Cell Reports, Volume *22*

Supplemental Information

Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation

Sebastian N. Kieper, Cristóbal Almendros, Juliane Behler, Rebecca E. McKenzie, Franklin L. Nobrega, Anna C. Haagsma, Jochem N.A. Vink, Wolfgang R. Hess, and Stan J.J. Brouns

Supplemental figures

Figure S1. High sensitivity spacer detection PCR, Related to Figure 1 D

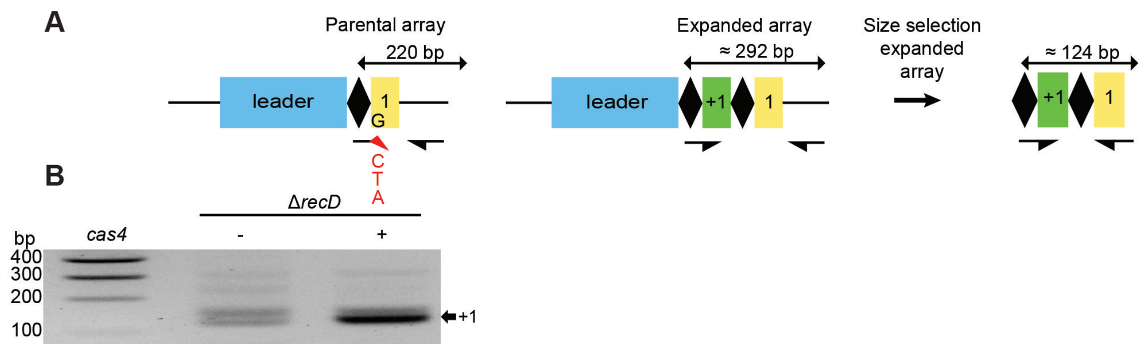

Figure S1 A. Following the first round of amplification, the hypothetical amplicon of expanded arrays of ~292 bp (not visible on gel in **Fig. 1D**) is size selected using the BluePippin system (3% agarose cassette, SageScience). The extracted band is then subjected to a second PCR using the same forward degenerate primer mix (Heler et al. 2015) with a reverse primer annealing in spacer1 of the parental array. **B** By applying the PCR approach described in **A** spacer integration is observed in the *E. coli* K12 Δ*recD* background in the presence or absence of *cas4*. Band intensity is not a quantitative measure for integration efficiency, but rather a binary result (yes/no), because PCR product input resulting from automatized size selection cannot be normalized. The amplicon corresponding to expanded arrays is indicated with a black arrow.

Figure S2. PAM and size distribution of new spacers, Related to Figure 2.

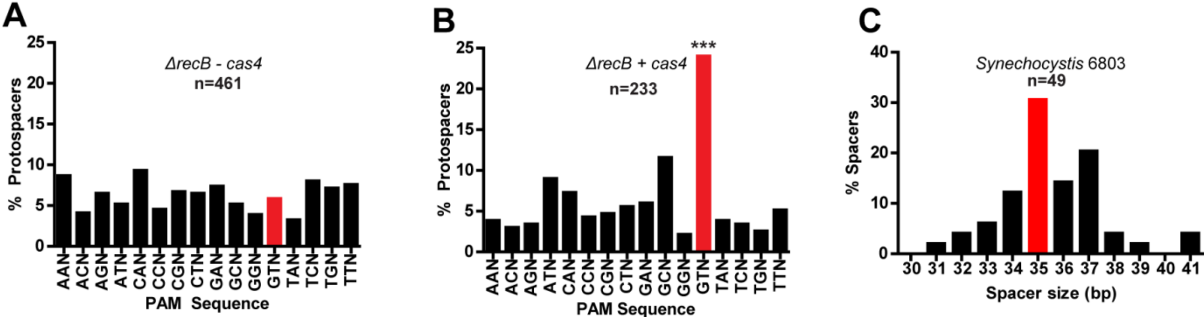

Figure S2, A Percentage of protospacers adjacent to indicated PAM matched by spacers acquired in **A** *ΔrecB* without *cas4* and **B** with *cas4*. **C** Spacer size distribution in the native type I-D host Synechocystis. n= number of analyzed spacers, significance level α = 0.001.

Figure S3. Genomic spacer mapping, Related to Figure 2.

