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

Joshua R. Elmore,† Yuusuke Yokooji,‡ Takaaki Sato,§ Sara Olson,¶ Claiborne V.C. Glover III,¶ Brenton R. Graveley,¶ Haruyuki Atomi,‡ Rebecca M. Terns,** and Michael P. Terns†,‡,§,*

†Department of Molecular Biology and Biochemistry; University of Georgia; Athens, GA U.S.A.; ‡Department of Genetics; University of Georgia; Athens, GA U.S.A.; §Department of Microbiology; University of Georgia; Athens, GA U.S.A.; ¶Department of Synthetic Chemistry and Biological Chemistry; Graduate School of Engineering; Kyoto University; Katsura, Nishikyo-ku, Kyoto, Japan; ‡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 kodakarense* (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. 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"). Transcription of CRISPR loci and subsequent processing of primary CRISPR transcripts produce mature CRISPR (cr)RNAs that each contain an invader-derivable (guide) sequence flanked by various amounts of the CRISPR repeat sequence at the 5′ and/or 3′ end. 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.

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. At least 10 unique sets of Cas protein genes are found among prokaryotic genomes, forming distinct CRISPR-Cas systems. 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. 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. 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. The Type III-B Cmr system cleaves invading target RNAs.

*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
CRISPRs are transcribed from a leader sequence of ~100–500 base-pairs found directly upstream of the CRISPR. Putative promoters have been identified within the leader sequence in several species. 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. This species is used directly by Type I-E Cse, I-F Csy and some I-A Csa systems, but is further processed at the 3’ end in other systems. In Type I-C Csd systems, Cas5d cleaves CRISPR transcripts 11 nts upstream of the guide sequence. In Type II Csm 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). 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. Production of these crRNAs also requires Cas1, Cas2 and two Cas6 proteins. 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.

**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 interspersed 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. 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.

The *P. furiosus* repeat RNA is unstructured in solution and specifically recognized and cleaved by the Cas6 endonuclease (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 Tnej; found in a subset of both archaea and bacteria). Between the Type I-A and Type I-B *cas* genes, there 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). 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. 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.
in P. furiosus as a result of site-specific endonucleolytic Cas6 cleavage of CRISPR transcripts at this site; 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 with 22 ~60 nts downstream of the 5’ tag corresponds with the expected repeat sequence at the 3’ end. The potential cluster of 3’ ends (35–37 nts of a typical 38-nt guide) and no 3’ repeat sequence crRNA species typically includes nearly all of the guide sequence (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 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). 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.
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 protospece**r. 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 protospececr 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 *chiA*Δ*Δ* gene, which encodes the C-terminal endochitinase domain of *Tko* chitinase65 (Fig. 7A and B). Strain KC2 is derived from *Tko* KUW1 (Δ*pyrF* Δ*trpE*), and lacks the entire *chiA* gene, *chiA*. Northern analysis indicates that anti-*chiA* crRNAs A and E

*www.landesbioscience.com*  

**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 archael 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.
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 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.
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 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.
Features of Tko crRNAs. crRNAs guide the CRISPR-Cas immune response and vary in composition among CRISPR-Cas pathways.\textsuperscript{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).\textsuperscript{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.\textsuperscript{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.\textsuperscript{22,38} This 5’ tag is a direct product of cleavage of the CRISPR transcripts by Cas6 or Cas6-like endonucleases,\textsuperscript{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\textsuperscript{68–70} and upon recruitment to target DNA via crRNA-protein complex (crRNP) interactions in vivo.\textsuperscript{13} 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 Cst5 (Cas5 superfamily) proteins (in addition to the universal Cas proteins Cas1 and Cas2). The Cst system (being studied in Tko (this study) and P. furiosus (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. mari- paludis).\textsuperscript{35,47} Evidence suggests that the Csh system of H. volcanii can target invading plasmids at the DNA level,\textsuperscript{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\textsuperscript{49})
and in vivo [in Sulfolobus solfataricus and Pyrococcus furiosus (our unpublished results)]. Evidence indicates that Csa2 (Cas7 superfamily) and Csa5a (Cas5 superfamily) may be central components that play a role in crRNA binding and DNA binding.9 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, Csa9/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.54,75 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 crRNAP 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.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

**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 in complex growth medium (ASW-AA). 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-S0) for competitive interference, or with 2 mL/L polysulfide solution (10 g of Na2S 9H2O and 3 g of sulfur flowers in 15 ml of H2O) for 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.

![Diagram of CRISPR-Cas interference](image)
recovery after transformation before plating. Preparation of solid medium and cultivation on plates was performed as previously described. 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 electrophoresis 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) to 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 MgCl2 (Roche). Details of procedure have been described previously.

