Fluorescent CRISPR Adaptation Reporter for rapid quantification of spacer acquisition

Lina Amlinger, Mirthe Hoekzema, E. Gerhart H. Wagner, Sanna Koskiniemi & Magnus Lundgren

CRISPR-Cas systems are adaptive prokaryotic immune systems protecting against horizontally transferred DNA or RNA such as viruses and other mobile genetic elements. Memory of past invaders is stored as spacers in CRISPR loci in a process called adaptation. Here we developed a novel assay where spacer integration results in fluorescence, enabling detection of memory formation in single cells and quantification of as few as 0.05% cells with expanded CRISPR arrays in a bacterial population. Using this fluorescent CRISPR Adaptation Reporter (f-CAR), we quantified adaptation of the two CRISPR arrays of the type I-E CRISPR-Cas system in *Escherichia coli*, and confirmed that more integration events are targeted to CRISPR-II than to CRISPR-I. The f-CAR conveniently analyzes and compares many samples, allowing new insights into adaptation. For instance, we show that in an *E. coli* culture the majority of acquisition events occur in late exponential phase.

CRISPR-Cas are prokaryotic adaptive immune systems that defend against e.g. bacteriophages\(^1\). They consist of a clustered regularly interspaced short palindromic repeats (CRISPR) array and CRISPR-associated (cas) genes that encode proteins required for immunity. CRISPR arrays are composed of repeats separated by short unique sequences called spacers, and are often preceded by a leader sequence\(^2\). Canonical CRISPR-Cas immunity is divided into three functional stages; adaptation, expression, and interference. During adaptation, new spacers acquired from e.g. phages and plasmids are integrated into the CRISPR array, enabling recognition of new targets\(^3\). During expression, the precursor CRISPR RNA (pre-crRNA) is processed and retained by Cas proteins, in some cases aided by additional factors\(^4, 5\). Finally, during interference the CRISPR RNA (crRNA) guides Cas proteins to complementary target sequences, which are subsequently cleaved or degraded\(^6, 7\).

Adaptation can be naïve, *i.e.* spacers are acquired from a sequence that is not already targeted by a spacer, which does not require the interference machinery\(^8\). The alternative is primed adaptation, where the interference machinery guides adaptation to a sequence that fully or partially matches a crRNA\(^10, 11\). Frequency of spacer acquisition in a population may be affected by environmental regulatory cues and other factors\(^12, 13\). The key proteins required for adaptation in all studied systems are Cas1 and Cas2. For details of the adaptation mechanism, see recent review\(^14\).

Here we report the development and characterization of a novel method for studying adaptation, fluorescent CRISPR Adaptation Reporter (f-CAR), using the type I-E CRISPR-Cas system of *Escherichia coli* MG1655\(^5\) as a model. MG1655 encodes two CRISPR loci, CRISPR-I carries 13 spacers and an adjacent set of cas genes whereas CRISPR-II (with six spacers) lacks neighboring cas genes\(^19\). The two CRISPR arrays differ in their respective leader sequences but have well conserved repeat sequences. The leader-proximal repeat sequence of the two CRISPR arrays only differs by one nucleotide. Naïve type I-E adaptation requires Cas1 and Cas2, 40–60 bp of the leader and one CRISPR repeat. Integration of a new 32 bp spacer results in duplication of the 29 bp leader-proximal repeat and the array is thus expanded by 61 bp\(^9\).

Several different methods to study CRISPR-Cas adaptation have been developed. In the most common approach, spacer integration is detected by PCR-amplification of the CRISPR arrays followed by gel electrophoresis, where the longer products generated from expanded arrays can be distinguished from the short products of unexpanded arrays\(^9, 10, 11, 20, 21\). However, detection of integration by PCR may be affected by amplification biases such as sequence preference and large differences in template abundance, and could result in false
negative results\textsuperscript{11}. PCR products or genomic DNA has also been analyzed by high-throughput DNA sequencing to quantify spacer integration\textsuperscript{17, 22, 23}. PCR and sequencing methods assay populations of bacteria and require substantial sample processing before obtaining the results. Other assays are based on functionality, where new spacers mediate plasmid curing\textsuperscript{10, 13} or survival of phage infection\textsuperscript{10, 20, 21, 24}. However, such methods do not detect non-functional or self-targeting adaptation events, as they may result in lack of phage protection or cytotoxicity\textsuperscript{25}.

