RNA Biology 10:5, 828–840; May 2013; © 2013 Landes Bioscience

Programmable plasmid interference by the CRISPR-Cas system in *Thermococcus kodakarensis*

Joshua R. Elmore,[1,†] Yuusuke Yokooji,[4,†] Takaaki Sato,[4] Sara Olson,[5] Claiborne V.C. Glover III,[1] Brenton R. Graveley,[5] Haruyuki Atomi,[4] Rebecca M. Terns[1,*] and Michael P. Terns[1,2,3,*]

[1]Department of Molecular Biology and Biochemistry; University of Georgia; Athens, GA USA; [2]Department of Genetics; University of Georgia; Athens, GA USA; [3]Department of Microbiology; University of Georgia; Athens, GA USA; [4]Department of Synthetic Chemistry and Biological Chemistry; Graduate School of Engineering; Kyoto University; Katsura, Nishikyo-ku, Kyoto, Japan; [5]Department of Genetics and Developmental Biology; University of Connecticut Institute for Systems Genomics; University of Connecticut Stem Cell Institute; University of Connecticut Health Center; Farmington, CT USA

[†]These authors contributed equally to this work.

Keywords: CRISPR, Cas, archaea, Thermococcus, hyperthermophile, immune, RNA, DNA, silencing, interference

CRISPR-Cas systems are RNA-guided immune systems that protect prokaryotes against viruses and other invaders. The CRISPR locus encodes crRNAs that recognize invading nucleic acid sequences and trigger silencing by the associated Cas proteins. There are multiple CRISPR-Cas systems with distinct compositions and mechanistic processes. *Thermococcus kodakarensis (Tko)* is a hyperthermophilic euryarchaeon that has both a Type I-A Csa and a Type I-B Cst CRISPR-Cas system. We have analyzed the expression and composition of crRNAs from the three CRISPRs in *Tko* by RNA deep sequencing and northern analysis. Our results indicate that crRNAs associated with these two CRISPR-Cas systems include an 8-nucleotide conserved sequence tag at the 5' end. We challenged *Tko* with plasmid invaders containing sequences targeted by endogenous crRNAs and observed active CRISPR-Cas-mediated silencing. Plasmid silencing was dependent on complementarity with a crRNA as well as on a sequence element found immediately adjacent to the crRNA recognition site in the target termed the PAM (protospacer adjacent motif). Silencing occurred independently of the orientation of the target sequence in the plasmid, and appears to occur at the DNA level, presumably via DNA degradation. In addition, we have directed silencing of an invader plasmid by genetically engineering the chromosomal CRISPR locus to express customized crRNAs directed against the plasmid. Our results support CRISPR engineering as a feasible approach to develop prokaryotic strains that are resistant to infection for use in industry.

Introduction

CRISPR-Cas immune systems are recently discovered, RNA-guided adaptive immune systems that provide sequence-specific protection to prokaryotic hosts from invasive genetic elements such as phages and plasmids.[1-8] CRISPRs (clustered regularly interspaced short palindromic repeats) are genetic loci found in prokaryotic genomes containing arrays of short direct repeat sequences (~30–40 base-pairs) interspaced by similarly-sized "spacer" sequences that are acquired from invader DNA (where the same sequence is known as a "proto-spacer").[9-15] Transcription of CRISPR loci and subsequent processing of primary CRISPR transcripts produce mature CRISPR (cr)RNAs that each contain an invader-derived (guide) sequence flanked by various amounts of the CRISPR repeat sequence at the 5' and/or 3' end.[16-23] Individual crRNAs interact with CRISPR-associated (Cas) proteins to form ribonucleoprotein effector complexes that can identify (via base-pairing) and silence the DNA or RNA of the invader.[16,22,24-30]

Multiple CRISPR-Cas systems have been identified and each appears to counteract viral and plasmid invaders using distinct sets of Cas proteins, crRNA species and molecular mechanisms.[2,4,7,8,31,32] At least 10 unique sets of Cas protein genes are found among prokaryotic genomes, forming distinct CRISPR-Cas systems.[2,31] The members of each group of co-segregating Cas protein genes are named for a prototypical organism where the set of genes is found, such as the Cas subtype *E. coli* or Cse protein genes.[31] Cas1 and Cas2 are the only proteins common to all CRISPR-Cas systems. Recently, CRISPR-Cas systems have been categorized into three general types (I, II and III) and some of the Cas proteins have been organized into superfamilies.[2] The three broad types are further classified into 10 subtypes (I-A to I-F, II-A and II-B and III-A and III-B) that generally correspond with the previously defined subtypes (e.g., the Type I-E/Cse system). Most of the CRISPR-Cas systems that have been characterized silence through the destruction of the DNA of the invader.[10,26,27,33-37] The Type III-B Cmr system cleaves invading target RNAs.[22,25,38]

*Correspondence to: Rebecca M. Terns, Email: rterns@bmb.uga.edu; Michael P. Terns, Email: mterns@bmb.uga.edu
Submitted: 12/14/12; Revised: 02/20/13; Accepted: 02/22/13
http://dx.doi.org/10.4161/rna.24084

crRNAs guide the silencing of invaders by CRISPR-Cas systems.[16,22,23,25,29,33,36,37,39] CRISPRs are transcribed from a leader sequence of ~100–500 base-pairs found directly upstream of the CRISPR.[40] Putative promoters have been identified within the leader sequence in several species.[22,41-45] Primary CRISPR transcripts are cleaved within each repeat sequence to generate unit crRNAs containing an invader-derived guide sequence and a portion of the repeat sequence at both the 5' and 3' end. In most Type I and III systems, this processing is performed by Cas6 or Cas6-like endoribonucleases, which cleave within the repeat leaving 8 nts of the repeat upstream of the guide sequence and the remainder of the repeat on the 3' end.[16,17,19,21,46] This species is used directly by Type I-E Cse, I-F Csy and some I-A Csa systems,[16,29,39] but is further processed at the 3' end in other systems.[20,22,24,38,41,43,44,47] In Type I-C Csd systems, Cas5d cleaves CRISPR transcripts 11 nts upstream of the guide sequence.[21,48] In Type II Csn systems, the primary transcript is processed by RNase III, in concert with a non-coding RNA with partial complementarity to the repeat sequence (tra-crRNA).[23,26] Further processing of the crRNA products at the 5' end removes the repeat and some of the guide sequence. The mature Type II Csn system crRNAs retain a 3' repeat sequence tag.[23,26,27] Production of these crRNAs also requires Csn1 (Cas9 superfamily).[23] The introduction of new crRNA guide sequences (e.g., by reconstitution of complexes, expression from plasmids, or exposure to invaders) can direct various CRISPR-Cas systems to silence new targets.[11-13,16,22,37,49,50] Programmable CRISPR-Cas-mediated resistance to invaders has the potential to protect valuable industrial strains from common invaders.

The CRISPR loci themselves contain target sequences for the crRNAs, necessitating a mechanism to protect the genome from CRISPR-Cas systems that target DNA. Base-pairing between the repeat-derived sequence at the 5'-end of the crRNAs and the repeat sequence in the CRISPR array protects the host genome from the Type III-A Csm system of *Staphylococcus epidermidis*.[51] Invaders, which lack the repeat sequence adjacent to the target sequence recognized by the crRNA, are silenced. Alternatively, a short (2–5 nt) sequence called a protospacer adjacent motif or PAM that is found immediately downstream or upstream of the target sequence marks invader DNA for destruction by some Type I and Type II systems, respectively.[11,12,34-36,50,52] Single base changes within the PAM prevent silencing of a potential target DNA with an otherwise perfect match to the crRNA guide sequence.[11,34-36,50]

In this study, we characterize the previously unexplored CRISPR-Cas systems of *Thermococcus kodakarensis* (*Tko*), a hyperthermophilic euryarchaeon (optimal growth temperature ~85°C) that was initially isolated from a shallow marine solfatara off the coast of Kodakara Island, Japan.[53] We have delineated the features of *Tko* crRNAs by both RNA-Seq and northern analysis. In addition, we show effective silencing of plasmid invaders containing sequences complementary to *Tko* crRNAs, presumably via DNA degradation. Finally, we demonstrate that it is possible to engineer a CRISPR locus to induce resistance to a previously untargeted invader.

Results

Encoded CRISPR and Cas components. The genome of *T. kodakarensis* contains three CRISPR loci encoding 74 potential crRNAs (**Fig. 1A**). CRISPRs 1, 2 and 3 share a common nt repeat sequence interspaced by 15, 23 and 36 unique spacers, respectively. The majority of spacer sequences are 37 or 38 nt in length but spacers range in size from 34–47 nt. None of the spacer sequences map to currently known Thermococcales viruses or plasmids.[54,55] The sequence of the Tko CRISPR repeat (**Fig. 1A**) is nearly identical to that of Pyrococcus furiosus. Both repeat sequences are members of the CRISPR repeat type 6 family.[56] The P. furiosus repeat RNA is unstructured in solution and specifically recognized and cleaved by the Cas6 endonuclease[17,18,54] (The equivalent Cas6 recognition and cleavage sites are indicated in **Fig. 1A**). Adjacent to each CRISPR array is a ~390 nt leader region, which has > 89% sequence identity among the three CRISPR loci.

The *Tko* genome also contains a single *cas* gene cluster encoding 20 proteins (**Fig. 1A and B**). These include core Cas proteins implicated in new spacer acquisition (Cas1, Cas2 and two Cas4 proteins) as well as crRNA biogenesis (two Cas6 proteins). Moreover, this gene cluster encodes proteins of two predicted effector immune complexes: Type I-A Csa (Cas subtype Apern; found in archaea) and Type I-B Cst (Cas subtype Tneap; found in a subset of both archaea and bacteria).[2,31] Between the Type I-A and Type I-B *cas* genes, are three non-*cas* genes that appear to encode a toxin-antitoxin system (ORFs TK0456 and TK0457) and an ATPase (ORF TK0459).

CRISPR RNA expression. To characterize the profiles of crRNA expression from the three CRISPRs in *Tko*, we deep-sequenced small RNAs isolated from the wild-type strain (KOD1) using the Illumina sequencing platform. Of the 2,415,599 uniquely mapped reads, 1,131,578 (46%) mapped to the three CRISPR loci (**Fig. 2A**). All three CRISPR loci are constitutively expressed and are produced in the sense orientation relative to the leader of each CRISPR. We identified canonical BRE/TATA core promoter elements in each leader (**Fig. 2B**).[58] The 5' ends of leader region transcripts from all three loci are located 23 nt downstream of the putative promoter, suggesting that the promoters identified in the leader regions are guiding transcription of the CRISPR arrays by RNA polymerase.

