

# CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of *Escherichia coli*

César Díez-Villaseñor,<sup>1</sup> Noemí M. Guzmán,<sup>1</sup> Cristóbal Almendros,<sup>1</sup> Jesús García-Martínez<sup>1</sup> and Francisco J.M. Mojica<sup>1,\*</sup>

<sup>1</sup>Departamento de Fisiología, Genética y Microbiología; Universidad de Alicante; Alicante, Spain

**Keywords:** Protospacer adjacent motif, *cas* genes, Cascade, CRISPR-spacer acquisition, reporter plasmids, RNA-guided immunity, spacer orientation, ruler mechanism, *Escherichia coli* K12, O157:H7

**Abbreviations:** CRISPR, clustered regularly interspaced short palindromic repeat; Cas, CRISPR-associated sequence; Cascade, CRISPR-associated complex for antiviral defense; PAM, protospacer adjacent motif; TIM, target interference motif; SAM, spacer acquisition motif; PAME, protospacer adjacent motif end; leader-A, type A leader of *E. coli* CRISPR2.1 arrays; leader-T, type T leader of *E. coli* CRISPR2.1 arrays; *cas*-EK, *cas* genes of the *E. coli* K12 CRISPR-Cas I-E system; *cas*-EO, *cas* genes of the *E. coli* O157:H7 CRISPR-Cas I-E system; ECOR, *Escherichia coli* reference

Prokaryotes immunize themselves against transmissible genetic elements by the integration (acquisition) in clustered regularly interspaced short palindromic repeats (CRISPR) loci of spacers homologous to invader nucleic acids, defined as protospacers. Following acquisition, mono-spacer CRISPR RNAs (termed crRNAs) guide CRISPR-associated (Cas) proteins to degrade (interference) protospacers flanked by an adjacent motif in extrachromosomal DNA. During acquisition, selection of spacer-precursors adjoining the protospacer motif and proper orientation of the integrated fragment with respect to the leader (sequence leading transcription of the flanking CRISPR array) grant efficient interference by at least some CRISPR-Cas systems. This adaptive stage of the CRISPR action is poorly characterized, mainly due to the lack of appropriate genetic strategies to address its study and, at least in *Escherichia coli*, the need of Cas overproduction for insertion detection. In this work, we describe the development and application in *Escherichia coli* strains of an interference-independent assay based on engineered selectable CRISPR-spacer integration reporter plasmids. By using this tool without the constraint of interference or *cas* overexpression, we confirmed fundamental aspects of this process such as the critical requirement of Cas1 and Cas2 and the identity of the CTT protospacer motif for the *E. coli* K12 system. In addition, we defined the CWT motif for a non-K12 CRISPR-Cas variant, and obtained data supporting the implication of the leader in spacer orientation, the preferred acquisition from plasmids harboring *cas* genes and the occurrence of a sequential cleavage at the insertion site by a ruler mechanism.

## Introduction

CRISPR-Cas is an adaptive RNA-mediated immune system discovered in the bacteria<sup>1</sup> and soon after also reported in the archaea.<sup>2</sup> The primary nucleic acid components of the system are the CRISPR, clustered short palindromic repeats regularly interspaced by non-redundant sequences (referred to as spacers)<sup>3,4</sup> that often derive from fragments (termed protospacers) of infecting DNA.<sup>5–7</sup> Cas are proteins functionally linked to CRISPR and the encoding *cas* genes are usually arranged in operons located in close proximity to repeat-spacer arrays.<sup>4</sup> For a recent review of the CRISPR-Cas systems, see refs. 8–10.

Three CRISPR-Cas types (I, II and III), each including several subtypes, are defined by the presence of signature *cas* genes.<sup>11</sup> This work deals with the CRISPR-Cas subtype I-E of diverse *E. coli* strains. This system consists of up to three CRISPR-spacer

arrays (i.e., CRISPR2.1, CRISPR2.2 and CRISPR2.3)<sup>12</sup> made of type 2 repeats.<sup>13</sup> Next to CRISPR2.1 and CRISPR2.3 arrays there is an AT-rich sequence known as the leader,<sup>4</sup> here named leader2.1 and 2.3, respectively, which contains the promoter that directs transcription of the adjoining array.<sup>14–16</sup> Two main repeat variants, hereinafter named CRISPR-A and CRISPR-T, are found across the species, differing by the presence of adenine or thymine, respectively, at the position next to the leader-proximal end of the repeat (see Fig. S1) and matching the consensus sequence 5'-CGGTTTATCC CCGCTGACGC GGGGAACWC-3' (W = A or T). Albeit, this is the repeated sequence of the CRISPR arrays, it has been recently proposed<sup>17</sup> and almost simultaneously demonstrated<sup>18,19</sup> that the cytosine nucleotide at the leader-distal end (Fig. S1) is inserted along with the protospacer, being part of the spacer. After this evidence, we will refer to the CRISPR duplicon in line with Goren et al.<sup>19</sup> as the 28 nt repeated sequence

\*Correspondence to: Francisco J.M. Mojica; Email: fmojica@ua.es  
Submitted: 12/04/12; Revised: 02/09/13; Accepted: 02/15/13  
<http://dx.doi.org/10.4161/rna.24023>

5'-GGTTTATCCC CGCTGACGCG GGGAACWC-3'. Thus, duplicons are interspaced by 33–34 nt spacers, 33 nt being the most frequently observed (about 95%).<sup>12,20–22</sup> Remarkably, the leader associated with CRISPR2.1 arrays that are composed mainly of CRISPR-A duplicons (hereinafter referred to as leader-A) differs from the one linked to CRISPR-T-enriched arrays (here named leader-T).<sup>23</sup> Eight genes, altogether referred to as *cas*-E, constitute the typical *cas* repertoire of the CRISPR-Cas I-E systems: *cas1*, *cas2*, *cas3*, *cse1*, *cse2*, *cas7*, *cas5* and *cas6e*,<sup>11,24</sup> the last five encoding the proteins that make CRISPR-associated complex for antiviral defense or Cascade.<sup>25</sup> Immunity is achieved by Cas3-catalyzed endonucleotidic cleavage of DNA targeted by the spacer sequence in mono-spacer CRISPR RNA (crRNA) molecules generated by Cascade.<sup>14,25–28</sup> The efficiency of interference is influenced by the sequence adjoining one end of the protospacer.<sup>17,18,28</sup> Connected to this, the consensus protospacer adjacent motif (PAM) CWT (5'-protospacerC-WT-3'; see Fig. S1) has been defined by the alignment of putative protospacers of this CRISPR-Cas system<sup>23</sup> and efficient interference is observed when this motif is present; notwithstanding that alternative sequences in the PAM region, defined as target interference motifs or TIMs,<sup>29,30</sup> also uphold CRISPR immunity.<sup>17,18,25,28</sup>

Insertion of new spacers within a CRISPR array (a process known as acquisition or adaptation; for a specific review, see ref. 31) may occur during infection by plasmids or viruses, hereinafter immunizing the host (therefore referred to as adapted cell) against invaders carrying a spacer-matching protospacer sequence.<sup>17,18,32</sup> Spacer addition is accompanied by a repeat duplication, so that spacers remain flanked by CRISPR duplicons. The leader-CRISPR boundary has been shown to be the preferential,<sup>33–38</sup> albeit not exclusive,<sup>39–42</sup> insertion site for several CRISPR systems. Although knowledge on the acquisition process is limited, recent contributions on the CRISPR-Cas I-E system of *E. coli* K12 have granted substantial advance. Notably, evidences have been provided supporting that (1) Cas1 and Cas2 are the only Cas proteins required for spacer uptake,<sup>32</sup> (2) the new repeat generated during spacer integration is a replica of the preexisting CRISPR adjoining the leader,<sup>32</sup> (3) a single repeat is sufficient to elicit adaptation,<sup>32</sup> (4) most protospacers are selected among sequences next to the tri-nucleotide CTT,<sup>17,18</sup> (5) the leader sequence is directly involved in the expansion of the adjacent CRISPR array, independently of its transcription,<sup>32</sup> (6) resident spacers promote addition of new spacers matching the target molecule (this activity is termed primed acquisition and requires all Cas proteins)<sup>18</sup> and (7) the DNA strand targeted during interference is selected for subsequent protospacer uptake in primed acquisition.<sup>17</sup>

