CRISPR-Cas Adaptation in *Escherichia coli* requires RecBCD helicase but not nuclease activity, is independent of homologous recombination, and is antagonised by 5' ssDNA exonucleases

Marin Radovčić[1]†, Tom Killelea[2]†, Ekaterina Savitskaya[3, 4], Lukas Wettstein[2], Edward L. Bolt[2*], Ivana Ivančić-Baće[1*].

[1] Department of Biology, Faculty of Science, University of Zagreb, Croatia.

[2] School of Life Sciences, University of Nottingham, U.K.

[3] Center for Life Sciences, Skolkovo Institute of Science and Technology, Moscow 143028, Russia.

[4] Institute of Molecular Genetics, Russian Academy of Sciences, Moscow 123182, Russia.

† Equal contributions.

* Correspondence to ed.bolt@nottingham.ac.uk or ivana.ivancic.bace@biol.pmf.hr

ABSTRACT

Prokaryotic adaptive immunity is established against mobile genetic elements (MGEs) by "naïve adaptation" when DNA fragments from a newly encountered MGE are integrated into CRISPR-Cas systems. In *E. coli*, DNA integration catalysed by Cas1-Cas2 integrase is well understood in mechanistic and structural detail but much less is known about events prior to integration that generate DNA for capture by Cas1-Cas2. Naïve adaptation in *E. coli* is thought to depend on the DNA helicase-nuclease RecBCD for generating DNA fragments for capture by Cas1-Cas2. The genetics presented here show that naïve adaptation does not require RecBCD nuclease activity but that helicase activity may be important. RecA loading by RecBCD inhibits adaptation explaining previously observed adaptation phenotypes that implicated RecBCD nuclease activity. Genetic analysis of other *E. coli* nucleases and naïve adaptation revealed that 5' ssDNA tailed DNA molecules promote new spacer acquisition. We show that purified *E. coli* Cas1-Cas2 complex binds to and nicks 5' ssDNA tailed duplexes and propose that *E. coli* Cas1-Cas2 nuclease activity on such DNA structures supports naïve adaptation.

INTRODUCTION

CRISPR-Cas is a prokaryotic adaptive immune system against mobile genetic elements (MGEs) in bacteria and archaea (1,2). Immunity is acquired through capture of MGE DNA fragments ("protospacers") and their site-specific integration into a CRISPR array as "spacers" positioned between repeat DNA sequences. These processes are called Adaptation and are catalysed by Cas1-Cas2 integrase from host CRISPR-Cas systems aided by other host proteins, reviewed recently in (3). "Naïve adaptation" relies on Cas1-Cas2 for cells to establish new immunity against an MGE that has not been previously encountered by integration of new spacer DNA into CRISPR arrays (4). Immunity is effected by transcription of the CRISPR array and transcript processing into shorter RNA molecules (crRNAs) that comprise a single spacer sequence. Assembly of crRNA into a ribonucleoprotein complex is used to recognize complementary MGE DNA "protospacer" sequence by base pairing with crRNA, beginning processes of CRISPR "interference". In *E. coli*, interference R-loops are formed by Cascade (CRISPR-associated complex for antiviral defence) after detecting MGE DNA through a protospacer adjacent motif (PAM) sequence (5,6). Cascade R-loop formation recruits Cas3 nuclease/helicase for degradation of the MGE DNA thus completing the immunity response (7-9).

Adaptation processes that generate prokaryotic immunity to an MGE can be separated into three major stages: MGE DNA capture, transport to a CRISPR array, and DNA integration into the CRISPR array followed by DNA gap filling to duplicate the associated repeat (10). Cas1 and Cas2 proteins encoded within CRISPR-Cas systems catalyse these processes aided by other host cell nucleic acid processing proteins. In *E. coli* there is substantial mechanistic detail known about how Cas1-Cas2 bound to MGE DNA recognizes CRISPR and subsequently integrates the DNA. A Cas1-Cas2 complex comprising Cas1 dimers held together by a Cas2 dimer is essential for adaptation in *E. coli* (11-13) binding to a short DNA duplex with flayed ssDNA ends in an adaptation "capture complex" (11,14). The Cas1-Cas2 capture complex is guided to the CRISPR array by DNA structures formed by binding of *E.*

coli integration host factor (IHF) to a conserved sequence motif within the promoter ("leader") sequence of CRISPR (15,16). The 3'OH groups of DNA in the capture complex direct nucleophilic attack of the CRISPR array catalysed by Cas1. This generates a half-site DNA intermediate from the first nucleophilic attack at the leader/promoter-end of CRISPR and then full site integration following the second nucleophilic attack at the repeat-spacer boundary (13,17-19). Host DNA repair gap-fills the integration site (20), completing adaptation by incorporation of a new spacer and new DNA repeat.

DNA pre-processing that leads to capture by Cas1-Cas2 is much less well understood than DNA integration. The Cas1-Cas2-DNA capture complex has been identified at the point of integration (17,19) but the genesis of DNA leading to capture is unclear. Pre-spacers should originate from MGE DNA, to avoid lethal autoimmunity, and their processing should be at specific position relative to PAM. Cas1 monomers contain a PAM-sensing region and Cas1 mediated processing of pre-spacers creates the 3'OH ends required for nucleophilic attack (Nuñez et al. 2014, Wang et al. 2015). Naïve adaptation requires active DNA replication or active transcription and majority of protospacers are non-randomly distributed with many acquired around the origin of replication (*oriC*), terminus (*ter*), CRISPR, rDNA loci, R-loops – specific regions known to experience DNA nicking or double-strand breaks. *E. coli* naïve adaptation is stimulated by RecBCD enzyme during the repair of double-stranded breaks (DSB) that may arise from stalled replication forks (21). RecBCD is thought to aid naïve adaptation by generating single-stranded DNA (ssDNA) intermediates from helicase and nuclease activities before reaching a *Chi* site (5'-GCTGGTGG-3') that attenuates these activities. In this model ssDNA generated by RecBCD nuclease re-anneals into partial duplex that is a substrate for Cas1-Cas2 (21). During naïve adaptation integration of host fragments as new spacers occurs but spacer integration from a plasmid MGE is more frequent (21,22) (23). The frequency of new MGE DNA spacers derived from the *E. coli* chromosome were about 10-fold higher in *recB*, *recC* and *recD* mutants compared to the *wt* strain suggesting that RecBCD also helps in self/non-self discrimination, or that DNA

substrates generated in these mutant backgrounds are particular targets for capture during adaptation. In this work, we analysed involvement of RecBCD and other host nucleases in naïve adaptation using genetic analysis. This indicated that (a) nuclease activity of RecBCD is not required for adaptation, (b) helicase, or other, activity of RecBCD promotes adaptation, and (c) recombination by RecA that is stimulated by RecBCD inhibits adaptation. We also show that purified Cas1-Cas2 complex can act as a nuclease with specificity for a 5' ssDNA tailed duplexes, substrates that genetics implied are important for stimulating adaptation.

MATERIALS AND METHODS

Strains, plasmids, media and general methods

E. coli strains used are described in Supplementary Table 1. Mutant bacterial strains were made by P1 *vir* transduction and selected for the appropriate antibiotic resistance. Antibiotic resistance genes were eliminated using pCP20 (24). Bacteria were grown at 37 °C in LB broth (10 g/L bacto-tryptone, 5 g/L yeast extract, 10 g/l NaCl) and on LB agar plates (supplemented with 15g of agar for solid media). When required appropriate antibiotics were added to LB plates at final concentrations: ampicillin at 100 μg/ml, kanamycin at 40 μg/ml, apramycin 30 μg/ml, tetracycline 10 μg/ml, spectinomycin 100 μg/ml, trimethoprim 100 μg/ml and chloramphenicol at 15 μg/ml. Plasmids used were pBad-HisA (Invitrogen) as an empty plasmid vector control and pEB628 for arabinose inducible expression of Cas1-Cas2 from pBad-HisA described in (20).

Naïve adaptation assay and plasmid instability

New spacer acquisition into a CRISPR locus by naïve adaptation was assessed by the procedure described in (4,20,25). Cells lacking chromosomally encoded Cas3, Cascade and Cas1-Cas2 were transformed by pEB628 (pCas1-Cas2) or pBad-HisA and individual transformants were inoculated in LB broth. Expression of Cas1-Cas2 was induced by addition of 0.2% (w/v) L-arabinose. Cells were aerated at 37 °C for 16 hours and then sub-

cultured (“passaged”) up to three times by diluting 1:300 the previous overnight culture into fresh LB with arabinose. Spacer acquisition was monitored by PCR using primers detailed in (20) followed by agarose gel electrophoresis on 2 % agarose gels stained using sybr safe. Template DNA was prepared from bacterial cultures by boiling in water. Relative band intensities for spacer acquisition quantification were measured using Kodak 1D Image Analysis Software v. 3.6.0. This software detected bands containing no spacer automatically, while the spacer containing bands were manually marked by a rectangle. The rectangle was used to mark all of the bands, including the bands of the negative control lanes, i.e. the PCR products of strains transformed with the empty vector pBad. In this way, the relative intensity values of bands were calculated by subtracting values with pBad from the corresponding values of the same strain with pCas1-Cas2. At least two independent experiments were done for each strain.

Each passage of naïve adaptation was also analysed for instability of pBad or pEB628 by viability “spot” tests of cell survival on ampicillin agar. Cells were serially diluted in 67 mM phosphate buffer (pH = 7.0) and 10 µl aliquots were spotted onto LB and LB with ampicillin plates for incubation overnight at 37 °C. Cells having lost the plasmid gave lower viable counts on ampicillin plates in comparison to LB plates. We also studied the plasmid presence in cells grown to log phase (OD_{600} = 0.5) in the presence of L-arabinose and antibiotic ampicillin. Cells were also serially diluted and analysed as above.

Spacer acquisition analysis and mapping

Spacer aquisition experiments for strains IIB1165 (*wt*), IIB1214 (*recB1080*) and IIB1245 (*recD recA)* were assessed from cells grown as described above. Cells were “passaged” two times for strains IIB1165 and IIB1245 and only once for IIB1214 (two biological replicas). PCR products that correspond to expanded CRISPR array were gel purified with Promega Wizard SV Gel and PCR Clean Up System. Sequencing was performed on Illumina Miniseq platform in 2x150 paired end mode. R packages ShortRead and BioString were utilized for

reads pre-processing and downstream analysis, mapping and mapping visualization. During pre-processing reads with Phred quality score of less than 20 were trimmed, and reads with two or more CRISPR repeats were filtered. Sequences between two CRISPR repeats determined with two mismatches allowed were extracted as spacers. Spacers were mapped first to the plasmid (unique mapping for plasmid locations) and those that did not match the plasmid were mapped to the genome, non-unique matches were discarded. Disregarding quantities (every spacer counts only once) were applied for statistical analysis of spacer distribution.

Protein Purification

Cas1 and Cas2 proteins were over-expressed individually according to the method described in (20) generating Cas1 with an N-terminal $(His)_6$-tag and untagged Cas2. Cell biomass for over-expression was thawed, sonicated and clarified. The resulting lysates were combined and mixed for 2 hours at 4 ^{o}C. This allows purification of stable Cas1-Cas2 complex that is identifiable in gel filtration and elutes separately from either Cas1 or Cas2 alone (Supplementary Figure S1A), and which is active *in vitro* for catalysing half- and full-site integration of duplex DNA into a CRISPR locus (Supplementary Figure S1B). Cas1-Cas2 was bound to a 5 ml HiTrap Chelating column (GE Healthcare) charged with Nickel. Unbound protein was washed with buffer A (20 mM Tris pH7.5, 500 mM NaCl, 20 mM imidazole, 10 % Glycerol) with bound protein eluted using a linear gradient of 20 – 500 mM Imadazole over 25 ml. Following dialysis in buffer B (20 mM Tris pH7.5, 150 mM NaCl, 1 mM DTT, 10 % Glycerol) Cas1-Cas2 was further purified using a 1 ml HiTrap Heparin HP column (GE Healthcare), washed with buffer B and eluted using a linear gradient 150 mM – 1M NaCl. Separation of Cas1-Cas2 from unbound Cas1 was achieved by elution from an Superdex 200 Increase 10/300 GL (GE Healthcare) using Buffer C (20 mM Tris pH 7.5, 150 mM KCl, 20 % Glycerol, 1 mM DTT) prior to storage at -80 ^{o}C.

Genes encoding *E.coli* IHF α and β subunits were PCR amplified using the primers listed in supplementary data for cloning into pACYCduet using sites for restriction endonucleases

BamHI/NotI and XhoI/AvrII respectively. Co-expression of IHF subunits was in *E. coli* BL21AI cells grown at 37 ^{o}C to O.D.600 of 0.6 followed by induction with 0.2 % L-arabinose and 0.5 mM IPTG with growth continued overnight at 18 ^{o}C. Harvested cells were resuspended in Buffer J (500 mM KCl, 20 mM Hepes pH7.5, 20 mM Imidazole, 0.1 % Triton x-100, 10 % glycerol) plus 1 x protease inhibitor cocktail tablet (EDTA free) (Roche). IHF subunits were co-purified using a 5 ml HiTrap Chelating column (GE Healthcare) charged with Nickel. Unbound protein was washed with Buffer J and bound protein eluted in an isocratic elution Buffer J plus 500 mM Imidazole. Eluted protein was dialysed overnight at 4 ^{o}C in Buffer K (150 mM KCl, 20 mM Hepes pH7.5, 0.1 % Triton x-100, 10 % glycerol), followed by further purification using a 1 ml HiTrap Heparin HP column (GE Healthcare), washed with Buffer K and eluted using a linear gradient of 150 mM – 1M KCl. Fractions containing both subunits were pooled and flash frozen for storage at -80 ^{o}C.

DNA substrates and Cas1-Cas2 EMSA and DNA nicking assays

Sequences of DNA oligonucleotides and the substrates generated for this work are presented in Supplementary Figure S2. Substrates were 5'- Cy5-end labelled for visualisation in gels. EMSAs to assess binding of Cas1-Cas2 to tailed duplex DNA molecules were in 5% acrylamide TBE gels, after mixing at 37°C for 30 minutes Cas1-Cas2 and DNA (20 nM) in Buffer HB (20 mM Tris.HCl pH 8.0, 100 μg/mL bovine serum albumin, 7% glycerol) and loaded directly onto the gels. Gels were electrophoresed for 1.5 hours at 120 volts. DNA cutting activity of Cas1-Cas2 was assessed in 15% TBE gels containing 8 M urea. Cas1-Cas2 was mixed with 20 nM DNA and buffer HB with addition of magnesium chloride (10 mM) for incubation at 37°C for 60 min. Reactions were stopped by adding proteinase K and EDTA for loading heated samples onto urea gels in formamide loading buffer.

RESULTS

Genetic analysis of RecBCD nuclease activity in naïve adaptation.