The f-CAR method allows rapid, easy-to-use, low-cost, sensitive, and quantifiable detection of adaptation. The method also enables detection of spacer acquisition in real-time and in single cells, which previously described methods do not. Our f-CAR system is based on a previously published system where spacer integration results in expression of a chloramphenicol resistance gene\textsuperscript{26}, which allows for selection of cells that have acquired a spacer. We developed this system further by instead using a fluorescent read-out, enabling \textit{in vivo} detection and quantification of adaptation. Here, the reporter is \textit{Yfp}, referred to as \textit{Yfp-CAR}, but f-CAR could be developed for use with other fluorescent reporter proteins. We characterize the developed reporter system and use it to investigate biological questions on type I-E CRISPR-Cas adaptation.

Results and Discussion

Design and construction of \textit{Yfp-CAR}. To construct \textit{Yfp-CAR}, the synthetic constitutive promoter \textit{pJ23101}\textsuperscript{27}, a Shine-Dalgarno sequence and a translational start codon (ATG) were placed upstream of a partial CRISPR array originating from either of the two CRISPR arrays in \textit{E. coli} MG1655, CRISPR-I and CRISPR-II. These two constructs are referred to as \textit{Yfp-CAR}\textsuperscript{CR-I} and \textit{Yfp-CAR}\textsuperscript{CR-II}. The arrays consist of 69 bp of the respective leader, including the sequence elements essential for adaptation\textsuperscript{9, 28}, and one spacer flanked by two repeats. The \textit{yfp} sequence lacking translational start signals was inserted downstream of the CRISPR array, such that upstream initiation results in out-of-frame translation (Fig. 1). In addition, translational stop codons (TAA) in the leader sequence are in frame with the ATG. For details and sequences of \textit{Yfp-CAR} constructs see Supplementary Table S1. The reporter is designed for constitutive transcription but translation terminates within the leader sequence of the RNA. As adaptation lengths the array by one 61 bp spacer-repeat unit, \textit{yfp} is moved into frame and the stop codons in the leader out of frame, resulting in \textit{Yfp} production and fluorescence (Fig. 1). Additional insertion events would move \textit{yfp} out of frame again. The entire construct was inserted into the \textit{galK} locus of the \textit{E. coli} chromosome in a strain deleted for \textit{cas} genes and both native CRISPR arrays. The latter was done to direct all integration events to \textit{Yfp-CAR} and allow accurate quantification of spacer acquisition. The constructed strain cannot perform interference based on acquired spacers, ensuring that cells are not lost from the population due to self-targeting. Absence of Cascade and Cas3 prevents primed adaptation, hence all spacer acquisitions are naïve.

Spacer acquisition assays were done in liquid culture with (or without) induction of Cas1 and Cas2 expression, followed by a period of regrowth with repression of \textit{cas} and \textit{cas} before measuring fluorescence (for details see Materials and Methods).

\section*{Spacers can be inserted into \textit{Yfp-CAR}.} \textit{Yfp-CAR} was first tested by performing a spacer acquisition assay followed by conventional detection of integration by PCR (Fig. 2a). When the Cas1-Cas2 complex was expressed, PCR products corresponding to unexpanded and expanded arrays were detected for both \textit{Yfp-CAR} CRISPR arrays (Fig. 2b). This demonstrated that the reporter arrays are targeted for integration. As expected, no expansion of the arrays was detected with non-functional Cas1 (Fig. 2b).