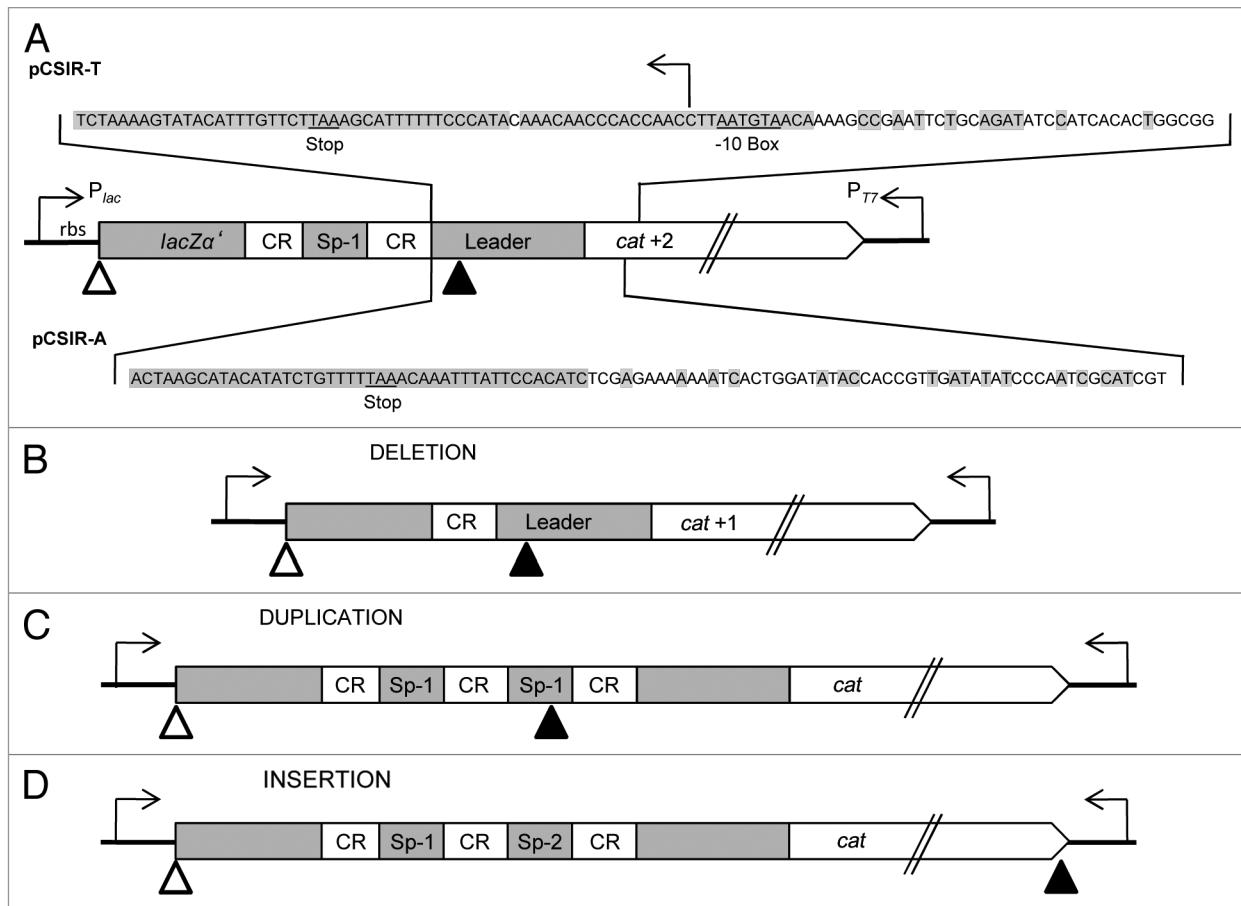
It is worth mentioning that Cascade genes, together with *cas1* and *cas2*, form an operon whose expression is repressed under normal laboratory growth by the transcriptional regulator H-NS, at least in *E. coli* K12.<sup>14–16,43</sup> In consequence, neither interference nor adaptation by this system has been observed in the non-manipulated wild-type strain.<sup>14,17,18,25,28,32</sup> Indeed, in order to detect acquisition, overproduction of at least Cas1 and Cas2 proteins<sup>17,18,32</sup> and/or the selection of CRISPR-Cas-induced cells showing interference provided by the integrated spacer<sup>17,18</sup>

had to be used. Both approaches carry intrinsic concerns. In the former, high protein concentration may affect cell fitness and/or produce aberrant behavior. In the latter approach, only spacers capable of providing efficient interference are detected. In order to circumvent these constraints, we implemented a reporter system capable of revealing acquisition even at very low CRISPR-Cas expression and independent of the immune phenotype. Using this tool, we have verified previous results obtained with the K12 I-E system and also extended them to *E. coli* strains with an alternative variant of *cas*-E genes, expanding conclusions to the species. Interestingly, instead of CTT triplet exhibited by the K12 system, the protospacer-associated motif recognized for spacer acquisition (named SAM for spacer acquisition motif)<sup>30</sup> by strains carrying the non-K12 *cas*-E variant is CWT, in agreement with the bioinformatically predicted PAM. Furthermore, the implication of the leader on spacer's orientation was revealed. A marked preference for, or exclusion of particular replicons as spacer donors, was evidenced. Finally, we hypothesize on the so-far-enigmatic mechanism of spacer insertion.

## Results

**CRISPR-spacer integration reporter plasmids.** With the aim of setting up a strategy for investigating spacer integration by the CRISPR-Cas I-E system of *E. coli*, we developed an assay based on positive selection of clones that become resistant to chloramphenicol upon expansion of a CRISPR array within recombinant plasmids (pCSIR-A and pCSIR-T; see Fig. 1). Both plasmids carry an insertion cassette made of a two-repeat CRISPR2 array followed by a fragment of the associated leader (proximal 43 and 69 bp in pCSIR-A and pCSIR-T, respectively; see Fig. 1A). Thus, the -10 TATA box required for transcription of the array<sup>14,15</sup> is present in pCSIR-T but absent in pCSIR-A. Repeats and leader in pCSIR-A belong to variant A (CRISPR-A and leader-A, respectively), while pCSIR-T carries T variants of both elements (CRISPR-T and leader-T). The CRISPR-spacer-CRISPR-leader cassette is translationally fused to the beginning of *lacZ*- $\alpha$  (*lacZ*- $\alpha'$ ) and to an out-of-frame (+2 from the *lacZ*- $\alpha$  start codon) gene (*cat*) for the chloramphenicol acetyltransferase (or CAT) protein (Fig. 1A). Translation initiated at the AUG codon of *lacZ*- $\alpha$  stops at the leader. Similarly, in case of deletion due to recombination between the two CRISPR units, translation would stop at another triplet of the leader (see Fig. 1B). In order to prevent CAT translation due to duplication (rather than new insertion) of a spacer-CRISPR unit, the carried spacer contains an out-of-frame stop tri-nucleotide that would become in frame in the duplicated spacer (Fig. 1C).

Insertion experiments were performed with several *E. coli* strains carrying different combinations of pCSIR and *cas*-containing plasmids. In order to avoid the possibility of perverted acquisition due to gross overexpression of *cas* genes (IPTG-inducible T7-*lac* promoters are located upstream of the *cas* operons and the *lacI* gene encoding the *lac* repressor is included in the constructions; see Fig. 2), all assays were performed in the absence of IPTG. Hence, at most low levels of *cas* expression are expected (e.g., in case of promoter leakage).