In current models of naïve adaptation in *E. coli* RecBCD nuclease activities that promote DNA repair by homologous recombination also generate DNA for capture by Cas1-Cas2, leading to adaptation (21). In previous work (20) it was demonstrated that *recB* was required for wild-type levels of naïve adaptation but *recA* was not, indicating that naïve adaptation is independent of RecA catalysed recombination. To better understand this, given that a major role for RecBCD in DNA repair is to load RecA, we carried out detailed genetic analysis using multiple alleles of RecBCD and assessed naïve adaptation. Naïve adaptation was detected by expansion of the CRISPR-1 locus in an *E. coli* K-12 strain that lacks functioning chromosomal Cas proteins (Supplementary Table S1) but has the chromosomal CRISPR-1 locus and expresses Cas1 and Cas2 from an inducible plasmid, summarised in Figure 1A. Acquisition of new spacer DNA was clearly visible in wild type cells after three passages of growth.

Compared with wild type *E. coli* cells, naïve adaptation was severely reduced or undetectable in cells inactivated for *recD* or *recB* in end point assays (Figure 1A) or when tested over three growth passages (Figure 1B and additional data in Supplementary Results S2). These results are in agreement with a model in which RecBCD nuclease activity is important for naive adaptation in *E. coli* (21) because neither *recB* or *recD* cells possess RecBCD nuclease activity. However, two further genetic traits of *recB* and *recD* cells were assessed, the effect of RecA loading onto DNA and plasmid stability, because they potentially impact on naive adaptation.

RecBC enzyme in cells inactivated for *recD* is a nuclease-free helicase that constitutively loads RecA onto 3’ ssDNA to initiate recombination (26). We observed that naïve adaptation was restored to measurable levels similar to wild type when *recA* was also removed to generate a *recD recA* double mutant background (Figure 1B and 1C and Supplementary Results Table S3). As established in previous work (20), deletion of *recA* alone has no discern-

able effect on naïve adaptation. Interestingly, in these assays naïve adaptation was not readily restored to *recB recA* cells (Figure 1B) that lack both RecBCD nuclease and helicase activity. Analysis of adaptation in *recB recA* cells using further iteration of PCR did detect some new spacer product but at significantly reduced efficiency compared to wild type cells (Supplementary Results Table S2). However, reduced adaptation associated with *recB recA* cells suggested that helicase activity, unlike nuclease activity, of RecBCD does promote naïve adaptation. Analysis of naïve adaptation in cells carrying the RecBCD allele *recB1080* further supported that RecBCD nuclease activity is dispensable for naïve adaptation (Figure 1C). This mutation encodes RecB^{1080A}CD protein that lacks nuclease activity and RecA loading, but helicase activity is retained (27,28). Spacer acquisition in *recB1080* cells after a single passage was comparable to wild type cells (Figure 1C and Supplementary Table S3) but dropped away in passages two and three due to plasmid instability compared to wild type cells (Table S4). In summary, the genetic analyses indicate that cells lacking RecBCD nuclease activity are proficient at naïve adaptation.

These assays for naïve adaptation were measured over three passages to account for plasmid instability that is associated with *recBCD* mutations in *E. coli* (29). Elimination of the Cas1-Cas2 plasmid results in loss of adaptation over time in these genetic backgrounds, for example as was observed in the third passage of *recD recA* cells (Figure 1B and 1C and Supplementary data Table S3). Full measurements of plasmid instability correlating to adaptation are presented in Supplementary data Table S4. It is significant that naïve adaptation in *recD recA* cells was readily detectable in passage 2 even though instability of plasmid expressing Cas1-Cas2 resulted in its loss with >200 – fold greater frequency compared to in wild type cells (Supplementary Table S4).

High throughput sequencing of DNA in extended CRISPR arrays identified that newly acquired spacers mapped to plasmid and genomic DNA and that no strand bias was detected, as expected for naive adaptation. Our analysis identified that most spacers (79 - 90%) originated from the *E. coli* chromosome in wild type and RecBCD/RecA mutant strains

compared to acquistion from plasmid pEB628 that was used for expression of Cas1-Cas2 (Supplementary Figure S3A). Close examination of the pattern of spacer mapping onto the chromsome highlighted that in all cells analysed 3 – 4 times more newly acquired spacers originated from Origin (*ori*) and termination (*ter*) regions of the chromosome relative to the reference genomic region spanning the same distance (670 kb, Figure 1D). *recB1080* cells were associated with >10 times more new spacers being acquired from *ter* sites, an effect not observed for *recD recA* cells (Figure 1D). These observations might be explained by loss of RecBCD functionality triggering accumulation of aberrant or unprocessed intermediate DNA structures arising during replication termination or recombination (29-31). Information for accessing raw DNA sequencing data underlying these results is given at the end of this manuscript.

The effect of exonucleases on naïve adaptation in *E. coli*.

We investigated if naïve adaptation was supported by nucleases other than RecBCD by testing if new spacer acquisition was affected by inactivating *E. coli* exonucleases that promote genome stability (32,33). Inactivation of individual 3' to 5' ssDNA exonucleases SbcB (also called ExoI), ExoVII (XseA subunit of XseAB complex), SbcCD or ExoX did not impinge on adaptation over three passages (Figure 2Ai, 2B and Supplementary Table S3) and combining these with inactivation of *recD* deletion gave cells that remained unable to acquire new spacers like the *recD* deletion alone (Supplementary Figure S3). Restoration of adaptation in *recD recA* cells (Figure 1) was used to assess if any of the 3' to 5' ssDNA nucleases are required for adaptation, which would manifest as reduced spacer acquisition by inactivating the nuclease in *recD recA* cells. Deletion of *xseA* (*exoVII*) in *recD recA* cells had little effect on adaptation over three passages compared to *recD recA* cells (Figure 2Aii and Supplementary Table S3) and plasmid instabilities associated with these strains were similar (Supplementary Table S4), indicating no effect of *xseA* in this context. Deletion of *sbcB, sbcD* or *exoX* in *recD recA* cells all gave significantly reduced adaptation compared to *recD recA* cells in all passages (Figure 2Aii and Figure 2C), but this correlated to 10-fold increased

plasmid instability (Supplementary Table S4). Therefore, it is likely that reduced adaptation by inactivation of these nucleases is caused by loss of Cas1-Cas2 encoding plasmids in these assays. To determine if these exonucleases are required for adaptation when RecBCD enzyme is functional we inactivated them in combination with the *recA* mutation only. Adaptation was not affected in *sbcD recA, exoX recA* or *sbcB recA* cells compared to wild type cells (Figure 2Aiii), and these cells showed much improved plasmid stability (Supplementary Table S4). Overall these results indicate that naïve adaptation does not require these 3' ssDNA exonucleases.

We investigated if 5' to 3' ssDNA exonuclease activities of RecJ and ExoVII (encoded by *xseAB*) influence naïve adaptation in *E. coli*. Adaptation was proficient after inactivation of *recJ* or *xseA* or both (Figure 3A and Supplementary Table S3) but could not be detected in *recD recJ/xseA* cells, as expected because of the dominant negative effect of the *recD* mutation (Supplementary Figure S3C). In contrast to results from the 3' ssDNA exonucleases, when *recA recD* cells were used to unmask any effect on adaptation of 5' to 3' exonucleases we observed that inactivation of *recJ* and *xseA (xseA recJ recD recA* cells) significantly increased new spacer acquisition compared to wild type and *xseA recJ recA* cells (Figure 3). This suggested that functioning RecJ and ExoVII have a negative effect on naïve adaptation that is alleviated by removing them, implying that DNA molecules with 5' ssDNA tails stimulate naïve adaptation.

Cas1-Cas2 complex binds to and nicks 5'-tailed partial duplexes

Genetic analyses implied that DNA duplexes with 5' ssDNA tails promote naïve adaptation. We used purified *E. coli* Cas1-Cas2 complex (Supplementary Figure S1A) that is proficient in catalysing new spacer integration *in vitro* (Supplementary Figure S1B), for investigating binding and processing of ssDNA tailed substrates in potential DNA capture events (Figures 4 and 5). Previous work showed that Cas1-Cas2 stably bound to fork and other branched DNA molecules that might be explained by their resemblance to half-site intermediates formed during Cas1-Cas2 catalysed integration reactions but which may not be relevant to

DNA capture (20). Cas1-Cas2 binding and catalysis was therefore assessed on duplex DNA molecules with ssDNA tails that cannot undergo spacer integration reactions.

Cas1-Cas2 bound to 3'- and 5'-ssDNA tailed molecules with 10-base-pair duplex regions and 40 nucleotides of ssDNA, but not to a corresponding fully base-paired duplex (Figure 4A). Binding of Cas1-Cas2 to tailed duplexes in these EMSAs included significant protein-DNA aggregation in gel wells, but a stable protein-DNA complex could be discerned from binding to the 5'-ssDNA tailed 10 bp duplex ("DNA-10") in addition to protein aggregates (Complex-1 in Figure 4A lanes 2 and 3). This Cas1-Cas2 binding pattern with DNA-10 was also seen in control reactions binding Cas1-Cas2 to a duplex DNA that was previously optimised for productive integration reactions (Supplementary Figure S4)((11,12)). However, Cas1-Cas2 complex formation in EMSAs was significantly improved by increasing the length of the duplex region of the 5' ssDNA tailed duplexes to 14 base pairs (Figure 4B, "DNA-14").

Interestingly, Cas1-Cas2 cut the DNA backbone in the same 5' ssDNA substrates that were bound in EMSAs, summarised in Figure 5A for substrates DNA-13, -14 and -15 that gave maximal activity of Cas1-Cas2 (up to 14% of DNA cut). Cas1 protein alone did not cut DNA-14, on which Cas1-Cas2 was most active (Figure 5B) indicating that active adaptation "capture complex" (12) is needed for DNA cutting. The equivalent 3' ssDNA substrate was not cut by Cas1-Cas2 complex (Supplementary Figure S5). Major products of Cas1-Cas2 DNA cutting DNA-10, -13, -14, or -15 (products A and B) were mapped to within ssDNA one nucleotide from AAC sequence (Figure 5B and Supplementary Figure S6), which is recognised as an *E. coli* PAM (34). To determine if this sequence was prerequisite for DNA cutting by Cas1-Cas2 we altered it to TTT in DNA-14, but this had little effect on product formation (Figure 5C). The results suggest that DNA structure (ssDNA and position of cut site relative to duplex DNA) may be important dictating efficacy of DNA cutting in these substrates. The *in vitro* activity of purified Cas1-Cas2 complex is compatible with observation from genetics that 5' ssDNA tailed duplexes are important as substrates for adaptation and may be bound and cut by Cas1-Cas2 for DNA capture.

DISCUSSION

CRISPR-Cas immunity in *E. coli* is established by naïve adaptation that involves capture of DNA fragments for integration into CRISPR loci by the Cas1-Cas2 enzyme complex. Molecular processes that pre-process DNA leading to its capture by Cas1-Cas2 are poorly understood but require DNA repair systems, including activities of RecBCD nuclease-helicase. Genetic analysis presented here challenges the current model that nuclease functions of RecBCD generate DNA that can be captured by Cas1-Cas2 (21). The genetic data show that *recB1080* and *recD recA* cells that lack RecBCD nuclease activity were proficient at acquiring new spacers, even in the face of plasmid instability associated with these *recBCD* genotypes. Removing RecA from *recD* cells unmasked the adaptation proficiency by removing the inhibitory effect of recombination on adaptation. Interestingly, *recB recA* cells acquired new spacers much less well than wild type cells, implicating an alternative activity of RecBCD in supporting naïve adaptation, most likely helicase function but in agreement that RecBCD is required in some way for naïve adaptation in *E. coli*.

RecBCD binds preferentially to duplex DNA ends (35,36), resects them into DNA fragments depending on prevailing buffer conditions (e.g. availability to the nuclease active site of metal ions and DNA) and on the translocation rate of helicase sub-units, but RecBCD helicase and nuclease activities are not dependent on one another (37,38). Helicase and nuclease functions are modulated when RecBCD encounters *Chi* DNA sequence, and together these events promote DNA repair by homologous recombination because they initiate RecA loading by RecBCD onto 3' tailed ssDNA (39). However, the genetic data presented here suggest that functions of RecBCD in DNA repair by recombination are separate from how it promotes naïve adaptation: Critically, removal of RecA from cells, therefore removing the loading role of RecBCD in recombination, restored naïve adaptation.

It was significant that cells expressing *recB1080* (27) were adaptation proficient further indicating that RecBCD nuclease activity is not needed for naïve adaptation. $RecB^{1080}CD$ is a proficient helicase that translocates DNA with dual directionality 3' to 5' (RecB) and 5' to 3' (RecD) (40). The adaptation phenotypes associated with *recBCD* might indicate that DNA

pre-processing and capture for naïve adaptation requires DNA translocation unwinding associated with RecB. RecBCD is a powerful translocase that can clear DNA of RNA polymerase, nucleosome and other DNA bound proteins (41,42). We propose that RecBCD helicase-translocase activities are required for adaptation to disrupt or displace nucleoprotein complexes present at DNA capture sites to provide access to DNA for Cas1-Cas2 and generate substrates that can be acted on by Cas1-Cas2 for DNA capture (Figure 6). We observed that the majority of new spacers were acquired from the *E. coli* chromosome compared to the Cas1-Cas2 plasmid whether in cells with fully functional RecBCD or in RecBCD compromised cells. This differs from a previous study (21) in which spacers were mainly derived from plasmid depending on whether or not Cas1-Cas2 protein expressed was induced or not. The previous study used BL-21AI strain (*E. coli* B) while we used *E. coli* K-12, which could be the reason for the observed difference. Another study (23) reported that *P. furiosus* cells acquired 96-99% of the unique spacers from the chromosome compared to 1 – 4% of new spacers derived from a plasmid expressing Cas proteins.