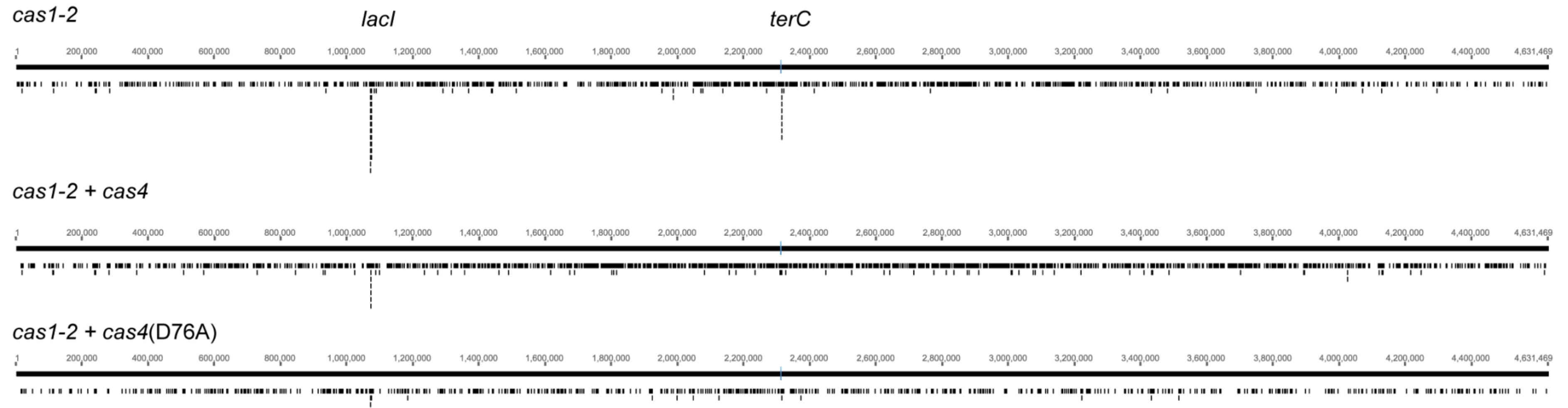

Figure S3,. Origin of spacers acquired from the WT *E. coli* K12 genome. In the absence of *cas4* the Cas1-2 integration complex acquires spacers with preference for the *lacI* sequence and *terC* site. The preferential uptake of spacers derived from the *terC* site is lost when supplying Cas4 WT or $Cas4^{D76A}$.

Supplementary Tables

Table S1. Plasmids used in this study, Related to Figure 1, 2, 3.

Name in this study	Name	Insert	Vector	Resistance	Source
pCas2	pTU084	Synechocystis PCC6803 Type I-D *cas2* (deltaCas1)	pET-T7	Amp	This study
pCas1	pTU085	Synechocystis PCC6803 Type I-D *cas1* (deltaCas2)	pET-T7	Amp	This study
pCas4^{D76A}	pTU086	Synechocystis PCC6803 Type I-D *cas4* (D76A)	pET-T7	Spec	This study
pCas4	pTU130	Synechocystis PCC6803 Type I-D *cas4*	pET-T7	Spec	This study
pCRISPR	pTU134	Synechocystis PCC6803 Type I-D Leader-R-S1	pACYCDuet1	Cm	This study
pCas1-2	pTU70	Synechocystis PCC6803 Type I-D *cas1-cas2*	pET-T7	Amp	This study
pEmp	pTU116	NA	pET-T7	Spec	Addgene Plasmid #48329
pVZ322	pNT	NA	pVZ322	Gent	This study
pVZ322 pT-GTA	pT-GTA	GTA-protospacer1-GentR	pVZ322	Gent	This study
pVZ322 pT-GTT	pT-GTT	GTT-protospacer1-GentR	pVZ322	Gent	This study
pVZ322 pT-GTC	pT-GTC	GTC-protospacer1-GentR	pVZ322	Gent	This study
pVZ322 pT-GTG	pT-GTG	GTG-protospacer1-GentR	pVZ322	Gent	This study

Table S2. Oligonucleotides used in this study, Related to Figure 1, 2, 3.