Based on the number of sequence reads that mapped to each guide sequence, crRNA sequences encoded near the leader of locus 1 and 2 are more abundant than those encoded in downstream regions (**Fig. 2A**). A comparable gradient of declining crRNA expression levels from the leader to the trailer end of CRISPR loci has been observed in other organisms.[24,25,43,44,59] Interestingly, a significant 5' to 3' gradient of expression is not observed for locus 3 (**Fig. 2A**).

The sequencing data also reveal that a series of size forms exist for each encoded crRNA. To investigate crRNA composition, we analyzed the 5' and 3' ends of the sequenced RNAs (**Fig. 3**). More than 65% of sequenced crRNAs contain 8 nts of CRISPR repeat-derived sequence at the 5' end (5'-AUUGAAAB-3'). An identical 5' tag sequence is found on the 5' end of mature crRNAs

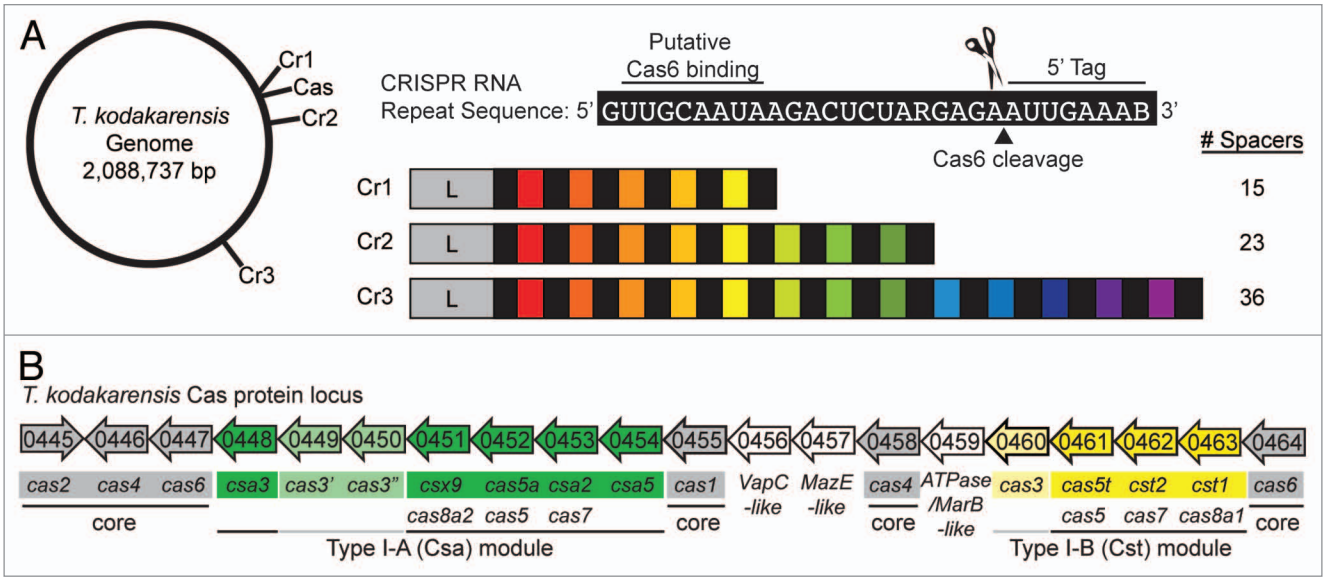

Figure 1. The CRISPR-Cas system of Thermococcus kodakarensis. (**A**) Relative locations of the CRISPRs (Cr1, Cr2, Cr3) and *cas* gene cluster (Cas) are indicated on the *Tko* genome. The consensus crRNA repeat sequence found in the three CRISPR loci is shown, and the proposed Cas6 binding and cleavage sites as well as the 8-nt 5' tag sequence (5'-AUUGAAAB-3'; where B = C, G, T) found on the mature crRNAs are highlighted. The relative size and spacer counts for each of the CRISPR loci are indicated and CRISPRs are arranged with leader regions (L) preceding repeat (black) and guide/spacer (colored) elements. (**B**) The genome organization and annotations of the predicted *cas* genes were adapted from the NCBI database (http://www.ncbi.nlm.nih.gov/). Type I-A Csa, Type I-B Cst and core *cas* genes are indicated by green, yellow and gray, respectively with recent *cas* gene superfamily designations[2] indicated below relevant *csa* and *cst* genes.[31] Three non-*cas* genes are found within the Cas protein locus. TK0459 is annotated as an archaeal ATPase with a MarB-like DNA-binding domain. TK0456 and TK0457 appear to be a potential toxin/anti-toxin pair, annotated as a VapC-like (toxin) protein and a MazE-like (anti-toxin) protein.

in P. furiosus as a result of site-specific endonucleolytic Cas6 cleavage of CRISPR transcripts at this site;[17,18,22,24] it is likely that the 5' ends of *Tko* crRNAs are also defined by Cas6 cleavage. The 3' ends of the *Tko* crRNAs are heterogenous; however, a pattern emerges from the analysis across the three loci. A ladder of species of diminishing abundance with ~36, ~43, ~50 (and perhaps ~55 and ~60) nts downstream of the 8-nt 5' repeat tag is observed for crRNAs from all three loci (**Fig. 3**). The two most prominent crRNA species are diagrammed in **Figure 3**. The most abundant crRNA species typically includes nearly all of the guide sequence (35–37 nts of a typical 38-nt guide) and no 3' repeat sequence (**Fig. 3**). The second most abundant species includes ~5 nts of repeat sequence at the 3' end. The potential cluster of 3' ends ~60 nts downstream of the 5' tag corresponds with the expected initial product of Cas6 cleavage (i.e., the 1X crRNA[17] with 22 nts of repeat downstream of a typical 38-nt guide sequence). The lengths of the crRNA species appear to be generally independent of the length of the encoded guide sequence. When analyzed independently, crRNAs with longer than average encoded guide regions (40 nts or greater) are also found primarily as species with ~36 and ~42 nts downstream of the 5' repeat tag (analysis not shown).

Tko crRNAs were also analyzed by northern analysis. Total *Tko* RNA was analyzed with DNA probes complementary to guide regions of representative crRNAs from each of the three CRISPR loci (**Fig. 4A–C**; crRNA 1.06 refers to the sixth crRNA encoded in CRISPR locus 1, etc.) as well as the repeat sequence, common to all three CRISPRs (**Fig. 4D**). Consistent with the sequencing results (**Figs. 2 and 3**), multiple discrete size forms are detected for each of the individual crRNA species by northern analysis (**Fig. 4**). No significant evidence of antisense RNAs was detected in either the sequencing data (**Fig. 2**) or northern analysis (**Fig. 4A–C**, see sense probe panels, lanes 2 and 6). A northern probe against the CRISPR repeat detected two prominent species of RNAs of ~70 and ~140 nts (**Fig. 4D**) that likely correspond to intermediate products in the processing of primary CRISPR transcripts by Cas6: the 1X and 2X RNAs respectively, comprised of one and two crRNA units (see refs. 17 and 18).

Taken together, our data show that Tko crRNAs from all three CRISPR loci are present in a series of size forms with a common 8-nt 5' CRISPR repeat tag.

The CRISPR-Cas systems of *T. kodakarensis* target DNA in vivo. To test the function of the CRISPR-Cas system in *Tko*, we challenged cells with plasmids containing target sequences complementary to endogenous crRNAs from both CRISPR loci 1 and 2 (**Fig. 5**). Plasmids with inserted target sequences (1.01, 1.02, 1.06, 2.01 and 2.05) and the parental plasmid (no target) were transformed into *Tko* KUW1 (ΔpyrF ΔtrpE) and transformants were selected by tryptophan prototrophy conferred by the trpE cassette on the plasmid (**Fig. 5A**).[60,61] Transformation with the parental plasmid resulted in the formation of 6.6×10^2 colonies per μg of DNA on average (**Fig. 5B**). In contrast, transformation with plasmids bearing the crRNA targets produced few or no colonies. The results indicate that crRNAs from CRISPR 1 and 2 can defend *Tko* against plasmids with complementary sequences.

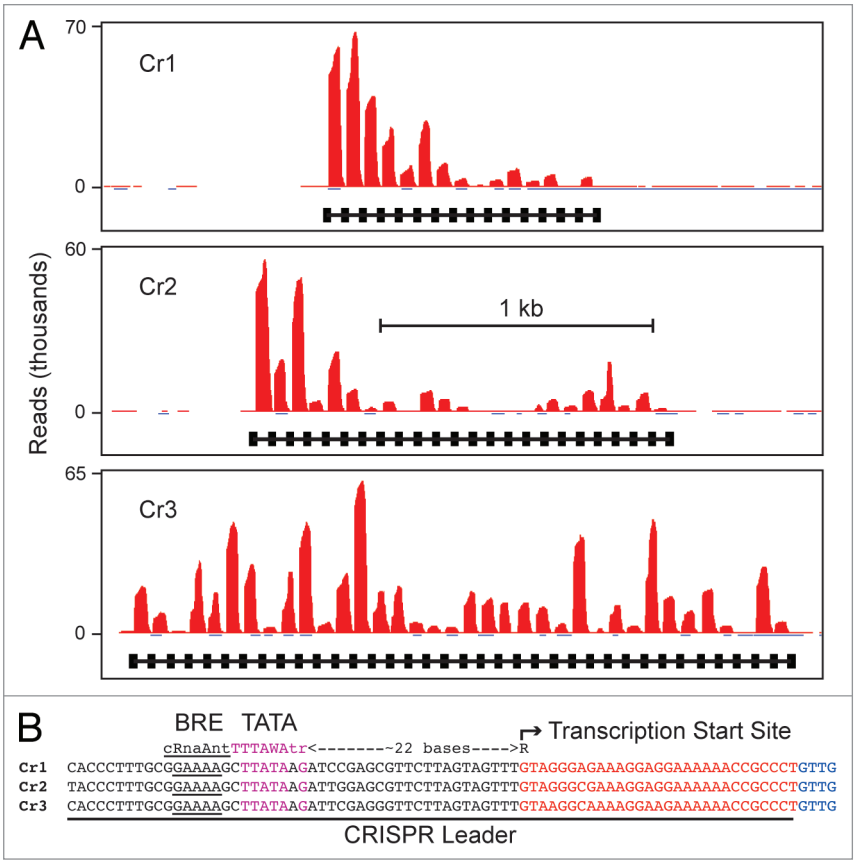

Figure 2. CRISPR RNAs are produced by all three CRISPRs in T. kodakarensis. (**A**) Small RNA sequencing reads mapped to the three CRISPR loci. CRISPRs are oriented with the leader on the left and the Y-axis represents thousands of reads. An X-axis scale of 1 kb is indicated and CRISPR repeats (black boxes) are indicated beneath the reads for each CRISPR locus. The RNAs transcribed from the leader region (red) are plotted above the X-axis while much rarer RNAs generated from the opposite strand (blue) are plotted below the X-axis. Images were generated using the UCSC archaeal genome browser.[83] (**B**) Predicted promoter elements found in the 3' end of the leader sequences. The sequences are aligned relative to the first leader-proximal repeat in each of the CRISPRs. Putative BRE and TATA box elements are indicated. Transcribed leader sequences (red) are indicated upstream of the first repeat (blue). Consensus BRE/TATA elements and predicted transcription start sites are shown above the sequence alignment.

