Supporting Online Material for

Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes

Stan J. J. Brouns, Matthijs M. Jore, Magnus Lundgren, Edze R. Westra, Rik J. H. Slijkhuis, Ambrosius P. L. Snijders, Mark J. Dickman, Kira S. Makarova, Eugene V. Koonin, John van der Oost*

*To whom correspondence should be addressed. E-mail: john.vanderoost@wur.nl

Published 15 August 2008, *Science* **321**, 960 (2008)
DOI: 10.1126/science.1159689

This PDF file includes:

Materials and Methods

Strains

E. coli K12 gene knockouts from the KEIO collection were kindly provided by the National BioResource Project (NBRP, NIG, Japan) (Table S2).

Gene cloning, protein production and purification

The *cas* genes and CRISPRs were PCR amplified from *E. coli* K12 W3110 (BW25113) genomic DNA (genomic locus: 2885900-2876300 nt), and directionally cloned into a compatible expression vector set consisting of pET-52b (Amp^R, 100 μg/ml), pCDF-1b (Str^R, 50 μg/ml), pRSF-1b (Kan^R, 50 μg/ml) and pACYCduet-1 (Cam^R, 34 μg/ml) (Novagen), or pIH1119 (Amp^R) (New England Biolabs) as indicated in Table S2 and S3. Primers for the *cas* genes were designed based on their predicted translation start sites according to their latest annotation (*S1*) (Table S2). Mutations were introduced using the QuikChange site-directed mutagenesis kit (Stratagene) (Table S2). Plasmids were transformed into *E. coli* BL21(DE3) (Novagen) which lacks endogenous *cas* genes, or *E. coli* DH5α in the case of pIH1119. Protein production was induced with 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) (Invitrogen) at an optical cell density at 600 nm of 0.6 for 16 h at 37 °C. Cells were harvested, resuspended in 20 mM Tris-HCl (pH 8.0) supplemented with 0.1 M NaCl, and disrupted using a French Pressure Cell. Tagged proteins were isolated using Strep-Tactin (IBA, Germany), HIS-Select (Sigma-Aldrich), or Amylose (New England Biolabs) affinity chromatography following manufacturer's

instructions. The identity of the proteins was determined by mass spectrometry, as described (*S2*).

Northern blotting

Total RNA from 11 ml of exponentially grown *E. coli* cells was isolated using the *mir*Vana miRNA isolation kit (Ambion). Strain *E. coli* BL21(DE3) containing various plasmid combinations (Table S3) was grown without IPTG induction, which resulted in low expression levels of the *cas* genes and CRISPR due to leakage of the expression system. Northern blots were performed by running 10 μg of RNA on a 9% polyacrylamide gels with 7 M urea in 0.5x TBE buffer (*S3*). The RNA was then transferred to a Genescreen Plus membrane (PerkinElmer) by semi-dry blotting using a Trans-blot SD (Bio-Rad). After 1 min of UV-crosslinking and baking (80 °C, 15 min), the membrane was hybridized with QuikHyb (Stratagene) at 42 °C. Blots were probed for 12 hours with a ^{32}P-5'-labelled DNA oligonucleotide of spacer 4 in the *E. coli* K12 CRISPR (oligonucleotide BG2349, Table S2). The blots were subsequently washed with 2x SSC buffer (*S3*) containing 0.1% SDS for 30 min, and 0.1x SSC buffer containing 0.1% SDS for 30 min. Blots were visualized using phosphorimaging with a Personal FX phosphorimager (Bio-Rad). RNA sizes were estimated by comparison with ^{32}P-labeled Decade RNA marker (Ambion).

Cleavage reactions

Internally radiolabelled transcripts were generated by *in vitro* transcription using the MAXIscript T7 kit (Ambion) and α-^{32}P-UTP (GE) (Table S2). Templates for *in vitro* transcription were generated by PCR using primers BG2559 and BG2374 for the *E. coli* K12 CRISPR, BG2462 and BG2463 for *E. coli* UTI89, and BG2452 and BG2461 for the non-CRISPR. Full-length RNA substrates were gel-isolated from denaturing 2% agarose gels as described (*S4*). Cleavage reactions were set up at 37 °C in 20 mM Tris-HCl (pH 8.0) supplemented with 0.1 M NaCl and 1 mM EDTA. Protein samples (Table S3) were treated with 10 mM EDTA prior to the cleavage assay. Assays were started by adding 0.3 μg of Cascade, or 0.1 μg of MalE-CasE to a reaction containing 10 ng of gel-purified internally ^{32}P labeled transcripts in a total volume of 10 μl. Samples were treated with 1 U of proteinase K for 5 min at 37 °C (Fluka) and acid-phenol extracted as described (*S5*). Reaction products were analyzed using 8% polyacrylamide gels containing 7 M urea.

RNA cloning

Protein-bound total RNA was isolated from Strep-Tactin-purified Cascade (Table S3) using the *mir*Vana miRNA isolation kit (Ambion). Approximately 6 μg of RNA was denatured for 10 min at 65 °C, and subsequently 3' polyadenylated for 75 min at 41 °C using Poly(A) polymerase (Ambion) as described (*S6*). RNA products were separated from unincorporated nucleotides and enzymes using a NucAway gel filtration spin column (Ambion), and reverse transcribed with the Superscript III kit (Invitrogen) for 50 min at 50 °C with anchored primer BG2164 (5'-GCCCGCCC**GGATCC**TTTTTTTTT-

TTTTTTTTTTTTTTTTTTTTTVN-3') (BamHI site in bold face) (*S6*). The RNA strand was degraded by RNase A (Fermentas) and RNase HI (Promega) for 15 min at 37 °C. Single-stranded cDNA was purified using the MinElute PCR purification kit (Qiagen). A 3' poly(dG) tail was added to the cDNA by 60 min incubation with Terminal Deoxynucleotidyl Transferase (Invitrogen) at 37 °C, followed by purification using MinElute reaction cleanup kit (Qiagen). The single stranded cDNA was used as a template in a PCR using Native Pfu polymerase (Stratagene) and primers BG2220 (5'-GCGCCCGC**GGATCC**CCCCCCCCDN-3') and BG2222 (5'-GCCCGCCC**GGATCC**-TT-3'). The PCR products were cloned into vector pUC29 and transformed into *E. coli* NEB5α (New England Biolabs).