\section*{Spacer insertion into \textit{Yfp-CAR} generates fluorescence.} When cells from a spacer acquisition assay were transferred to solid media, fluorescent colonies could be observed after overnight incubation. As expected, the majority of colonies were non-fluorescent, but fluorescent colonies that maintained fluorescence after restreaking were readily detected (Figs 2c, S1 and S2). To confirm spacer integration in fluorescent colonies, we performed colony PCR on \textit{Yfp} positive (\textit{Yfp}+) and negative (\textit{Yfp}−) colonies. All 27 tested fluorescent colonies
had expanded CRISPR arrays whereas the 24 tested non-fluorescent colonies were unexpanded (Fig. 2d, supplementary Fig. S2, data not shown). We conclude that spacer integration leads to Yfp expression. Sequencing of PCR products from four selected Yfp+ colonies identified acquired spacers and their origin (Supplementary Fig. S3), demonstrating another useful application for the f-CAR. Three spacers were unambiguously derived from plasmid. The fourth spacer was from \( \text{lacI} \), which is present on both plasmid and chromosome. However, the source is likely plasmid \( \text{lacI} \) as it is present in more copies than chromosomal \( \text{lacI} \). The sequenced spacers were all different and corresponded to individual spacer integration events that likely occurred in the liquid culture before plating.

To further characterize adaptation of the Yfp-CAR CRISPR arrays, a spacer acquisition assay was performed and PCR-amplified arrays from the experiment were analyzed by high throughput sequencing. Out of a total of 28,305 and 28,139 array sequences, 34 and 197 demonstrated expanded arrays for Yfp-CAR CR-I and Yfp-CAR CR-II, respectively (Table 1). Previous analyses indicated that approximately 95% of integrated spacers should be 32 bp\(^{19}\). Spacers with different length would not move the reporter gene into frame, but we detect only one spacer of aberrant length (33 bp) that, incidentally, also contained a 1 bp deletion in a repeat so the array was still extended by a total of 61 bp. In addition, newly incorporated spacers can contain in-frame stop codons resulting in translation termination and absence of fluorescence. We found that 53% of spacers inserted into Yfp-CAR CR-I and 68% into Yfp-CAR CR-II had no in-frame stop codons in any of the ten possible positions (Table 1), and would therefore permit Yfp expression after adaptation. These numbers correlate well with the calculation that 59% of spacer integration events should allow f-CAR fluorescence, on average\(^{26}\). It should be noted that the frequency of fluorescent cells (here Yfp+) is thus not equal to the absolute adaptation frequency. Relative differences are, however, still accurately measured. In line with previously observed bias in spacer acquisition from plasmids rather than the chromosome in \( E. \) coli\(^{9} \), the majority of acquired spacers were derived from the plasmid expressing \( \text{Cas1} \) and \( \text{Cas2} \), and only 6 and 18 spacers originating from the chromosome for Yfp-CAR CR-I and Yfp-CAR CR-II, respectively (Table 1, Supplementary Table S2). As mentioned above \( \text{lacI} \) spacers likely originated from the plasmid and are listed as such. Three spacers did not map to the plasmid or the chromosome and could be derived from sequences present in the organism but not represented in the available plasmid and genome sequences. The spacers origins were distributed over the entire plasmid and indicated no strand bias (Supplementary Fig. S4). Most sequenced spacers were unique, corresponding to individual spacer integration events, and not amplifications of spacers due to e.g. growth advantage or PCR bias.

**Yfp-CAR accurately and sensitively quantifies cells with expanded arrays.** To determine the detection limit and accuracy of Yfp-CAR, cells from a designed pre-expanded control strain (Yfp+) were mixed at different ratios with cells carrying the unexpanded array (Yfp-) and analyzed by PCR and flow cytometry (Fig. 3). For all flow cytometry analysis, gates were set so that less than 0.01% positive events were detected in the strain with an unexpanded array and a non-functional Cas1 (Fig. 3a). By PCR, spacer integration could be detected in a sample with 0.5% cells with expanded arrays, but not with 0.25% (Fig. 3b). This correlates well with

| Sample         | Total no. of sequences | Expanded sequences | 32 bp spacers | Spacers without in-frame stop codons | Expanded sequences supporting Yfp fluorescence | Unique spacers | Genome targeting spacers |
|----------------|------------------------|--------------------|---------------|--------------------------------------|-----------------------------------------------|----------------|--------------------------|
| Yfp-CAR CR-I   | 28,305                 | 34                 | 34            | 18                                   | 53%                                           | 33             | 6                        |
| Yfp-CAR CR-II  | 28,139                 | 197                | 196           | 133                                  | 68%                                           | 170            | 18                       |