Plasmid silencing by *Tko* CRISPR-Cas systems appears to occur at the DNA (rather than RNA) level. First, the region of the plasmid where the crRNA targets were inserted is not expected to be transcribed or to encode an essential RNA transcript. Second, and importantly, the target sequence was effectively silenced in either orientation (**Fig. 4B**, lanes 2.01 and 2.01R). Taken together, our results show that Tko contains a CRISPR-Cas system that is capable of crRNA homology-dependent DNA silencing in vivo.

CRISPR-Cas silencing depends on sequences adjacent to the protospacer. The plasmids used in the invader challenge experiments (**Fig. 5**) contain a potential PAM sequence adjacent to the crRNA target sequence. In other Type I (and Type II) CRISPR-Cas systems, silencing depends on the presence of a protospacer adjacent motif or PAM near the target sequence in the invader, and mutation of the PAM interferes with silencing.[11,12,34-36,50,52,62-64] This mechanism protects the CRISPR locus in the host genome, which lacks the PAM, from crRNA-guided destruction. In *Tko*, the bioinformatically predicted PAM sequence (on the DNA strand that base pairs with the crRNA or the target strand) is 5'-NGG-3'.[7,36,52] This PAM sequence is associated with both of the CRISPR-Cas systems (Types I-A and I-B) and with the CRISPR repeat sequence family found in Tko.[7,36,52]

To assess the requirement for a PAM in invaders targeted by the CRISPR-Cas systems in *Tko*, we altered the region immediately downstream of the 1.02 and 2.01 crRNA target sequences, from the predicted PAM (5'-CGG-3'), to a non-PAM sequence (5'-CCA-3') or a sequence predicted to interact with the 5' tag of the crRNAs (5'-GTTTCAAT-3') (**Fig. 6**). The target-containing plasmids and parental control (no target) were transformed into the *Tko* strain KUW1. Plasmids containing a predicted PAM and either 1.02 or 2.01 crRNA targets were subject to efficient CRISPR-mediated defense (**Fig. 6E**). In contrast, modification of the putative PAM to either the non-PAM or 5'-tag complement sequence (for both 1.02 and 2.01 crRNA targets) resulted in colony formation similar to that of the non-target bearing plasmid (**Fig. 6E**). These results suggest that DNA silencing in Tko is PAM-dependent.

Engineering CRISPR-Cas systems. Finally, we sought to engineer a *Tko* CRISPR locus to produce crRNAs that would effectively silence new targets. Using *Tko* KC2 as the host strain, we replaced the first five spacers of CRISPR locus 1 with five spacers (termed A-E) designed to target the *chiAΔh* gene, which encodes the C-terminal endochitinase domain of *Tko* chitinase[65] (**Fig. 7A and B**). Strain KC2 is derived from *Tko* KUW1 (Δ*pyrF* Δ*trpE*), and lacks the entire *Tko* chitinase gene, *chiA*. Northern analysis indicates that anti-chiA crRNAs A and E and downstream crRNA 1.06 are expressed in the new aChiA strain (Δ*pyrF* Δ*trpE* Δ*chiA* CRISPR1::engineered CRISPR1), and that crRNA 1.01 is not (**Fig. 7C**).

To test for silencing directed by the newly engineered crRNAs, we used the plasmid challenge assay described above. *Tko* strains KUW1 (wild-type CRISPR) and aChiA (engineered CRISPR) were challenged with plasmids containing the *chiAΔ4* gene (with targets for novel crRNAs A-E), the crRNA A or E target sequence alone, or the crRNA 1.01, 1.02 or 1.06 target sequence (i.e., plasmids used in **Fig. 5**). In the strain with the wild-type CRISPR (KUW1), plasmids containing targets for the endogenous crRNAs 1.01, 1.02 and 1.06 were silenced and failed to produce colonies (**Fig. 7D**). Plasmids with targets for the novel crRNAs produced numbers of colonies (per μg DNA) similar to the parent plasmid in KUW1 (**Fig. 7D**). In the aChiA strain, plasmids containing targets for crRNAs 1.01 and 1.02 (replaced in this strain) produced numbers of colonies (per μg DNA) similar to

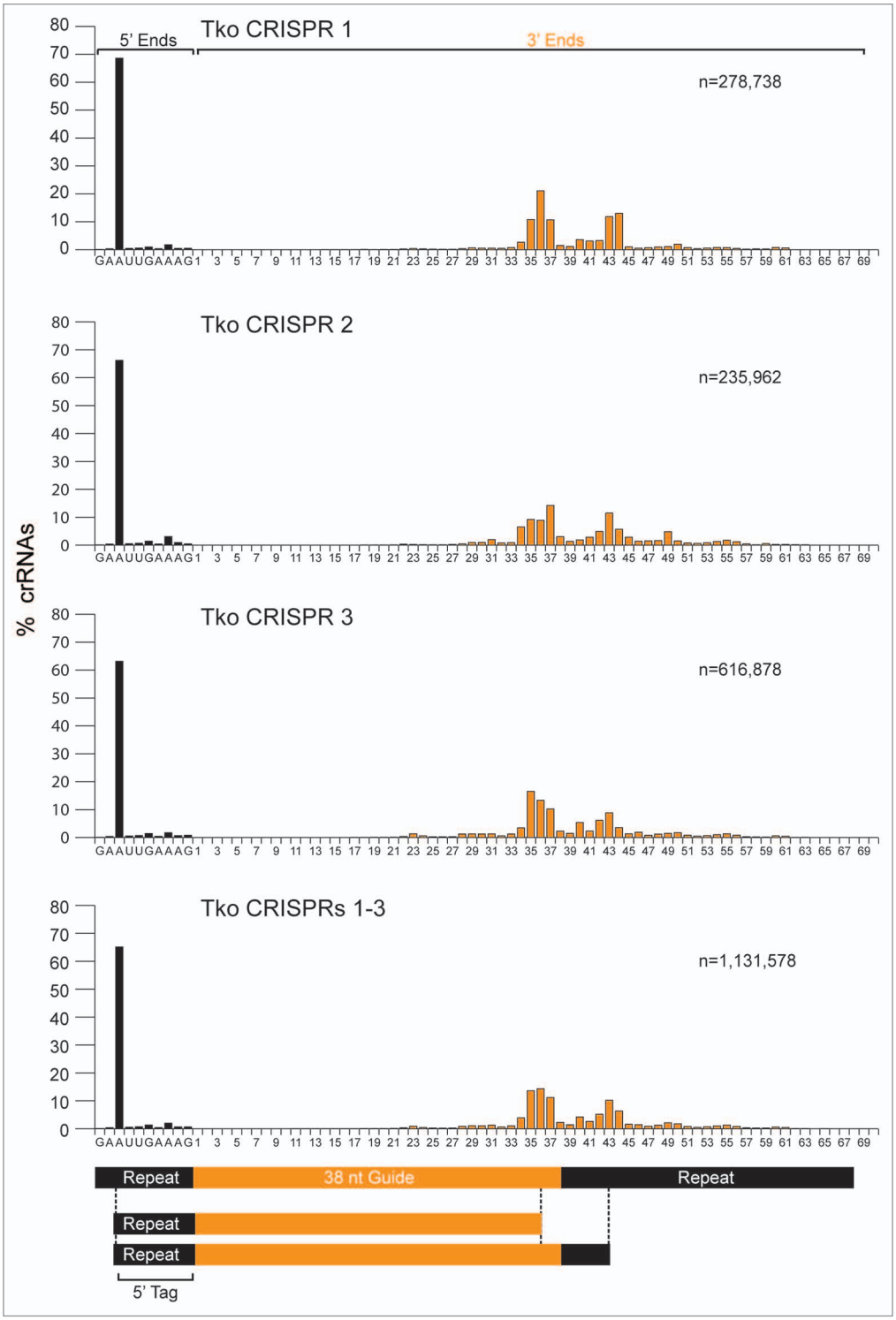

Figure 3. End analysis of sequenced crRNAs. The Y-axis represents the percentage of all crRNAs sequenced. The 5' and 3' ends of all crRNAs are mapped relative to the upstream repeat/guide junction for each of the three CRISPR loci as well as for all three. The 5' ends (black bars) and 3' ends (orange bars) of all crRNAs sequenced are indicated. Beneath the plot, the corresponding positions of a CRISPR repeat (black) and guide sequence (orange) are indicated. The model illustrates the most common *Tko* guide sequence length (38 nts). The two most common crRNA species (with 36 and 43 nts downstream of the 5' repeat tag) are diagrammed below.

the parent plasmid (**Fig. 7E**). Plasmids containing targets for the novel crRNAs (plasmids A, E and *chiAΔ4*) and the downstream crRNA 1.06 failed to produce colonies upon transformation in the aChiA strain (**Fig. 7E**). The results demonstrate successful engineering of a CRISPR locus in *Tko* to obtain specific homology-dependent silencing of an intended target plasmid in vivo.

