev/msy194

41. Mistry J, Chuguransky S, Williams L, et al. Pfam: The protein families database in 2021. Nucleic Acids Res 2021;49:D412-D419; doi: 10.1093/nar/gkaa913

42. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596:583–589; doi: 10.1038/s41586-021-03819-2

43. Pagès H AP, Gentleman R, DebRoy S. Biostrings: Efficient manipulation of biological strings. R package version 2520 2019; doi: 10.18129/B9.bioc.Biostrings

44. Crooks GE, Hon G, Chandonia JM, et al. WebLogo: A sequence logo generator. Genome Res 2004;14:1188–1190; doi: 10.1101/gr.849004

45. Goddard TD, Huang CC, Meng EC, et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci 2018;27:14–25; doi: 10.1002/pro.3235

46. Rouillon C, Athukoralage JS, Graham S, et al. Control of cyclic oligoadenylate synthesis in a type III CRISPR system. Elife 2018;7;e36734; doi: 10.7554/eLife.36734

47. Koonin EV, Makarova KS. Evolutionary plasticity and functional versatility of CRISPR systems. PLoS Biol 2022;20:e3001481; doi: 10.1371/journal.pbio.3001481

48. Lau RK, Ye Q, Birkholz EA, et al. Structure and mechanism of a cyclic trinucleotide-activated bacterial endonuclease mediating bacteriophage immunity. Mol Cell 2020;77:723–733 e726; doi: 10.1016/j.molcel.2019.12.010

49. Dy RL, Przybilski R, Semeijn K, et al. A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism. Nucleic Acids Res 2014;42:4590–4605; doi: 10.1093/nar/gkt1419

50. Hampton HG, Smith LM, Ferguson S, et al. Functional genomics reveals the toxin-antitoxin repertoire and AbiE activity in Serratia. Microb Genom 2020;6:mgen000458; doi: 10.1099/mgen.0.000458

51. Makarova KS, Timinskas A, Wolf YI, et al. Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antivirus defense. Nucleic Acids Res 2020;48:8828–8847; doi: 10.1093/nar/gkaa635

52. Cai Y, Usher B, Gutierrez C, et al. A nucleotidyltransferase toxin inhibits growth of Mycobacterium tuberculosis through inactivation of tRNA acceptor stems. Sci Adv 2020;6:eabb6651; doi: 10.1126/sciadv.abb6651

53. Cocozaki AI, Ramia NF, Shao Y, et al. Structure of the Cmr2 subunit of the CRISPR-Cas RNA silencing complex. Structure 2012;20:545–553; doi: 10.1016/j.str.2012.01.018

54. Liu TY, Iavarone AT, Doudna JA. RNA and DNA targeting by a reconstituted thermus thermophilus type III-A CRISPR-Cas system. PLoS One 2017;12:e0170552; doi: 10.1371/journal.pone.0170552

55. Zhang J, Graham S, Tello A, et al. Multiple nucleic acid cleavage modes in divergent type III CRISPR systems. Nucleic Acids Res 2016;44:1789–1799; doi: 10.1093/nar/gkw020

56. Sridhara S, Rai J, Whyms C, et al. Structural and biochemical characterization of in vivo assembled Lactococcus lactis CRISPR-Csm complex. Commun Biol 2022;5:279; doi: 10.1038/s42003-022-03187-1

57. Vaisman A, Ling H, Woodgate R, et al. Fidelity of Dpo4: Effect of metal ions, nucleotide selection and pyrophosphorolysis. EMBO J 2005;24:2957–2967; doi: 10.1038/sj.emboj.7600786

58. Kondratick CM, Washington MT, Prakash S, et al. Acidic residues critical for the activity and biological function of yeast DNA polymerase eta. Mol Cell Biol 2001;21:2018–2025; doi: 10.1128/MCB.21.6.2018-2025.2001

59. Wong JH, Fiala KA, Suo Z, et al. Snapshots of a Y-family DNA polymerase in replication: Substrate-induced conformational transitions and implications for fidelity of Dpo4. J Mol Biol 2008;379:317–330; doi: 10.1016/j.jmb.2008.03.038

60. Ling H, Boudsocq F, Woodgate R, et al. Crystal structure of a Y-family DNA polymerase in action: A mechanism for error-prone and lesion-bypass replication. Cell 2001;107:91–102; doi: 10.1016/s0092-8674(01)00515-3

61. Makarova KS, Aravind L, Wolf YI, et al. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. Biol Direct 2011;6:38.

62. Makarova KS, Haft DH, Barrangou R, et al. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol 2011;9:467–477; doi: 10.1038/nrmicro2577

63. Burroughs AM, Zhang D, Schaffer DE, et al. Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. Nucleic Acids Res 2015;43:10633–10654; doi: 10.1093/nar/gkv1267

64. Garcia-Doval C, Schwede F, Berk C, et al. Activation and self-inactivation mechanisms of the cyclic oligoadenylate-dependent CRISPR ribonuclease Csm6. Nat Commun 2020;11:1596; doi: 10.1038/s41467-020-15334-5