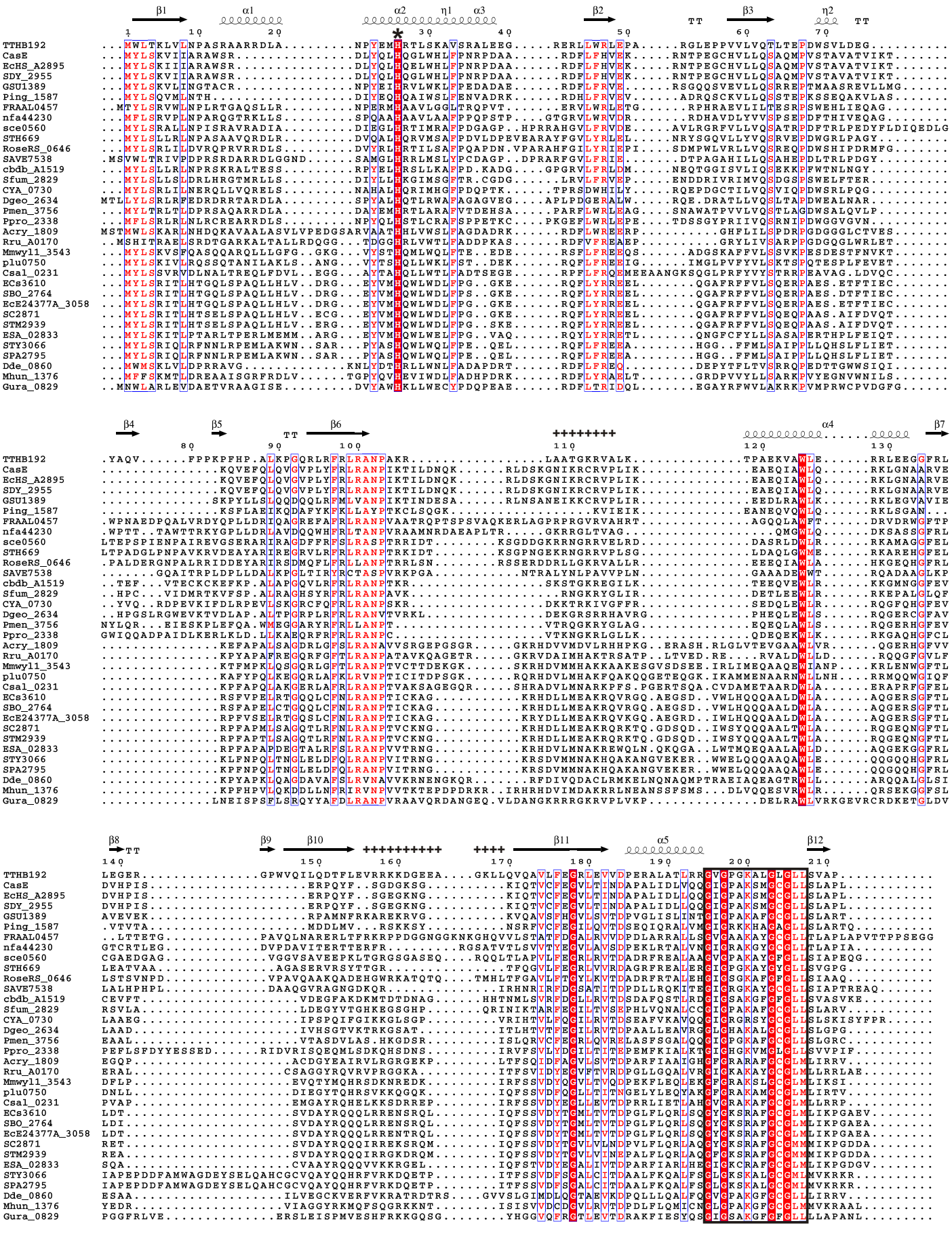
Phage studies

Host sensitivity to phages was tested using a virulent variant of phage Lambda (λ_{vir}) (*S7*) obtained from Centraalbureau voor Schimmelcultures (Utrecht, Netherlands) and *E. coli* BL21-AI (Invitrogen) as a host (Table S2 and S3). Strains were grown in 2YT-Lambda (2YTL) media (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, 10 mM $MgSO_4$, 0.2 % maltose) until the optical density ($OD_{600\ nm}$) reached 0.3. Cas protein and pre-crRNA production was then induced for 30 min by adding a final concentration of 0.2 % L-arabinose (Sigma-Aldrich) and 0.1 mM IPTG. Cells were spun down and resuspended in 10 mM $MgSO_4$, before being used in plaque assays according to standard procedures (*S3*). Plaque assays were performed in triplicate. Plates and top-agar contained 2YTL and above mentioned concentrations of inducers. The sensitivity of the host to phage

infection was calculated as the efficiency of plaquing (*S8*), which is the plaque count ratio of a strain containing an anti-λ_{vir} CRISPR to that of a strain containing a CRISPR with non-targeting spacers (N). Error-bars were calculated as one standard deviation. Anti-λ CRISPRs were designed by randomly picking proto-spacer sequences in four genes of the λ genome (Fig.S3). The artificial anti-λ CRISPR design did not take any *S. thermophilus* CRISPR motifs into account (*S9, S10*). The motifs are conserved nucleotide sequences located downstream of proto-spacers on the virus genome, which are important for the phage resistant phenotype in *S. thermophilus*. No CRISPR motif could be identified for the *E. coli* K12 CRISPR/*cas* system using the flanking regions of the four known proto-spacers in phage P1 and plasmid F (*S11*). The anti-λ CRISPRs were synthesized by Geneart AG (Regensburg, Germany), and subcloned into vector pACYCduet-1 vector (Novagen) ($C_{1\text{-}4}$, $T_{1\text{-}4}$, Table S2) using restriction sites NcoI and Acc65I. CRISPRs with single targeting spacers (C_1, C_2, C_3, C_4, T_1, T_2, T_3, T_4) were obtained by exchange of single non-targeting spacers of the N CRISPR with the corresponding single targeting spacers of $C_{1\text{-}4}$ and $T_{1\text{-}4}$ CRISPRs using restriction enzyme pairs NcoI and EcoRI, EcoRI and BamHI, BamHI and NsiI, NsiI and Acc65I (Fig.S3). The CRISPR sequences are provided in Figure S4.

Figure S1

A


TTHB192
CasE
EcHS_A2895
SDY_2955
GSU1389
Ping_1587
FRAAL0457
nfa44230
sce0560
STH669
RoseRS_0646
SAVE7538
cbdb_A1519
Sfum_2829
CYA_0730
Dgeo_2634
Pmen_3756
Ppro_2338
Acry_1809
Rru_A0170
Mmwyl1_3543
plu0750
Csal_0231
ECs3610
SBO_2764
EcE24377A_3058
SC2871
STM2939
ESA_02833
STY3066
SPA2795
Dde_0860
Mhun_1376
Gura_0829

B

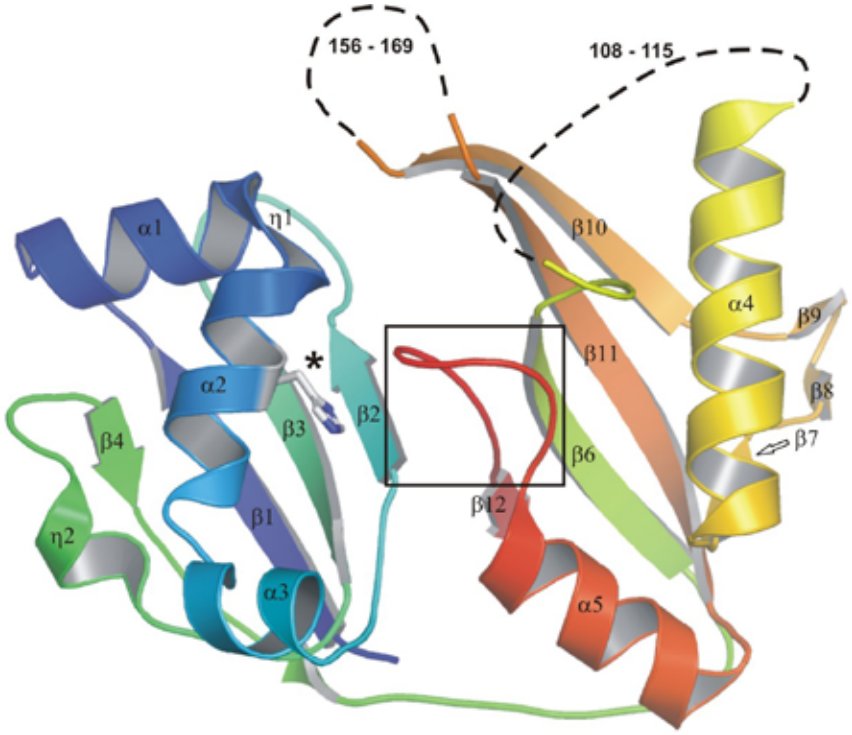


Figure S1

A. Multiple sequence alignment of CasE homologs. Sequences were aligned with TCoffee (*S12*) and aligned to the structure of TTHB192 (PDB ID: 1WJ9) (*S13*) using ESPript (*S14*). TTHB192, *Thermus thermophilus* HB8; CasE, *Escherichia coli* K12 W3110; EcHS_A2895, *E. coli* HS; SDY_2955, *Shigella dysenteriae* Sd197; GSU1389, *Geobacter sulfurreducens* PCA; Ping_1587, *Psychromonas ingrahamii* 37; FRAAL0457, *Frankia alni* ACN14a; nfa44230, *Nocardia farcinica* IFM 10152; sce0560, *Sorangium cellulosum* 'So ce 56'; STH669, *Symbiobacterium thermophilum* IAM 14863; RoseRS_0646, *Roseiflexus* sp. RS-1; SAVE7538, *Streptomyces avermitilis* MA-4680; cbdb_A1519, *Dehalococcoides* sp. CBDB1; Sfum_2829, *Syntrophobacter fumaroxidans* MPOB; CYA_0730, *Synechococcus* sp. JA-3-3Ab; Dgeo_2534, *Deinococcus geothermalis* DSM 11300; Pmen_3756, *Pseudomonas mendocina* ymp; Ppro_2338, *Pelobacter propionicus* DSM 2379; Acry_1809, *Acidiphilium cryptum* JF-5; Rru_A0170, *Rhodospirillum rubrum* ATCC 11170; Mmwyl1_3543, *Marinomonas* sp. MWYL1; plu0750, *Photorabdus luminescens* subsp. laumondii TT01; Csal_0231, *Chromohalobacter salexigens* DSM 3043; Ecs3610, *E. coli* 0157:H7 str. Sakai; SBO_2764, *Shigella boydii* Sb227; EcE24377A_3058, *E. coli* E24377A; SC2871, *Salmonella enterica* subsp. Enterica serovar Choleraesuis str. SC-B67; STM2939, *Salmonella typhimurium* LT2; ESA_02833, *Enterobacter sakazakii* ATCC BAA-894; STY3066, *S. enterica* subsp. Enterica serovar Typhi str. CT18; SPA2795, *S. enterica* subsp. Enterica serovar Paratyphi A str. ATCC9150; Dde_0860, *Desulfovivrio desulfuricans* G20; Mhun_1376, *Methanospirillum hungatei* JF-1; Gura_0829, *Geobacter uraniireducens* Rf4. Secondary structural elements and amino acid numbering follows the TTHB192 structure and sequence. Disordered loops are indicated with a plus (+), and the highly conserved residue His26 is marked with an asterisk (*). The highly conserved C-terminal glycine-rich loop, which is the hallmark of this protein family, is highlighted with a box.

B. Ribbon diagram of the structure of TTHB192, a CasE homolog from *Thermus thermophilus* HB8 (PDB ID: 1WJ9) (*S13*). Structural features are indicated as in **A**. Structurally disordered residues 108 to 115 and 156 to 169 are depicted by dashed lines. Note that the highly conserved glycine-rich loop between secondary structure elements α5 and β12 is spatially close to His26.

Figure S2

Sequences of Cascade-bound RNA.