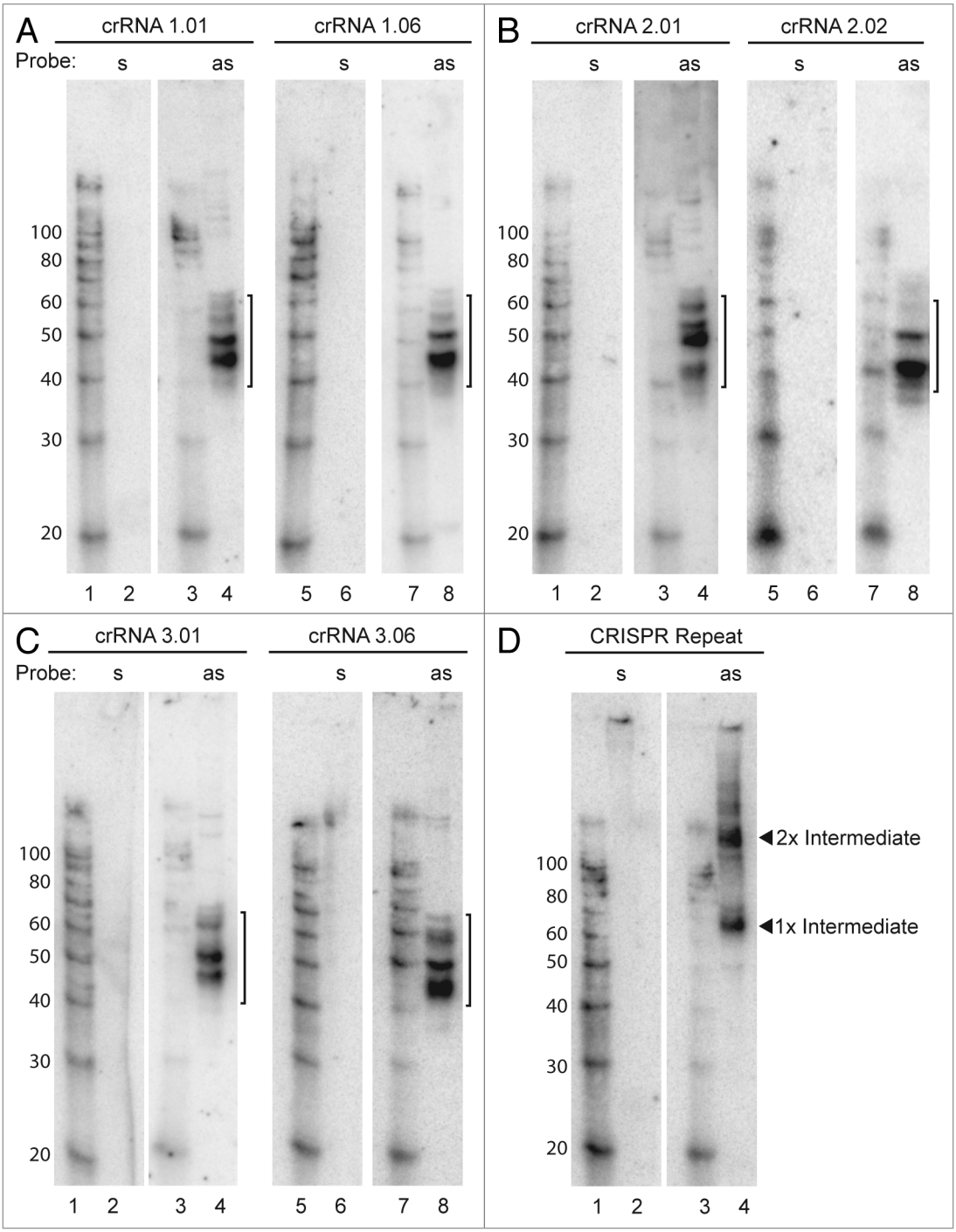

Figure 4. Northern analysis of crRNA expression. Blots were probed for two representative crRNAs from each of the three CRISPR loci (**A–C**) as well as the CRISPR repeat sequence (**D**). For each panel, both sense (s) and antisense (as) probes were tested. Odd numbered lanes contain the Decade Marker RNA, and even numbered lanes indicate lanes containing 10 μg of *Tko* total RNA. The crRNA species are indicated (brackets, **A–C**). In (**D**), RNAs corresponding in size to the predicted Cas6 cleavage products (1x: ~70 nt and 2x: ~140 nt) are indicated.

Discussion

Two largely uncharacterized CRISPR-Cas systems co-exist in T. kodakarensis: the Type I-A Csa system and Type I-B Cst system. In this initial investigation, we have characterized the crRNAs present in *Tko* (**Figs. 2–4**) and determined that at least one of these CRISPR-Cas systems can silence invading plasmids by a mechanism that functions at the DNA level (**Fig. 5**). In addition, we demonstrated that *Tko* CRISPR loci can be engineered to express functional crRNAs that effectively direct silencing of a corresponding target (**Fig. 7**).

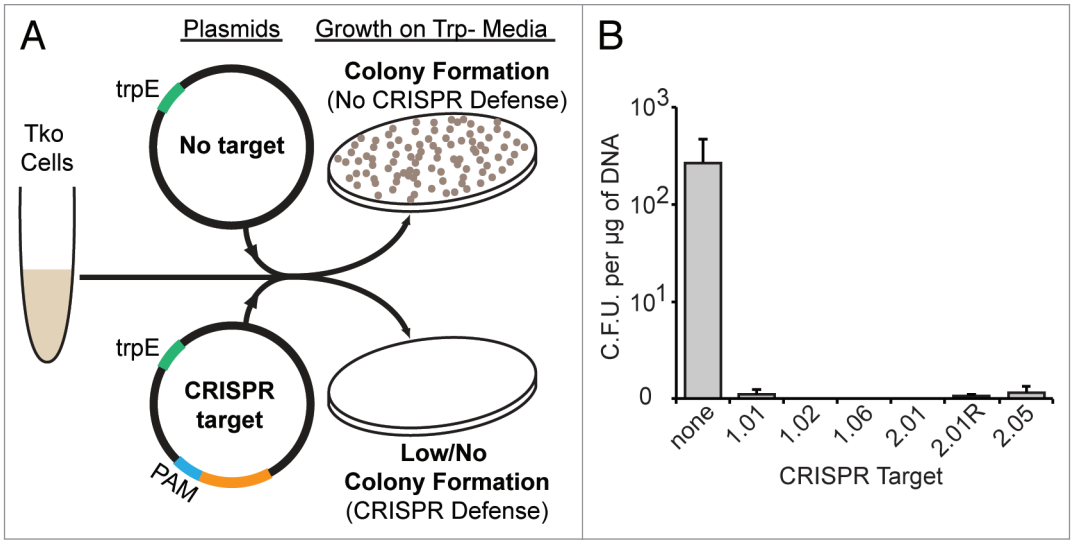

Figure 5. T. kodakarensis CRISPR-Cas systems silence plasmid DNA. (**A**) Graphic representation of the plasmid challenge assay. In the absence of CRISPR-Cas defense, the pLC64 plasmid, bearing a trpE marker gene cassette, is able to restore tryptophan prototrophy when transformed into *Tko* strain KUW1 (ΔpyrFΔtrpE). However, when derivatives of pLC64 bearing a CRISPR-Cas target sequence and a PAM are transformed, CRISPR-Cas silencing results in reduced or no colony formation on tryptophan-deficient medium. When silencing occurs by targeting DNA, the relative orientation of the target on the plasmid has no effect on silencing. (**B**) Plasmid challenge assay results. Colony formation per µg of plasmid DNA transformed is plotted on the Y-axis with error bars showing the standard error in five assay replicates. The target sequence of each plasmid is displayed on the X-axis. Nomenclature for the CRISPR targets is as follows: 2.01 indicates the protospacer target derived from the first leader proximal spacer of CRISPR locus 2, etc. The 2.01R target sequence is identical to the 2.01 target sequence, but its relative orientation on the plasmid is reversed. The assay was performed in two separate labs with a total of five replicates, using two different plasmid preparations.

Features of *Tko* crRNAs. crRNAs guide the CRISPR-Cas immune response and vary in composition among CRISPR-Cas pathways.[7,8,32] The mature crRNAs found in *Tko* exhibit 3' end heterogeneity but include a common, 8-nt CRISPR repeat-derived sequence (5' tag) at their 5' termini (**Fig. 3**). The observed 3' end heterogeneity may reflect the presence of distinct crRNA species associated with each the Csa and Cst systems, which may be resolved upon isolation of individual effector complexes (as we showed previously in P. furiosus, another organism with multiple CRISPR-Cas systems).[22,24,25] Interestingly, most of the analyzed crRNAs that co-purified with an affinity-tagged Csa2 (Cas7) protein overexpressed in *Sulfolobus solfataricus* were unprocessed 1X RNAs (with an 8-nt 5' tag and the full remaining repeat sequence at the 3' end). However, in *Tko*, the two most prominent species are crRNAs with no 3' repeat sequence (missing 1–3 nts of the guide region) and crRNAs with 5 nts of the conserved repeat downstream of the guide region (**Fig. 3**). The 1X crRNA is a relatively minor species in *Tko* (see northern analysis with guide region probes, **Fig. 4**). The presence of an 8 nt 5' tag is a common feature of studied Type I and Type III CRISPR-Cas systems.[16,19,24,33,38,43,47,48,59,66] The 5' tag provides crRNA identity and has been demonstrated to be essential for the function of the Type III-B Cmr system.[22,38] This 5' tag is a direct product of cleavage of the CRISPR transcripts by Cas6 or Cas6-like endonucleases,[16,17,19,21,47,67] suggesting that *Tko* crRNAs are generated by one of the two Cas6 proteins encoded in the *cas* gene cluster (**Fig. 1B**). The enzymes that trim crRNA 3' ends subsequent to Cas6 cleavage have not been identified.

Silencing by Type I-A Csa and Type I-B Cst CRISPR-Cas systems. Our findings indicate that the Type I-A Csa system or Type I-B Cst system, or both, can silence invaders by a mechanism that functions at the DNA level (**Fig. 5**). Very little is known about either of these CRISPR-Cas systems. Both the Csa and Cst systems include Cas3 superfamily proteins, the hallmark component of Type I systems. Cas3 has been observed to cleave DNA both in the absence of other components in vitro[68-70] and upon recruitment to target DNA via crRNA-protein complex (crRNP) interactions in vivo.[34] In addition to nuclease activity, Cas3 proteins include a conserved helicase domain (which is sometimes encoded as a separate polypeptide, e.g., Cas3' in the case of Csa; see **Fig. 1B**). Cas3 is a strong candidate for the effector nuclease that mediates invader silencing by the Csa and/or Cst system in *Tko*.

There are currently no reports on the structure or function of other components of Type I-B Cst CRISPR-Cas systems, which include Cst1 (Cas8a1 superfamily), Cst2 (Cas7 superfamily) and Cas5t (Cas5 superfamily) proteins (in addition to the universal Cas proteins Cas1 and Cas2). The Cst system [being studied in *Tko* (this study) and P. furious (our unpublished work)] is a member of the Type I-B class, which also includes the Csh system (being studied in H. volcanii, C. thermocellum and M. maripaludis).[35,47] Evidence suggests that the Csh system of H. volcanii can target invading plasmids at the DNA level,[35] however the extent to which the Cst and Csh CRISPR-Cas systems are structurally and mechanistically related is not yet known.