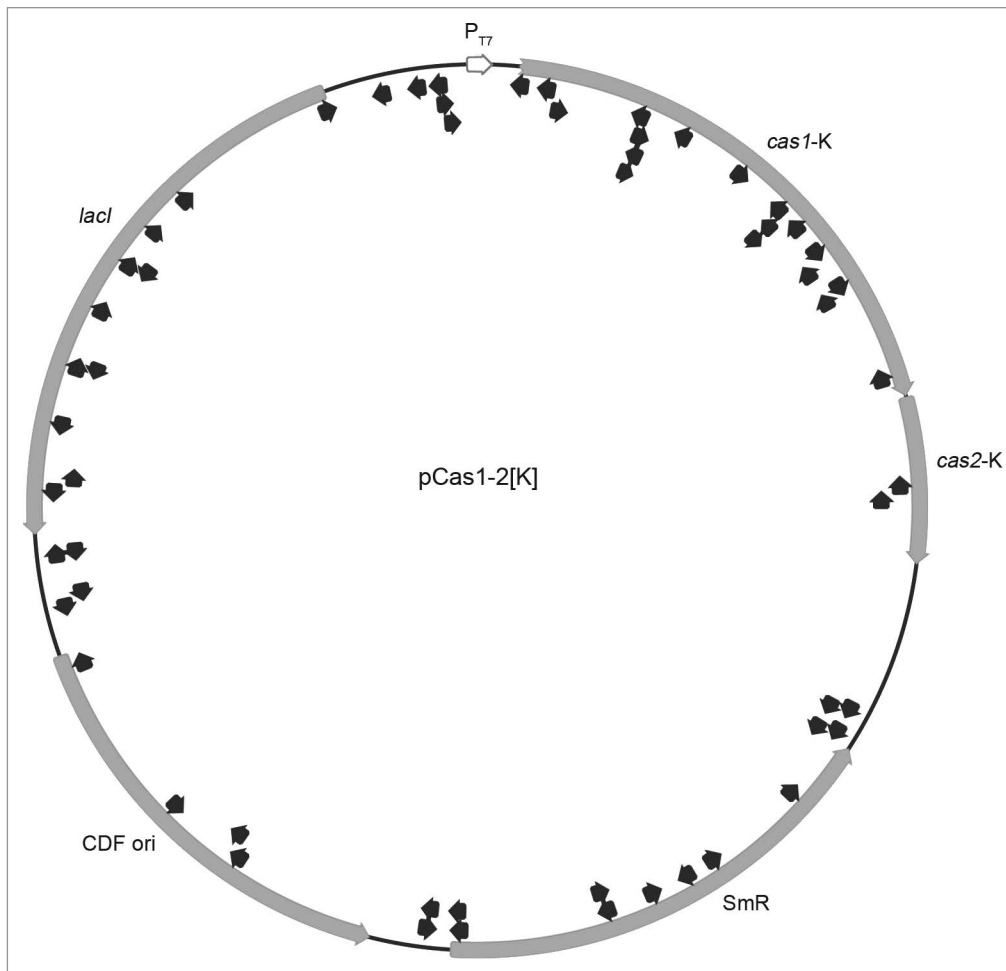


**Figure 1.** Schematic representation of the integration reporter cassettes cloned in the pCSIR plasmids and strategy used for specific detection of integration events. **(A)** Integration cassette in pCSIR-A or pCSIR-T consists of a leader and two CRISPR duplicons (CR) interspaced by a spacer (Sp-1) fused upstream to a fragment of *lacZ-α* gene (*lacZα'*) and, downstream, to the complete coding sequence of an out-of-frame *cat* gene (*cat +2*). The ribosome-binding site (rbs) and the translation initiation codon of *lacZ-α* (white arrowhead) are indicated. Translation of transcripts generated from the *lac* promoter ( $P_{lac}$ , denoted with an arrow) will stop at the leader (black arrowhead). A T7 promoter ( $P_{T7}$ ) is located downstream of *cat* gene, promoting transcription in the opposite direction (arrow pointing toward the direction of transcription). The sequences of fragments in the leader-*cat* region of pCSIR-T and pCSIR-A with the size of the typical T and A variants respectively of the leader are shown; nucleotides matching the corresponding leader variant are shadowed and stop codons as well as, when present, the -10 box underlined and accordingly labeled (initiation site is indicated with an arrow pointing toward the direction of transcription). In order to avoid stop codons, thymine was replaced with cytosine at position 42 and with adenine at position 70 of the leader in pCSIR-T (see primer TLR in **Table S7**). Translation of the transcript generated from *Plac* will end at the leader in case of deletion (*cat* would remain out of frame, *cat +1*) **(B)**, at the duplicated spacer (Sp-1) in case of CRISPR-spacer duplication **(C)** or at the end of the in frame *cat* coding sequence in case of new spacer (Sp-2) insertion **(D)**.

After the observation that elevated expression of *cas1* and *cas2* genes produces acquisition in the absence of Cascade and Cas3,<sup>18,32</sup> initial insertion assays were performed with strain BL21-AI carrying pCSIR-T and plasmid pCas1-2(K) (containing *cas1* and *cas2* genes of K12). BL21-derivative strains lack resident *cas* genes,<sup>12</sup> and although BL21-AI harbors a T7 RNA polymerase gene under the control of an arabinose-inducible promoter, arabinose was not added to the growth medium. Under these conditions, one adapted clone (spacer sequence AAATACACAG ACACGGAGAA TCACTATGTT TAC; identical to a chromosomal sequence adjacent to CTT) was obtained after multiple trials and prolonged cultivation (six growth cycles, see Materials and Methods). However, when plasmid pWUR399 [same vector as pCas1-2(K) but carrying Cascade genes transcriptionally fused to *cas1-cas2*] was used

instead of pCas1-2(K), nine adapted clones were isolated after 12 h (one growth cycle) incubation (**Table S1**). These combined results indicate that *cas* genes are transcribed in the two constructs despite the absence of induction, and although they also suggest that Cascade might facilitate acquisition mediated by Cas1 and Cas2, further investigation is required to substantiate this implication.

**The identity of the spacer acquisition motif (SAM) depends on the CRISPR-Cas I-E variant.** In a previous comparative analysis of leader vs. PAMs of 15 putative CRISPR2 protospacers of *E. coli*,<sup>23</sup> a preference for either CAT or CTT was proposed for CRISPR arrays associated with leader-A or -T, respectively. Now we have confirmed this bias by a similar in silico study performed with 76 putative protospacers showing over 90% identity to CRISPR2 spacers of ECOR strains<sup>44</sup> and sequenced *E. coli*



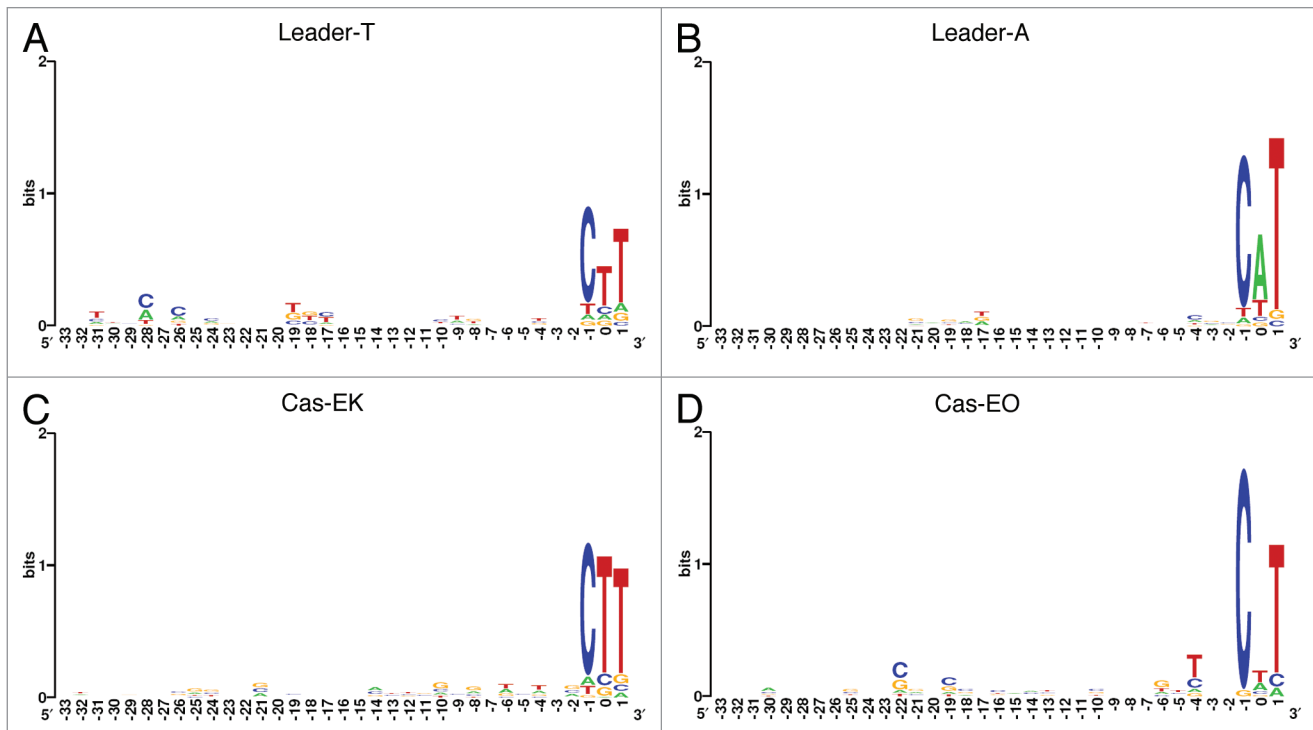
**Figure 2.** Distribution in pCas1-2(K) of sequences matching distinct spacers sampled in acquisition assays. Spacers are shown as black arrows pointing toward the leader as they were inserted in the integration cassette. The origin of replication (CDF ori), the genes encoding the *lac* operon repressor (*lacI*) and a streptomycin resistance protein (SmR) as well as the T7-*lac* promoter ( $P_{T7}$ ) leading transcription of K-variant *cas1-2* genes (*cas1-K* and *cas2-K*) are shown.