Table 1. Summary of analysis of expanded arrays by SMRT sequencing. Spacer acquisition assay using Yfp-CAR CR-I and Yfp-CAR CR-II was analyzed by SMRT sequencing of the CRISPR arrays in the two populations.
the previously reported PCR detection limit of 0.4% expanded arrays. In contrast, flow cytometry reproducibly detected expanded arrays when present in as few as 0.05% cells (Fig. 3c). Furthermore, samples with 0.5, 0.75 and 1% expanded arrays generated PCR bands of similar intensity, obscuring quantitative differences between these samples, whereas the expected percentages could be reproducibly measured by flow cytometry (Fig. 3b,c). We conclude that Yfp-CAR is at least 10-fold more sensitive than PCR-based assays, and that reliable quantification of adaptation events can be obtained even at low frequencies.

**Quantification and detection of adaptation using Yfp-CAR.**Spacer integration in Yfp-CARCR-I and Yfp-CARCR-II after spacer acquisition assays was quantified by flow cytometry. With Yfp-CARCR-I, 0.6% of the cells were fluorescent (Fig. 4b) whereas 3.0% of the cells with Yfp-CARCR-II were fluorescent (Fig. 4b), suggesting that spacer integration is about five times more frequent in Yfp-CARCR-II than in Yfp-CARCR-I. Congruent with this, a 5.8-fold preference for adaptation into Yfp-CARCR-II was observed by comparing the number of expanded sequences obtained by high throughput sequencing (Table 1). The qualitative difference is in line with a previous report where more integration events were detected in CRISPR-II than in CRISPR-I. Adaptation could also be detected in Yfp-CARCR-II in a strain with the two wildtype CRISPR arrays and the endogenous chromosomal cas genes (data not shown), demonstrating that integration occurs into Yfp-CAR even in the presence of native arrays. It should be noted that the endogenous E. coli cas genes are silenced by H-NS, so potentially lethal targeting of chromosome, or plasmid with selection marker, by CRISPR-Cas (which would preclude detection of spacer acquisition) does not occur under these conditions.

Without induction of Cas1 and Cas2 proteins, the percentage of fluorescent cells was dramatically reduced, but still above that in the negative controls (Fig. 4b). This is probably due to low-level leaky expression of Cas1 and Cas2.

Cells from assayed strains were also examined by microscopy. As expected, fluorescent cells were detected in a fraction of the population (Fig. 4c). Spacer integration in cells carrying Yfp-CARCR-II was also investigated by manual time-lapse microscopy. Analysis after 24 h of Cas1 and Cas2 protein expression detected the presence of fluorescent cells. These cells continued to produce fluorescent offspring (Supplementary Fig. S5), suggesting that no new spacers are integrated during imaging. Some cells were non-fluorescent at the beginning of regrowth but later became fluorescent (Supplementary Fig. S5). They likely represent integration events that occurred late in stationary phase, without enough time for detectable Yfp expression before initial imaging, but could also correspond to real-time observations of spacer integration.