If RecBCD nuclease activity is not needed for naïve adaptation, how is DNA fragmented for capture? The genetic data presented here and in previous work suggest that neither 3' ssDNA exonucleases nor 5' ssDNA exonucleases have significant roles in DNA pre-processing for adaptation by Cas1-Cas2. Instead, we propose that Cas1-Cas2 nuclease activity when targeted to DNA end structures with PAM sequences can result in protospacer DNA capture prior to new spacer integration. Nuclease activity of *E. coli* Cas1 has been detected previously on a variety of model branched DNA substrates (20,43). The observation from genetics that deletion of 5' ssDNA exonucleases in *recD recA* cells caused a significant improvement to naïve adaptation suggested that substrates for these enzymes (5' ssDNA tailed DNA) may resemble those targeted by Cas1-Cas2. Purified Cas1-Cas2 complex was able to bind and nick these substrates without a requirement for PAM sequence, in this case AAC, being present. Although DNA structures present at DNA replication termini are not determined, broken replication forks processed at DNA ends by RecBCD for repair by recombination inhibit adaptation. If recombination is unable to occur

because of mutations in RecBCD or RecA, or because Chi sequences are unavailable in foreign DNA then processing of DNA ends by alternative nucleases to RecBCD might promote Cas1-Cas2 activity at these sites, leading to DNA capture. Such an effect could explain the enrichment of new spacers acquired from replication termination sequences during naïve adaptation in *E. coli* (21). In *wt* cells, processing of broken replication forks involves asymmetric degradation of *ter*-oriented DNA ends (44) that may explain enrichment of new spacers from *ter* in these cells. In wild type cells Chi sequences place limits on spacer acquisition at *ter* regions (21) that seem to be released in *recBCD* mutants (Figure 1D). Structures of phage genomes during late rolling circle DNA replication form linear concatamers of DNA that include 5' ssDNA tailed regions for lagging strand DNA synthesis. These may be important for targeting by Cas1-Cas2 for DNA capture as part of establishing CRISPR immunity to a newly encountered MGE. Similarly, events at DNA replication termination sites potentially generate DNA ends and 5' ssDNA tailed DNA structures that are processed as part of the normal cell cycle by genome stability enzymes, including RecBCD (Figure 6). The 3' to 5' polarity of Cas3 DNA translocase activity would also generate 5' ssDNA tailed DNA if it acts as a helicase, which may be important for DNA capture in the context of CRISPR interference reactions (45). Further work will be needed to determine the molecular mechanisms of DNA capture during adaptation, in particular using *in vitro* reactions with defined components that couple DNA replication, DNA repair and CRISPR adaptation.

Acknowledgements

This work was supported from Croatian Science Foundation Grant IP-2016-06-8861, The Croatian Academy of Sciences and Arts, Biotechnology and Biological Sciences Research Council grant BB/M020541-1 to ELB and Russian Foundation for Basic Research grant 16-04-00767 to ES. We are grateful to The ERASMUS + mobility scheme for funding for LW, and to Dr. Christian Rudolph and Dr. Davor Zahradka for strains. We also thank students Valentina Petanjek, Vanja Jurić, Lara Šamadan and Dorotea Pali for practical assistance.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MR, TK, IIB, ES and LW performed the experiments, analysed and interpreted data. IIB and ELB designed the study, IIB and EB supervised experiments and IIB and EB wrote the manuscript. All authors read and approved the final version of the manuscript.

Data Availability

Updated DNA sequencing data for identifying newly acquired spacers is freely available from authors' ResearchGate pages:

https://www.researchgate.net/profile/Ekaterina_Savitskaya

https://www.researchgate.net/profile/Edward_Bolt

https://www.researchgate.net/profile/Ivana_Ivancic_Bace

And is also available as supplementary material to this manuscript.

References

1. Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I. and Koonin, E.V. (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct*, **1**, 7.
2. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709-1712.
3. Sternberg, S.H., Richter, H., Charpentier, E. and Qimron, U. (2016) Adaptation in CRISPR-Cas Systems. *Molecular cell*, **61**, 797-808.
4. Yosef, I., Goren, M.G. and Qimron, U. (2012) Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic acids research*, **40**, 5569-5576.
5. Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R. *et al.* (2012) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature structural & molecular biology*, **18**, 529-536.
6. Krivoy, A., Rutkauskas, M., Kuznedelov, K., Musharova, O., Rouillon, C., Severinov, K. and Seidel, R. (2018) Primed CRISPR adaptation in Escherichia coli cells does not depend on conformational changes in the Cascade effector complex detected in Vitro. *Nucleic acids research*, **46**, 4087-4098.
7. Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V. and van der Oost, J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960-964.

8. Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. and Siksnys, V. (2011) Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *Embo J*, **30**, 1335-1342.
9. Hochstrasser, M.L., Taylor, D.W., Bhat, P., Guegler, C.K., Sternberg, S.H., Nogales, E. and Doudna, J.A. (2014) CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 6618-6623.
10. Jackson, S.A., McKenzie, R.E., Fagerlund, R.D., Kieper, S.N., Fineran, P.C. and Brouns, S.J. (2017) CRISPR-Cas: Adapting to change. *Science*, **356**.
11. Nunez, J.K., Harrington, L.B., Kranzusch, P.J., Engelman, A.N. and Doudna, J.A. (2015) Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature*.
12. Nunez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W. and Doudna, J.A. (2014) Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nature structural & molecular biology*, **21**, 528-534.
13. Nunez, J.K., Lee, A.S., Engelman, A. and Doudna, J.A. (2015) Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature*, **519**, 193-198.
14. Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M. and Wang, Y. (2015) Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell*.
15. Nunez, J.K., Bai, L., Harrington, L.B., Hinder, T.L. and Doudna, J.A. (2016) CRISPR Immunological Memory Requires a Host Factor for Specificity. *Molecular cell*, **62**, 824-833.
16. Yoganand, K.N., Sivathanu, R., Nimkar, S. and Anand, B. (2017) Asymmetric positioning of Cas1-2 complex and Integration Host Factor induced DNA bending guide the unidirectional homing of protospacer in CRISPR-Cas type I-E system. *Nucleic acids research*, **45**, 367-381.
17. Wright, A.V., Liu, J.J., Knott, G.J., Doxzen, K.W., Nogales, E. and Doudna, J.A. (2017) Structures of the CRISPR genome integration complex. *Science*, **357**, 1113-1118.
18. Rollie, C., Schneider, S., Brinkmann, A.S., Bolt, E.L. and White, M.F. (2015) Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *Elife*, **4**.
19. Xiao, Y., Ng, S., Nam, K.H. and Ke, A. (2017) How type II CRISPR-Cas establish immunity through Cas1-Cas2-mediated spacer integration. *Nature*, **550**, 137-141.
20. Ivancic-Bace, I., Cass, S.D., Wearne, S.J. and Bolt, E.L. (2015) Different genome stability proteins underpin primed and naive adaptation in E. coli CRISPR-Cas immunity. *Nucleic acids research*, **43**, 10821-10830.
21. Levy, A., Goren, M.G., Yosef, I., Auster, O., Manor, M., Amitai, G., Edgar, R., Qimron, U. and Sorek, R. (2015) CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature*, **520**, 505-510.
22. Staals, R.H., Jackson, S.A., Biswas, A., Brouns, S.J., Brown, C.M. and Fineran, P.C. (2016) Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. *Nature communications*, **7**, 12853.
23. Shiimori, M., Garrett, S.C., Chambers, D.P., Glover, C.V.C., 3rd, Graveley, B.R. and Terns, M.P. (2017) Role of free DNA ends and protospacer adjacent motifs for CRISPR DNA uptake in Pyrococcus furiosus. *Nucleic acids research*, **45**, 11281-11294.
24. Cherepanov, P.P. and Wackernagel, W. (1995) Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene*, **158**, 9-14.
25. Yosef, I., Goren, M.G., Kiro, R., Edgar, R. and Qimron, U. (2011) High-temperature protein G is essential for activity of the Escherichia coli clustered regularly interspaced short palindromic repeats (CRISPR)/Cas system. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 20136-20141.
26. Churchill, J.J., Anderson, D.G. and Kowalczykowski, S.C. (1999) The RecBC enzyme loads RecA protein onto ssDNA asymmetrically and independently of chi,

resulting in constitutive recombination activation. *Genes & development*, **13**, 901-911.
27. Anderson, D.G. and Kowalczykowski, S.C. (1997) The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a chi-regulated manner. *Cell*, **90**, 77-86.
28. Ivancic-Bace, I., Vlasic, I., Salaj-Smic, E. and Brcic-Kostic, K. (2006) Genetic evidence for the requirement of RecA loading activity in SOS induction after UV irradiation in Escherichia coli. *Journal of bacteriology*, **188**, 5024-5032.
29. Wendel, B.M., Courcelle, C.T. and Courcelle, J. (2014) Completion of DNA replication in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 16454-16459.
30. Yao, N.Y. and O'Donnell, M.E. (2018) Replication fork convergence at termination: A multistep process. *Proceedings of the National Academy of Sciences of the United States of America*, **115**, 237-239.
31. Dimude, J.U., Midgley-Smith, S.L., Stein, M. and Rudolph, C.J. (2016) Replication Termination: Containing Fork Fusion-Mediated Pathologies in Escherichia coli. *Genes*, **7**.
32. Dermic, E., Zahradka, D., Vujaklija, D., Ivankovic, S. and Dermic, D. (2017) 3'-Terminated Overhangs Regulate DNA Double-Strand Break Processing in Escherichia coli. *G3*, **7**, 3091-3102.
33. Lovett, S.T. (2011) The DNA Exonucleases of Escherichia coli. *EcoSal Plus*, **4**.
34. Leenay, R.T., Maksimchuk, K.R., Slotkowski, R.A., Agrawal, R.N., Gomaa, A.A., Briner, A.E., Barrangou, R. and Beisel, C.L. (2016) Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. *Molecular cell*, **62**, 137-147.
35. Taylor, A. and Smith, G.R. (1985) Substrate specificity of the DNA unwinding activity of the RecBC enzyme of *Escherichia coli*. *J. Mol. Biol.*, **185**, 431-443.
36. Roman, L.J. and Kowalczykowski, S.C. (1989) Characterization of the helicase activity of the *Escherichia coli* recBCD enzyme using a novel helicase assay. *Biochemistry*, **28**.
37. Taylor, A. and Smith, G.R. (1980) Unwinding and rewinding of DNA by the RecBC enzyme. *Cell*, **22**, 447-457.
38. Telander-Muskavitch, K.M. and Linn, S. (1982) A unified mechanism for the nuclease and unwinding activities of the *recBC* enzyme of *Escherichia coli*. *J. Biol. Chem.*, **257**, 2641-2648.
39. Dillingham, M.S. and Kowalczykowski, S.C. (2008) RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiology and molecular biology reviews : MMBR*, **72**, 642-671, Table of Contents.
40. Dillingham, M.S., Spies, M. and Kowalczykowski, S.C. (2003) RecBCD enzyme is a bipolar DNA helicase. *Nature*, **423**, 893-897.
41. Terakawa, T., Redding, S., Silverstein, T.D. and Greene, E.C. (2017) Sequential eviction of crowded nucleoprotein complexes by the exonuclease RecBCD molecular motor. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, E6322-E6331.
42. Finkelstein, I.J., Visnapuu, M.L. and Greene, E.C. (2010) Single-molecule imaging reveals mechanisms of protein disruption by a DNA translocase. *Nature*, **468**, 983-987.
43. Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., Binkowski, A. *et al.* (2011) A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. *Mol Microbiol*, **79**, 484-502.
44. Wiktor, J., van der Does, M., Buller, L., Sherratt, D.J. and Dekker, C. (2018) Direct observation of end resection by RecBCD during double-stranded DNA break repair in vivo. *Nucleic acids research*, **46**, 1821-1833.
45. Kunne, T., Kieper, S.N., Bannenberg, J.W., Vogel, A.I., Miellet, W.R., Klein, M., Depken, M., Suarez-Diez, M. and Brouns, S.J. (2016) Cas3-Derived Target DNA

Degradation Fragments Fuel Primed CRISPR Adaptation. *Molecular cell*, **63**, 852-864.

Figure Legends

Figure 1. Genetic analysis of RecBCD in naïve adaptation. (A). Agarose gels summarizing PCR-based detection of *E. coli* CRISPR-1 expansion after integration of a new spacer (C+1) during naïve adaptation. Strains are indicated above each panel (wt, wild type) as are plasmids either pBad-HisA (ev, empty vector) or pEB628 for arabinose inducible Cas1-Cas2 (pCas1-2). Results from the third passage are presented. **(B).** Agarose gels summarizing CRISPR expansion in the *E. coli* strains indicated in three passages (p1 – p3) in all cases containing plasmid encoding inducible Cas1-Cas2 (pCas1-2). **(C).** Measurements of new spacer acquisition detectable as expansion of CRISPR-1 (C+1) using PCR of chromosomal DNA from the strains indicated. See also Table S3. Percentage spacer acquisition refers to intensity of C +1 DNA/(C+1 DNA + C DNA). Each strain indicated below the x-axis has three histograms representing measured adaptation in passage one (black), two (light grey) and three (light grey). **(D).** The relative quantities of spacers mapped to specified chromosomal regions. The spacers mapped onto 670 kb area spanning either Terminus (Ter) regions, CRISPR arrays (Cr) or Origin (Ori) regions were added and normalized to the number of spacers mapped to the *E. coli* chromosomal region spanning 0-670 kb.

Figure 2. Analysis of 3'→5' ssDNA exonucleases in naïve adaptation. (A). Graph summarizing measurements of new spacer acquisition in the strains indicated detectable as expansion of CRISPR-1 (C+1) using PCR of chromosomal DNA from the strains indicated. See also Table S3. Percentage spacer acquisition refers to intensity of C +1 DNA/(C+1 DNA + C DNA). Each strain indicated below the x-axis has three histograms representing measured adaptation in passage one (black), two (light grey) and three (light grey). (B and C). Agarose gel slices summarizing naïve adaptation effects shown for strains selected from

the graph. All strains contained the plasmid encoding inducible Cas1-Cas2 complex (pCas1-2).

Figure 3. Analysis of 5'→3' ssDNA exonucleases in naïve adaptation. (A). Graph summarizing measurements of new spacer acquisition in the strains indicated detectable as expansion of CRISPR-1 (C+1) using PCR of chromosomal DNA from the strains indicated. See also Table S3. Percentage spacer acquisition refers to intensity of C +1 DNA/(C+1 DNA + C DNA). Each strain indicated below the x-axis has three histograms representing measured adaptation in passage one (black), two (light grey) and three (light grey). (B). Agarose gels summarizing CRISPR expansion in the *E. coli* strains indicated in three passages (p1 – p3) in all cases containing plasmid encoding inducible Cas1-Cas2 (pCas1-2).

Figure 4. Comparative mobility shift analysis of Cas1-Cas2 binding to DNA substrates. (A). Electrophoretic mobility shift analysis of increasing concentrations of Cas1-Cas2 binding to 5' overhang, 3' overhang and duplex DNA as indicated. Oligonucleotide sequences used to prepare the substrates are shown in Supplementary Figure S2. Cy5 end labeled DNA substrates (20 nM) were incubated with 0, 31.25, 62.5, 125, 250, 500 nM Cas1-Cas2 complex for 30 minutes at 37 ^{o}C followed by analysis on a 5 % native acrylamide gel and imaged using a FLA3000 (FujiFilm). The graph shows quantified binding of Cas1-2 to 5' overhang (■), 3' overhang (▲) and duplex DNA (●) DNA substrates. Band quantification was carried out using ImageJ (NIH) as a normalized value of bound substrate as a percentage of total Cy5 fluorescence per lane, with error bars showing the standard error (n=3). **(B).** EMSAs comparing Cas1-Cas2 complex formation with 5'ssDNA tailed duplexes of varying lengths, as indicated. Assay conditions were the same as used in EMSAs in part A.