genomes (Fig. S2): 11 out of 21 protospacers associated with leader-T showed the PAM CTT vs. 1 CAT (Fig. S2), holding the CTT PAM sequence (Fig. 3A), and 34 out of 55 predicted protospacers associated with leader-A flanked a CAT sequence whereas nine showed the triplet CTT (Fig. S2), confirming the CWT PAM consensus (Fig. 3B).

Although a connection between PAM and leader is apparent, suggesting that the latter might take part in the recognition of the specific motif during spacer uptake, additional elements of the CRISPR-Cas system could also be involved. As shown in Figure S3, in CRISPR2.1 arrays of *E. coli*, type A repeats co-occur with leader-A, and CRISPR-T with leader-T. Moreover, in a separate study (unpublished data), we noticed that each leader type correlates to a specific variant of *cas-E* genes, differentiating two groups within the species defined by the variant that here we call *cas-EK* (after K12 strain), linked to CRISPR/leader of type T, and the variant *cas-EO* (after O157:H7) associated with type A, respectively.

In order to experimentally identify the actual SAM sequence recognized by each CRISPR/Leader/Cas variant, we performed

spacer integration assays with natural combinations of the respective elements in strains harboring equivalent CRISPR-Cas systems (this way interactions between CRISPR-Cas elements of different variants were prevented). First, insertion was assessed in K12 carrying an integration reporter plasmid with leader-T and CRISPR-T (pCSIR-T), and a second recombinant plasmid with either *cas1* (pCas1K), *cas2* (pCas2K) or *cas1* and *cas2* [pCas1-2(K)] of this strain. Efficient spacer integration (59 insertions) was observed with pCas1-2(K) (Table S2) within 12 h of incubation, but not with pCas1K or pCas2K after six growth cycles, proving that both *cas* genes are required in a multicopy plasmid in order to detect acquisition in K12 with our system. Moreover, the analysis of putative precursors of the new spacers showed that pCSIR-T plasmids tended to acquire spacers from sequences flanking the triplet CTT (42 out of 59 insertions; note that CAT was not detected; Table S2). This preference obtained further support from the experiments described in the previous section performed in BL21-AI carrying pCSIR-T and pWUR399: six out of nine insertions had the motif CTT and no CAT was observed (see Fig. 3C; Table S1).



**Figure 3.** WebLogos generated by the alignment of protospacer regions of *E. coli* CRISPR2 spacers. Sequences containing protospacers (positions -1 to -33) and the adjoining two nucleotides (positions 0 and 1) of the PAM region were equally oriented with respect to the corresponding spacer in the CRISPR array (3' end toward the leader) and aligned in stacks. Logos were obtained for sequences in non-CRISPR loci showing over 90% identity to spacers detected in *E. coli* genomes, associated with leader-T (A) or leader-A (B), as well as for protospacers matching spacers sampled in acquisition experiments performed in strains harboring Cas-EK (C) or Cas-EO (D) encoding genes.

Remarkably, when acquisition was assayed in strain O157:H7 carrying pCSIR-A and pCas1-2(O) (containing *cas1* and *cas2* of this strain), in contrast with K12, both CTT and CAT protospacer motifs were revealed (detected for 26 and 11, respectively, out of 45 insertions; see Fig. 3D; Table S3).

Hence, our data demonstrate that CTT is the dominant SAM (CAT was not detected) recognized by the CRISPR-Cas subtype I-E of K12 but CTT and CAT by the Cas-EO proteins of O157:H7, concurring with the PAMs revealed by the alignment of putative precursors of spacers found in CRISPR2 arrays of the species (compare Fig. 3A with Fig. 3C and Fig. 3B with Fig. 3D, respectively). These results disclose that both system variants may differ mechanistically, at least at the CRISPR expansion stage.

**Spacer donor molecules.** Albeit the content of DNA per cell (see Fig. S4) and CTT tri-nucleotide per Kb (Table S4) is higher for pCSIR-T than for pWUR399, all nine spacers acquired in BL21-AI cells harboring both plasmids were identical to sequences present in pWUR399 and absent in pCSIR-T (Table S1). Similarly, although chromosomal DNA is much more abundant than pWUR399 sequences (Fig. S4), none of these spacers matched a chromosomal fragment (Table S1).

As well as with BL21-AI, experiments performed with K12 or O157:H7 strains revealed a preference for plasmids carrying *cas* genes as source of spacers. A ratio of 39/104 inserted spacers were identical to chromosomal sequences, just four of them not present in pCas1-2 plasmids (Tables S2 and S3). Also, although

the DNA content per cell of *cas*-harboring plasmids was lower than that of integration-reporter plasmids (the four constructs have similar size and both pCas1-2 plasmids and pWUR399 have a CDF origin of replication; see Fig. S4 and Table S4), seven insertions were identical to pCSIR fragments vs. 94 to pCas1-2 sequences (Tables S2 and S3). Moreover, the frequency of CWT tri-nucleotides in the constructs does not justify the observed underrepresentation of protospacers in the reporter plasmids (CTT is more frequent in pCSIR than in pCas1-2 plasmids and CAT incidence is similar for all of them; Table S4). In this context, the akin CTT/CAT ratio in pCas1-2 plasmids (0.93–0.99; Table S4) provides further support to the preference for CTT acquisition motif by *cas*-EK and CWT by *cas*-EO systems. Thus, either some replicons are less prone to contribute spacers compared with others or there is a preference for certain molecules as spacer donors (see Discussion).

**Implication of CRISPR-Cas elements in spacer orientation.** Spacers within a CRISPR locus tend to be equally oriented with respect to the PAM.<sup>23</sup> Moreover, this conservation in orientation applies to the CRISPR-Cas type: spacer ends equivalent to the protospacer edges adjacent to the PAM (termed PAMes) of CRISPR-Cas type I are oriented toward the leader.<sup>11,23,40,45</sup> This implies that acquisition proteins (Cas1 and/or Cas2) together with element(s) at the spacer integration region (either the leader and/or the adjoining CRISPR sequence) are involved in determining the direction of insertion.





As in the assays performed with cognate CRISPR/leader and Cas pairs, the majority of new spacers (17/18) matched pCas1-2 sequences (seven of them were also present in the chromosome), and just one was exclusive of the chromosome, while no spacer originated from pCSIR-A (Table S5).

## Discussion

**Utility of integration reporter plasmids.** CRISPR-spacer integration reporter plasmids allow sensitive detection of insertions without the requirement of an elevated production of Cas proteins or the selection of adapted cells relying on the interference phenotype provided by the new spacer. Yet, compared with the native situation, the presence in our system of multiple copies of the CRISPR locus may lead to altered behavior. Moreover, detectable additions are restricted to sequences leading to the synthesis of a functional CAT protein. Thus, 34 nt spacers, observed in CRISPR arrays in a proportion of about 4%,<sup>12,20-22</sup> would be excluded as only integration of a 3n+1 bp fragment, such as a CRISPR-spacer unit with the canonical length of 61 bp, can potentially be selected (Fig. 1D). About 62% of them will be detected due to the absence of stop codons in phase with the reporter gene (see Materials and Methods). Although multiple insertions (3n+1) might be observed, only single additions were sampled. This could be explained considering that integrations must be sequential; a plasmid carrying multiple new spacers would be outnumbered with those harboring a lower number (i.e., due to more replication cycles of the latter). The increased chance of stop codons (probability of 0.38 for one insertion vs. 0.85 for four; see Materials and Methods) could also be a reason. Taking into account these considerations, solid conclusions on the adaptation process can be drawn.