Protein components of the Type I-A Csa system have been studied and found to interact in vitro (in Thermoproteus tenax[44])

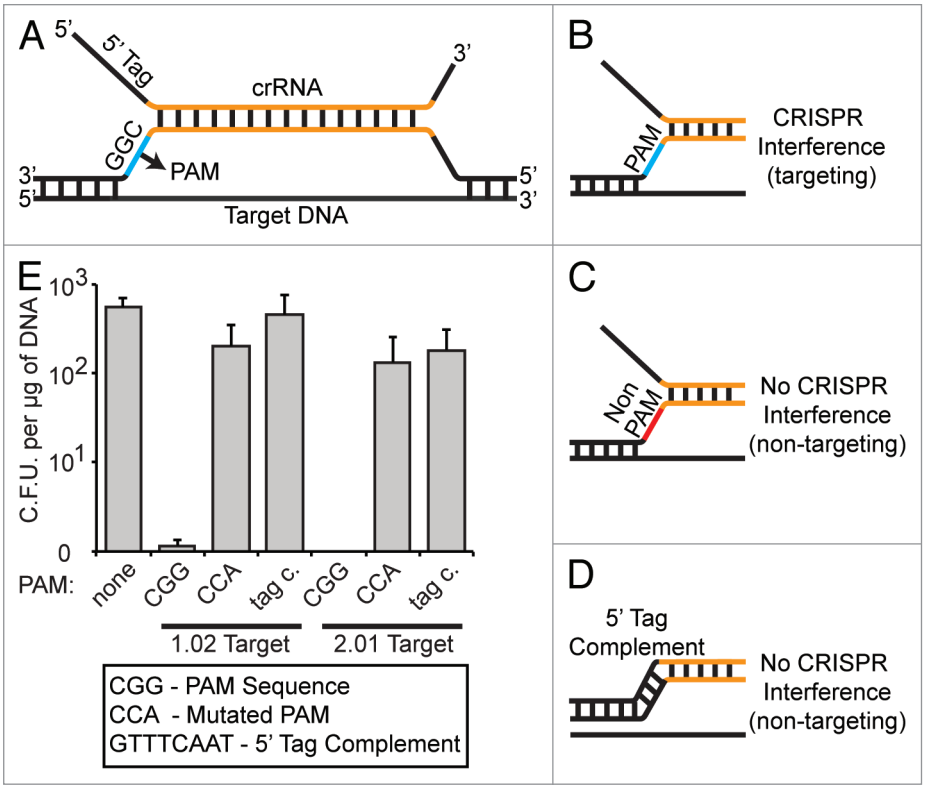

Figure 6. CRISPR-Cas plasmid interference is Protospacer Adjacent Motif (PAM)-dependent. (**A**) Diagram of crRNA base-pairing with targeted DNA molecule. The guide sequence of the crRNA (orange), derived from the spacer, is shown base pairing to its complement strand in the target (orange). Repeat sequences of the crRNA (black) and the PAM region (blue) are indicated (**B–D**). Diagrams for comparing three constructs with different protospacer flanking sequences used to test the mode of self vs. non-self determination in *Tko* CRISPR-Cas defense. (**B**) Targeting via detection of a PAM (tested with a predicted PAM sequence 5'-CGG-3', blue). (**C**) Avoidance of CRISPR-Cas targeting via mutation of the PAM to a non-PAM sequence (tested with predicted non-PAM sequence 5'-CCA-3', red). (**D**) Avoidance of CRISPR-Cas targeting via base-pairing between the 8 nt of the 5'-tag sequence and the sequence immediately downstream of the region complementary to the guide (tested with the 5'-tag complement sequence 5'-GTTTCAAT-3', black). (**E**) Plasmid challenge assay results. Colony formation per µg of plasmid DNA transformed is plotted on the Y-axis with error bars displaying the standard error in six assay replicates, using two different plasmid preparations. The target sequence and the flanking region (PAM) of each plasmid are displayed on the X-axis.

and in vivo [in Sulfolobus solfataricus[29] and Pyrococcus furiosus (our unpublished results)]. Evidence indicates that Csa2 (Cas7 superfamily) and Cas5a (Cas5 superfamily) may be central components that play a role in crRNA binding and DNA binding.[29] The X-ray structure of S. solfataricus Csa3 suggests that it is a potential DNA binding protein and regulator of *cas* gene expression.[71] The roles of Csa5, Csx9/Csa4 (Cas8a2 superfamily), Csa3 and Csa1 (Cas4 superfamily; not found in *Tko*) remain to be determined.

Most DNA targeting CRISPR-Cas systems require the presence of a PAM near the target sequence in the invader DNA. Our data corroborate the prediction that an NGG PAM sequence is required for Type I-A and/or I-B CRISPR-Cas systems.[7,52,56] We found that a plasmid with a CGG PAM was silenced in *Tko*, and that mutation to CCA prevented silencing (**Fig. 6**). The molecular mechanism by which NGG PAMs mark DNA for targeting by CRISPR-Cas systems is not yet known. Recently, it was found that a different PAM sequence increases the affinity of the Type I-E Cse crRNP effector complex for a target sequence.[34] Moreover, PAM recognition and target DNA binding required a type-specific Cas protein (Cse1) in Type I-E systems.[28,72-74] By analogy, it will be of interest to determine if particular Csa and Cst proteins specifically interact with NGG PAM sequences during CRISPR defense.

CRISPR loci can be engineered to silence novel target DNAs. In this study, we successfully engineered a CRISPR in situ to program silencing of a novel target plasmid. Engineering of a CRISPR to provide resistance to phage infection has also been demonstrated for the Type I-E Cse system in *E. coli*.[45,75] Programmable CRISPR-Cas targeting is of great interest for the protection of industrial microbes (used to produce food, biofuels, pharmaceuticals, etc.) against viruses that spoil fermentations and cause significant economic losses.[1,76] Other potential approaches to direct CRISPR-Cas systems against specific DNA targets include the introduction into an organism of a plasmid containing a CRISPR with sequences that target the DNA and inoculation of an organism with the target DNA to induce CRISPR adaptation,[11-13,16,37,49,50] however these approaches have significant limitations. The use of plasmid-encoded CRISPRs requires constant selective pressure to maintain the plasmid, such as growth in the presence of antibiotics or in defined media lacking a nutrient. Phage/plasmid challenge to induce CRISPR-Cas adaptation and immunity has only been successful in very few organisms. In this work, we showed that intentional engineering of a CRISPR locus can provide CRISPR-Cas-mediated resistance against an invader in *Tko*.

Materials and Methods

Strains and growth conditions. The strains and plasmids used in this study are listed in **Table S1**. T. kodakarensis KOD1 and derivative strains were grown under strict anaerobic conditions at 85°C either in complex growth medium (ASW-YT) or defined medium (ASW-AA).[77] ASW-YT growth medium was supplemented with 5 g/L sodium pyruvate (ASW-YT-Pyr) for RNA isolation, with 2 g/L elemental sulfur (ASW-YT-S^0) for competent cell preparation, or with 2 mL/L polysulfide solution (10 g of Na_2S $9H_2O$ and 3 g of sulfur flowers in 15 ml of H_2O) for