Name	Sequence	Description
BG7615	TTATGGAGTTGGGATCTTATTAGATAATAAT ACTACCAGGTTTTTCTGGTTG	*cas2* rv
BG8223	TACTTCCAATCCAATGCAATGGATGATTATT TACCTTTAGC	*cas4* fw
BG8224	TTATCCACTTCCAATGTTATTATTTAGTAAGT TTTTTTAATTCTTTCGG	*cas4* rv
BN015	CGTCCATGGGAAGTCATTCTTCAAATTTTGG C	leader fw
BN016	TACAAGCTTAGGCATTGAAAGCGACC	Sp1 Rv (for degenerate PCR)
BN114	TTTAAGAAGGAGATATAGATCATGTCTACAC TTTACTTGACTCAACC	*cas1* fw
BN135	GCTGCAGTGGAAGAAAGTG	Type I-D *cas4* mutagenesis D76A Fw
BN136	AATAATCCCTTTAACTTTTAGGCG	Type I-D *cas4* mutagenesis D76A Rv
BN143	GCGATCGGGACTGAAACT	Degenerated Fw1
BN144	GCGATCGGGACTGAAACA	Degenerated Fw2
BN145	GCGATCGGGACTGAAACC	Degenerated Fw3
BN156	AGGCATTGAAAGCGACC	Degenerate PCR Rv (internal. Sp1)
BN172	AGATCTGCCATATGTATATCTCCTTC	pAcyc backbone Rv (for degenerate PCR)
BN212	GATCTATATCTCCTTCTTAAAGTTAAAC	*cas1* deletion Rv
BN213	CAGTTATCAGTTGTGTTTTGAC	*cas1* deletion Fw
BN214	TTTTTAGTCGTCAAAACACAAC	*cas2* deletion Rv
BN215	TAATAAGATCCCAACTCCATAAG	*cas2* deletion Fw
GentaR_pUC 19_fwd	CGGTGATGACGGTGAGATTCCATTTTTACAC TGATGAATGTTCCGTTGCG	Gentamicin resistance cassette with overlaps to pUC19
GentaR_pUC 19_rev	CGCCTTTGAGTGAGCTCCCGGCATTCGCTG CGCT	Gentamicin resistance cassette with overlaps to pUC19
CRISPR1_G TG_S1_fwd	GTGGATTGTTGTGCCCCTGGCGGTCGCTTT CAATGCCTTTAACAATTCGTTCAAGCCGAGA TC	GTG PAM motif spacer 1
CRISPR1_G TG_S1_rev	AAGGCATTGAAAGCGACCGCCAGGGGCAC AACAATCCACGGTGGCGGTACTTGGGTC	GTG PAM motif spacer 1
CRISPR1_G TA_S1_fwd	GTAGATTGTTGTGCCCCTGGCGGTCGCTTT CAATGCCTTTAACAATTCGTTCAAGCCGAGA TC	GTA PAM motif spacer 1
CRISPR1_G TA_S1_rev	AAGGCATTGAAAGCGACCGCCAGGGGCAC AACAATCTACGGTGGCGGTACTTGGGTC	GTA PAM motif spacer 1
CRISPR1_G TT_S1_fwd	GTTGATTGTTGTGCCCCTGGCGGTCGCTTT CAATGCCTTTAACAATTCGTTCAAGCCGAGA TC	GTT PAM motif spacer 1
CRISPR1_G TT_S1_rev	AAGGCATTGAAAGCGACCGCCAGGGGCAC AACAATCAACGGTGGCGGTACTTGGGTC	GTT PAM motif spacer 1
CRISPR1_G TC_S1_fwd	GTCGATTGTTGTGCCCCTGGCGGTCGCTTT CAATGCCTTTAACAATTCGTTCAAGCCGAGA TC	GTC PAM motif spacer 1
CRISPR1_G TC_S1_rev	AAGGCATTGAAAGCGACCGCCAGGGGCAC AACAATCGACGGTGGCGGTACTTGGGTC	GTC PAM motif spacer 1

CRISPR1_AGC_S1_fwd	AGCGATTGTTGTGCCCCTGGCGGTCGCTTTCAATGCCTTTAACAATTCGTTCAAGCCGAGATC	AGC mock-PAM motif spacer 1
CRISPR1_AGC_S1_rev	AAGGCATTGAAAGCGACCGCCAGGGGCACAACAATCGCTGGTGGCGGTACTTGGGTC	AGC mock-PAM motif spacer 1
GentaR_pVZ322_fwd	TCTGCTCCTGCAGGTCGACTGATTCCATTTTTACACTGATGAATGTTCCGTTGCGCTGCCC	Gentamicin resistance cassette with overlaps to pVZ322
GentaR_pVZ322_rev	CCCGGCATTCGCTGCGCTTATGGCAGAGCA	Gentamicin resistance cassette with overlaps to pVZ322

Table S3. Spacer origin analysis of unique spacers, Related to Figure 2

Strain	Cas4	Genome		pCas4/pEmp		pCas1-2		pCRISPR		% Protospacers		% Protospacers with GTN PAM
		Fw	Rv	Fw	Rv	Fw	Rv	Fw	Rv	G	P	
WT	+	852	835	89	98	10	11	44	30	85.7	14.3	17.1
WT	-	642	627	262	247	27	26	123	97	61.9	38.1	4.7
WT	D76A	351	353	25	26	2	1	10	12	90.3	9.7	4.1
ΔrecB	+	49	51	46	47	4	6	16	14	42.9	57.1	24
ΔrecB	-	125	138	51	73	8	13	25	28	57.0	43.0	5.9

Table S3. Total number of unique spacers acquired from the *E. coli* K12 genome, the *cas4* expression plasmid and corresponding empty vector control, the *cas1-2* expression plasmid and the minimalized type I-D array plasmid in different strains and either presence or absence of *cas4*. Strand orientation indicated with Forward (Fw) or Reverse (Rv). The two last columns represent the percentage of protospacers that match the genome (G) or plasmids (P) and spacers matching a protospacer with the GTN PAM.

Supplemental reference

Heler, R., Samai, P., Modell, J.W., Weiner, C., Goldberg, G.W., Bikard, D., and Marraffini, L.A. (2015). Cas9 specifies functional viral targets during CRISPR-Cas adaptation. Nature *519*, 199-202.