To highlight, the dominant SAM revealed in our assays with the K12 CRISPR-Cas system was CTT (73.5%, 64/87, of motifs; see Table S6), in agreement with predictions from the comparative analysis of protospacer regions and with interference-reliant acquisition studies performed under increased levels of Cas proteins.<sup>17,18</sup> Conversely, in an interference-unrelated analysis with overexpressed *cas1* and *cas2* genes of K12, as few as 36 out of 94 protospacers (38.3%) adjoined CTT,<sup>32</sup> perhaps as a consequence of the elevated Cas1 and Cas2 levels utilized. Strikingly, as from the analysis of data reported by Datsenko et al.<sup>18</sup> and Swarts et al.,<sup>17</sup> we observed that CAT motif was apparently excluded (not sampled in assays with *cas*-EK; Table S6). The reason for this omission is intriguing, notably since efficient interference against targeted protospacers flanking CAT has been reported.<sup>28,46</sup> Otherwise, in contrast to the K12 system and concurrent with in silico predictions, we have identified the CWT SAM triplet for the CRISPR-Cas I-E system of *E. coli* O157:H7, revealing mechanistic differences among variants of a CRISPR-Cas subtype within a given species. Yet, CAT tri-nucleotide is the most frequent PAM observed for spacers present in leader-A-associated CRISPR arrays, but CTT is preferentially selected by Cas-EO proteins. Then, our combined bioinformatic and experimental results demonstrate that specific SAMs are required for protospacer recognition independent of their interference

consequence, and suggest that the more efficient spacers (perhaps those targeting sequences with the CAT motif in the case of the *cas*-EO variant) are subsequently selected during the arms race with target-invading DNAs.

In another context, we also verified<sup>17-19</sup> that the nucleotide formerly considered as the first of the CRISPR (distal with respect to the leader) invariably matched the 33<sup>rd</sup> position of the protospacers, whether the SAM was conserved or not (i.e., C, A, T or G were at the first nucleotide of the motif 119, 5, 5 and 3 times respectively; see Tables S1–3 and S5). Also, some of these non-consensus sequences showed either CTT (6) or CAT (1) shifted one position with respect to the SAM location (Tables S1–3 and S5), most likely as a result of inaccurate protospacer cleavage after recognition of the motif.

Yosef et al.<sup>32</sup> elegantly demonstrated that the first 40–60 bp of the leader associated with the CRISPR2.3 of BL21-AI are required for efficient expansion of the associated repeat-spacer cassette, suggesting that its transcription is not necessary as this fragment does not include promoter.<sup>14,15</sup> The detection of insertions within the CRISPR cassette of pCSIR-A, which contains just the first 43 bp of a leader2.1 (Fig. 1A), expands the observation made by Yosef et al. with leader2.3 to the leader2.1 in a shorter sequence segment.

Apart from the above discussed results, our developed genetic strategy also yielded data supporting rejection/preference for certain replicons as spacer-donors and the occurrence of a ruler mechanism for insertion-site cleavage during the spacer integration process. Both aspects are discussed in the following sections.

**Spacers are preferentially obtained from plasmids carrying *cas1* and *cas2* genes.** Sequences in *cas*-containing plasmids were overrepresented in the protospacers sampled (120/132). The fact that the majority of putative chromosomal protospacers also exist in these plasmids (43/48), even though shared sequences are less than 0.1% of the chromosome, further supports the plasmidic origin of these spacers. At first glance, interference guided by new spacers targeting the chromosome could explain this rejection. However, such activity is not expected in the conditions assayed, neither in BL21-AI nor in K12. Indeed, resident *cas* genes are absent in BL21-derivative strains and the *cas*-containing plasmids used lack Cas3 protein required for target cleavage. Furthermore, H-NS repression of CRISPR arrays and resident *cas* genes in wild-type K12 impedes interference,<sup>14-16,25</sup> and only *cas1* and *cas2* have been cloned in the assays performed with this strain. In addition, the association of the CWT protospacer motif, capable of eliciting interference,<sup>17,18,28</sup> with chromosomal sequences matching new spacers (34/48), also argues against the occurrence of CRISPR-mediated degradation activity. Once interference is discarded, the apparent rejection of the chromosome as spacer donor observed here, and also by Yosef et al.<sup>32</sup> could be explained by its disruption during protospacer excision.

A different reason must account for the underrepresentation of spacers from integration reporters vs. *cas*-containing plasmids (7:120) as they are multicopy replicons and, thus, disruption of the donor molecule does not imply plasmid loss and besides, antibiotics were used during the acquisition assays to select cells maintaining both plasmids (see Materials and Methods). Likewise,

the involvement of primed acquisition<sup>17,18</sup> as a way to favor integration from molecules targeted by a preexisting spacer can be dismissed as *cas*-containing plasmids lack sequences similar to chromosomal or pCSIR spacers (the only sequences matching pre-existing spacers are in the chromosomal and pCSIR CRISPR arrays, where interference is anticipated to be prevented).<sup>47,48</sup>

As a tentative explanation, we propose that the presence of *cas* genes involved in acquisition in the preferred spacer-donor replicon could account for a more efficient uptake of sequences close to their coding regions, through transcription-translation coupling.<sup>49</sup> This possibility is substantiated by the detection of spacers in CRISPR arrays, as well as RNAs in ribonucleoprotein Cascade complexes, which sequences match regions in close proximity to CRISPR-Cas loci, including *cas* genes.<sup>12,25,50</sup> Alternatively, CRISPR-harboring molecules (including the chromosome) could be less prone to contribute spacers as consequence of an unforeseen discerning mechanism. This matter requires further investigation.

The large set of spacers acquired from pCas1-2(K) sequences allowed us to assess whether a preference for particular regions was apparent. Protospacers were scattered along the plasmid and located on both strands in coding and intergenic regions (Fig. 2), corroborating previous reports that dismiss an influence of transcription or the direction of replication on protospacer selection.<sup>17,18,32</sup>

**Proposal of a ruler mechanism for insertion-site cleavage.** When an unconventional CRISPR/leader/*cas* combination was used (i.e., CRISPR-A/leader-A/*cas*-EK), we observed anomalous insertions (four events of reversed spacer orientation and three atypical CRISPR-spacer sequences; see Table S5), most probably due to wrong leader/Cas interactions. Strikingly, the three atypical sequences shared the same odd features: (1) the di-nucleotide GG at the spacer-adjacent end of the duplicated CRISPR was missing (Fig. 4) and (2) CRISPR intervening sequences were 35 bp long (instead of 33) containing AC at the leader-distal end (Fig. 4), independently of the sequence at the corresponding positions in the putative protospacer (see Table S5). These observations fit with integration between the second and third nucleotide of the leader (A<sub>1</sub>C<sub>2</sub>-T<sub>3</sub>) instead of at the CRISPR-leader junction (C-A<sub>1</sub>; Figs. 1A and 4A). Accordingly, note that the CRISPR-leader boundary in CRISPR2.1 arrays associated with *cas*-EK genes is CT (see Figs. 4A; Fig. S5). As a consequence of the erroneous insertion, the leader-proximal end of the distal CRISPR will adjoin the extra di-nucleotide AC (Fig. 4A). The fact that the new CRISPR is trimmed at the spacer-adjacent end suggests that following an initial cleavage at the leader, the new duplcon was generated after a subsequent nick within the adjacent CRISPR at a fixed distance (28 nt from the first cut) unrelated to a putative recognition of the CRISPR end sequence (gain of AC by the leader-distal duplcon is compensated with GG loss by the proximal CRISPR; see Fig. 4A). Yet, as insertion of a 63 bp CRISPR-spacer unit in the pCSIR plasmids would not be detected (it would not restore *cat* frame), a second cleavage at this distance from the first nick cannot be dismissed. Nevertheless, the three additions observed demonstrate that the leader-distal CRISPR edge is dispensable for cleavage, and the occurrence of

61 bp inserts but not of 58 or 64 bp supports the ruler mechanism hypothesis. Hence, taking into account these considerations, we can conclude that the initial step of integration consists of sequential cleavages at the insertion sites. The first cut, at the CRISPR/leader boundary, is sequence-dependent (probably recognized by either Cas1 or Cas2).<sup>51,52</sup> The second nick takes place at the leader-distal end of the CRISPR and is independent of the CRISPR sequence, most likely determined by the distance from the previous cleavage (Fig. 4B). Apart from Cas1 and Cas2, acquisition must require non-Cas proteins with activities like DNA ligation and polymerization. This could be achieved by the implication of DNA repair mechanisms, providing a new meaning to the interactions of Cas1 with key components of repair systems, including RecB, RecC and RuvB.<sup>51</sup> Although we have considered insertion of a double-stranded spacer sequence, integration of a single strand from the protospacer and its subsequent replication cannot be dismissed.