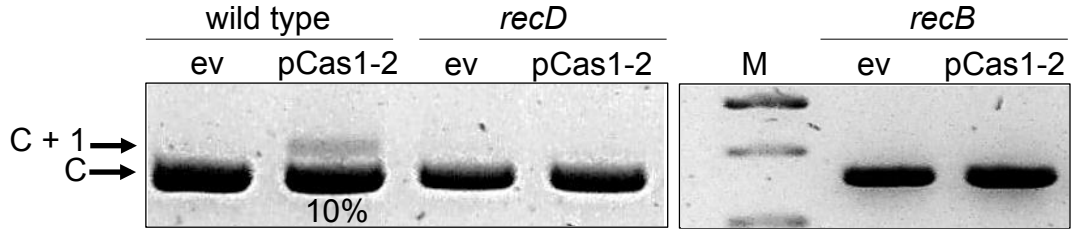
Figure 5. Nicking of DNA substrates by purified Cas1-Cas2 complex. (A). A summary of Cas1-Cas2 nicking activity on 5'-ssDNA tailed DNA duplexes, as indicated. Oligonucleotide

sequences used to prepare the substrates are shown in Supplementary Figure S2. Marker ssDNA nucleotide lengths are given to the left of the gel panel. Cy5 end labeled DNA substrates (20 nM) were incubated with 0 or 250 nM Cas1-Cas2 complex for 60 minutes at 37 ^{o}C, followed by analysis on a 15 % denaturing acrylamide gel and imaged using a FLA3000 (FujiFilm). Arrows indicate the major nicking products (A and B) generated by Cas1-Cas2. The graph shows cutting activity of Cas1-Cas2 complex (250 nM of total protein) on 5'-ssDNA tailed DNA duplexes (20 nM) as indicated, as a function of time. Reactions were in duplicate and error bars represent standard deviation from the mean values. Details of each substrate are given in Supplementary Figure S2. **(B).** Nuclease activity on DNA-14 (20 nM) of Cas1-Cas2 complex (0, 62.5, 125 and 250 nM) compared to the same assays containing only Cas1 at the same concentrations. The three DNA marker fragments are the same as Figure 5A and the major cutting product B is indicated. **(C).** Illustration of Cas1-Cas2 cutting sites identified in substrates (see also Supplementary Figure S6). The graph compares Cas1-Cas2 (250 nM) cutting activity, as a function of time, when mixed with DNA-14 and DNA-14-TTT, as indicated, and is plotted as means of two independent assays with standard deviation displayed as error bars

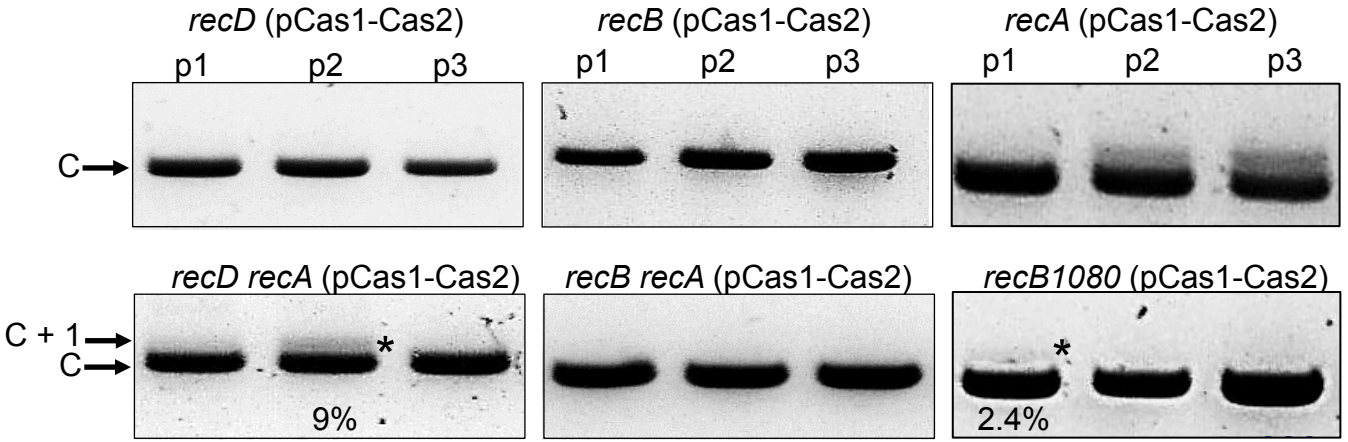
Figure 6. A model summarizing one way in which 5' ssDNA tailed duplexes can arise in an area of the genome (*Ter* sites) that is targeted for new spacer acquisition during CRISPR-Cas adaptation reactions, and is processed by RecBCD and other enzymes during the normal cell cycle. The role of RecBCD during replication termination is unclear but its helicase activity may contribute to removal of nucleoprotein roadblocks in this context. Similar DNA structures that may be targeted by Cas1-Cas2 could also arise during global DNA and repair of replication forks, and during lagging strand replication of phage in the later stages of its replicative cycle.

Figure 1

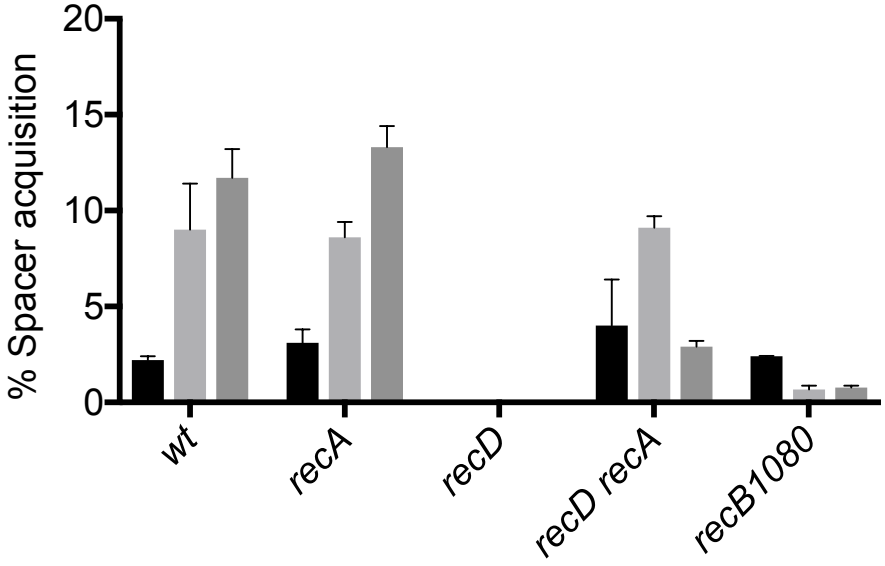
A.


wild type
recD
recB
ev
pCas1-2
ev
pCas1-2
M
ev
pCas1-2
C + 1
C
10%

B.


recD (pCas1-Cas2)
recB (pCas1-Cas2)
recA (pCas1-Cas2)
p1
p2
p3
C
recD recA (pCas1-Cas2)
recB recA (pCas1-Cas2)
recB1080 (pCas1-Cas2)
C + 1
C
*
9%
2.4%

C.


% Spacer acquisition
20
15
10
5
0
wt
recA
recD
recD recA
recB1080

D.

Figure 1

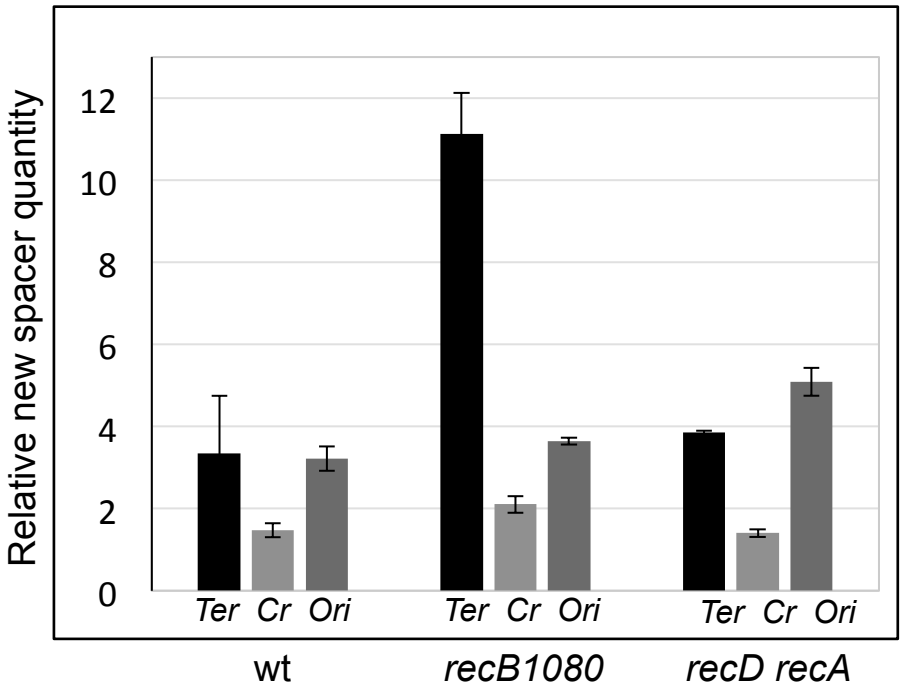

Relative new spacer quantity
12
10
8
6
4
2
0
Ter Cr Ori
Ter Cr Ori
Ter Cr Ori
wt
recB1080
recD recA

Figure 2

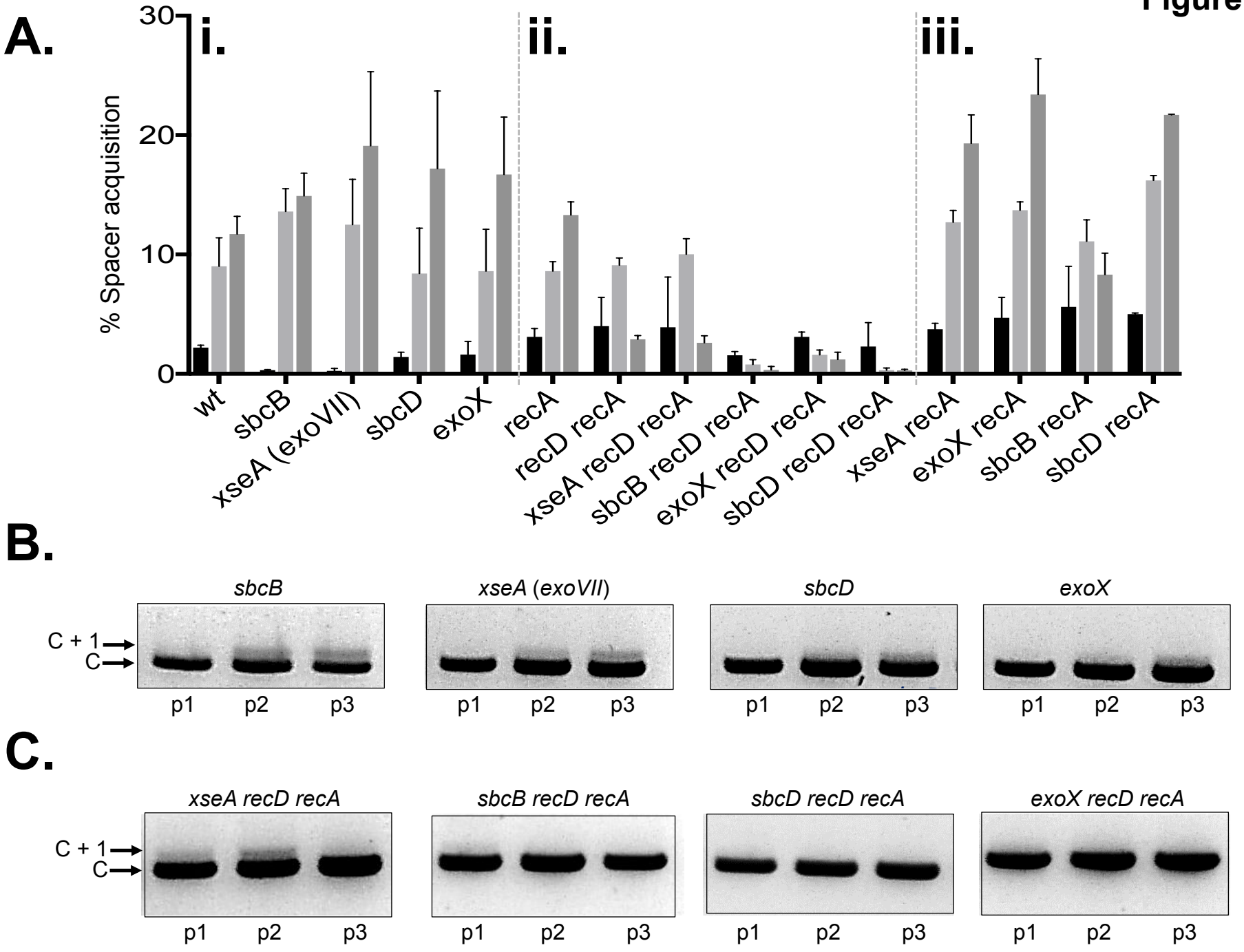
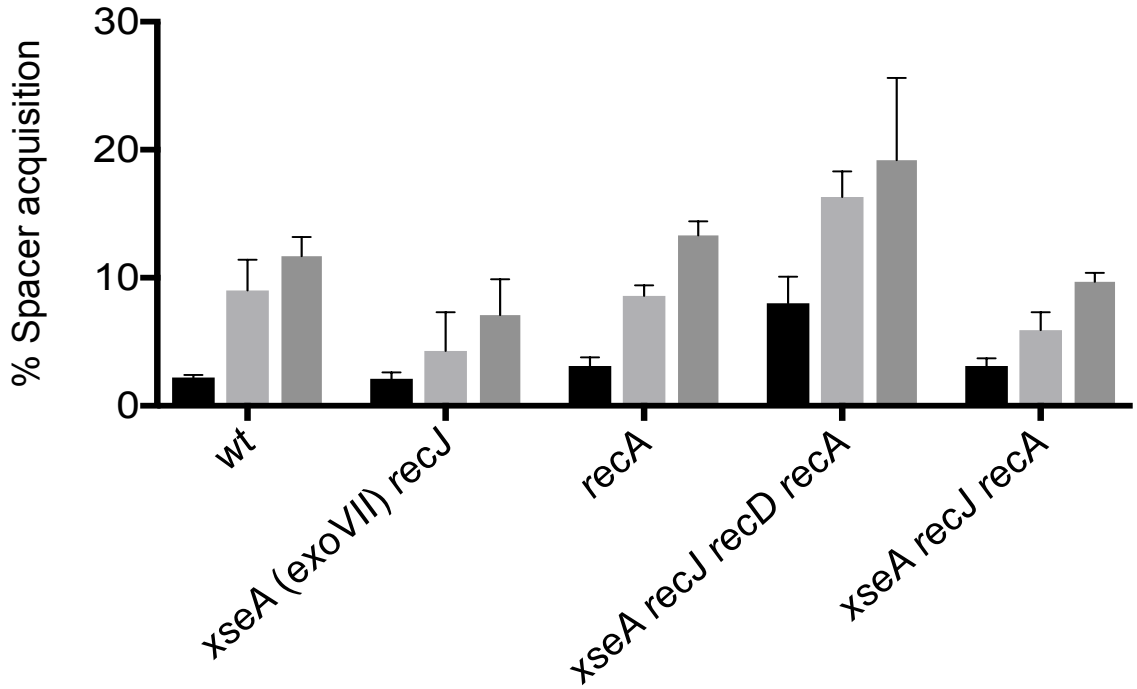

A.
i.
ii.
iii.
% Spacer acquisition
30
20
10
0
wt
sbcB
xseA (exoVII)
sbcD
exoX
recA
recD recA
xseA recD recA
sbcB recD recA
exoX recD recA
sbcD recD recA
xseA recA
exoX recA
sbcB recA
sbcD recA
B.
sbcB
xseA (exoVII)
sbcD
exoX
C + 1
C
p1
p2
p3
C.
xseA recD recA
sbcB recD recA
sbcD recD recA
exoX recD recA

Figure 3

A.