>RNA1-SP6
ATAAACCGTCAGCTTTATAAATCCGGAGATACGGAAACTAGAGTTCCCCG

>RNA2_D-TAGATOSE-1,6-DISPHOSPHATE ALDOLASE GATY
TTGTACGTGGTATCGACAAAGCAGATGCTGAACAACGCACAGCGCGGCGGTTATG

>RNA3-SP3
ATAAACCGGGCTCCCTGTCGGTTGTAATTGATAATGTTGAGAGTTCCCCG

>RNA4-SP3
ATAAACCGGGCTCCCTGTCGGTTGTAATTGATAATGTTGAGAGTTCCCC

>RNA5-SP2
ATAAACCGCAGCCGAAGCCAAAGGTGATGCCGAACACGCTGAGTTCCCCG

>RNA6-SP5
ATAAACCGCGAATCGCGCATACCCTGCGCGTCGCCGCCTGCGAGTTCCCCG

>RNA7-SP8
ATAAACCGCTGCTGGAGCTGGCTGCAAGGT

>RNA8-SP3
ATAAACCGGGCTCCCTGTCGGTTGTAATTGATAATGTTGAGAGTTCCCCGCGCCAGCG

>RNA9-SP3
ATAAACCGGGCTCCCTGTCGGTTGTAATTGATAATGTTGAGAGTTCCCCG

>RNA10-SP8
ATAAACCGCTGCTGGAGCTGGCTGCAAG

>RNA11-SP2
ATAAACCGCAGCCGAAGCCAAAGGTGATGCCGAACACGCTGAGTTCCCCG

>RNA12-SP8
ATAAACCGCTGCTGGAGCTGGCTG

>RNA13-K12-2876563-2876522, INTERGENIC
ATAAGGAAATGTTACATTAAGGTTGGTGGGTTGTTTTTATGG

>RNA14-SP1 EN 2
ATAAACCGCTTTCGCAGACGCGCGGCGATACGCTCACGCAGAGTTCCCCGCGCCAGCGGGGA
TAAACCGCAGCCGAAGCCAAAGGTGATGCCGAACACGCTGAGTTCCCCG

>RNA15-16S
AAAACTGGAGGAAGGTGGGGATGACGTCAAGTCATCATG

>RNA16-SP4

ATAAACCGTTTGGATCGGGTCTGGAATTTCTGAGCGGTCGCGAGTTCCCCGCGCCAGC

>RNA17-SP2
CCGAAGCCAAAGGTGATGCCGAACACGCTGAGTTCCCCG

>RNA18-SP4
ATAAACCGTTTGGATCGGGTCTGGAATTTCTGAGCGGTCGCGAGTTCCCC

>RNA19-23S
TCGCGGATGGAGCTGG

>RNA20-SP2
ATAAACCGCAGCCGAAGCCAAAGGTGATGCCGAACACGCTGAGTTCCCTG

>RNA21-SP8
ATAAACCGCTGCTGGAGCTGGCTGC

>RNA22-16S
AAATTGAAGAGTTTGATCATGGCTCAGATTGAACGCTGGCGGCAGGCCTAACACATGCAAGT
CGAACGGTAACAG

>RNA23-SP4
ATAAACCGTTTGGATCGGGTCTGGAATTTCTGAGCGGTCGCGAGTTCCCCGCGCCAGCGG

>RNA24-INTERGENIC K12 2876539-2876477
GGTGGGTTGTTTTTATGGGAAAAAATGCTTTAAGAACAAATGTATACTTTTAGAGAGTTCCC
C

>RNA25-SP2
AGCGGGGATAAACCGCAGCCGAAGCCAAAGGTGATGCCGAACACGCTGAGTTCCCCGCGCCA
GCGG

>RNA26-LEADER-REPEAT ONE
TGGTGGGTTGTTTTTATGGGAAAAAATGCTTTAAGAACAAATGTATACTTTTAGAGAGTTCC
CCGT

>RNA27-YGCL
TATCGTAATAATCAAGCATCTATTCTTGAACGGCGTCATGATGTGTTGATGT

>RNA28-YGCL
ACATCAACACATCATGACGCCGTTCAAGAATAGATGCTTGATTATTACGATA

>RNA29-INTERGENIC-K12-2876396-2876357
GCAGCCGAAGCCAAAGGTGATGCCGAACACGCTGAGTTCC

>RNA30-PROQ PUTATIVE SOLUTE/DNA COMPETENCE EFFECTOR
AATAGCAGTAAAGAAGTAATCGCGTTTCTGGCCGAACGTTTTCCCCACTGTTTCAGTGCGGA
AGGTGAAGCGCGTCCGCTG

>RNA31-SP2
ATAAACCGCAGCCGAAGCCAAAGGTGATGCCGAACACGCTGAGTTCCCCG

>RNA32-SP1
ATAAACCGCTTTCGCAGACGCGCGGCGATACGCTCACGCAGAGTTCC

>RNA33-SP8
ATAAACCGCTGCTGGAGCTGGCTGCAAGG

>RNA34-SP4
ATAAACCGTTTGGATCGGTTTTGGAATTTTTGATCGTTTCCGAGTTCCCCG

>RNA35-SP5
ATAAACCGCGAATCGCGCATACCCTGCGCGTCGCCGCCTGCGAGTTCCCCGCGC

>RNA36-SP7
GATAAACCGGACTCACCCCGAAAGAGATTGCCAGCCAGCTTGAGTT

>RNA37-SP8
ATAAACCGCTGCTGGAGCTGGCTGCAAGGCAAGC

>RNA38-16S
ACGACCAGGGCTACACACGTGCTACAATGGCGCATACAAAGAGAAGCGACCTCGCGAGAGCA
AGCGGACCTCATAAAGTGCGTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTC
GGAATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAATACGT

>RNA39-23S
AACTTCGGGAGAAGGCACGCTGATATGTAGGT

>RNA40-SP1
ATAAACCGCTTTCGCAGACGCGCGGCGATACGCTCACGCAGAGTTCC

>RNA41-SP4
ATAAACCGTTTGGATCGGGTCTGGAATTTCTGAGCGGTCGCGAGTTC

>RNA42-SP4
ATAAACCGTTTGGATCGGGTCTGGAATTTCTGAGC

>RNA43-SP4
ATAAACCGTTTGGATCGGGTCTGGAATTTCTGAGCGGTCGCGAGTTCCCCG

>RNA44-LACI
TTCTCCCATGAAGACGGTACGCGACTGGGCGTGGAGCAC

>RNA45-SP2
ATAAACCGCAGCCGAAGCCAAAGGTGATGCCGAACACGCTGAGTTCCCCG

>RNA46-SP5
ATAAACCGGGAATCGCGCATACCCTGCGCGTCGCCGCCTGCGAGTTCCCCG

>RNA47-SP4
ATAAACCGTTTGGATCGGGTCTGGAATTTCTGAGCGGTCGCGAGTTCCCCG

>RNA48-SP1
ATAAACCGCTTTCGCAGACGCGCGGCGATACGCTCACGCAGAGTTCCCC

>RNA49-SP5
ATAAACCGCGAATCGCGCATACCCTGCGCGTCGCCGCCTGCGAGTTCCCCG

>RNA50-SP8
ATAAACCGCTGCTGGAGCTGGCTGCAAGG

>RNA51-SP2
ATAAACCGCAGCCGAAGCCAAAGGTGATGCCGAACACGCTGAGTTCCCCG

>RNA52-SP7
ATAAACCGGACTCACCCCGAAAGAGATTGCCAGCCAGCTTGAGTTCCC

>RNA53-SP2
TAAACCGCAGCGGAAGCCAAAGGTGATGCGGAACACGCTGAGTTCCC

>RNA54-SP4
ATAAACCGTTGGGATCGGTTCTGGAATTTTTGAGCGGTCGGGAGTTCCCC

>RNA55-SP2
ATAAACCGCAGCGGAAGCCAAAGGTGATGCCGAACACGTTCGAGTTCCCC

>RNA56-THR tRNA
GTAATGCGAAGGTCGTAGGTTCGACTCC

>RNA57-SP4
ATAAACCGTTTGGATCGGGTCTGGAATTTCTGAGCGGTCGCGAGTTCCCCGCGCCAGCG

>RNA58-SP4
ATAAACCGTTTGGATCGGGTCTGGAATTTCTGAGCGGTCGCGAGTTCCCCG

>RNA59-YGBF
TGTATCCGCAAAAATTCGTGAAATGATCTGGGAACAAATAGCTGGACTGGCGGAAGGGGGC

>RNA60-SP8
AAACCGCTGCTGGAGCTGGCT

>RNA61-23S
TCAAGGTTGAGGCGTGATG

>RNA62-SP8
ATAAACCGCTGCTGGAGCTGGCTGCAAGGCAAGCCGCCC

>RNA63-SP7
ATAAACCGGACTCACCCCGAAAGAGATTGCCAGCCAGCTTGAGTTC

>RNA64-SP4
ATAAACCGTTTGGATCGGTTCTGGAATTTTTGAGCGGTCGGGAGTTCCCCGCGC

>RNA65-YGCH
ATCTTGTACAGCAAGGTATTGGGCCAGCTAAATCGATGGGATGTGGCTTGGTATCTTT

>RNA66-SP1
ATAAACCGCTTTCGCAGACGCGCGGCGATACGCTCACGCAGAGTTCCCCG

>RNA67-SP8
ATAAACCGCTGCTGGAGCTGGCTGCAAG

>RNA68-SP8
ATAAACCGCTGCTGGAGCTGGCTGCAAG

>RNA69-SP4
GTCTGGAATTTTTGAGCGGTCGCGAGTTCCCCG

>RNA70-LEADER
ATAAGGAAATGTTACATTAAGGTTGGTGGGTTGTTTTTATGGGAAAAAATGCTTTAAGAAC

>RNA71-SP8
ATAAACCGCTGCTGGAGCTGGCTG

>RNA72-SP3
ATGTTGAGAGTTCCCC

>RNA73-SP4
ATAAACCGTTTGGATCGGGTCTGGAATTTCTGAGCGGTCGCAAGTTCCCCG

>RNA74-SP3
GGGCTCCCTGTCGGTTGTAATTGATAATGTTGAGAGTTCCCCGCGCC

>RNA75-SP2
ATAAACCGCAGCCGAAGCCAAAGGTGATGCCGAACACGCTGAGTTCCCC

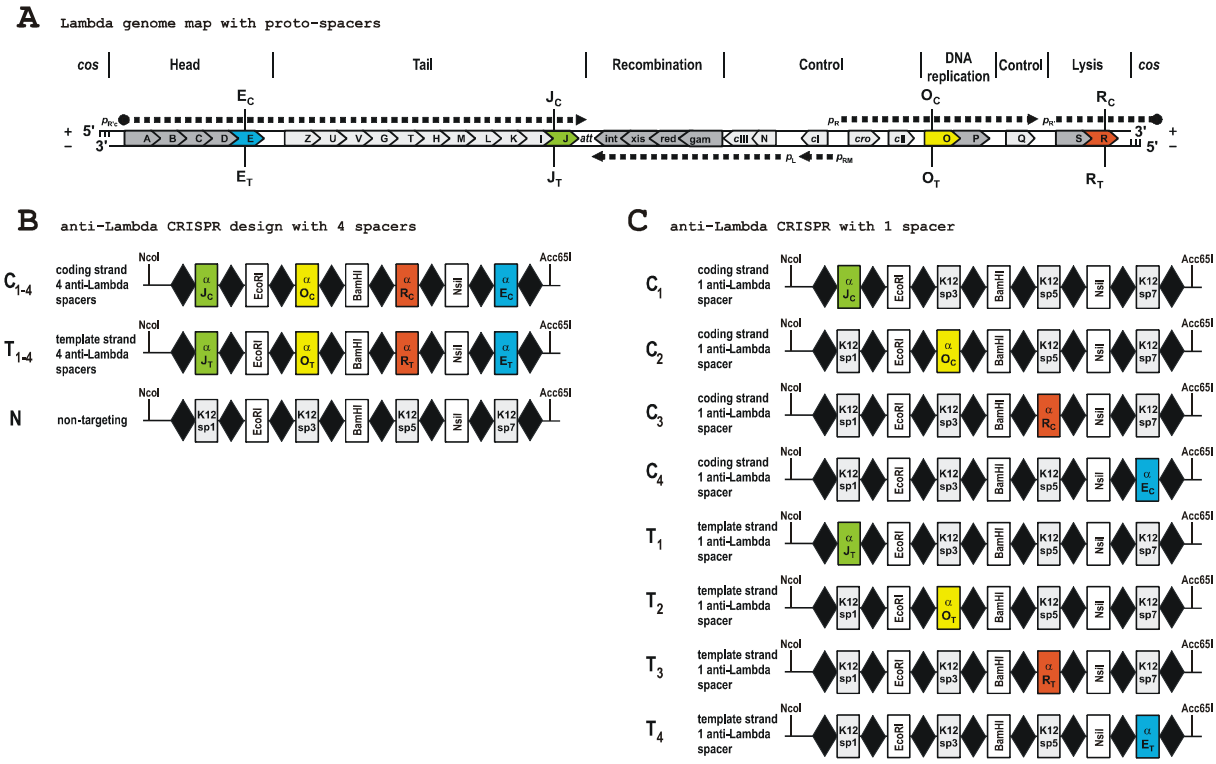
>RNA76-SP4
ATAAACCGTTTGGATCGGTTCTGGAATTTTTGAGCGGTCGCGAGTTCCCC

>RNA77-SP2
ATAAACCGCAGCGGAAGCCAAAGGTGATGCCGAACACGCTGAGTTCCCC

>RNA78-SP1
ATAAACCGCTTTCGCAGACGCGCGGCGATACGCTCACGCAGAGTTCCCC

>RNA79-SP4
ATAAACCGTTTGGATCGGGTCTGGAATTTCAGAGCGGTCGCGAGTTCCCCGCG

Figure S3



A Phage λ genome map indicating the main genes and transcripts (dotted arrows), and the positions of the proto-spacers on the coding or template strand. **B** Design of two anti-λ CRISPRs (repeats: diamonds, spacers: rectangles) producing crRNAs complementary to (indicated by α) either the coding strand ($C_{1\text{-}4}$) of the four genes J, O, R and E (*i.e.* mRNA and plus strand of the viral genome), or the template strand ($T_{1\text{-}4}$) of these four genes (*i.e.* minus strand). A third CRISPR (N) was designed as a non-targeting control containing the naturally occurring spacers 1, 3, 5, and 7 from *E. coli* K12, which have no homology to any known phage. The number of plaque forming units obtained in the presence of this CRISPR was used to calculate the efficiency of plaquing (Fig.4). Restriction sites were introduced in spacer 2, 4 and 6 (EcoRI, BamHI and NsiI, respectively). **C** CRISPRs with single targeting spacers (C_1, C_2, C_3, C_4, T_1, T_2, T_3, T_4) were obtained by exchange of single non-targeting spacers of the N CRISPR with the corresponding single targeting spacers of $C_{1\text{-}4}$ and $T_{1\text{-}4}$ CRISPRs using restriction enzyme pairs NcoI and EcoRI, EcoRI and BamHI, BamHI and NsiI, NsiI and Acc65I.

Figure S4

Sequences of the CRISPRs used in this study

Coding CRISPR (C_{1-4}), pWUR479

GGCGCGCCATGGAAACAAAGAATTAGCTGATCTTTAATAATAAGGAAATGTTACATTAAGGTTGGTGGGTTGTTT
TTATGGGAAAAAATGCTTTAAGAACAAATGTATACTTTTAGAGAGTTCCCCGCGCCAGCGGGGATAAACCGGGCC
CTTCGCTGATGGCATCGATCACACTCAGGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCCGAAGCCAAAGAAT
TCGCCGAACACGCTGAGTTCCCCGCGCCAGCGGGGATAAACCGGTTTGCTGGCTGTCACGCCTGCCTGTTGCTTG
GAGTTCCCCGCGCCAGCGGGGATAAACCGTTTGGATCGGGTCTGGATCCTCTGAGCGGTCGGAGTTCCCCGCGCC
AGCGGGGATAAACCGTTCAGGCCAAGCTGCTTGCGGTAGGCATCCCAGAGTTCCCCGCGCCAGCGGGGATAAACC
GTCAGCTTTATAAATATGCATATACGGAAACTAGAGTTCCCCGCGCCAGCGGGGATAAACCGTATCGTCGGTCGG
GTCATACGTGGACTTGTCAGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCTCCCATTTTCAAACCCATCAAGA
CGCGGTACCTTAATTAA

Template CRISPR (T_{1-4}), pWUR478

GGCGCGCCATGGAAACAAAGAATTAGCTGATCTTTAATAATAAGGAAATGTTACATTAAGGTTGGTGGGTTGTTT
TTATGGGAAAAAATGCTTTAAGAACAAATGTATACTTTTAGAGAGTTCCCCGCGCCAGCGGGGATAAACCGCTGA
GTGTGATCGATGCCATCAGCGAAGGGCCGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCCGAAGCCAAAGAAT
TCGCCGAACACGCTGAGTTCCCCGCGCCAGCGGGGATAAACCGCAAGCAACAGGCAGGCGTGACAGCCAGCAAAC
GAGTTCCCCGCGCCAGCGGGGATAAACCGTTTGGATCGGGTCTGGATCCTCTGAGCGGTCGGAGTTCCCCGCGCC
AGCGGGGATAAACCGTGGGATGCCTACCGCAAGCAGCTTGGCCTGAAGAGTTCCCCGCGCCAGCGGGGATAAACC
GTCAGCTTTATAAATATGCATATACGGAAACTAGAGTTCCCCGCGCCAGCGGGGATAAACCGTGACAAGTCCACG
TATGACCCGACCGACGATAGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCTCCCATTTTCAAACCCATCAAGA
CGCGGTACCTTAATTAA