A ruler mechanism has also been proposed for protospacer excision by another CRISPR-Cas type I system.<sup>42</sup> As a consequence of this common strategy directing the two stages of the acquisition process (i.e., protospacer excision and spacer integration), CRISPR-spacer periodicity is conserved.

## Materials and Methods

**Strains.** *E. coli* K12 str. MG1655, *E. coli* O157:H7 str. EDL931 (Colección Española de Cultivos Tipo, CECT 4267) and *E. coli* str. BL21-AI (Novagen) were used as hosts in integration assays. The three strains harbor CRISPR2.2 and CRISPR2.3-leader2.3 loci. K12 strains also carry a CRISPR2.1 array made of CRISPR-T duplcons associated with a leader of type T and a complete set of *cas*-EK genes. The CRISPR2.1 locus of *E. coli* O157:H7 is made of type A repeats and leader, and adjoins a complete set of *cas*-EO genes. The CRISPR2.1 locus of BL21-derivative strains consists of CRISPR-A duplcons and both leader and *cas* genes are absent.<sup>12</sup> An arabinose-inducible T7 RNA polymerase is encoded in BL21-AI, whereas homologous genes are not present in the genome of the other two strains.

The CRISPR2.1 locus of *E. coli* str. ECOR69,<sup>53</sup> used as template for PCR amplification of the integration-reporter cassette cloned in pCSIR-A, has CRISPRs and leader of type A.<sup>12</sup>

**Plasmids.** Plasmids developed in this study and details of their construction are indicated in Table S4 and the respective primers used are shown in Table S7. The 9 Kb pWUR339 plasmid (provided by John van der Oost's laboratory) derives from the low-copy number pCDF-1b vector (Novagen) and carries the Cascade-*cas1-cas2* operon of *E. coli* K12 under a T7-*lac* promoter.<sup>25</sup> Constructions were performed by cloning DNA fragments (Fig. S5) obtained by PCR amplification with primers containing or not (cloning in 3'-T overhangs) restriction sites (see Tables S4 and S7).

The pCR2.1-Cm plasmid was obtained by cloning the chloramphenicol acetyltransferase (*cat*) gene from pKK232-8<sup>54</sup> in the high-copy number (pUC origin) pCR2.1 vector (Invitrogen). pCSIR-A derived from pCR2.1-Cm by insertion of a sequence carrying 43 bp of the leader2.1 and the adjacent



CRISPR-spacer-CRISPR from ECOR69 strain. pCSIR-T was constructed by cloning in the 3'-T overhangs of the linearized pCR<sup>®</sup> 2.1 vector (TA Cloning<sup>®</sup> Kit from Invitrogen) a segment containing 69 bp of the leader2.1 and the adjacent CRISPR-spacer-CRISPR amplified from *E. coli* K12 str. MG1655, and a subsequent insertion of an amplicon containing the *cat* gene as in pCR2.1-Cm. pCas1-2(K) and pCas1-2(O) were obtained by insertion in the pCDF-1b vector of amplicons containing both *cas1* and *cas2* genes from pWUR399 and *E. coli* O157:H7 str. EDL931, respectively.

For cloning into the 3'-T overhangs of the linearized pCR<sup>®</sup> 2.1 vector, PCR products were obtained with recombinant *Taq* polymerase (Invitrogen). DNA amplicons to be cloned in plasmids other than linear pCR<sup>®</sup> 2.1 vector (Table S4) were obtained with the Expand High Fidelity PCR System (Roche). A Mastercycler Gradient (Eppendorf) thermal cycler was used for PCR incubations. PCR products were purified and afterwards digested with Thermo Scientific enzymes according to Double Digest web tool recommendations ([www.thermoscientificbio.com/webtools/doubledigest/](http://www.thermoscientificbio.com/webtools/doubledigest/)). Ligations were performed with T4 DNA Ligase (New England Biolabs). Following ligation, pCR2.1 and pCDF-1b-derived plasmids were transformed in TOP10 (Invitrogen) or NovaBlue (Novagen) strains, respectively, and constructions were verified by sequencing with primers T7 (pCR2.1 and pCDF-1b) and M13R (pCR2.1) or T7t (pCDF-1b).

**Detection of CRISPR, leader and PAM sequences.** Leaders of ECOR strains were determined by sequencing PCR amplification products as described by Diez-Villaseñor et al.<sup>12</sup> CRISPR and leaders in available *E. coli* genomes were identified with the CRISPRfinder program ([www.crispr.u-psud.fr/Server/](http://www.crispr.u-psud.fr/Server/)).<sup>55</sup>

For the identification of *E. coli* CRISPR2 PAMs associated with A or T CRISPR-leader variants, protospacer regions with at least 30 nt identity to spacers of CRISPR-Cas I-E systems of *E. coli* strains were searched with the BLASTn program<sup>56</sup> run against the Nucleotide Collection database ([www.blast.ncbi.nlm.nih.gov/Blast.cgi](http://www.blast.ncbi.nlm.nih.gov/Blast.cgi)) and aligned using the leader-proximal end as a reference. Similarly, the experimental SAMs were determined by the alignment of sequences matching (100% identity) spacers acquired in the integration assays and detected in the replicons of the corresponding host using BLASTn or Genious R6 created by Biomatters Ltd. (available from [www.geneious.com/](http://www.geneious.com/)). The DNA strands carrying the protospacer nucleotides complementary to the corresponding spacer sequence in the crRNA were aligned using WebLogo ([www.weblogo.berkeley.edu/logo.cgi/](http://www.weblogo.berkeley.edu/logo.cgi/))<sup>57</sup> to obtain sequence logos as described elsewhere.<sup>23</sup>

**DNA purification and sequencing.** Genomic and plasmid DNA were purified from cells grown in LB medium (10 g/l tryptone, 5 g/l yeast extract and 10 g/l NaCl) using Wizard<sup>®</sup> Genomic DNA Purification Kit (Promega) and High Pure Plasmid Isolation Kit (Roche), respectively. PCR products and restriction fragments were purified with GFX<sup>™</sup> PCR DNA and Gel Band Purification Kit (GE Healthcare).

Sequencing was performed with the Big Dye Terminator Cycle Sequencing kit in an ABI PRISM 310 DNA Sequencer (Applied Biosystems) following the manufacturer's instructions (Servicios Técnicos de Investigación, Universidad de Alicante).

**Transformation procedure.** Transformations were performed by electroporation (2.45 KV, 25  $\mu$ F, 200  $\Omega$ ) using an Electroporator 2510 (Eppendorf). Electrocompetent cells were prepared following the procedure described by Shi et al.<sup>58</sup> Transformant colonies were selected on LB agar plates containing the appropriate antibiotics.