**Both leader and repeat sequences affect adaptation frequency.** While the leader sequence of CRISPR-I and CRISPR-II differ substantially, their leader-proximal repeats display only one nucleotide difference (Fig. 4a). To investigate the sequence elements affecting adaptation frequencies and causing the observed differences in integration into Yfp-CARCR-I and Yfp-CARCR-II, minimal Yfp-CAR reporters were constructed with CRISPRs consisting only of the leader sequence and the leader-proximal repeat (Fig. 4d). Four combinations of leader and repeat sequences were tested for integration (Fig. 4d,e). Higher integration frequencies were observed for Yfp-CARCR-I, II vs. II (corresponding to CRISPR-II) than for Yfp-CARCR-I vs. CRISPR-I, similar to previous results. Average adaptation frequencies into the minimal reporter arrays were comparable to those of the longer arrays, 0.5% for Yfp-CARCR-I, II vs. Yfp-CARCR-II, II (from Fig. 4b and e), indicating that arrays with only one repeat are as functional as the longer arrays. Interestingly, the arrays with combinations of leader and repeat from different arrays exhibited intermediate adaptation frequencies (Fig. 4e), suggesting that both leader and repeat sequences contribute to adaptation efficiency. The sequence of the leader-repeat boundary has previously been shown to affect spacer integration frequencies, both in vitro and in vivo, but within this region only one base pair differs between the arrays. This single base pair difference in the repeat sequence decreases adaptation frequency by almost half, possibly due to a critical position in the insertion site for the new spacer. The leader sequences are less conserved (Fig. 4a) and further studies are required to elucidate the sequences within the leader that affect adaptation frequencies.

**New spacers are primarily integrated during late exponential phase.** As a further demonstration of the utility of f-CAR, the percentage of Yfp fluorescent cells was monitored over time during an acquisition assay. This time course assay addressed whether adaptation preferentially occurs during a specific growth phase. As integration is more frequent into CRISPR-II than into CRISPR-I, Yfp-CARCR-II was used. Samples were taken at regular time intervals, and the percentage of Yfp+ cells was measured after a short regrowth, thus cancelling out differences in fluorescence due to e.g. different Yfp-CAR promoter activity or Yfp maturation in different growth phases. During early exponential phase (Fig. 4f, white area), the percentage of Yfp+ cells remained unchanged, while it rapidly increased during late exponential phase (Fig. 4f, light grey area). The fraction of Yfp+ cells continued to increase in early stationary phase and plateaued during late stationary phase (Fig. 4f, dark grey area) to the same level observed in the previous acquisition assays with Yfp-CARCR-II (Fig. 4b and f). We conclude that, under the experimental conditions used, adaptation occurs during late exponential/early stationary phase, with very little or no adaptation during exponential growth and late stationary phase. Since Cas protein expression was induced we cannot exclude that the induction kinetics affect the observed growth phase dependence. It is however tempting to speculate that chromosome topology, replication or presence/absence of other factors may also play a role. As an example the IHF protein, which is required for adaptation, has been demonstrated to be essential for chromosome replication or presence/absence of other factors may also play a role. As an example the IHF protein, which is required for adaptation, has been demonstrated to be essential for chromosome replication or presence/absence of other factors may also play a role. As an example the IHF protein, which is required for adaptation, has been demonstrated to be essential for chromosome replication or presence/absence of other factors may also play a role.

**Conclusion and Outlook**
The fluorescent reporter for adaptation (f-CAR) described in this communication enables real-time, quantitative detection of acquisition events by the CRISPR-Cas system in E. coli. The method allow several options for...
measuring spacer integration by fluorescence; in single cells by e.g. flow cytometry or microscopy as well as in colonies. The system could be developed for other organisms and different types of CRISPR-Cas systems as long as the repeat-spacer unit is not (always) a multiple of three base pairs. Other possible developments are addition of a full set of cas genes to the system which, together with expression of pre-crRNA, could allow investigation of interference and primed adaptation. Also, detection of multiple spacer insertions could be achieved by further shifting the frame of yfp or by inserting multiple fluorescent reporter genes in different frames.

To demonstrate this read-out system's advantageous properties to address CRISPR-Cas biology, we performed several experiments. We confirm that adaptation is more frequent for CRISPR-II than for CRISPR-I, and demonstrate that both the leader and repeat sequences of CRISPR-II contribute to it being more favored for insertion than CRISPR-I. Furthermore, we show that adaptation primarily takes place in late exponential/early stationary phase.

In conclusion, f-CAR is a fluorescent reporter system that provides new and significantly improved possibilities for detection and quantification of CRISPR-Cas adaptation. The f-CAR method will hopefully assist further investigations of both fundamental aspects and details of the adaptation process of CRISPR-Cas immune systems.