% Spacer acquisition
30
20
10
0
wt
xseA (exoVII) recJ
recA
xseA recJ recD recA
xseA recJ recA

B.

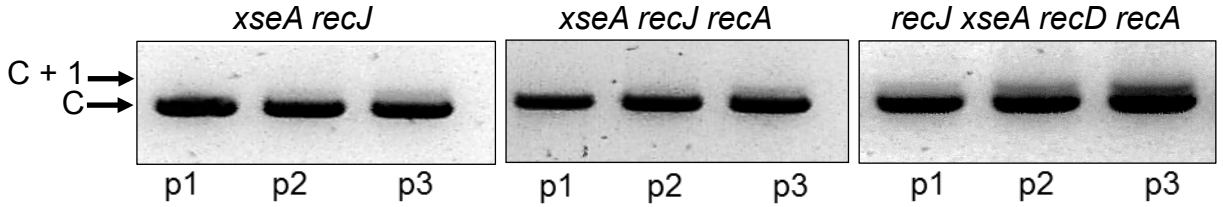
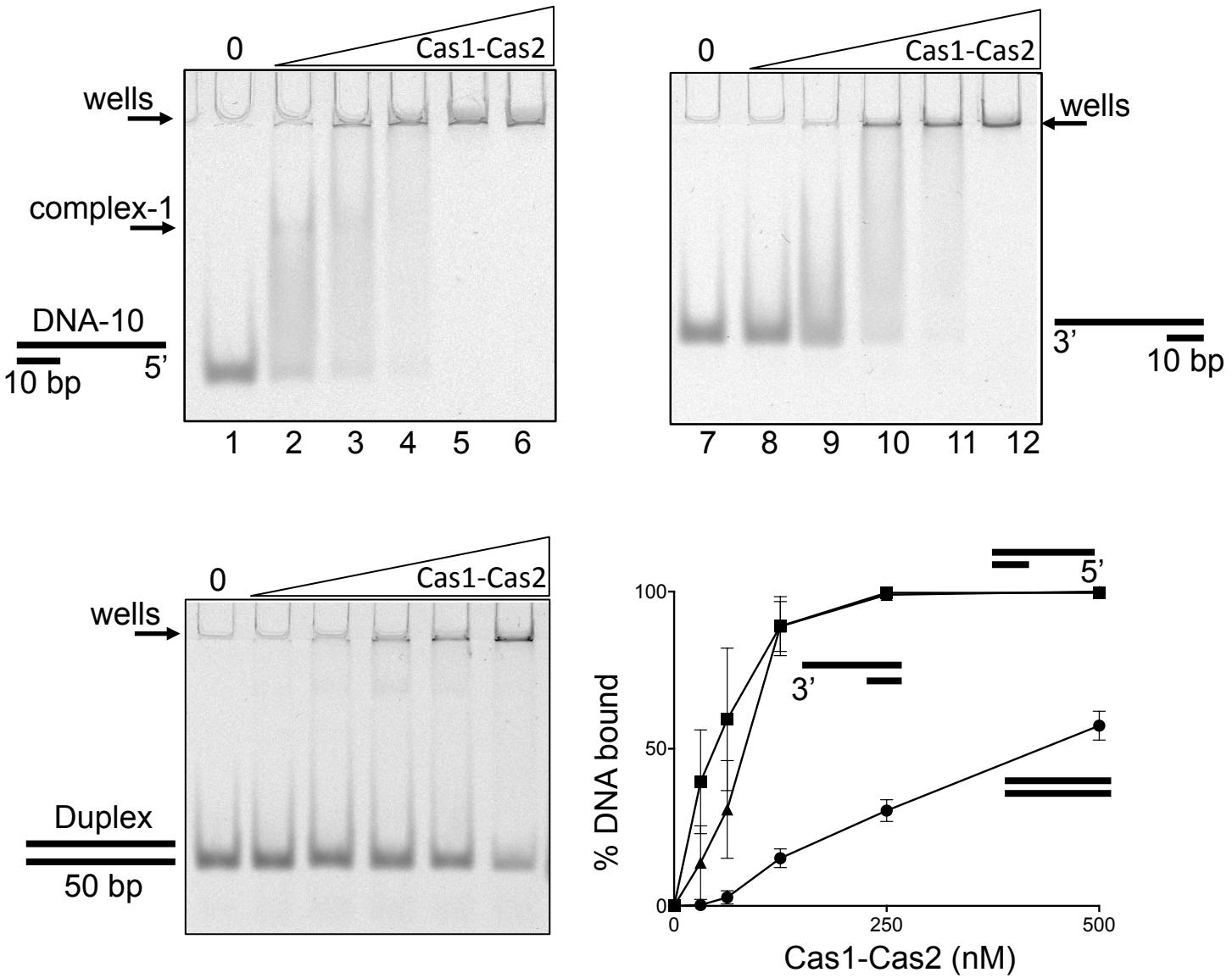

xseA recJ
xseA recJ recA
recJ xseA recD recA
C + 1
C
p1
p2
p3
p1
p2
p3
p1
p2
p3

Figure 4

A.


0
Cas1-Cas2
wells
complex-1
DNA-10
10 bp
5'
1 2 3 4 5 6
0
Cas1-Cas2
wells
3'
10 bp
7 8 9 10 11 12
0
Cas1-Cas2
wells
Duplex
50 bp
5'
3'
100
50
0
% DNA bound
0
250
500
Cas1-Cas2 (nM)

B.

Figure 4

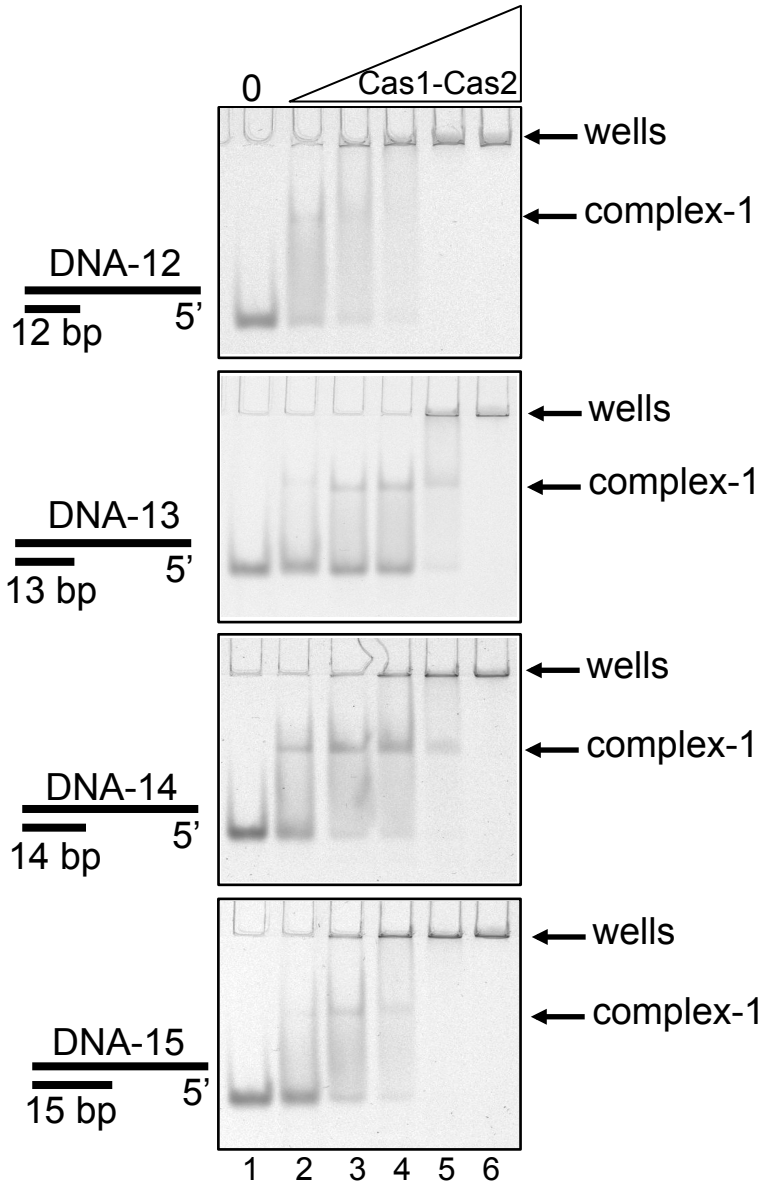
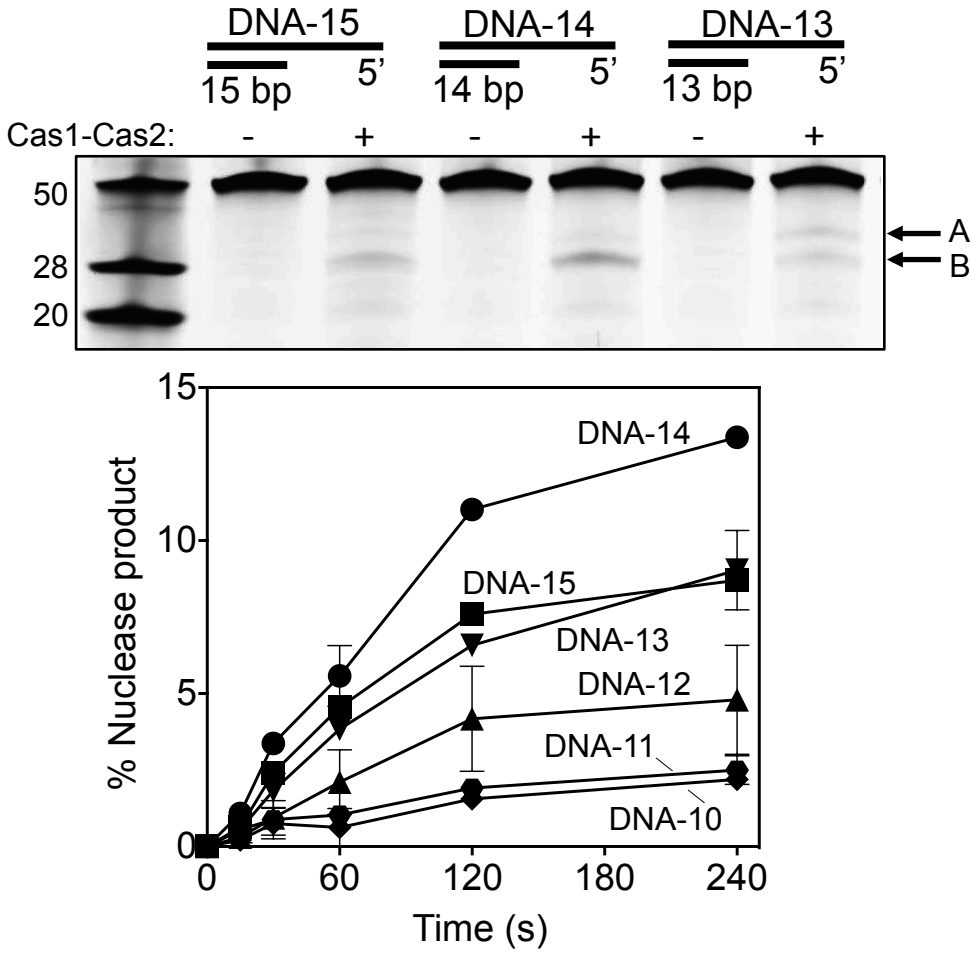

0
Cas1-Cas2
wells
complex-1
DNA-12
12 bp
5'
wells
complex-1
DNA-13
13 bp
5'
wells
complex-1
DNA-14
14 bp
5'
wells
complex-1
DNA-15
15 bp
5'
1 2 3 4 5 6

Figure 5

A.


DNA-15
15 bp
5'
DNA-14
14 bp
5'
DNA-13
13 bp
5'
Cas1-Cas2:
- + - + - +
50
28
20
A
B
% Nuclease product
Time (s)
0
5
10
15
0
60
120
180
240
DNA-14
DNA-15
DNA-13
DNA-12
DNA-11
DNA-10

B.

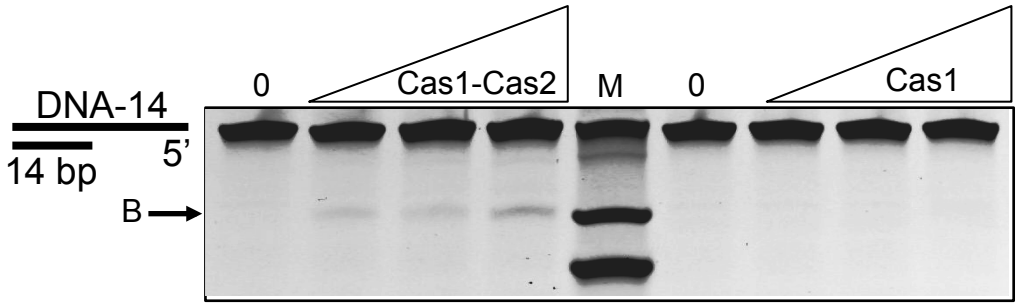
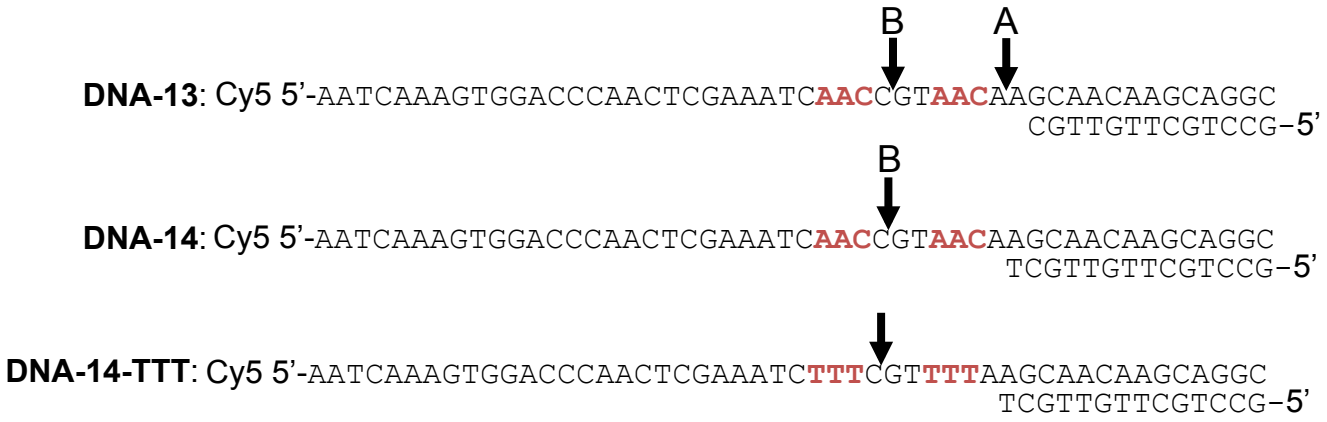

DNA-14
14 bp
5'
B
0
Cas1-Cas2
M
0
Cas1

Figure 5

C.