**Spacer acquisition assay.** As a preliminary step of the acquisition assays, in order to confirm sensitivity to chloramphenicol, colonies of *E. coli* carrying integration reporter and *cas*-containing plasmids were streaked on LB plates supplemented with 25  $\mu$ g/ml chloramphenicol (Sigma). For each acquisition assay, a different sensitive clone was grown for 12 h at 37°C with shaking (150 rpm) in LB liquid medium supplemented with the antibiotic necessary for plasmid selection (100  $\mu$ g/ml ampicillin, Sigma, or 20  $\mu$ g/ml streptomycin, Sigma, for pCSIR and *cas*-containing plasmids, respectively). Additional (up to six) 12 h cycles of growth were performed (i.e., 1:300 dilutions of the culture in fresh medium and incubation under the same conditions), until chloramphenicol-resistant colonies (expectedly carrying an extra spacer-CRISPR unit) were detected by spreading culture samples on solid LB medium containing 25  $\mu$ g/ml chloramphenicol. In order to avoid sampling colonies derived from a single adapted cell, only one resistant colony from each assay was selected for further analysis. Insertion was primarily assessed by PCR amplification of the integration cassette with primers T7 and M13R matching flanking sequences in the vector (see Table S7). Usually, PCR fragments with increased size were observed along with non-extended amplicons, the latter likely due to extant copies of the original construction. In these cases, in order to select cells enriched with the plasmid carrying the insertion, three additional cycles of growth (1:5,000 dilutions, incubated at 37°C with shaking for 24 h) were performed in LB medium supplemented with increasing concentrations of chloramphenicol (75  $\mu$ g/ml, 150  $\mu$ g/ml and 200  $\mu$ g/ml). Finally, plasmids were extracted and sequenced with M13R primer.

The first growth cycle in which adapted clones were detected provides an estimation of the efficiency of acquisition.

**Probability of detection of integrated spacers.** To calculate the probability of integrations not eliciting *cat* translation due to the presence in the spacer of stop codons in phase with the ATG start codon in the reporter plasmids, an equivalent frequency of the four nucleotides in the donor DNA was considered. Then, the probability of stop triplets (TAA, TAG or TGA) is  $3 \times 1/4^3 = 3/64$  and of a triplet not being a stop codon  $1 - 3/64 = 61/64$ . For each 33 nt spacer inserted there are 10 tri-nucleotides that could produce termination. For example, in the case of a single insertion, the remaining three nucleotides would complete triplets in the adjoining CRISPR: the nucleotide at the leader-distal end of the spacer contributes the triplet TCN and the opposite two nucleotides the NCG codon at the edges of the flanking CRISPR duplicons (underlined nucleotides). Hence, the probability of *n* insertions without translation interruption is  $(61/64)^{10n}$  (i.e., 0.62 for a single insertion). For a single insertion ( $n = 1$ ), taking into account that only 95% of naturally occurring spacers are 33 nt long, the probability of integration detection is  $(61/64)^{10} \times 0.95 = 0.59$ .

## Perspectives

The robustness of our results and their consistency with bioinformatically predicted and other experimentally supported data validate the integration reporter plasmids we have developed as a powerful and reliable genetic tool for characterizing acquisition by the CRISPR-Cas I-E system of *E. coli*. For instance, the occurrence of spacer insertion into these constructions and its efficiency under diverse conditions can serve to identify growth parameters and additional proteins affecting this process. Furthermore, the same approach could be used to develop equivalent strategies and instruments applicable to alternative systems and strains.

### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

### References

1. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol* 1987; 169:5429-33; PMID:3316184.
2. Mojica FJM, Juez G, Rodríguez-Valera F. Transcription at different salinities of *Haloflex mediterranei* sequences adjacent to partially modified *PstI* sites. *Mol Microbiol* 1993; 9:613-21; PMID:8412707; <http://dx.doi.org/10.1111/j.1365-2958.1993.tb01721.x>.
3. Mojica FJM, Díez-Villaseñor C, Soria E, Juez G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol* 2000; 36:244-6; PMID:10760181; <http://dx.doi.org/10.1046/j.1365-2958.2000.01838.x>.
4. Jansen R, Embden JD, Gastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* 2002; 43:1565-75; PMID:11952905; <http://dx.doi.org/10.1046/j.1365-2958.2002.02839.x>.
5. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 2005; 151:2551-61; PMID:16079334; <http://dx.doi.org/10.1099/mic.0.28048-0>.
6. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 2005; 60:174-82; PMID:15791728; <http://dx.doi.org/10.1007/s00239-004-0046-3>.
7. Pourcel C, Salvignol G, Vergnaud G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* 2005; 151:653-63; PMID:15758212; <http://dx.doi.org/10.1099/mic.0.27437-0>.
8. Westra ER, Swarts DC, Staals RH, Jore MM, Brouns SJJ, van der Oost J. The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. *Annu Rev Genet* 2012; 46:311-39; PMID:23145983; <http://dx.doi.org/10.1146/annurev-genet-110711-155447>.
9. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 2012; 482:331-8; PMID:22337052; <http://dx.doi.org/10.1038/nature10886>.
10. Mojica FJM, Garrett RA. Discovery and seminal developments in the CRISPR field. In: Barrangou R, van der Oost J, eds. *CRISPR-Cas systems: RNA-mediated adaptive immunity in bacteria and archaea*. Berlin-Heidelberg: Springer, 2013:1-32.

## Acknowledgments

This work was funded by the Ministerio de Economía y Competitividad (BIO2011-24417). The University of Alicante (Vicerrectorado de Investigación, Desarrollo e Innovación) supported the use of its research technical services. We are grateful to John van der Oost and Stan J.J. Brouns from the University of Wageningen for providing the plasmid pWUR399. Manuel Martínez-García (University of Alicante) helped with figures. We thank Rafael Maldonado (University of Alicante) for critical reading of the manuscript. C.D.-V. and N.M.G. contributed similarly to this work.

### Supplemental Material

Supplemental material may be found here:

[www.landesbioscience.com/journals/rnabiology/article/24023](http://www.landesbioscience.com/journals/rnabiology/article/24023)

11. Makarova KS, Haft DH, Barrangou R, Brouns SJJ, Charpentier E, Horvath P, et al. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 2011; 9:467-77; PMID:21552286; <http://dx.doi.org/10.1038/nrmicro2577>.
12. Díez-Villaseñor C, Almendros C, García-Martínez J, Mojica FJM. Diversity of CRISPR loci in *Escherichia coli*. *Microbiology* 2010; 156:1351-61; PMID:20133361; <http://dx.doi.org/10.1099/mic.0.036046-0>.
13. Kunin V, Sorek R, Hugenholtz P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol* 2007; 8:R61; PMID:17442114; <http://dx.doi.org/10.1186/gb-2007-8-4-r61>.
14. Pougach K, Semenova E, Bogdanova E, Datsenko KA, Djordjevic M, Wanner BL, et al. Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. *Mol Microbiol* 2010; 77:1367-79; PMID:20624226; <http://dx.doi.org/10.1111/j.1365-2958.2010.07265.x>.
15. Pul U, Wurm R, Arslan Z, Geissen R, Hofmann N, Wagner R. Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. *Mol Microbiol* 2010; 75:1495-512; PMID:20132443; <http://dx.doi.org/10.1111/j.1365-2958.2010.07073.x>.
16. Westra ER, Pul U, Heidrich N, Jore MM, Lundgren M, Stratmann T, et al. H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. *Mol Microbiol* 2010; 77:1380-93; PMID:20659289; <http://dx.doi.org/10.1111/j.1365-2958.2010.07315.x>.
17. Swarts DC, Mosterd C, van Passel MW, Brouns SJJ. CRISPR interference directs strand specific spacer acquisition. *PLoS One* 2012; 7:e35888; PMID:22558257; <http://dx.doi.org/10.1371/journal.pone.0035888>.
18. Datsenko KA, Pougach K, Tikhonov A, Wanner BL, Severinov K, Semenova E. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat Commun* 2012; 3:945; PMID:22781758; <http://dx.doi.org/10.1038/ncomms1937>.
19. Goren MG, Yosef I, Auster O, Qimron U. Experimental definition of a clustered regularly interspaced short palindromic dupleon in *Escherichia coli*. *J Mol Biol* 2012; 423:14-6; PMID:22771574; <http://dx.doi.org/10.1016/j.jmb.2012.06.037>.
20. Grissa I, Vergnaud G, Pourcel C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 2007; 8:172; PMID:17521438; <http://dx.doi.org/10.1186/1471-2105-8-172>.
21. Touchon M, Rocha EP. The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia coli* and *Salmonella*. *PLoS One* 2010; 5:e11126; PMID:20559554; <http://dx.doi.org/10.1371/journal.pone.0011126>.
22. Touchon M, Charpentier S, Clermont O, Rocha EPC, Denamur E, Branger C. CRISPR distribution within the *Escherichia coli* species is not suggestive of immunity-associated diversifying selection. *J Bacteriol* 2011; 193:2460-7; PMID:21421763; <http://dx.doi.org/10.1128/JB.01307-10>.
23. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 2009; 155:733-40; PMID:19246744; <http://dx.doi.org/10.1099/mic.0.023960-0>.
24. Haft DH, Selengut J, Mongodin EF, Nelson KE. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* 2005; 1:e60; PMID:16292354; <http://dx.doi.org/10.1371/journal.pcbi.0010060>.
25. Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 2008; 321:960-4; PMID:18703739; <http://dx.doi.org/10.1126/science.1159689>.
26. Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* 2011; 18:529-36; PMID:21460843; <http://dx.doi.org/10.1038/nsmb.2019>.
27. Sashital DG, Wiedenheft B, Doudna JA. Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol Cell* 2012; 46:606-15; PMID:22521690; <http://dx.doi.org/10.1016/j.molcel.2012.03.020>.
28. Westra ER, van Erp PB, Künne T, Wong SP, Staals RH, Seegers CL, et al. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol Cell* 2012; 46:595-605; PMID:22521689; <http://dx.doi.org/10.1016/j.molcel.2012.03.018>.
29. Almendros C, Guzmán NM, Díez-Villaseñor C, García-Martínez J, Mojica FJM. Target motifs affecting natural immunity by a constitutive CRISPR-Cas system in *Escherichia coli*. *PLoS One* 2012; 7:e50797; PMID:23189210; <http://dx.doi.org/10.1371/journal.pone.0050797>.
30. Shah SA, Erdmann S, Mojica FJM, Garrett RA. Protospacer recognition motifs: Mixed identities and functional diversity. *RNA Biol* 2013; 10: In press; PMID:23403393; <http://dx.doi.org/10.4161/rna.23764>.
31. Fineran PC, Charpentier E. Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology* 2012; 434:202-9; PMID:23123013; <http://dx.doi.org/10.1016/j.virol.2012.10.003>.