Non-targeting CRISPR (N), pWUR477

GGCGCGCCATGGAAACAAAGAATTAGCTGATCTTTAATAATAAGGAAATGTTACATTAAGGTTGGTGGGTTGTTT
TTATGGGAAAAAATGCTTTAAGAACAAATGTATACTTTTAGAGAGTTCCCCGCGCCAGCGGGGATAAACCGCTTT
CGCAGACGCGCGGCGATACGCTCACGCAGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCCGAAGCCAAAGAAT
TCGCCGAACACGCTGAGTTCCCCGCGCCAGCGGGGATAAACCGGGCTCCCTGTCGGTTGTAATTGATAATGTTGA
GAGTTCCCCGCGCCAGCGGGGATAAACCGTTTGGATCGGGTCTGGATCCTCTGAGCGGTCGGAGTTCCCCGCGCC
AGCGGGGATAAACCGCGAATCGCGCATACCCTGCGCGTCGCCGCCTGGAGTTCCCCGCGCCAGCGGGGATAAACC
GTCAGCTTTATAAATATGCATATACGGAAACTAGAGTTCCCCGCGCCAGCGGGGATAAACCGGACTCACCCCGAA
AGAGATTGCCAGCCAGCTTGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCTCCCATTTTCAAACCCATCAAGA
CGCGGTACCTTAATTAA

Coding CRISPR (C_1), pWUR491

GGCGCGCCATGGAAACAAAGAATTAGCTGATCTTTAATAATAAGGAAATGTTACATTAAGGTTGGTGGGTTGTTT
TTATGGGAAAAAATGCTTTAAGAACAAATGTATACTTTTAGAGAGTTCCCCGCGCCAGCGGGGATAAACCGGGCC
CTTCGCTGATGGCATCGATCACACTCAGGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCCGAAGCCAAAGAAT
TCGCCGAACACGCTGAGTTCCCCGCGCCAGCGGGGATAAACCGGGCTCCCTGTCGGTTGTAATTGATAATGTTGA
GAGTTCCCCGCGCCAGCGGGGATAAACCGTTTGGATCGGGTCTGGATCCTCTGAGCGGTCGGAGTTCCCCGCGCC
AGCGGGGATAAACCGCGAATCGCGCATACCCTGCGCGTCGCCGCCTGGAGTTCCCCGCGCCAGCGGGGATAAACC
GTCAGCTTTATAAATATGCATATACGGAAACTAGAGTTCCCCGCGCCAGCGGGGATAAACCGGACTCACCCCGAA
AGAGATTGCCAGCCAGCTTGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCTCCCATTTTCAAACCCATCAAGA
CGCGGTACCTTAATTAA

Coding CRISPR (C_2),pWUR492

GGCGCGCCATGGAAACAAAGAATTAGCTGATCTTTAATAATAAGGAAATGTTACATTAAGGTTGGTGGGTTGTTT
TTATGGGAAAAAATGCTTTAAGAACAAATGTATACTTTTAGAGAGTTCCCCGCGCCAGCGGGGATAAACCGCTTT
CGCAGACGCGCGGCGATACGCTCACGCAGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCCGAAGCCAAAGAAT
TCGCCGAACACGCTGAGTTCCCCGCGCCAGCGGGGATAAACCGGTTTGCTGGCTGTCACGCCTGCCTGTTGCTTG
GAGTTCCCCGCGCCAGCGGGGATAAACCGTTTGGATCGGGTCTGGATCCTCTGAGCGGTCGGAGTTCCCCGCGCC
AGCGGGGATAAACCGCGAATCGCGCATACCCTGCGCGTCGCCGCCTGGAGTTCCCCGCGCCAGCGGGGATAAACC
GTCAGCTTTATAAATATGCATATACGGAAACTAGAGTTCCCCGCGCCAGCGGGGATAAACCGGACTCACCCCGAA
AGAGATTGCCAGCCAGCTTGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCTCCCATTTTCAAACCCATCAAGA
CGCGGTACCTTAATTAA

Coding CRISPR (C_3), pWUR493

GGCGCGCCATGGAAACAAAGAATTAGCTGATCTTTAATAATAAGGAAATGTTACATTAAGGTTGGTGGGTTGTTT
TTATGGGAAAAAATGCTTTAAGAACAAATGTATACTTTTAGAGAGTTCCCCGCGCCAGCGGGGATAAACCGCTTT
CGCAGACGCGCGGCGATACGCTCACGCAGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCCGAAGCCAAAGAAT
TCGCCGAACACGCTGAGTTCCCCGCGCCAGCGGGGATAAACCGGGCTCCCTGTCGGTTGTAATTGATAATGTTGA
GAGTTCCCCGCGCCAGCGGGGATAAACCGTTTGGATCGGGTCTGGATCCTCTGAGCGGTCGGAGTTCCCCGCGCC
AGCGGGGATAAACCGTTCAGGCCAAGCTGCTTGCGGTAGGCATCCCAGAGTTCCCCGCGCCAGCGGGGATAAACC
GTCAGCTTTATAAATATGCATATACGGAAACTAGAGTTCCCCGCGCCAGCGGGGATAAACCGGACTCACCCCGAA
AGAGATTGCCAGCCAGCTTGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCTCCCATTTTCAAACCCATCAAGA
CGCGGTACCTTAATTAA

Coding CRISPR (C_4), pWUR494

GGCGCGCCATGGAAACAAAGAATTAGCTGATCTTTAATAATAAGGAAATGTTACATTAAGGTTGGTGGGTTGTTT
TTATGGGAAAAAATGCTTTAAGAACAAATGTATACTTTTAGAGAGTTCCCCGCGCCAGCGGGGATAAACCGCTTT
CGCAGACGCGCGGCGATACGCTCACGCAGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCCGAAGCCAAAGAAT
TCGCCGAACACGCTGAGTTCCCCGCGCCAGCGGGGATAAACCGGGCTCCCTGTCGGTTGTAATTGATAATGTTGA
GAGTTCCCCGCGCCAGCGGGGATAAACCGTTTGGATCGGGTCTGGATCCTCTGAGCGGTCGGAGTTCCCCGCGCC
AGCGGGGATAAACCGCGAATCGCGCATACCCTGCGCGTCGCCGCCTGGAGTTCCCCGCGCCAGCGGGGATAAACC
GTCAGCTTTATAAATATGCATATACGGAAACTAGAGTTCCCCGCGCCAGCGGGGATAAACCGTATCGTCGGTCGG
GTCATACGTGGACTTGTCAGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCTCCCATTTTCAAACCCATCAAGA
CGCGGTACCTTAATTAA

Template CRISPR (T_1), pWUR487

GGCGCGCCATGGAAACAAAGAATTAGCTGATCTTTAATAATAAGGAAATGTTACATTAAGGTTGGTGGGTTGTTT
TTATGGGAAAAAATGCTTTAAGAACAAATGTATACTTTTAGAGAGTTCCCCGCGCCAGCGGGGATAAACCGCTGA
GTGTGATCGATGCCATCAGCGAAGGGCCGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCCGAAGCCAAAGAAT
TCGCCGAACACGCTGAGTTCCCCGCGCCAGCGGGGATAAACCGGGCTCCCTGTCGGTTGTAATTGATAATGTTGA
GAGTTCCCCGCGCCAGCGGGGATAAACCGTTTGGATCGGGTCTGGATCCTCTGAGCGGTCGGAGTTCCCCGCGCC
AGCGGGGATAAACCGCGAATCGCGCATACCCTGCGCGTCGCCGCCTGGAGTTCCCCGCGCCAGCGGGGATAAACC
GTCAGCTTTATAAATATGCATATACGGAAACTAGAGTTCCCCGCGCCAGCGGGGATAAACCGGACTCACCCCGAA
AGAGATTGCCAGCCAGCTTGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCTCCCATTTTCAAACCCATCAAGA
CGCGGTACCTTAATTAA

Template CRISPR (T_2), pWUR488

GGCGCGCCATGGAAACAAAGAATTAGCTGATCTTTAATAATAAGGAAATGTTACATTAAGGTTGGTGGGTTGTTT
TTATGGGAAAAAATGCTTTAAGAACAAATGTATACTTTTAGAGAGTTCCCCGCGCCAGCGGGGATAAACCGCTTT
CGCAGACGCGCGGCGATACGCTCACGCAGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCCGAAGCCAAAGAAT
TCGCCGAACACGCTGAGTTCCCCGCGCCAGCGGGGATAAACCGCAAGCAACAGGCAGGCGTGACAGCCAGCAAAC
GAGTTCCCCGCGCCAGCGGGGATAAACCGTTTGGATCGGGTCTGGATCCTCTGAGCGGTCGGAGTTCCCCGCGCC
AGCGGGGATAAACCGCGAATCGCGCATACCCTGCGCGTCGCCGCCTGGAGTTCCCCGCGCCAGCGGGGATAAACC
GTCAGCTTTATAAATATGCATATACGGAAACTAGAGTTCCCCGCGCCAGCGGGGATAAACCGGACTCACCCCGAA
AGAGATTGCCAGCCAGCTTGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCTCCCATTTTCAAACCCATCAAGA
CGCGGTACCTTAATTAA

Template CRISPR (T_3), pWUR489

GGCGCGCCATGGAAACAAAGAATTAGCTGATCTTTAATAATAAGGAAATGTTACATTAAGGTTGGTGGGTTGTTT
TTATGGGAAAAAATGCTTTAAGAACAAATGTATACTTTTAGAGAGTTCCCCGCGCCAGCGGGGATAAACCGCTTT
CGCAGACGCGCGGCGATACGCTCACGCAGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCCGAAGCCAAAGAAT
TCGCCGAACACGCTGAGTTCCCCGCGCCAGCGGGGATAAACCGGGCTCCCTGTCGGTTGTAATTGATAATGTTGA
GAGTTCCCCGCGCCAGCGGGGATAAACCGTTTGGATCGGGTCTGGATCCTCTGAGCGGTCGGAGTTCCCCGCGCC
AGCGGGGATAAACCGTGGGATGCCTACCGCAAGCAGCTTGGCCTGAAGAGTTCCCCGCGCCAGCGGGGATAAACC
GTCAGCTTTATAAATATGCATATACGGAAACTAGAGTTCCCCGCGCCAGCGGGGATAAACCGGACTCACCCCGAA
AGAGATTGCCAGCCAGCTTGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCTCCCATTTTCAAACCCATCAAGA
CGCGGTACCTTAATTAA