**Illumina sequencing 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. 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 a ChiA) using Trizol (Invitrogen). Ten μg of total RNA samples were separated on 15% TBE-Urea polyacrylamide gels (Criterion, Bio-Rad) beside ∆g of total RNA samples were separated on 15% TBE-Urea polyacrylamide gels (Criterion, Bio-Rad). Membranes were baked at 80°C for 1 h before prehybridization in a prehybridization buffer containing 5x SSC, 7% SDS, 20 mM NaPO₄ (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-[32P]-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 SSC, 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 pLC64.64 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.64 A 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 of 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 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 phenotypic traits. 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 Cabonova and John Reeve (Ohio State University) for sharing technical assistance and helpful advice regarding T. kodakarensis. We thank Anisha Hedge and Megan Tibbits (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
12. Swarts DC, Mosterod C, van Pasell 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 adaption process in Escherichia coli. Nuclear 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, Pluci ski MM, Defen VJ, Thomas BC, Horvath P, et al. Percolating viral sequences shape microbial CRISPR-based immunity. PLoS Comput Biol 2012; 8:e1002475; 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 RNA guide nucleases guide defense in prokaryotes. Science 2008; 321:560-4; PMID:18703739; http://dx.doi.org/10.1126/science.1152609.
17. Carre J, Wang R, Li H, Terns RM, Terns MP. Cas9 is an endonuclease that generates guide RNA for invader defense in prokaryotes. Dev Genes 2008; 22:3489-96; PMID:19141480; http://dx.doi.org/10.1016/j. gait.1742988.
18. Carre J, Pfeifer NT, Compton MM, Terns RM, Terns MP. Binding and cleavage by CRISPR RNA by Cas9. RNA 2010; 16:2181-8; PMID:20888478; http://dx.doi.org/10.1261/rna.229101.
19. Haurwitz RE, Jinek M, Wiedenheft B, Zhou K, Doudna JA. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. Science 2010; 330:1355-8; PMID:20829488; http://dx.doi.org/10.1126/science.1192277.
20. Harout-Ayan A, Maniv I, Marraffini LA. Nature clustered, regularly interspaced, short palindromic repeats (CRISPR)-associated 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:21660969; http://dx.doi.org/10.1073/pnas.1112823108.
21. Garisde EL, Schellenberg MJ, Genev EM, Bonanno JB, Saurier JM, Butley SK, et al. Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonuclease. RNA 2012; 18:2020-8; PMID:23006625; http://dx.doi.org/10.1016/j. rna.033100.112.
22. Hale CR, Majumdar S, Elmore J, Pfeifer N, Compton M, Olson S, et al. Essential features and rational design of CRISPR RNAs that function with the Cas3-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.

Supplemental Material
Supplemental material may be found here: www.landesbioscience.com/journals/rnabiology/article/24084
detect multiple protospacer adjacent motifs (PAMs) to guide of 45 CRISPR-associated (Cas) protein families. Proc Natl Acad Sci USA 2011; 108:10092-7; PMID:21536913; http://dx.doi.org/10.1111/j.1365-2958.2010.07452.x. 4. Backofen R, Randau L. Characterization of CRISPR RNAs: recognition and cleavage. Structure 2010; 18:688-92; PMID:21572444; http://dx.doi.org/10.1016/j.str.2010.04.006. 5. Richter H, Zoepf J, Schermuly J, Maticic D, Backofen R, Saul L. Characterization of CRISPR RNA processing in Clostridium thermocellum and Methanococcus maripaludis. Nucleic Acids Res 2012; 40:9878-96; PMID:22877397; http://dx.doi.org/10.1093/nar/gkt737. 6. Nam KH, Harajima 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/D TypeVIII CRISPR-Cas system. Structure 2012; 20:1574-84; PMID:22844129; http://dx.doi.org/10.1016/j.str.2012.06.016. 7. Cady KC, Bondy-Denomy J, Heusler 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:22882597; http://dx.doi.org/10.1128/JB.01844-12. 8. Deveau H, Barrangou R, Garneau JE, Labonté J, Fremeaux B, 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. 9. Marraffini LA, Sontheimer EJ. Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature 2010; 465:568-71; PMID:20702129; http://dx.doi.org/10.1038/nature09070. 10. 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/micro.0.03960-0. 11. Motokawa T, Iwasa Y, Rashid N, Hoaki T, Imanaka T. Purification and characterization of a thermostable thiold protease from a newly isolated hyperthermo-phile Pyrococcus sp. Appl Environ Microbiol 1994; 60:4556-66; PMID:7684259; http://dx.doi.org/10.1128/AEM.60.10.4556-4566. 12. Gissi L, Vergnaud G, Pourcel C. 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. 13. Rouxset C, Genta R, Le Romancer M, Nicolas J, CRISPR: a CRISPR interactive database. Bioinformatics 2009; 25:3317-8; PMID:19846435; http://dx.doi.org/10.1093/bioinformatics/btp586. 14. Kunin V, Sorek R, Hugenholtz P. Evolutionary conser-vation 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. 15. Wang R, Prewpmlue G, Terns MP, Terns RM, Li H. Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. Structure 2011; 19:257-64; PMID:21100293; http://dx.doi.org/10.1016/j.str.2011.01.014. 16. 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:10520030; http://dx.doi.org/10.1073/pnas.96.24.13668. 17. Randau L. RNA processing in the minimal organism Nanocystis eugametos. Genome Biol 2012; 13:R63; PMID:22809431; http://dx.doi.org/10.1186/gb-2012-13-7-r63.
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.C111.379503.

74. Westra ER, Nilges B, van der Oost J, Dame RT, Boumans 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, Pal U, Heidrich N, Jore MM, Lundgren M, Sträumann 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. Takeda R, Yokooji Y, Yamatsu A, Atomi H, Imanaka T. Thermococcus kodakaraensis as a host for gene expression and protein secretion. Appl Environ Microbiol 2011; 77:2592-8; 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; 284:28157-45; 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:16730855; 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.