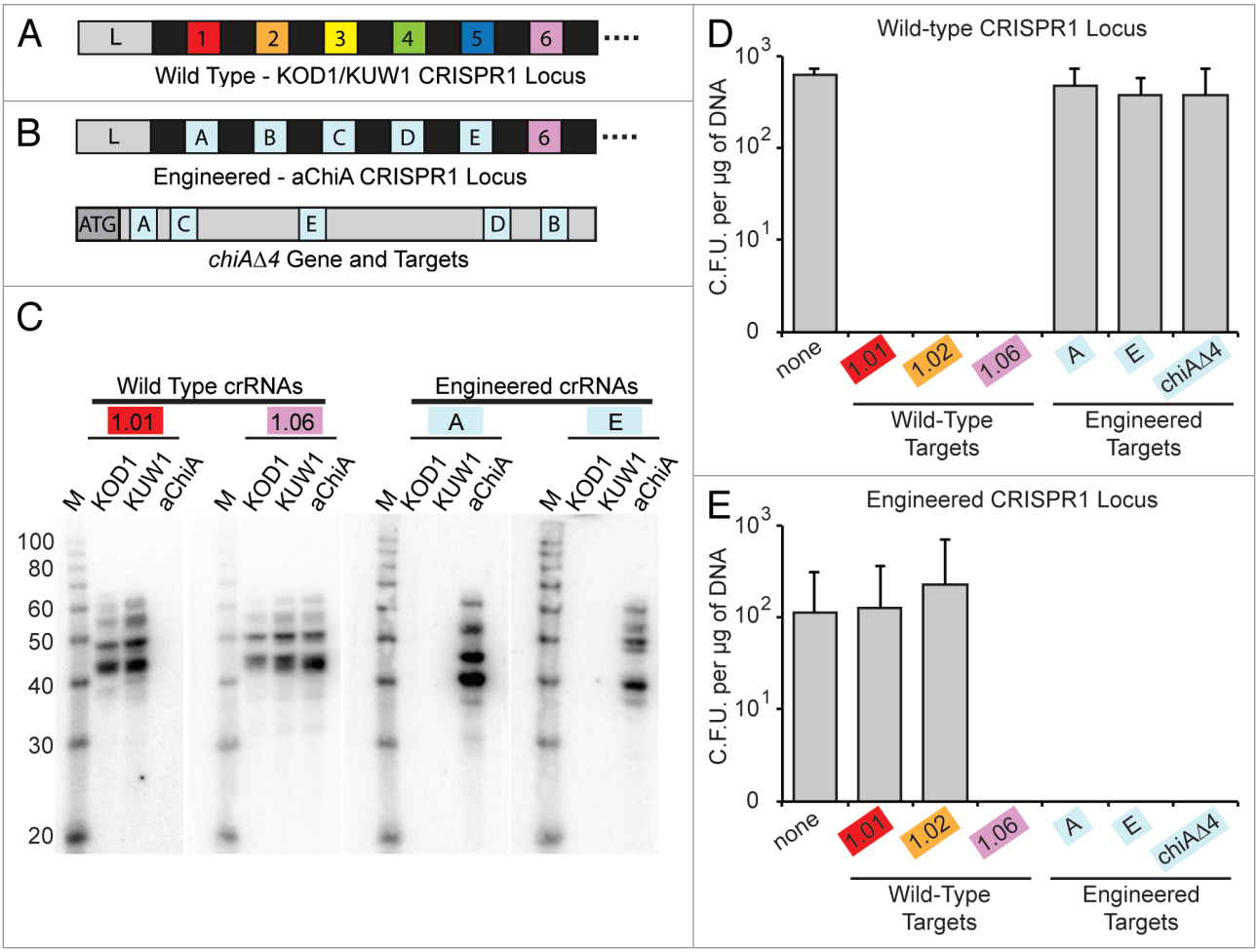

Figure 7. Engineered CRISPR1 locus guides CRISPR-Cas to target plasmid. Graphic representation of the 5' ends of the wild-type CRISPR1 locus found in strains KOD1 and KUW1 (**A**) and the engineered CRISPR1 locus found in the strain aChiA, and the *chiAΔ4* gene (**B**). Engineered spacers and their corresponding targets within the *chiAΔ4* gene are indicated by light blue boxes. (**C**) Northern analysis of wild-type crRNA (1.01, 1.06) or engineered crRNA (**A and E**) expression. Lanes contain either Decade Marker RNA (M) or 10 μg of total RNA isolated from the wild-type CRISPR1 strains (KOD1 and KUW1) and engineered CRISPR1 strain (aChiA). Colored boxes above each blot indicate the crRNA guide sequence probed in each experiment. Plasmid challenge assay results using the KUW1 (WT CRISPR1) strain (**D**) or the aChiA (Engineered CRISPR1) strain (**E**). Colony formation per μg of plasmid DNA transformed is plotted on the Y-axis with error bars show the standard error in four replicate experiments performed by two separate labs. Wild-type CRISPR and engineered CRISPR plasmid targets are indicated on the X-axis, with colored boxes indicating the origin of the crRNA guiding the potential plasmid silencing. *ChiAΔ4* target plasmids contain the *chiAΔ4* expression cassette.

recovery after transformation before plating. Preparation of solid medium and cultivation on plates was performed as previously described.[78] Specific conditions applied to select transformants are described below. *Escherichia coli* strains Top10 and DH5α were used for general plasmid DNA manipulation. Cultures were grown at 37°C in Luria-Bertani (LB) broth supplemented with ampicillin (100 μg/mL).

Small RNA library preparation. Total RNA samples were isolated from ~50 mg *Tko* KOD1 cells using Trizol (Invitrogen) following manufacturer's protocol. Total RNA (50 μg) was separated by denaturing PAGE gel electrophorhesis and small RNAs (~18–75 nt) were isolated by gel extraction. RNAs were dephosphorylated with thermosensitive alkaline phosphatase (Promega). Dephosphorylated RNAs were ligated to a 3' adaptor (3'-Adaptor Solexa, **Table S2**) with truncated T4 RNA Ligase (NEB). Ligated RNAs were isolated by gel extraction. Gel extracted RNAs were 5' phosphorylated with T4 polynucleotide kinase (Ambion). RNAs were then ligated with a 5' adaptor (5'-Adaptor Solexa, **Table S2**) by T4 ssRNA Ligase 1 (NEB). 5'- and 3'-ligated RNAs were subjected to reverse transcription using the Solexa RT primer (**Table S2**) with Superscript III reverse transcriptase (Invitrogen). cDNAs were cleared of RNAs by RNase H (NEB) treatment. PCR amplification of cDNA libraries was performed using primers Solexa RT and Solexa PCR (**Table S2**) with Crimson Taq (NEB) using High Fidelity Buffer with $MgCl_2$ (Roche). Details of procedure have been described previously.[22]

Illumina xequencing and analysis. cDNA libraries were visualized by electrophoresis, quantitated by Agilent Bioanalyzer and NanoDrop, diluted to 2 pM and subject to 76 cycles of sequencing on an Illumina Genome Analyzer IIx. Sequence reads were trimmed of the 3' linker, and reads 18–76 nt in length were aligned to the *Tko* genome using Bowtie.[79] The 5' and 3' ends

of reads mapping to the CRISPR loci were calculated using a custom PERL script available upon request.

Northern analysis. Total RNA samples were isolated from ~50 mg of *Tko* cells (KOD1, KUW1 or aChiA) using Trizol (Invitrogen). Ten μg of total RNA samples were separated on 15% TBE-Urea polyacrylamide gels (Criterion, Bio-Rad) beside a (^{32}P)-5' end-labeled Decade Marker RNA (Ambion). RNAs were transferred to nylon Zeta-Probe membranes (Bio-Rad) using a Trans-Blot SD Semi-Dry Cell (Bio-Rad). Membranes were baked at 80°C for 1 h before prehybridization in a ProBlot hydridization oven (LabNet) for 1 h at 42°C. Prehybridization and hybridization was performed in hybridization buffer containing 5x SSC, 7% SDS, 20 mM $NaPO_4$ (pH 7.0) and 1x Denhardt's solution. Deoxyribonucleotide probes (MWG/Operon) (10 pmol) were 5' end-labeled with T4 Polynucleotide Kinase (Ambion) and γ-(^{32}P)-ATP (specific activity > 7,000 Ci/mmol, MP Biomedicals) using standard protocols. Labeled probes (1 million cpm/mL) were added to prehybridization buffer, followed by hybridization at 42°C overnight. Probed membranes were washed twice in 2x SCC, 0.5% SDS for 30 min at 42°C. Blots were exposed to a PhosphoImager screen for 4–24 h and scanned. Probe sequences used are listed in **Table S2**.

General DNA manipulation and construction of shuttle vector plasmids. Plasmid DNAs for sequencing and routine analysis were isolated from Top10 and DH5α using the QIAprep Spin Miniprep Kit (Qiagen) or Quantum Prep Plasmid Miniprep kit (BIO-RAD). Large-scale plasmid DNA isolation from *E. coli* strains was performed for the construction of target-bearing plasmids or for use in *Tko* plasmid transformation assays. These isolations were done using the QIAfilter Plasmid Midi Kit (Qiagen) or Quantum Prep Plasmid Miniprep kit (BIO-RAD). Routine PCR screening was performed with Crimson Taq (NEB) or KOD plus (Toyobo).

Plasmids containing CRISPR target inserts (**Table S1**) were constructed by ligating annealed 5'-phosphorylated oligos (**Table S2**) into the BamHI site of the shuttle vector plasmid pLC64.[61] The plasmid pLC64 was linearized with either BamHI-HF (NEB) or BamHI (Toyobo) and dephosphorylated with thermosensitive alkaline phosphatase (Promega) or calf intestine alkaline phosphatase (Toyobo). 5'-Phosphorylated CRISPR target DNA oligo pairs (**Table S2**) purchased (MWG/Operon) or prepared using T4 polynucleotide kinase (Toyobo), were annealed in 10 mM Tris (pH 8.0), 1 mM EDTA with or without 500 mM NaCl. Annealed oligo pairs were ligated with BamHI-linearized pLC64 using T4 DNA ligase (NEB) or Ligation high (Toyobo). *E. coli* transformants bearing plasmids with the CRISPR target insert were identified by digestion of isolated plasmids with BamHI or by PCR. Plasmids containing inserts were sequenced to confirm insert sequence and orientation. Oligo pairs used for the construction of target-bearing plasmids are indicated in **Table S2**, with the "+" oligo being annealed with the cognate "-" oligo to generate a double-stranded target for ligation. pLC64, a plasmid that autonomously replicates in *Tko*, was a kind gift from Tom Santangelo and John Reeve of Ohio State University.[61] An expression cassette of chiAΔ4, which encodes the C-terminal endochitinase domain of the chitinase from *Tko* (ChiAΔ4), was amplified using pSecChiA as a template.[80] Inverse PCR using the primer set ChiAF1/ChiAR1 was performed to remove the secretion signal, and the amplified fragment was self-ligated. A second PCR was performed using this DNA as a template with the primers ChiAF2/ChiAR2. The amplified fragment was digested with NotI and EcoRV, and inserted into the NotI/EcoRV sites of pLC64, resulting in pLC64-ChiAΔ4.

Plasmid transformation assay in T. kodakarensis. T. kodakarensis strains KUW1 (ΔpyrFΔtrpE) and aChiA (ΔpyrFΔtrpEΔchiA CRISPR1::engineered CRISPR1) were used for plasmid transformation assays. All steps were done under strict anaerobic conditions. Transformations were performed as follows. Cells were harvested after anaerobic growth to mid- to late-log phase by centrifugation, resuspended into 200 μL of 0.8 x ASW and incubated on ice for 30 min. After addition of 2 μg of plasmid, cells were incubated on ice for an additional hour. Following recovery in 3 mL of ASW-YT media for 2 h at 85°C, cells were harvested by brief centrifugation, and resuspended in 200 μL 0.8 x ASW prior to plating on ASW-AAW$^-$ media supplemented with 10 μg/mL uracil. Plates were incubated anaerobically at 85°C for 3 d. After incubation, colonies per plate were enumerated.

CRISPR-Cas escape mutant analysis. In order to examine the sequences of the CRISPR loci of the escape mutants, DNA fragments including the entire CRISPR loci 2 were amplified by PCR using the primer set TkL2F1/TkL2R1 and genomic DNA from each of the mutant strains as template DNA. Primers used for sequencing were TkL2F1-F3 and TkL2R1-R3.