B
A
DNA-13: Cy5 5'-AATCAAAGTGGACCCAACTCGAAATCAACCGTAACAAGCAACAAGCAGGC
CGTTGTTCGTCCG-5'
B
DNA-14: Cy5 5'-AATCAAAGTGGACCCAACTCGAAATCAACCGTAACAAGCAACAAGCAGGC
TCGTTGTTCGTCCG-5'
DNA-14-TTT: Cy5 5'-AATCAAAGTGGACCCAACTCGAAATCTTTCGTTTTAAGCAACAAGCAGGC
TCGTTGTTCGTCCG-5'

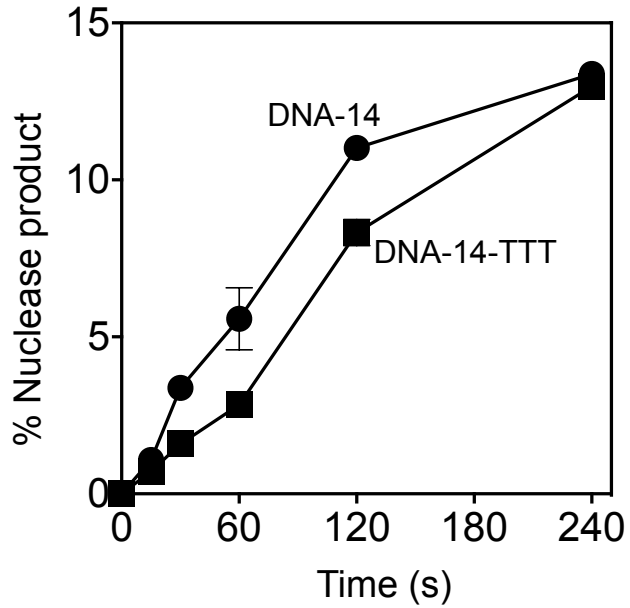
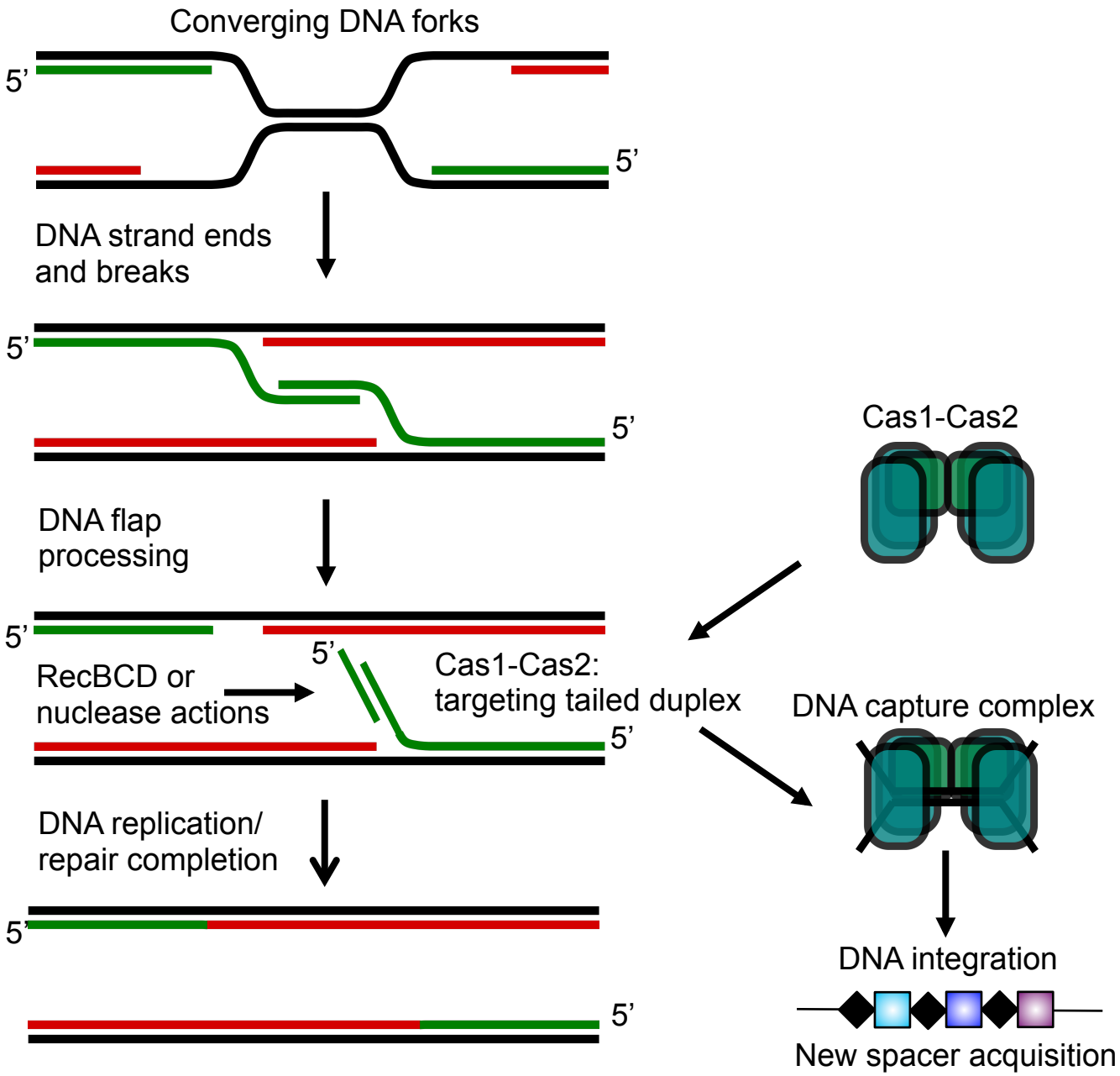

15
10
5
0
% Nuclease product
DNA-14
DNA-14-TTT
0
60
120
180
240
Time (s)

Figure 6


Converging DNA forks
5’
5’
DNA strand ends
and breaks
5’
5’
DNA flap
processing
5’
5’
RecBCD or
nuclease actions
Cas1-Cas2:
targeting tailed duplex
5’
Cas1-Cas2
DNA capture complex
DNA replication/
repair completion
5’
5’
DNA integration
New spacer acquisition

Supplementary Data

Table S1. *E. coli* strains used in this study listed below were derived from BW25113.

Bacterial strain	Relevant genotype	Source or reference
MG1655	+ F^- *rec*$^+$	Bachmann 1996
SLM1023	+ *xseA*::*dhfr*	C. Rudolph
IIB1151	+ *recD1903*::mini-Tn*10*	Laboratory collection
AM1986	+ Δ*recA1921*::*spec*	C. Rudolph
TH446	+ *recA*::*cam*	Laboratory collection
IIB360	+ *recB1080 argA*::Tn*10*	Ivančić-Baće et al. 2006
LMM1032	+ *recJ2052*::Tn*10kan*	D. Zahradka
LMM1247	+ *sbcD*::*kan*	D. Zahradka
N5288	+ *exoX1*::*npt*	C. Rudolph
JW1993-1	+ *sbcB780*::*kan*	D. Zahradka
BW25113	Δ(*araD-araB*)*567* Δ*lacZ4787*(::*rrnB-3*) λ$^-$ *rph*-1, Δ(*rhaD-rhaB*)*568 hsdR514*	Wanner BL
	Strains related to BW25113	
IIB892	+ Δ*cas3*::*apra* Δ*casC760*::*kan*^{S}	Ivančić-Baće et al. 2015
BW39183	+ Δ*cas1*::*kan*	Keio collection, E. Semenova
IIB1156	+ Δ*cas1*::*kan* Δ*cas3*::*apra* Δ*casC760*::FRT	P1. IIB892 x BW39183
IIB1157	+ Δ*cas1*::*kan* Δ*cas3*::*apra* Δ*casC760*::FRT *recD1903*::mini-Tn*10*	P1. IIB1151 x IIB1156
IIB1165	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT)	Removal of kan by pCP20
IIB1192	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *recD1903*::mini-Tn*10*	P1. IIB1151 x IIB1157
IIB1195	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *recD1903*::mini-Tn*10 recJ2052*::Tn*10kan*	P1. LMM1032 x IIB1192
IIB1199	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *recJ2052*::Tn*10kan*	P1. LMM1032 x IIB1165
IIB1207	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *recD1903*::mini-Tn*10 sbcD*::*kan*	P1. LMM1247 x IIB1192
IIB1208	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *recB268*::Tn*10*	P1. N3071 x IIB1165
IIB1211	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *recD1903*::mini-Tn*10 xseA*::*dhfr*	P1. SLM1023 x IIB1192
IIB1213	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *recD1903*::mini-Tn*10 sbcD*::*kan recA*::*cam*	P1. TH446 x IIB1207
IIB1214	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *recB1080 argA*::Tn*10*	P1. IIB360 x IIB1165
IIB1215	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *recD1903*::mini-Tn*10 exoX1*::*npt*	P1. N5288 x IIB1192
IIB1218	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *recD1903*::mini-Tn*10 xseA*::*dhfr recA*::*cam*	P1. TH446 x IIB1211
IIB1221	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *recB1080 argA*::Tn*10 recA*::*cam*	P1. TH446 x IIB1214
IIB1222	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *recD1903*::mini-Tn*10 exoX1*::*npt recA*::*cam*	P1. TH446 x IIB1215
IIB1226	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *xseA*::*dhfr*	P1. SLM1023 x IIB1165
IIB1227	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *recD1903*::mini-Tn*10 xseA*::*dhfr*	P1. LMM1032 x IIB1211

	recJ2052::Tn*10kan*	
IIB1228	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *sbcD*::*kan*	P1. LMM1247 x IIB1165
IIB1229	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *exoX1*::*npt*	P1. N5288 x IIB1165
IIB1235	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *xseA*::*dhfr* *recJ2052*::Tn*10kan*	P1. LMM1032 x IIB1227, Tc^S
IIB1136	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *recD1903*::mini-Tn*10 xseA*::*dhfr* *recJ2052*::Tn*10kan recA*::*cam*	P1. TH446 x IIB1227
IIB1239	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *sbcB780*::*kan*	P1. JW1993-1 x IIB1165
IIB1240	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *recD1903*::mini-Tn*10 sbcB780*::*kan*	P1. JW1993-1 x IIB1192
IIB1242	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *recD1903*::mini-Tn*10 sbcB780*::*kan recA*::*cam*	P1. TH446 x IIB1240
IIB1244	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *recB268*::Tn*10 recA*::*cam*	P1. TH446 x IIB1242
IIB1245	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *recD1903*::mini-Tn*10 recA*::*cam*	P1. TH446 x IIB1192
IIB1248	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *xseA*::*dhfr* *recJ2052*::Tn*10kan recA*::*cam*	P1. TH446 x IIB1235
IIB1252	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *recA*::*cam*	P1. TH446 x IIB1165
IIB1253	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *xseA*::*dhfr* *recA*::*cam*	P1. TH446 x IIB1226
IIB1254	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *exoX1*::*npt* *recA*::*cam*	P1. TH446 x IIB1229
IIB1255	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *sbcB780*::*kan* *recA*::*cam*	P1. TH446 x IIB1239
IIB1258	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *sbcD*::*kan* *recA*::*cam*	P1. TH446 x IIB1228
IIB1259	+ Δ*cas3*::*apra* Δ(*casC-cas1*::FRT) *recJ2052*::Tn*10kan recA*::*cam*	P1. TH446 x IIB1199

Genotypes of newly created strains were confirmed by PCR using the following primers:

ygcJ-3: 5'-GGATGTTGACCTGGTGG

ygcJ-4: 5'-GCACACTCTCTGATAACG

cas1del-F: 5' CAGCTAAATCGATGGGATGTG 3'

cas1del-R: 5' GATGGCTAATCTGCCTCGTAAG 3'

apra 1 (R): 5' CCA GAA TGT GTC AGA GAC AAC 3'

upcas3 (F): 5' CGA TAT TTA TGA GCA GCA TC 3'

The *sbcD* mutation was verified by comparing the plaqing efficiencies of mutant λ*pal* phage on *wt* and *sbcCD* mutants.

Table S2.

Strain (passage)	**% of spacer acquisition +/- standard deviation from the mean**
IIB1165 (*wt*) + pBad empty plasmid (3^{rd})	0.00: control used to give baseline zero reading.
IIB1165 (*wt*) + pEB628 (3^{rd})	47.9 +/- 10.7
IIB1192 (*recD*) + pEB628 (3^{rd})	3.0 +/- 0.5
IIB1208 (*recB*) + pEB628 (2^{nd})	6.0 +/- 4.0
IIB1208 (*recB*) + pEB628 (3^{rd})	1.0 +/- 1.7
IIB1244 (*recB recA*) + pEB628 (2^{nd})	18.5 +/- 3.9

Naïve adaptation in strains *recD*, *recB* and *recB recA* was not readily detectable in agarose gels of PCR across CRISPR-1 in data presented in Figure 1. Gel areas corresponding to gel "blank space", where new spacer DNA would be expected to migrate if present but undetectable by that method, was extracted and used as a template for a further iteration of PCR using primers annealing to spacer 3 and leader-repeat 1 border:

CRISPR-NGS-F: 5'-TGCTTTAAGAACAAATGTATACTTT-3'

CRISPR-NGS-R: 5'-CAACATTATCAATTACAACCGA-3'

Outcomes from this second PCR are a 217 bp DNA product for no new spacer detection or 278 bp if new spacer was detectable. Percentage of any detectable spacer acquisition was obtained by measuring relative band intensities using Kodak 1D Image Analysis Software v. 3.6.0. Wild type strain transformed with pEB628 was used as a positive control.

Table S3. Efficiency of spacer acquisition expressed as relative band intensity; data is for strains in which new spacer acquistion was detectable at greater than zero.