Template CRISPR (T_4), pWUR490

GGCGCGCCATGGAAACAAAGAATTAGCTGATCTTTAATAATAAGGAAATGTTACATTAAGGTTGGTGGGTTGTTT
TTATGGGAAAAAATGCTTTAAGAACAAATGTATACTTTTAGAGAGTTCCCCGCGCCAGCGGGGATAAACCGCTTT
CGCAGACGCGCGGCGATACGCTCACGCAGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCCGAAGCCAAAGAAT

TCGCCGAACACGCTGAGTTCCCCGCGCCAGCGGGGATAAACCGGGCTCCCTGTCGGTTGTAATTGATAATGTTGA
GAGTTCCCCGCGCCAGCGGGGATAAACCGTTTGGATCGGGTCTGGATCCTCTGAGCGGTCGGAGTTCCCCGCGCC
AGCGGGGATAAACCGCGAATCGCGCATACCCTGCGCGTCGCCGCCTGGAGTTCCCCGCGCCAGCGGGGATAAACC
GTCAGCTTTATAAATATGCATATACGGAAACTAGAGTTCCCCGCGCCAGCGGGGATAAACCGTGACAAGTCCACG
TATGACCCGACCGACGATAGAGTTCCCCGCGCCAGCGGGGATAAACCGCAGCTCCCATTTTCAAACCCATCAAGA
CGCGGTACCTTAATTAA

Table S1. Overview of different Cas systems. This table focuses on a subset of the Cas proteins: well conserved Cas proteins (Cas1-4), and subunits of the *E. coli* K12 Cascade complex (CasA-E); for more extensive comparative analyses, see (*S15, S16*). *E. coli* strains discussed in this study are in bold.

Cas protein[a]	Family[b]		Cas system (CASS)[c]								
			none	CASS1	CASS2	CASS3	CASS4	CASS5	CASS6	CASS7	Function predicted (p)[a], or demonstrated experimentally (e)
		Species Strain	***Escherichia coli*** **BL21(DE3)**	*Bacillus halodurans* C-125	***Escherichia coli*** **K12**	***Escherichia coli*** **UTI89**	*Streptococcus thermophilus* CNRZ1066	*Sulfolobus solfataricus* P2	*Nitrosomonas europaea* ATCC 19718	*Archaeoglobus fulgidus* VC-16	
Cas1	COG1518		-	BH0341	b2755 (YgbT)	C0890	str0658	SSO1450	NE0111	AF1878	nuclease/integrase (p)
Cas2	COG1343		-	BH0342	b2754 (YgbF)	C0891[f]	str0659	SSO8090	NE0112	AF1876	RNA endonuclease (e) (*S17*)
Cas3	COG1203		-	BH0336	b2761 (YgcB)	C0891[f]	-	SSO1440	-	AF1874	DEAD-box helicase, often fused to HD nuclease (p)
Cas4	COG1468		-	BH0340	-	-	-[d]	SSO1451	-	AF1877	RecB-like nuclease (p), often has C-terminal Zn clusters
Cascade complex[e]											
CasA (Cse1)	YgcL		-	-	b2760 (YgcL)	-	-	-	-	-	Zn-finger containing protein (p)
CasB (cse2)	YgcK		-	-	b2759 (YgcK)	-	-	-	-	-	α-helical protein (p)
CasC (Cse4)	YgcJ (COG1857)		-	BH0339 (COG3649)	b2758 (YgcJ)	C0893 (y1725)	-	SSO1442	-	AF1871	α/β protein, nuclease (p)
CasD (Cas5/5e)	YgcI (COG1688)		-	BH0337	b2757 (YgcI)	C0893/894 (y1726)	-	SSO1441	-	AF1872	RAMP (p)
CasE (Cse3)	YgcH		-	-	b2756 (YgcH)	C0896 (y1727)	-	-	-	-	RAMP (p), crRNA endonuclease (e)[e]

[a] Cas protein nomenclature and functional prediction according to (*S15, S16, S18*)
[b] Family nomenclature according to (*S16*), family may contain several COGs
[c] CASS nomenclature according to (*S16*)
[d] Cas4 is absent in CASS4 (*S. thermophilus*), but is present in CASS4a (*S16*)
[e] this study
[f] Fusion of Cas2 and Cas3

Table S2

Strains, plasmids and primers used in this study.

Strains	Description	Source, reference
Escherichia coli K12 W3110	BW25113, F^- λ^- rph-1 INV(rrnD, rrnE)	(*S1*)
E. coli K12 W3110 Δ*casA*	JW2730	(*S19*)
E. coli K12 W3110 Δ*casB*	JW2729	(*S19*)
E. coli K12 W3110 Δ*casC*	JW2728	(*S19*)
E. coli K12 W3110 Δ*casD*	JW5844	(*S19*)
E. coli K12 W3110 Δ*casE*	JW2726	(*S19*)
E. coli K12 W3110 Δ*uidA*	JW1609, β-glucuronidase knockout	(*S19*)
E. coli BL21(DE3)	F^- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])	Novagen
E. coli BL21-AI	F^- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ araB::T7RNAP-tetA	Novagen
E. coli HB101	F^- mcrB mrr $hsdS20(r_B^- m_B^-)$ recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20(Sm^R) glnV44 λ^-	(*S20*)
E. coli DH5α	F^- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, $hsdR17(r_K^- m_K^+)$, λ–	
E. coli NEB5α	fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80Δ (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	New England Biolabs

Plasmids	Description and order of genes (5'-3')	Restriction sites	Primers	Source
pET-52b(+)	T7 RNA polymerase based expression vector, Amp^R			Novagen
pRSF-1b	T7 RNA polymerase based expression vector, Kan^R			Novagen
pCDF-1b	T7 RNA polymerase based expression vector, Str^R			Novagen
pACYCduet-1	T7 RNA polymerase based expression vector, Cam^R			Novagen
pZErO-1.1	Cloning plasmid, Zeo^R			Invitrogen
pIH1119	Expression vector for N-terminal MalE fusion proteins, Amp^R			New England Biolabs
pWUR384	*casE* in pET-52b with both StrepII- tag (N-term) and His_{10}-tag (C-term)	BamHI/NotI	BG2253 + BG2254	This study
pWUR385	*casD* in pET-52b with both StrepII- tag (N-term) and His_{10}-tag (C-term)	KpnI/NotI	BG2316 + BG2252	This study
pWUR386	*casC* in pET-52b with both StrepII- tag (N-term) and His_{10}-tag (C-term)	BamHI/NotI	BG2249 + BG2250	This study
pWUR387	*casB* in pET-52b with both StrepII- tag (N-term) and His_{10}-tag (C-term)	BamHI/NotI	BG2245 + BG2246	This study
pWUR388	*casA* in pET-52b with both StrepII- tag (N-term) and His_{10}-tag (C-term)	Acc65I/NotI	BG2247 + BG2248	This study
pWUR396	*E. coli* K12 CRISPR in pACYCDuet-1	NcoI/Acc65I	BG2459 + BG2460	This study
pWUR397	*cas3* in pRSF-1b, no tags	NcoI/NotI	BG2371 + BG2372	This study
pWUR399	*casA-casB-casC-casD-casE-cas1-cas2* in pCDF-1b, no tags	NcoI/NotI	BG2394 + BG2375	This study
pWUR400	*casA-casB-casC-casD-casE* in pCDF-1b, no tags	NcoI/NotI	BG2394 + BG2481	This study
pWUR401	*casB-casC-casD-casE* in pCDF-1b, no tags	NcoI/NotI	BG2464 + BG2481	This study
pWUR402	*casC-casD-casE* in pCDF-1b, no tags	NcoI/NotI	BG2465 + BG2481	This study
pWUR403	*casD-casE* in pCDF-1b, no tags	NcoI/NotI	BG2466 + BG2481	This study
pWUR404	*casE* in pCDF-1b, no tags	NcoI/NotI	BG2480 + BG2481	This study
pWUR405	*casA-casB-casC-casD* in pRSF-1b, no tags	NcoI/NotI	BG2394 + BG2482	This study
pWUR406	*casA-casB-casC* in pRSF-1b, no tags	NcoI/NotI	BG2394 + BG2483	This study
pWUR407	*casA-casB* in pRSF-1b, no tags	NcoI/NotI	BG2394 + BG2484	This study
pWUR408	*casA* in pRSF-1b, no tags	NcoI/NotI	BG2394 + BG2485	This study
pWUR472	pWUR400 with H20A mutation in CasE		BG2584 + BG2585	This study
pWUR473	pWUR404 with H20A mutation in CasE		BG2584 + BG2585	This study
pWUR477	non targeting CRISPR in pACYCDuet-1 (*N*)			Geneart, Germany
pWUR478	template CRISPR in pACYCDuet-1 (T_{1-4})			Geneart, Germany
pWUR479	coding CRISPR in pACYCDuet-1 (C_{1-4})			Geneart, Germany
pWUR480	*casB* with StrepII-tag (N-term)-*casC-casD* in pET52b		BG2573 + BG2482	This study
pWUR481	*casE* in pIH1119 generating MalE-CasE fusion protein		BG2586 + BG2587	This study
pWUR487	Template CRISPR (T_1), pWUR477 with template spacer 1, subcloned from pWUR 478	NcoI/EcoRI		This study
pWUR488	Template CRISPR (T_2), pWUR477 with template spacer 2, subcloned from pWUR 478	EcoRI/BamHI		This study
pWUR489	Template CRISPR (T_3), pWUR477 with template spacer 3, subcloned from pWUR 478	BamHI/NsiI		This study
pWUR490	Template CRISPR (T_4), pWUR477 with template spacer 4, subcloned from pWUR 478	NsiI/Acc65I		This study
pWUR491	Coding CRISPR (C_1), pWUR477 with coding spacer 1, subcloned from pWUR 479	NcoI/EcoRI		This study
pWUR492	Coding CRISPR (C_2), pWUR477 with coding spacer 2, subcloned from pWUR 479	EcoRI/BamHI		This study
pWUR493	Coding CRISPR (C_3), pWUR477 with coding spacer 3, subcloned from pWUR 479	BamHI/NsiI		This study
pWUR494	Coding CRISPR (C_4), pWUR477 with coding spacer 4, subcloned from pWUR 479	NsiI/Acc65I		This study