Construction of engineered CRISPR strain. The plasmid pUD3, a derivative of pUC118 containing a pyrF marker gene cassette cloned into its ApaI site, was used for the construction of two genome modification plasmids.[81] The plasmid pJE2 was designed to replace the first five spacers of CRISPR1 locus in *Tko* with five spacers targeting the DNA sequence of an established reporter gene, the C-terminal domain of TK1765 (*chiAΔ4*).[65] Spacers were designed to target regions of *chiAΔ4* with flanking sequences lacking complementarity with the putative 5' tag sequence (5'-AUUGAAAC-3') of the engineered *Tko* crRNAs. To facilitate proper homologous recombination into the genome, we designed a synthetic construct containing the engineered CRISPR1 locus (with spacers 1–5 replaced), an additional 1 kilobase of 5' and 3' flanking sequence identical to the flanking sequence of CRISPR1 in the *Tko* genome, and a short polylinker region with multiple restriction sites just beyond the 5' and 3' flanking sequences. The polylinker regions included a KpnI site on the leader-proximal end and a BamHI site on the leader-distal end of the construct. GeneArt was contracted to synthesize and clone the engineered CRISPR construct into the BamHI/KpnI sites of the pUD3 plasmid, creating the plasmid pJE2. A *Tko* strain carrying a disruption of the chiA gene (KC2) was used as the host strain. KC2 was constructed as described previously using *Tko* KUW1 as the parent strain.[82] *Tko* KC2 cells grown in ASW-YT-S^0 until the late log-phase were harvested, washed and resuspended in 200 μL of 0.8 x ASW and kept on ice for 30 min. After treatment with 3.0 μg of pJE2 and further incubation on ice for 1 h, cells were cultivated in

ASW-AA medium (without uracil) with elemental sulfur for 2–5 d at 85°C. Cells were consecutively cultivated under the same conditions to enrich the desired transformants that display uracil prototrophy. Cells were then spread onto ASW-YT solid medium supplemented with 2 mL/L polysulfide solution, 10 g/L 5-fluoroorotic acid (5-FOA) and 60 mM NaOH. Only cells that have undergone a second recombination that removes the pyrF gene can grow in the presence of 5-FOA. Cells were grown for 2 d at 85°C for colony formation, and transformants were isolated and cultivated in ASW-YT-S^0. The genotypes of the transformants were analyzed by PCR using the primer set TkL1F1/TkL1R1. A transformant whose first five spacers of CRISPR1 locus were properly exchanged was chosen and designated as *Tko* aChiA.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Thomas Santangelo, Lubomira Cubonova and John Reeve (Ohio State University) for sharing technical assistance and helpful advice regarding *T. kodakarensis*. We thank Anisha Hedge and Megan Tibbitts (Terns lab) for help with northern blot experiments, and Barney Whitman, Felipe Sarmiento, Nancy Garrison and Ron Garrison (University of Georgia) for help with setting up anaerobic culturing equipment. This work was supported by National Institutes of Health (NIH) grant RO1GM54682 (including ARRA funds) to M.P.T. and R.M.T.

Supplemental Material

Supplemental material may be found here:
www.landesbioscience.com/journals/rnabiology/article/24084

References

1. Deveau H, Garneau JE, Moineau S. CRISPR/Cas system and its role in phage-bacteria interactions. Annu Rev Microbiol 2010; 64:475-93; PMID:20528693; http://dx.doi.org/10.1146/annurev.micro.112408.134123.
2. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, et al. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol 2011; 9:467-77; PMID:21552286; http://dx.doi.org/10.1038/nrmicro2577.
3. Marraffini LA, Sontheimer EJ. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet 2010; 11:181-90; PMID:20125085; http://dx.doi.org/10.1038/nrg2749.
4. Terns MP, Terns RM. CRISPR-based adaptive immune systems. Curr Opin Microbiol 2011; 14:321-7; PMID:21531607; http://dx.doi.org/10.1016/j.mib.2011.03.005.
5. Fineran PC, Charpentier E. Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. Virology 2012; 434:202-9; PMID:23123013; http://dx.doi.org/10.1016/j.virol.2012.10.003.
6. Garrett RA, Vestergaard G, Shah SA. Archaeal CRISPR-based immune systems: exchangeable functional modules. Trends Microbiol 2011; 19:549-56; PMID:21945420; http://dx.doi.org/10.1016/j.tim.2011.08.002.
7. Westra ER, Swarts DC, Staals RH, Jore MM, Brouns SJ, van der Oost J. The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. Annu Rev Genet 2012; 46:311-39; PMID:23145983; http://dx.doi.org/10.1146/annurev-genet-110711-155447.
8. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. Nature 2012; 482:331-8; PMID:22337052; http://dx.doi.org/10.1038/nature10886.
9. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science 2007; 315:1709-12; PMID:17379808; http://dx.doi.org/10.1126/science.1138140.
10. Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P, et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 2010; 468:67-71; PMID:21048762; http://dx.doi.org/10.1038/nature09523.
11. Datsenko KA, Pougach K, Tikhonov A, Wanner BL, Severinov K, Semenova E. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat Commun 2012; 3:945; PMID:22781758; http://dx.doi.org/10.1038/ncomms1937.
12. Swarts DC, Mosterd C, van Passel MW, Brouns SJ. CRISPR interference directs strand specific spacer acquisition. PLoS One 2012; 7:e35888; PMID:22558257; http://dx.doi.org/10.1371/journal.pone.0035888.
13. Yosef I, Goren MG, Qimron U. Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. Nucleic Acids Res 2012; 40:5569-76; PMID:22402487; http://dx.doi.org/10.1093/nar/gks216.
14. Erdmann S, Garrett RA. Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. Mol Microbiol 2012; 85:1044-56; PMID:22834906; http://dx.doi.org/10.1111/j.1365-2958.2012.08171.x.
15. Weinberger AD, Sun CL, Pluciński MM, Denef VJ, Thomas BC, Horvath P, et al. Persisting viral sequences shape microbial CRISPR-based immunity. PLoS Comput Biol 2012; 8:e1002475; PMID:22532794; http://dx.doi.org/10.1371/journal.pcbi.1002475.
16. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 2008; 321:960-4; PMID:18703739; http://dx.doi.org/10.1126/science.1159689.
17. Carte J, Wang R, Li H, Terns RM, Terns MP. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. Genes Dev 2008; 22:3489-96; PMID:19141480; http://dx.doi.org/10.1101/gad.1742908.
18. Carte J, Pfister NT, Compton MM, Terns RM, Terns MP. Binding and cleavage of CRISPR RNA by Cas6. RNA 2010; 16:2181-8; PMID:20884784; http://dx.doi.org/10.1261/rna.2230110.
19. Haurwitz RE, Jinek M, Wiedenheft B, Zhou K, Doudna JA. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. Science 2010; 329:1355-8; PMID:20829488; http://dx.doi.org/10.1126/science.1192272.
20. Hatoum-Aslan A, Maniv I, Marraffini LA. Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. Proc Natl Acad Sci USA 2011; 108:21218-22; PMID:22160698; http://dx.doi.org/10.1073/pnas.1112832108.
21. Garside EL, Schellenberg MJ, Gesner EM, Bonanno JB, Sauder JM, Burley SK, et al. Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases. RNA 2012; 18:2020-8; PMID:23006625; http://dx.doi.org/10.1261/rna.033100.112.
22. Hale CR, Majumdar S, Elmore J, Pfister N, Compton M, Olson S, et al. Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. Mol Cell 2012; 45:292-302; PMID:22227116; http://dx.doi.org/10.1016/j.molcel.2011.10.023.
23. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 2011; 471:602-7; PMID:21455174; http://dx.doi.org/10.1038/nature09886.
24. Hale C, Kleppe K, Terns RM, Terns MP. Prokaryotic silencing (psi)RNAs in Pyrococcus furiosus. RNA 2008; 14:2572-9; PMID:18971321; http://dx.doi.org/10.1261/rna.1246808.
25. Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, et al. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell 2009; 139:945-56; PMID:19945378; http://dx.doi.org/10.1016/j.cell.2009.07.040.
26. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 2012; 337:816-21; PMID:22745249; http://dx.doi.org/10.1126/science.1225829.
27. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci USA 2012; 109:E2579-86; PMID:22949671; http://dx.doi.org/10.1073/pnas.1208507109.
28. Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol 2011; 18:529-36; PMID:21460843; http://dx.doi.org/10.1038/nsmb.2019.
29. Lintner NG, Kerou M, Brumfield SK, Graham S, Liu H, Naismith JH, et al. Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). J Biol Chem 2011; 286:21643-56; PMID:21507944; http://dx.doi.org/10.1074/jbc.M111.238485.
30. Wiedenheft B, Lander GC, Zhou K, Jore MM, Brouns SJ, van der Oost J, et al. Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature 2011; 477:486-9; PMID:21938068; http://dx.doi.org/10.1038/nature10402.