32. Yosef I, Goren MG, Qimron U. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res* 2012; 40:5569-76; PMID:22402487; <http://dx.doi.org/10.1093/nar/gks216>.
33. Andersson AF, Banfield JF. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* 2008; 320:1047-50; PMID:18497291; <http://dx.doi.org/10.1126/science.1157358>.
34. Horvath P, Romero DA, Coûté-Monvoisin AC, Richards M, Deveau H, Moineau S, et al. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J Bacteriol* 2008; 190:1401-12; PMID:18065539; <http://dx.doi.org/10.1128/JB.01415-07>.
35. Tyson GW, Banfield JF. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ Microbiol* 2008; 10:200-7; PMID:17894817.
36. Horvath P, Coûté-Monvoisin AC, Romero DA, Boyaval P, Fremaux C, Barrangou R. Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *Int J Food Microbiol* 2009; 131:62-70; PMID:18635282; <http://dx.doi.org/10.1016/j.ijfoodmicro.2008.05.030>.
37. Pride DT, Sun CL, Salzman J, Rao N, Loomer P, Armitage GC, et al. Analysis of streptococcal CRISPRs from human saliva reveals substantial sequence diversity within and between subjects over time. *Genome Res* 2011; 21:126-36; PMID:21149389; <http://dx.doi.org/10.1101/gr.111732.110>.
38. Pride DT, Salzman J, Relman DA. Comparisons of clustered regularly interspaced short palindromic repeats and viromes in human saliva reveal bacterial adaptations to salivary viruses. *Environ Microbiol* 2012; 14:2564-76; PMID:22583485; <http://dx.doi.org/10.1111/j.1462-2920.2012.02775.x>.
39. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 2007; 315:1709-12; PMID:17379808; <http://dx.doi.org/10.1126/science.1138140>.
40. Lillestøl RK, Shah SA, Brügger K, Redder P, Phan H, Christiansen J, et al. CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. *Mol Microbiol* 2009; 72:259-72; PMID:19239620; <http://dx.doi.org/10.1111/j.1365-2958.2009.06641.x>.
41. Held NL, Herrera A, Cadillo-Quiroz H, Whitaker RJ. CRISPR associated diversity within a population of *Sulfolobus islandicus*. *PLoS One* 2010; 5:e12988; PMID:20927396; <http://dx.doi.org/10.1371/journal.pone.0012988>.
42. Erdmann S, Garrett RA. Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. *Mol Microbiol* 2012; 85:1044-56; PMID:22834906; <http://dx.doi.org/10.1111/j.1365-2958.2012.08171.x>.
43. Mojica FJM, Díez-Villaseñor C. The on-off switch of CRISPR immunity against phages in *Escherichia coli*. *Mol Microbiol* 2010; 77:1341-5; PMID:20860086; <http://dx.doi.org/10.1111/j.1365-2958.2010.07326.x>.
44. Ochman H, Selander RK. Standard reference strains of *Escherichia coli* from natural populations. *J Bacteriol* 1984; 157:690-3; PMID:6363394.
45. Semenova E, Nagornykh M, Pyatnitskiy M, Artamonova II, Severinov K. Analysis of CRISPR system function in plant pathogen *Xanthomonas oryzae*. *FEMS Microbiol Lett* 2009; 296:110-6; PMID:19459963; <http://dx.doi.org/10.1111/j.1574-6968.2009.01626.x>.
46. Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, et al. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci USA* 2011; 108:10098-103; PMID:21646539; <http://dx.doi.org/10.1073/pnas.1104144108>.
47. Marraffini LA, Sontheimer EJ. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* 2010; 463:568-71; PMID:20072129; <http://dx.doi.org/10.1038/nature08703>.
48. Wiedenheft B, Lander GC, Zhou K, Jore MM, Brouns SJJ, van der Oost J, et al. Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* 2011; 477:486-9; PMID:21938068; <http://dx.doi.org/10.1038/nature10402>.
49. Castro-Roa D, Zenkin N. In vitro experimental system for analysis of transcription-translation coupling. *Nucleic Acids Res* 2012; 40:e45; PMID:22210860; <http://dx.doi.org/10.1093/nar/gkr1262>.
50. Dyall-Smith ML, Pfeiffer F, Klee K, Palm P, Gross K, Schuster SC, et al. *Haloquadratum walsbyi*: limited diversity in a global pond. *PLoS One* 2011; 6:e20968; PMID:21701686; <http://dx.doi.org/10.1371/journal.pone.0020968>.
51. Babu M, Beloglazova N, Flick R, Graham C, Skarina T, Nocek B, et al. A dual function of the CRISPR-Cas system in bacterial antiviral immunity and DNA repair. *Mol Microbiol* 2011; 79:484-502; PMID:21219465; <http://dx.doi.org/10.1111/j.1365-2958.2010.07465.x>.
52. Nam KH, Ding F, Haitjema C, Huang Q, DeLisa MP, Ke A. Double-stranded endonuclease activity in *Bacillus halodurans* clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. *J Biol Chem* 2012; 287:35943-52; PMID:22942283; <http://dx.doi.org/10.1074/jbc.M112.382598>.
53. Selander RK, Caugant DA, Ochman H, Musser JM, Gilmour MN, Whittam TS. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl Environ Microbiol* 1986; 51:873-84; PMID:2425735.
54. Brosius J. Plasmid vectors for the selection of promoters. *Gene* 1984; 27:151-60; PMID:6327464; [http://dx.doi.org/10.1016/0378-1119\(84\)90136-7](http://dx.doi.org/10.1016/0378-1119(84)90136-7).
55. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 2007; 35(Web Server issue):W52-7; PMID:17537822; <http://dx.doi.org/10.1093/nar/gkm360>.
56. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; 25:3389-402; PMID:9254694; <http://dx.doi.org/10.1093/nar/25.17.3389>.
57. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res* 2004; 14:1188-90; PMID:15173120; <http://dx.doi.org/10.1101/gr.849004>.
58. Shi X, Karkut T, Alting-Mees M, Chamankhah M, Hemmingsen SM, Hegedus DD. Enhancing *Escherichia coli* electrotransformation competency by invoking physiological adaptations to stress and modifying membrane integrity. *Anal Biochem* 2003; 320:152-5; PMID:12895481; [http://dx.doi.org/10.1016/S0003-2697\(03\)00352-X](http://dx.doi.org/10.1016/S0003-2697(03)00352-X).