Strain (+ pCas1-Cas2)	**Genotype**	**Percentage of spacer acquisition (%)**		
		1 p	**2 p**	**3 p**
IIB1165	*wt*	2.2 ± 0.2	8.8 ± 3	11 ± 1.5
IIB1252	*recA*	3.1 ± 0.7	8.6 ± 0.8	13.3 ± 1.1
IIB1245	*recD recA*	4 ± 2.4	9.1 ± 0.6	2.9 ± 0.3
IIB1214	*recB1080*	2.4 ± 0.02	0.68 ± 0.2	0.77 ± 0.1
IIB1239	*sbcB*	0.27 ± 0.05	13.6 ± 1.9	14.9 ± 1.9
IIB1226	*xseA*	0.25 ± 0.2	12.5 ± 3.8	19.1 ± 6.2
IIB1228	*sbcD*	1.4 ± 0.4	8.4 ± 3.8	17.2 ± 6.5
IIB1229	*exoX*	1.6 ± 1.1	8.6 ± 3.5	16.7 ± 4.8
IIB1253	*xseA recA*	3.73 ± 0.5	12.69 ± 1	19.3 ± 2.4
IIB1254	*exoX recA*	4.7 ± 1.7	13.7 ± 0.7	23.4 ± 3
IIB1255	*sbcB recA*	5.6 ± 3.4	11.1 ± 1.8	8.3 ± 1.8
IIB1258	*sbcD recA*	5 ± 0.1	16.2 ± 0.4	21.7 ± 0.05
IIB1242	*recD sbcB recA*	1.56 ± 0.3	0.78 ± 0.4	0.32 ± 0.3
IIB1218	*recD xseA recA*	3.9 ± 4.2	10.02 ± 1.3	2.59 ± 0.6
IIB1222	*recD exoX recA*	3.1 ± 0.4	1.59 ± 0.4	1.2 ± 0.6
IIB1213	*recD sbcD recA*	2.28 ± 2	0.3 ± 0.19	0.28 ± 0.1
IIB1199	*recJ*	0.7 ± 0.47	10.2 ± 2.5	20.5 ± 1.5
IIB1235	*xseA recJ*	2.1 ± 0.5	4.3 ± 3	7.1 ± 2.8
IIB1246	*xseA sbcD*	1.8 ± 1.5	10.8 ± 2.2	15.9 ± 6.4
IIB1247	*xseA exoX*	0.7 ± 0.1	7.7 ± 1.1	13.6 ± 3.3
IIB1259	*recJ recA*	2.8 ± 0.4	13.9 ± 0.2	18.5 ± 2
IIB1209	*recD recJ recA*	3.29 ± 1.3	1.7 ± 1.4	1.5 ± 0.88

IIB1223	*recD xseA exoX recA*	2.7 ± 0.02	6.8 ± 3	0.75 ± 0.5
IIB1224	*recD xseA sbcD recA*	3.27 ± 0.5	6.48 ± 0.37	0.73 ± 0.5
IIB1236	*recD xseA recJ recA*	8 ± 2	16.3 ± 2	19.2 ± 6
IIB1248	*xseA recJ recA*	3.1 ± 0.6	5.9 ± 1.4	9.7 ± 0.7
IIB1256	*xseA sbcD recA*	1.8 ± 1.5	12.8 ± 7.3	19.6 ± 5.6
IIB1257	*xseA exoX recA*	3.8 ± 2.2	14.1 ± 1	20.3 ± 4.2

Plasmid stability measurements and adaptation

To better understand if poor naïve adaptation, which is stimulated by overexpression of Cas1-Cas2 from the plasmid, in certain mutants was caused by plasmid instability than mutation(s) itself (i.e. *recB* or *recD*) we determined the number of cells that kept the plasmid after three sub-cultivations (passages). Indeed, we noted that adaptation experiments were strongly influenced by the stability of the Cas1-Cas2 expressing plasmid (pEB628) in cells. Mutants of *recB* or *recD* are known for plasmid instabilities and defects in replication termination (Wendel et al. 2014). Overall, Cas1-Cas2 expressing plasmid was stable in *wt* and *recA* cells, moderately lost in *recD* (about 500-fold) and ≈ 10^3 fold in *recB* and completely lost in *recD recJ* cells after two sub-cultivations (data not shown). A similarly strong effect was also noticed in *recD exoX*, *recD sbcD*, *recD sbcB* and *recD xseA recJ* cells (data not shown), strains that have longer ssDNA tails that probably provoke recombination and generate plasmid multimers that are eventually removed from cells.

Table S4. Plasmid instability measurements of major strains referred to in the results measured during passage two during naive adaptation assays

Strain	**Genotype**	**Number of cells containing**	
		pBad (x 10^7)	**pEB628 (x 10^7)**
IIB1165	*wt*	233 ± 15	193 ± 10
IIB1208	*recB*	88 ± 25	0.1 ± 0.08
IIB1192	*recD*	17 ± 15	0.4 ± 0.3

IIB1252	*recA*	133 ± 18	13 ± 0.5
IIB1245	*recD recA*	47 ± 3	0.67 ± 0.46
IIB1244	*recB recA*	55 ± 28	0.008 ± 0.005
IIB1214	*recB1080*	55 ± 7	0.08 ± 0.08
IIB1221	*recB1080 recA*	18 ± 5	0.008 ± 0.002
IIB1242	*recD sbcB recA*	45 ± 12	-
IIB1218	*recD xseA recA*	61 ± 20	0.4 ± 0.3
IIB1222	*recD exoX recA*	17 ± 15	0.016 ± 0.02
IIB1213	*recD sbcD recA*	45 ± 14	-
IIB1253	*xseA recA*	55 ± 7	5.7 ± 1
IIB1254	*exoX recA*	57 ± 25	7 ± 2.4
IIB1255	*sbcB recA*	52 ± 10	0.65 ± 0.4
IIB1258	*sbcD recA*	85 ± 28	5.5 ± 2.3

Supplementary Figures and methods

Figure S1

(A). Elution profile of Cas1-Cas2 co-purification by Superdex S200 gel filtration showing Cas1-Cas2 complex formation and its separation from Cas1 alone, visualised by coomassie staining of fractions analysed by SDS-PAGE. **(B).** *In vitro* assay to detect new spacer integration ("spIN") into the *E. coli* CRISPR-1 DNA sequence (25 nM) catalysed by purified *E. coli* Cas1-Cas2 complex (250 nM). This reaction is optimised for integration by using a synthetic DNA protospacer made from annealing two ssDNA oligonucleotides of sequences described in (Nunez et al. 2015) and by adding purified *E. coli* Integrase Host Factor (IHF, 250 nM). CRISPR-1 DNA for integration comprised the leader and first two spacer-repeat pairs of the CRISPR-1 locus from *E. coli* MG1655. CRISPR-1 was generated with a 5' Cy5 end-label by PCR from CRISPR-1 cloned into pUC18 generating a plasmid (pJRW2), using the method described below.

IHF protein was made as described in the main results. PCR amplification of the genes encoding each IHF subunit used the following primers:

IHFα forward
5'-
ACGTCGGATCCGGAAAACCTGTATTTTCAGGGCTCCATGGCGCTTACAAAAGCTGAAAT
GTC
IHFα reverse
5' ACGTCGCGGCCGCTTACTCGTCTTTGGGCGAAGCG

IHFβ forward

5’ ACGTCCTCGAGACCAAGTCAGAATTGATAGAAAGACTTGCC

IHFβ reverse

5’ ACGTCCCTAGGTTAACCGTAAATATTGGCGCGATCGC

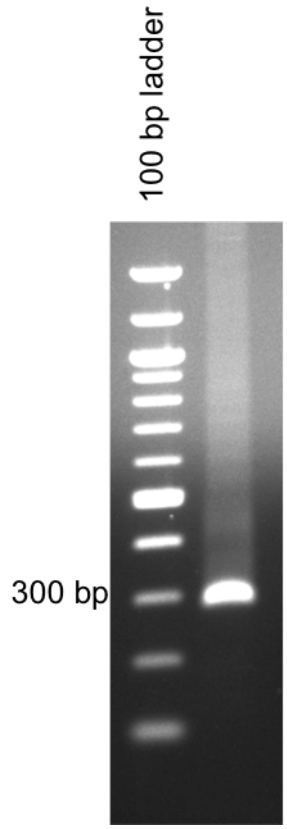
Cy-5 end labeled CRISPR-1 DNA for spacer integration (spIN) assays was generated by PCR of CRISPR-1 from *E. coli* MG1655 cloned into pUC19 (pJRW2) using the following primers:

Crispr1 F ; 5’ Cy5-AGAATTAGCTGATCTTTAATAATAAGG and Crispr1 R short;

5’ TCTCAACATTATCAATTACAACCG

The PCR reaction contained 1 ng of pJRW2 and the following reagents in a final volume of 50 µl using Vent DNA polymerase (NEB). PCR reactions were as follows:

95 ^{o}C – 5 min, followed by thirty cycles of 95 ^{o}C – 30 sec.; 71 ^{o}C – 30 sec.; 72 ^{o}C – 30 sec. And finally 72 ^{o}C for 5 min. CRISPR-1 DNA product (284 base pairs) was purified using gel extraction kits (QiaGen) and verified for size in an agarose gel:



The full sequence of CRISPR-1 DNA that was amplified for use in spIN reactions is:

AGAATTAGCTGATCTTTAATAATAAGGAAATGTTACATTAAGGTTGGTGGGTTGTTTTTAT
GGGAAAAAATGCTTTAAGAACAAATGTATACTTTTAGAGAGTTCCCCGCGCCAGCGGGG
ATAAACCGCTTTCGCAGACGCGCGGCGATACGCTCACGCAGAGTTCCCCGCGCCAGC

GGGGATAAACCGCAGCCGAAGCCAAAGGTGATGCCGAACACGCTGAGTTCCCCGCGC
CAGCGGGGATAAACCGGGCTCCCTGTCGGTTGTAATTGATAATGTTGAGA

Figure S2

Summary of DNA substrates used in this work.

Figure S3

(A). Histogram showing the ratio between numbers of plasmid (P) and chromosomal (C) dervied new spacers. (B and C). Agarose gels summarizing lack of CRISPR expansion in the *E. coli* strains lacking *recD* and ssDNA exonucleases indicated in three passages (p1 – p3) in all cases containing plasmid encoding inducible Cas1-Cas2 (pCas1-2).

Figure S4

EMSA showing Cas1-Cas2 complex formation and aggregation when mixed with an optimized protospacer DNA substrate that is used widely for analyzing Cas1-Cas2 catalyzed spacer integration into CRISPR DNA sequences (Nunez *et al*). Cy5 end labeled DNA substrate (20 nM, see also Supplementary Figure S2) was incubated with 0, 31.25, 62.5, 125, 250, 500 nM Cas1-Cas2 complex for 30 minutes at 37 ^{o}C followed by analysis on a 5 % native acrylamide gel and imaged using a FLA3000 (FujiFilm).

Figure S5

Denaturing (urea) gel analysis of DNA cutting of DNA-10 and a the equivalent 3' ssDNA tailed duplex (DNA-3') by Cas1-Cas2 complex or Cas1 alone. Proteins were used at 0, 31.25, 62.5, 125, 250 and 500 nM. Reactions were carried out at 37^{o}C for 60 minutes. Gel analysis was by 15 % acrylamide urea (8 M) denaturing gels that were imaged for Cy5 imaged using a FLA3000 (FujiFilm).

Figure S6

Denaturing (urea) gel showing in summary the cutting of DNA-10 (lanes 2 - 4) and DNA-11 (lanes 6 – 8) and its use to determine cut sites. Marker DNA lengths in nucleotides are shown alongside three independent reactions mixing Cas1-Cas2 (250 nM) with DNA (20 nM) as indicated. To determine the cut sites the migration distance of DNA bands from the marker DNA and cut products was measured from the base of the well. Migration distance (mm) of marker oligonucleotides was plotted against nucleotide length (Log10) using Prism (GraphPad software) with nuclease product size interpolated from the graph.

Supplementary Reference

Nunez, J. K., et al. (2015), 'Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity', *Nature,* 519 (7542), 193-8.

Figure S1

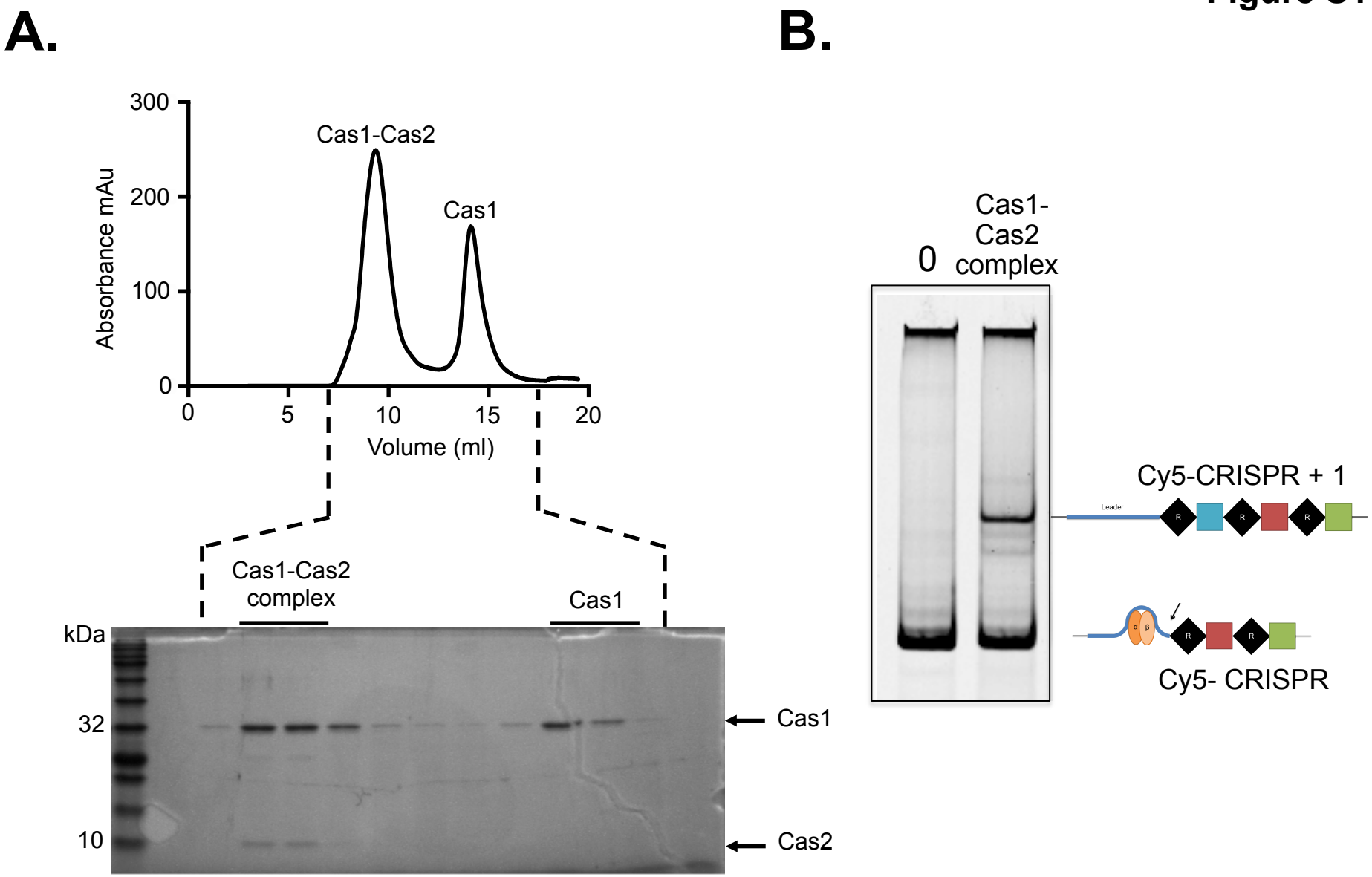
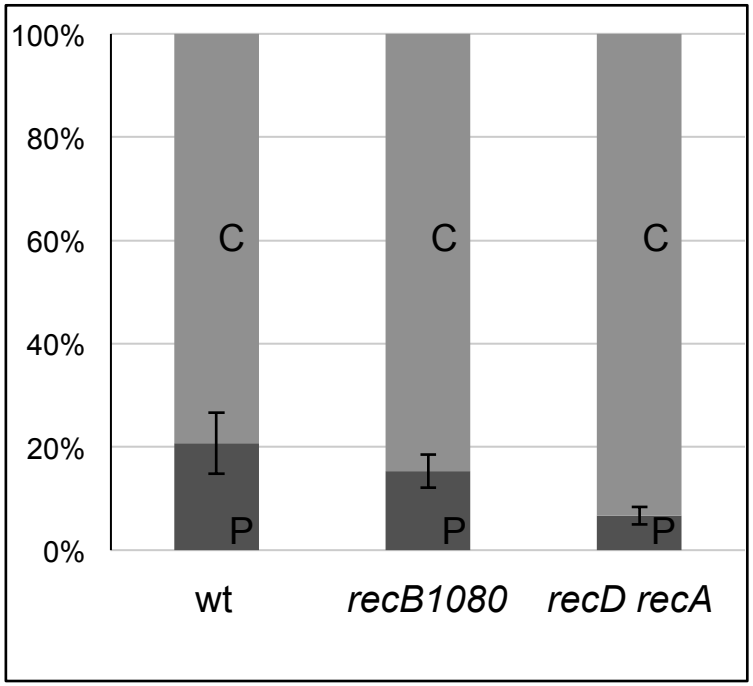