Experiment	Primer	Sequence (5'-3')	Description
Plasmid construction	BG2245	GCGCGGGATCCTATGGCTGATGAAATTGATGCAATG	*casB* + BamHI (fw)
	BG2246	GGCCCGCGGCCGCCGCATTTTTGTTTGTGGTCAATAC	*casB* + NotI (rv)
	BG2247	GCGCGGGTACCAGATGAATTTGCTTATTGATAACTGGAT	*casA* + Acc65I (fw)
	BG2248	GGCCCGCGGCCGCGCCATTTGATGGCCCTCCTTG	*casA* + NotI (rv)
	BG2249	GCGCGGGATCCTATGTCTAACTTTATCAATATTCATGT	*casC* + BamHI (fw)
	BG2250	GGCCCGCGGCCGCCGCCTCGCCATTATTACGAAC	*casC* + NotI (rv)
	BG2316	GCGCGGGTACCAGATGAGATCTTATTTGATCTTGCGG	*casD* + KpnI (fw)
	BG2252	GGCCCGCGGCCGCCTGAGATACATCCATACCTCCT	*casD* + NotI (rv)
	BG2253	GCGCGGGATCCTATGTATCTCAGTAAAGTCATCATTG	*casE* + BamHI (fw)
	BG2254	GGCCCGCGGCCGCCAGTGGAGCCAAAGATAGCAAG	*casE* + NotI (rv)
	BG2371	GCGCGCCATGGAACCTTTTAAATATATATGCCATTACT	*cas3* + NcoI (fw)
	BG2372	GGCCCGCGGCCGCTTATTTGGGATTTGCAGGGATGACT	*cas3* + NotI + stopcodon (rv)
	BG2375	GGCCCGCGGCCGCTCAAACAGGTAAAAAAGACACCAAC	*cas2* + NotI + stopcodon (rv)
	BG2394	GCGCGCCATGGCTAATTTGCTTATTGATAACTGGATCC	*casA* + NcoI (fw)
	BG2459	GCGTACCATGGCATAAGGAAATGTTACATTAAGGTTGG	*E. coli* K12 CRISPR + NcoI (fw)
	BG2460	GGTCCGGTACCCAGCGTCAGGCGTGAAATCTCACC	*E. coli* K12 CRISPR + Acc65I (rv)
	BG2464	GCGCGCCATGGCTATGGCTGATGAAATTGATGCAATG	*casB* + NcoI (fw)
	BG2465	GCGCGCCATGGCTATGTCTAACTTTATCAATATTCATGT	*casC* + NcoI (fw)
	BG2466	GCGCGCCATGGCTATGAGATCTTATTTGATCTTGCGG	*casD* + NcoI (fw)
	BG2480	GCGCGCCATGGCTATGTATCTCAGTAAAGTCATCATTG	*casE* + NcoI (fw)
	BG2481	GGCCCGCGGCCGCTCACAGTGGAGCCAAAGATAGC	*casE* + NotI + stopcodon (rv)
	BG2482	GGCCCGCGGCCGCTTACTGAGATACATCCATACCTCC	*casD* + NotI + stopcodon (rv)
	BG2483	GGCCCGCGGCCGCTCACGCCTCGCCATTATTACGA	*casC* + NotI + stopcodon (rv)
	BG2484	GGCCCGCGGCCGCTTACGCATTTTTGTTTGTGGTCAAT	*casB* + NotI + stopcodon (rv)
	BG2485	GGCCCGCGGCCGCTCAGCCATTTGATGGCCCTCC	*casA* + NotI + stopcodon (rv)
	BG2573	GCGCGGGTACCAGATGGCTGATGAAATTGATGCAATG	*casB* + Acc65I (fw)
	BG2586	GGCCCGCGGCCGCCTGCAGTCACAGTGGAGCCAAAGATAGCAAG	*casE* + NotI + PstI + stopcodon (rv)
	BG2587	GCGCGGAATTCATGTATCTCAGTAAAGTCATCATTG	*casE* + EcoRI (fw)
QuikChange mutation	BG2584	GAGCAGGGATCTTTACCAACTTGCGCAGGGATTATGGCATTTATTTCC	QuikChange primer *casE* H20A (fw)
	BG2585	GGAAATAAATGCCATAATCCCTGCGCAAGTTGGTAAAGATCCCTGCTC	QuikChange primer *casE* H20A (rv)
Cleavage reaction	BG2559	CTGACTTAATACGACTCACTATAGGATAAACCGCTTTCGCAGACGCGCGGCGA	T7 promotor + *E. coli* K12 CRISPR transcript (fw)
	BG2374	GGCCCGCGGCCGCCAGCGTCAGGCGTGAAATCTC	*E.coli* K12 CRISPR transcript (rv)
	BG2462	CTGACTTAATACGACTCACTATAGGGAGGATAAAAAAGGGTGGCAGCAG	T7 promotor + *E. coli* UTI89 CRISPR transcript (fw)
	BG2463	GGTCGCACTCTTCAACGTTCAG	*E. coli* UTI89 CRISPR transcript (rv)
	BG2452	GAATTGTAATACGACTCACTATAGG	T7 promotor + non-CRISPR transcript pZErO-1.1 (fw)
	BG2461	CTGAGATACCTACAGCGTGAGC	pZErO-1.1 non-CRISPR transcript (rv)
Northern blot	BG2349	AGCTTGCGACCGCTCAGAAATTCCAGACCCGATCCAAAG	*E. coli* K12 CRISPR spacer 4 (rv)
RNA cloning	BG2164	GCCCGCCCGGATCCTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN	Reverse transcription primer
	BG2220	GCGCCCGCGGATCCCCCCCCCCDN	cDNA amplification + BamHI (fw)
	BG2222	GCCCGCCCGGATCCTT	cDNA amplification + BamHI (rv)

Table S3

Extended legend presenting strains and plasmids used in each figure. Plasmid number and corresponding products are depicted in same color in each row. For strains, plasmids and primers see Table S2.