31. Haft DH, Selengut J, Mongodin EF, Nelson KE. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput Biol 2005; 1:e60; PMID:16292354; http://dx.doi.org/10.1371/journal.pcbi.0010060.
32. Bhaya D, Davison M, Barrangou R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu Rev Genet 2011; 45:273-97; PMID:22060043; http://dx.doi.org/10.1146/annurev-genet-110410-132430.
33. Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science 2008; 322:1843-5; PMID:19095942; http://dx.doi.org/10.1126/science.1165771.
34. Westra ER, van Erp PB, Künne T, Wong SP, Staals RH, Seegers CL, et al. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Mol Cell 2012; 46:595-605; PMID:22521689; http://dx.doi.org/10.1016/j.molcel.2012.03.018.
35. Fischer S, Maier LK, Stoll B, Brendel J, Fischer E, Pfeiffer F, et al. An archaeal immune system can detect multiple protospacer adjacent motifs (PAMs) to target invader DNA. J Biol Chem 2012; 287:33351-63; PMID:22767603; http://dx.doi.org/10.1074/jbc.M112.377002.
36. Gudbergsdottir S, Deng L, Chen Z, Jensen JV, Jensen LR, She Q, et al. Dynamic properties of the Sulfolobus CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. Mol Microbiol 2011; 79:35-49; PMID:21166892; http://dx.doi.org/10.1111/j.1365-2958.2010.07452.x.
37. Manica A, Zebec Z, Teichmann D, Schleper C. In vivo activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon. Mol Microbiol 2011; 80:481-91; PMID:21385233; http://dx.doi.org/10.1111/j.1365-2958.2011.07586.x.
38. Zhang J, Rouillon C, Kerou M, Reeks J, Brugger K, Graham S, et al. Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. Mol Cell 2012; 45:303-13; PMID:22227115; http://dx.doi.org/10.1016/j.molcel.2011.12.013.
39. Wiedenheft B, van Duijn E, Bultema JB, Waghmare SP, Zhou K, Barendregt A, et al. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. Proc Natl Acad Sci USA 2011; 108:10092-7; PMID:21536913; http://dx.doi.org/10.1073/pnas.1102716108.
40. Jansen R, Embden JD, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol 2002; 43:1565-75; PMID:11952905; http://dx.doi.org/10.1046/j.1365-2958.2002.02839.x.
41. Lillestøl RK, Shah SA, Brügger K, Redder P, Phan H, Christiansen J, et al. CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties. Mol Microbiol 2009; 72:259-72; PMID:19239620; http://dx.doi.org/10.1111/j.1365-2958.2009.06641.x.
42. Pul U, Wurm R, Arslan Z, Geissen R, Hofmann N, Wagner R. Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. Mol Microbiol 2010; 75:1495-512; PMID:20132443; http://dx.doi.org/10.1111/j.1365-2958.2010.07073.x.
43. Bernick DL, Cox CL, Dennis PP, Lowe TM. Comparative genomic and transcriptional analyses of CRISPR systems across the genus Pyrobaculum. Front Microbiol 2012; 3:251; PMID:22811677; http://dx.doi.org/10.3389/fmicb.2012.00251.
44. Plagens A, Tjaden B, Hagemann A, Randau L, Hensel R. Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon Thermoproteus tenax. J Bacteriol 2012; 194:2491-500; PMID:22408157; http://dx.doi.org/10.1128/JB.00206-12.
45. Pougach K, Semenova E, Bogdanova E, Datsenko KA, Djordjevic M, Wanner BL, et al. Transcription, processing and function of CRISPR cassettes in Escherichia coli. Mol Microbiol 2010; 77:1367-79; PMID:20624226; http://dx.doi.org/10.1111/j.1365-2958.2010.07265.x.
46. Gesner EM, Schellenberg MJ, Garside EL, George MM, Macmillan AM. Recognition and maturation of effector RNAs in a CRISPR interference pathway. Nat Struct Mol Biol 2011; 18:688-92; PMID:21572444; http://dx.doi.org/10.1038/nsmb.2042.
47. Richter H, Zoephel J, Schermuly J, Maticzka D, Backofen R, Randau L. Characterization of CRISPR RNA processing in Clostridium thermocellum and Methanococcus maripaludis. Nucleic Acids Res 2012; 40:9887-96; PMID:22879377; http://dx.doi.org/10.1093/nar/gks737.
48. Nam KH, Haitjema C, Liu X, Ding F, Wang H, DeLisa MP, et al. Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system. Structure 2012; 20:1574-84; PMID:22841292; http://dx.doi.org/10.1016/j.str.2012.06.016.
49. Cady KC, Bondy-Denomy J, Heussler GE, Davidson AR, O'Toole GA. The CRISPR/Cas adaptive immune system of Pseudomonas aeruginosa mediates resistance to naturally occurring and engineered phages. J Bacteriol 2012; 194:5728-38; PMID:22885297; http://dx.doi.org/10.1128/JB.01184-12.
50. Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, et al. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J Bacteriol 2008; 190:1390-400; PMID:18065545; http://dx.doi.org/10.1128/JB.01412-07.
51. Marraffini LA, Sontheimer EJ. Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature 2010; 463:568-71; PMID:20072129; http://dx.doi.org/10.1038/nature08703.
52. Mojica FJ, Díez-Villaseñor C, García-Martínez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 2009; 155:733-40; PMID:19246744; http://dx.doi.org/10.1099/mic.0.023960-0.
53. Morikawa M, Izawa Y, Rashid N, Hoaki T, Imanaka T. Purification and characterization of a thermostable thiol protease from a newly isolated hyperthermophilic Pyrococcus sp. Appl Environ Microbiol 1994; 60:4559-66; PMID:7811092.
54. Grissa I, Vergnaud G, Pourcel C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics 2007; 8:172; PMID:17521438; http://dx.doi.org/10.1186/1471-2105-8-172.
55. Rousseau C, Gonnet M, Le Romancer M, Nicolas J. CRISPI: a CRISPR interactive database. Bioinformatics 2009; 25:3317-8; PMID:19846435; http://dx.doi.org/10.1093/bioinformatics/btp586.
56. Kunin V, Sorek R, Hugenholtz P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol 2007; 8:R61; PMID:17442114; http://dx.doi.org/10.1186/gb-2007-8-4-r61.
57. Wang R, Preamplume G, Terns MP, Terns RM, Li H. Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. Structure 2011; 19:257-64; PMID:21300293; http://dx.doi.org/10.1016/j.str.2010.11.014.
58. Littlefield O, Korkhin Y, Sigler PB. The structural basis for the oriented assembly of a TBP/TFB/promoter complex. Proc Natl Acad Sci USA 1999; 96:13668-73; PMID:10570130; http://dx.doi.org/10.1073/pnas.96.24.13668.
59. Randau L. RNA processing in the minimal organism Nanoarchaeum equitans. Genome Biol 2012; 13:R63; PMID:22809431; http://dx.doi.org/10.1186/gb-2012-13-7-r63.
60. Sato T, Fukui T, Atomi H, Imanaka T. Improved and versatile transformation system allowing multiple genetic manipulations of the hyperthermophilic archaeon Thermococcus kodakaraensis. Appl Environ Microbiol 2005; 71:3889-99; PMID:16000802; http://dx.doi.org/10.1128/AEM.71.7.3889-3899.2005.
61. Santangelo TJ, Cubonová L, Reeve JN. Shuttle vector expression in Thermococcus kodakaraensis: contributions of cis elements to protein synthesis in a hyperthermophilic archaeon. Appl Environ Microbiol 2008; 74:3099-104; PMID:18378640; http://dx.doi.org/10.1128/AEM.00305-08.
62. Almendros C, Guzmán NM, Díez-Villaseñor C, García-Martínez J, Mojica FJ. Target motifs affecting natural immunity by a constitutive CRISPR-Cas system in Escherichia coli. PLoS One 2012; 7:e50797; PMID:23189210; http://dx.doi.org/10.1371/journal.pone.0050797.
63. Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, et al. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc Natl Acad Sci USA 2011; 108:10098-103; PMID:21646539; http://dx.doi.org/10.1073/pnas.1104144108.
64. Semenova E, Nagornykh M, Pyatnitskiy M, Artamonova II, Severinov K. Analysis of CRISPR system function in plant pathogen Xanthomonas oryzae. FEMS Microbiol Lett 2009; 296:110-6; PMID:19459963; http://dx.doi.org/10.1111/j.1574-6968.2009.01626.x.
65. Tanaka T, Fukui T, Imanaka T. Different cleavage specificities of the dual catalytic domains in chitinase from the hyperthermophilic archaeon Thermococcus kodakaraensis KOD1. J Biol Chem 2001; 276:35629-35; PMID:11468293; http://dx.doi.org/10.1074/jbc.M105919200.
66. Xu N, Li Y, Zhao YT, Guo L, Fang YY, Zhao JH, et al. Identification and characterization of small RNAs in the hyperthermophilic archaeon Sulfolobus solfataricus. PLoS One 2012; 7:e35306; PMID:22514725; http://dx.doi.org/10.1371/journal.pone.0035306.
67. Przybilski R, Richter C, Gristwood T, Clulow JS, Vercoe RB, Fineran PC. Csy4 is responsible for CRISPR RNA processing in Pectobacterium atrosepticum. RNA Biol 2011; 8:517-28; PMID:21519197; http://dx.doi.org/10.4161/rna.8.3.15190.
68. Sinkunas T, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. EMBO J 2011; 30:1335-42; PMID:21343909; http://dx.doi.org/10.1038/emboj.2011.41.
69. Beloglazova N, Petit P, Flick R, Brown G, Savchenko A, Yakunin AF. Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. EMBO J 2011; 30:4616-27; PMID:22009198; http://dx.doi.org/10.1038/emboj.2011.377.
70. Mulepati S, Bailey S. Structural and biochemical analysis of nuclease domain of clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 3 (Cas3). J Biol Chem 2011; 286:31896-903; PMID:21775431; http://dx.doi.org/10.1074/jbc.M111.270017.
71. Lintner NG, Frankel KA, Tsutakawa SE, Alsbury DL, Copié V, Young MJ, et al. The structure of the CRISPR-associated protein Csa3 provides insight into the regulation of the CRISPR/Cas system. J Mol Biol 2011; 405:939-55; PMID:21093452; http://dx.doi.org/10.1016/j.jmb.2010.11.019.
72. Sashital DG, Wiedenheft B, Doudna JA. Mechanism of foreign DNA selection in a bacterial adaptive immune system. Mol Cell 2012; 46:606-15; PMID:22521690; http://dx.doi.org/10.1016/j.molcel.2012.03.020.

73. Mulepati S, Orr A, Bailey S. Crystal structure of the largest subunit of a bacterial RNA-guided immune complex and its role in DNA target binding. J Biol Chem 2012; 287:22445-9; PMID:22621933; http://dx.doi.org/10.1074/jbc.C112.379503.
74. Westra ER, Nilges B, van Erp PB, van der Oost J, Dame RT, Brouns SJ. Cascade-mediated binding and bending of negatively supercoiled DNA. RNA Biol 2012; 9:1134-8; PMID:22954644; http://dx.doi.org/10.4161/rna.21410.
75. Westra ER, Pul U, Heidrich N, Jore MM, Lundgren M, Stratmann T, et al. H-NS-mediated repression of CRISPR-based immunity in Escherichia coli K12 can be relieved by the transcription activator LeuO. Mol Microbiol 2010; 77:1380-93; PMID:20659289; http://dx.doi.org/10.1111/j.1365-2958.2010.07315.x.
76. Barrangou R, Horvath P. CRISPR: new horizons in phage resistance and strain identification. Annu Rev Food Sci Technol 2012; 3:143-62; PMID:22224556; http://dx.doi.org/10.1146/annurev-food-022811-101134.
77. Sato T, Imanaka H, Rashid N, Fukui T, Atomi H, Imanaka T. Genetic evidence identifying the true gluconeogenic fructose-1,6-bisphosphatase in Thermococcus kodakaraensis and other hyperthermophiles. J Bacteriol 2004; 186:5799-807; PMID:15317785; http://dx.doi.org/10.1128/JB.186.17.5799-5807.2004.
78. Sato T, Fukui T, Atomi H, Imanaka T. Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon Thermococcus kodakaraensis KOD1. J Bacteriol 2003; 185:210-20; PMID:12486058; http://dx.doi.org/10.1128/JB.185.1.210-220.2003.
79. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009; 10:R25; PMID:19261174; http://dx.doi.org/10.1186/gb-2009-10-3-r25.
80. Takemasa R, Yokooji Y, Yamatsu A, Atomi H, Imanaka T. Thermococcus kodakarensis as a host for gene expression and protein secretion. Appl Environ Microbiol 2011; 77:2392-8; PMID:21278271; http://dx.doi.org/10.1128/AEM.01005-10.
81. Yokooji Y, Tomita H, Atomi H, Imanaka T. Pantoate kinase and phosphopantothenate synthetase, two novel enzymes necessary for CoA biosynthesis in the Archaea. J Biol Chem 2009; 284:28137-45; PMID:19666462; http://dx.doi.org/10.1074/jbc.M109.009696.
82. Endoh T, Kanai T, Sato YT, Liu DV, Yoshikawa K, Atomi H, et al. Cell-free protein synthesis at high temperatures using the lysate of a hyperthermophile. J Biotechnol 2006; 126:186-95; PMID:16730835; http://dx.doi.org/10.1016/j.jbiotec.2006.04.010.
83. Chan PP, Holmes AD, Smith AM, Tran D, Lowe TM. The UCSC Archaeal Genome Browser: 2012 update. Nucleic Acids Res 2012; 40(Database issue):D646-52; PMID:22080555; http://dx.doi.org/10.1093/nar/gkr990.