A.
300
200
100
0
Absorbance mAu
Cas1-Cas2
Cas1
0
5
10
15
20
Volume (ml)
Cas1-Cas2
complex
Cas1
kDa
32
10
Cas1
Cas2
B.
0
Cas1-
Cas2
complex
Cy5-CRISPR + 1
Leader
R
R
R
α
β
R
R
Cy5- CRISPR

Figure S2

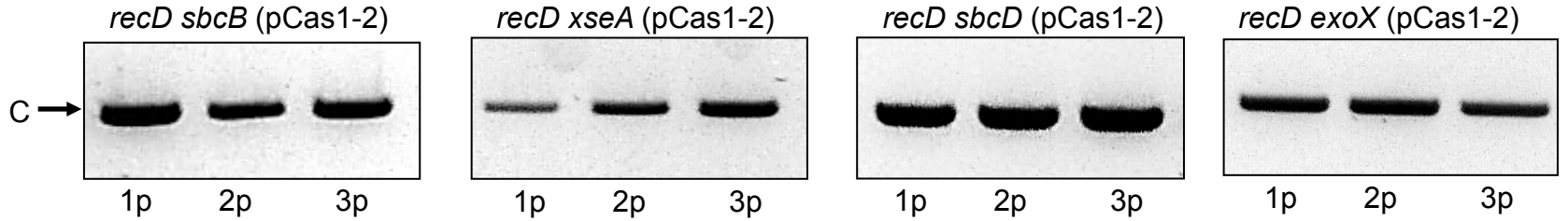

DNA-10
CGGACGAACAACGAACAATGCCAACTAAAGCTCAACCCAGGTGAAACTAA-Cy5 5′
5′GCCTGCTTGT
DNA-11
CGGACGAACAACGAACAATGCCAACTAAAGCTCAACCCAGGTGAAACTAA-Cy5 5′
5′GCCTGCTTGTT
DNA-12
CGGACGAACAACGAACAATGCCAACTAAAGCTCAACCCAGGTGAAACTAA-Cy5 5′
5′GCCTGCTTGTTG
DNA-13
CGGACGAACAACGAACAATGCCAACTAAAGCTCAACCCAGGTGAAACTAA-Cy5 5′
5′GCCTGCTTGTTGC
DNA-14
CGGACGAACAACGAACAATGCCAACTAAAGCTCAACCCAGGTGAAACTAA-Cy5 5′
5′GCCTGCTTGTTGCT
DNA-15
CGGACGAACAACGAACAATGCCAACTAAAGCTCAACCCAGGTGAAACTAA-Cy5 5′
5′GCCTGCTTGTTGCTT
DNA-14-TTT
CGGACGAACAACGTTTAATGCCTTTTAAAGCTCAACCCAGGTGAAACTAA-Cy5 5′
5′GCCTGCTTGTTGCT
Duplex
CGGACGAACAACGAACAATGCCAACTAAAGCTCAACCCAGGTGAAACTAA-Cy5 5′
5′ GCCTGCTTGTTGCTTGTTACGGTTGATTTCGAGTTGGGTCCACTTTGATT
DNA-3’
AATCAAAGTGGACCCAACTCGAAATCAACCGTAACAAGCAACAAGCAGGC-Cy5 5′
5′TGTTCGTCCG
Protospacer
TGCTCGCATCGACTCCGCTCCCCTGACG-Cy5 5′
5′CGTAGCTGAGGCGAGGGGACTGCTGGGC

Figure S3

A.



B.



C.

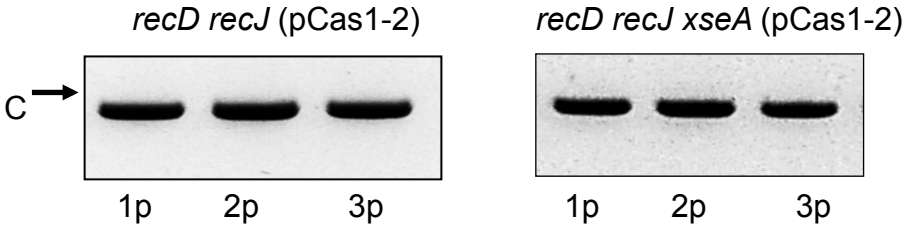


Figure S4

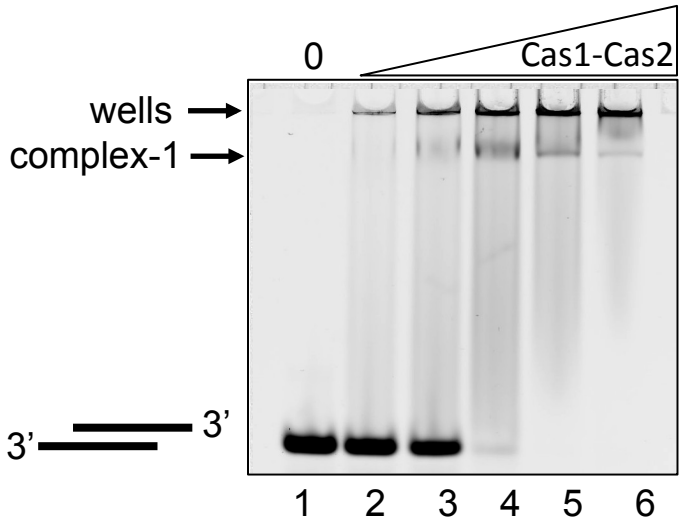


Figure S5

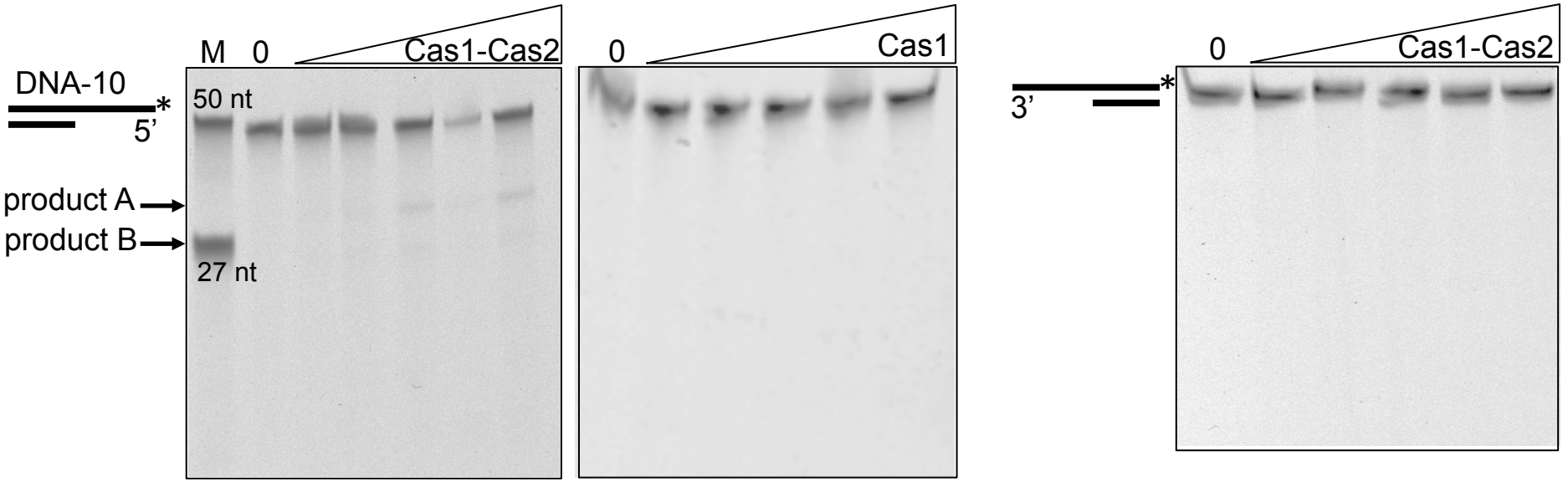
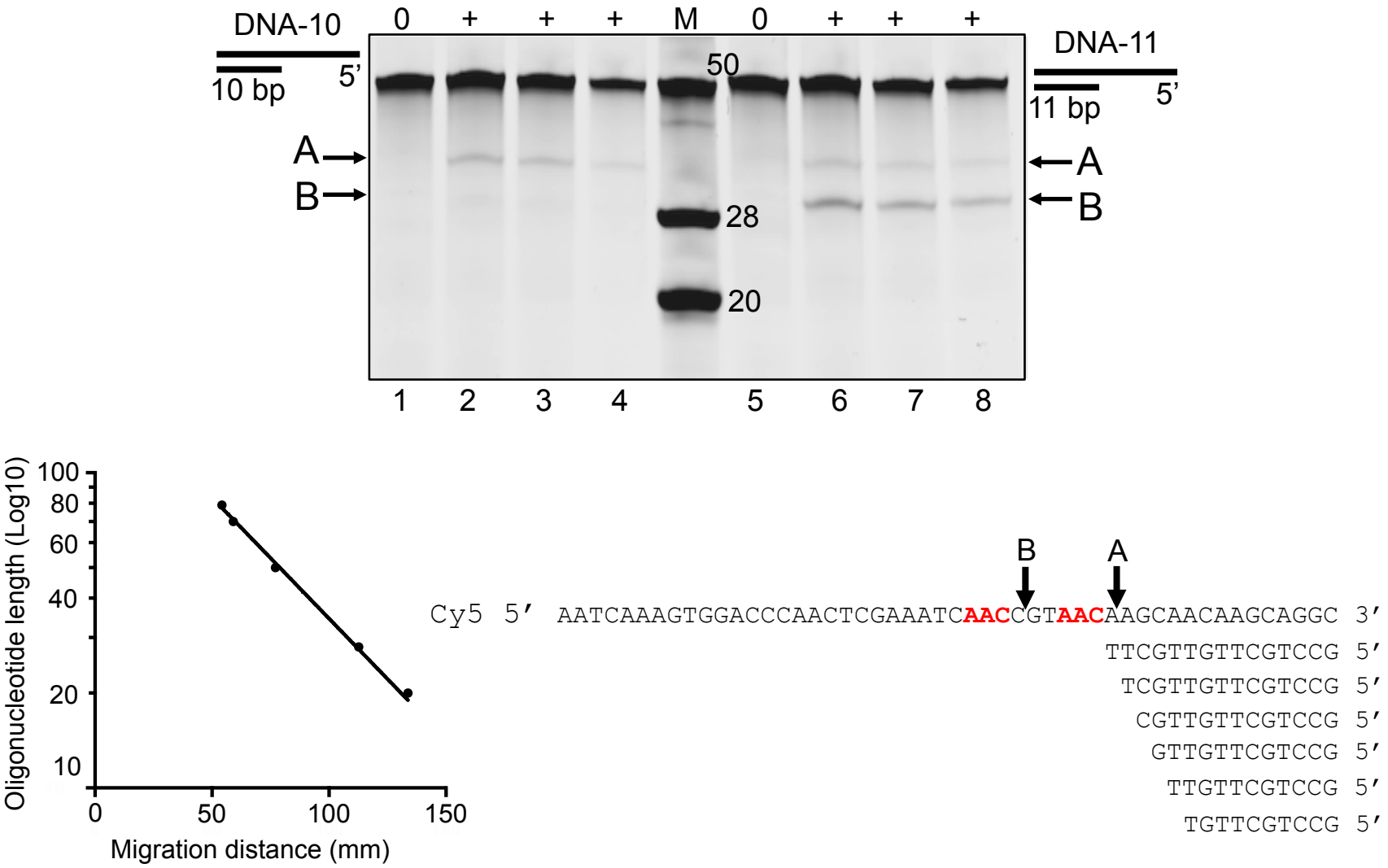

M
0
Cas1-Cas2
0
Cas1
0
Cas1-Cas2
DNA-10
*
5'
3'
50 nt
27 nt
product A
product B

Figure S6


DNA-10
0 + + + M 0 + + +
10 bp
5'
50
A
B
28
20
DNA-11
11 bp
5'
A
B
1 2 3 4 5 6 7 8
Oligonucleotide length (Log10)
100
80
60
40
20
10
0
50
100
150
Migration distance (mm)
B
A
Cy5 5′ AATCAAAGTGGACCCAACTCGAAATCAACCGTAACAAGCAACAAGCAGGC 3′
TTCGTTGTTCGTCCG 5′
TCGTTGTTCGTCCG 5′
CGTTGTTCGTCCG 5′
GTTGTTCGTCCG 5′
TTGTTCGTCCG 5′
TGTTCGTCCG 5′