Fig.	manuscript legend	lane(s)	strain	plasmids (pWUR-)	comments
1	The composition of the Cascade complex.				
1B	Coomassie Blue stained SDS-PAGE gel of the affinity purified protein complex using either the amino-terminal StrepII-tag (S) or carboxy-terminal His-tag (H) of each of the subunits CasB, CasC, CasD or CasE as bait. Asterisks indicate the 5.5 kDa heavier double-tagged subunits. Marker sizes in kDa on the left, and the location of untagged subunits on the right.	1	*E. coli* BL21(DE3), for protein overproduction	387 + 399 + 397	CasB with both StrepII-tag (N-term) and His_{10}-tag (C-term) + CasABCDE/Cas1/Cas2 + Cas3
		2			
		3		386 + 399 + 397	CasC with both StrepII-tag (N-term) and His_{10}-tag (C-term) + CasABCDE/Cas1/Cas2 + Cas3
		4			
		5		385 + 399 + 397	CasD with both StrepII-tag (N-term) and His_{10}-tag (C-term) + CasABCDE/Cas1/Cas2 + Cas3
		6			
		7		384 + 399 + 397	CasE with both StrepII-tag (N-term) and His_{10}-tag (C-term) + CasABCDE/Cas1/Cas2 + Cas3
		8			
2	Cascade cleaves CRISPR RNA precursors into small RNAs of 57 nucleotides (←).				
2A	Northern analysis of total RNA of wild type *E. coli* K12 (WT), a non-*cas* gene knockout (*Δu*, *uidA*, β-glucuronidase) and Cascade gene-knockouts using a single stranded spacer sequence as a probe.	WT	*E. coli* K12 W3110		
		Δ*uidA*	*E. coli* K12 W3110 Δ*uidA*		
		Δ*casA*	*E. coli* K12 W3110 Δ*casA*		
		Δ*casB*	*E. coli* K12 W3110 Δ*casB*		
		Δ*casC*	*E. coli* K12 W3110 Δ*casC*		
		Δ*casD*	*E. coli* K12 W3110 Δ*casD*		
		Δ*casE*	*E. coli* K12 W3110 Δ*casE*		
2B	Northern blot of total RNA from *E. coli* BL21(DE3) expressing the *E. coli* K12 pre-crRNA and either the complete or incomplete Cascade complex.	1	*E. coli* BL21(DE3)	396	*E. coli* K12 CRISPR
		2		400 + 396	CasABCDE + *E. coli* K12 CRISPR
		3		401 + 396	CasBCDE + *E. coli* K12 CRISPR
		4		408 + 402 + 396	CasA + CasCDE + *E. coli* K12 CRISPR
		5		407 + 403 + 396	CasAB + CasDE + *E. coli* K12 CRISPR
		6		406 + 404 + 396	CasABC + CasE + *E. coli* K12 CRISPR
		7		405 + 396	CasABCD + *E. coli* K12 CRISPR
2C	Activity assays with purified Cascade using *in vitro* transcribed α-^{32}P-UTP labeled pre-crRNA from *E. coli* K12 (repeat sequence: GAGUUCCCCGCCAGCGGGGAUAAACCG) and *E. coli* UTI89 (repeat sequence: GUUCACUGCCGUACAGGCAGCUUAGAAA) and non-crRNA as substrates.	1, 4, 7	*E. coli* BL21(DE3), for protein overproduction	400 + 386	CasABCDE + CasC with StrepII-tag (N-term) and His_{10}-tag (C-term), purified using StrepII-tag
		2, 5, 8			
		3, 6, 9			
2D	Activity assays as in 2C for 15 min using purified MalE-LacZa and MalE-CasE fusion proteins.	1, 4, 7			control lanes without protein
		2, 5, 8	*E. coli* DH5α, for protein overproduction	pIH1119	purified MalE-LacZα
		3, 6, 9		481	purified MalE-CasE
2E	Northern blot as in 2B with Cascade or Cascade-$CasE^{H20A}$.	1	*E. coli* BL21(DE3)	396	*E. coli* K12 CRISPR
		2		400 + 396	CasABCDE + *E. coli* K12 CRISPR
		3		472 + 396	CasABCDE with H20A mutation in CasE + *E. coli* K12 CRISPR
2F	Activity assays as in 2C for 30 min using purified Cascade or Cascade-$CasE^{H20A}$.	1			control lane without protein
		2	*E. coli* BL21(DE3), for overproduction of protein	408 + 480 + 404	CasA + CasB with StrepII-tag (N-term)/CasCD + CasE, purified using StrepII-tag
		3		408 + 480 + 473	CasA + CasB with StrepII-tag (N-term)/CasCD + CasE with H20A mutation, purified using StrepII-tag
3	Cleaved crRNAs remain bound by Cascade.				
3A	Denaturing polyacrylamide gel showing the RNA isolated from purified Cascade in the absence and presence of co-expressed pre-crRNA.	1	*E. coli* BL21(DE3), for protein and RNA overproduction	387 + 400	CasB with both StrepII-tag (N-term) and His_{10}-tag (C-term) + CasABCDE, purified using StrepII-tag
		2		387 + 400 + 396	CasB with both StrepII-tag (N-term) and His_{10}-tag (C-term) + CasABCDE + *E. coli* K12 CRISPR, purified using StrepII-tag
4	Engineered CRISPRs confer resistance to phage Lambda (λ) in the presence of Cascade and Cas3.				
4A	Effect of the presence of different sets of cas genes on the sensitivity of *E. coli* to phage λ_{vir}. Cells were equipped with two engineered CRISPRs containing four anti-λ spacers each (Fig.S3). The C_{1-4} CRISPR produces crRNA complementary to the coding strand of λ_{vir}, and the T_{1-4} CRISPR targets only the template strand. The sensitivity of each strain to phage λ_{vir} is represented as a histogram of the efficiency of plaquing, which is the plaque count ratio of the anti-λ CRISPR to that of the non-targeting control CRISPR (N).	Cascade + C_{1-4}	*E. coli* BL21-AI	400 + 479	CasABCDE + coding spacer 1-4
		Cascade + T_{1-4}		400 + 478	CasABCDE + template spacer 1-4
		Cas3 + C_{1-4}		397 + 479	Cas3 + coding spacer 1-4
		Cas3 + T_{1-4}		397 + 478	Cas3 + template spacer 1-4
		Cascade + Cas3 + C_{1-4}		400 + 397 + 479	CasABCDE + Cas3 + coding spacer 1-4
		Cascade + Cas3 + T_{1-4}		400 + 397 + 478	CasABCDE + Cas3 + template spacer 1-4
		Cascade-$CasE^{H20A}$ + Cas3 + C_{1-4}		472 + 397 + 479	CasABCDE with H20A mutation in CasE + Cas3 + coding spacer 1-4
		Cascade-$CasE^{H20A}$ + Cas3 + T_{1-4}		472 + 397 + 478	CasABCDE with H20A mutation in CasE + Cas3 + template spacer 1-4
		Cascade + Cas1 + Cas2 + C_{1-4}		399 + 479	CasABCDE/Cas1/Cas2 + coding spacer 1-4
		Cascade + Cas1 + Cas2 + T_{1-4}		399 + 478	CasABCDE/Cas1/Cas2 + template spacer 1-4
		Cascade + Cas 3 + Cas1 + Cas2 + C_{1-4}		399 + 397 + 479	CasABCDE/Cas1/Cas2 + Cas 3 + coding spacer 1-4
		Cascade + Cas 3 + Cas1 + Cas2 + T_{1-4}		399 + 397 + 478	CasABCDE/Cas1/Cas2 + Cas 3 + template spacer 1-4
4B	Effect of single anti-λ spacers (Fig.S3) on the sensitivity of *E. coli* to λ_{vir}.	Cascade + Cas3 + C_1	*E. coli* BL21-AI	400 + 397 + 491	CasABCDE + Cas3 + coding spacer 1
		Cascade + Cas3 + C_2		400 + 397 + 492	CasABCDE + Cas3 + coding spacer 2
		Cascade + Cas3 + C_3		400 + 397 + 493	CasABCDE + Cas3 + coding spacer 3
		Cascade + Cas3 + C_4		400 + 397 + 494	CasABCDE + Cas3 + coding spacer 4
		Cascade + Cas3 + T_1		400 + 397 + 487	CasABCDE + Cas3 + template spacer 1
		Cascade + Cas3 + T_2		400 + 397 + 488	CasABCDE + Cas3 + template spacer 2
		Cascade + Cas3 + T_3		400 + 397 + 489	CasABCDE + Cas3 + template spacer 3
		Cascade + Cas3 + T_4		400 + 397 + 490	CasABCDE + Cas3 + template spacer 4

References

S1. K. Hayashi *et al., Mol. Syst. Biol.* 2, 2006 0007 (2006).
S2. A. P. Snijders *et al., Proteomics* 6, 1518 (2006).
S3. J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular Cloning* (Cold Spring Harbor Laboratory Press, New York, ed. 2nd, 1989).
S4. J. Locker, *Anal. Biochem.* 98, 358 (1979).
S5. S. M. Elbashir, W. Lendeckel, T. Tuschl, *Genes Dev.* 15, 188 (2001).
S6. L. M. Botero *et al., Appl. Environ. Microbiol.* 71, 1267 (2005).
S7. F. Jacob, E. L. Wollman, *Ann. Inst. Pasteur (Paris)* 87, 653 (1954).
S8. R. Barrangou *et al., Science* 315, 1709 (2007).
S9. H. Deveau *et al., J. Bacteriol.* 190, 1390 (2008).
S10. P. Horvath *et al., J. Bacteriol.* 190, 1401 (2008).
S11. F. J. Mojica, C. Diez-Villasenor, J. Garcia-Martinez, E. Soria, *J. Mol. Evol.* 60, 174 (2005).
S12. O. Poirot, E. O'Toole, C. Notredame, *Nucleic Acids Res.* 31, 3503 (2003).
S13. A. Ebihara *et al., Protein Sci.* 15, 1494 (2006).
S14. P. Gouet, X. Robert, E. Courcelle, *Nucleic Acids Res.* 31, 3320 (2003).
S15. D. H. Haft, J. Selengut, E. F. Mongodin, K. E. Nelson, *PLoS Comput. Biol.* 1, e60 (2005).
S16. K. S. Makarova, N. V. Grishin, S. A. Shabalina, Y. I. Wolf, E. V. Koonin, *Biol. Direct* 1, 7 (2006).
S17. N. Beloglazova *et al., J. Biol. Chem.,* in press (available online at http://www.jbc.org/cgi/content/abstract/M803225200v1).
S18. R. Jansen, J. D. Embden, W. Gaastra, L. M. Schouls, *Mol. Microbiol.* 43, 1565 (2002).
S19. T. Baba, H. Mori, *Methods Mol. Biol.* 416, 171 (2008).
S20. H. W. Boyer, R. Rouland-Dussoix, *J. Mol. Biol.* 41, 459 (1969).