Methods
Reagents and growth conditions. Routine growth of E. coli was done in LB (5 g/L yeast extract, 10 g/L NaCl, 10 g/L tryptone), supplemented with the appropriate antibiotics (50 mg/L kanamycin, 100 mg/L ampicillin, 50 mg/L streptomycin, 15 mg/L chloramphenicol or 15 mg/L tetracycline). Media components were from Oxoid. All other chemicals, and all oligonucleotides, were from Sigma-Aldrich.
Strains and plasmids. For all experiments, a derivative of MG1655 deleted for all cas genes and both CRISPR arrays was constructed. The cas genes and CRISPR-I were deleted by Lambda Red recombination through insertion of a kanamycin cassette amplified from pKD4 with primers LML009 and LML010 (Supplementary Table S3). The kanamycin cassette was subsequently removed using the Flp recombinase encoded in CRISPR-I.

Figure 4. Quantification of adaptation using Yfp-CAR. (a) Alignment of the partial leader and leader-proximal repeat of the two CRISPR arrays in MG1655 used in this study. Identical bases are indicated by · in CRISPR-II. Repeat is highlighted in grey. (b) Flow cytometry detection of Yfp + cells in strains with Yfp-CAR after spacer acquisition assay and regrowth with or without induction of Cas1-Cas2 expression. Negative control strain (C-) with unexpanded Yfp-CAR and pCas1D221A + 2 shown for comparative purpose. Error bars: SD. N = 6 (Yfp-CAR,C-I), N = 3 (Yfp-CAR,C-II). (c) Imaging of indicated strains after spacer acquisition assay and subsequent regrowth with phase contrast and fluorescence microscopy. Scale bar is 1 µm. (d) Graphical representation of the minimal CRISPR arrays used in (e), consisting of partial leader (line) and one repeat (diamond) from CRISPR-I (blue) or CRISPR-II (orange). (e) Flow cytometry quantification of integration into minimal Yfp-CAR with leader sequences and repeat sequences in the indicated combinations after spacer acquisition assay and subsequent regrowth. Error bars: SD. N = 3. (f) Time course of adaptation using Yfp-CAR,C-II with induction of Cas1 and Cas2 expression at t = 0. Samples were withdrawn at the indicated time points and Yfp + cells were quantified by flow cytometry after a short regrowth. Non-shaded area: exponential phase (t = 0 to t = 4.5h), light grey area: late exponential phase (t = 4.5h to 9.5h), dark grey area: stationary phase. Error bars: SD. N = 3.
by pCP20 as previously described, creating a marker-less deletion of cas3-CRISPR-I. CRISPR-II was deleted by scar-less mutagenesis using the kan-sacB cassette. Briefly, the CRISPR-II array was replaced by a kan-sacB cassette by Lambda Red recombination and subsequent kanamycin selection. Next, a single-stranded DNA oligonucleotide, LA157, was used to remove the kan-sacB cassette in a second recombination step. Sucrose toxicity in cells expressing SacB was used to select for cells that had successfully replaced the kan-sacB cassette with the provided oligonucleotide.

Cas1 and Cas2 were expressed from pCas1 + 2 (Supplementary Table S4), where Cas1 and Cas2 are cloned under an IPTG-inducible T7-promoter. Cas1 with a point-mutation inactivating its DNase activity was used as a negative control; pCas1 D221A + 2 (Supplementary Table S4). The gene encoding the T7 polymerase was inserted in the araBAD-operand by P1 mediated transduction from MLS640. MLS640 is a BW25113 derivative, constructed by insertion of the T7 polymerase and tetracycline resistance gene in the araBAD-locus by Lambda Red recombination. The T7 polymerase and tetR were amplified from BL21AI using LA120 and LA121 primers. Primers and overhangs were chosen so that the intergenic region between araC and the araBAD promoter was reconstituted in the constructed BW25113-strain, but the araBAD operon was not. All relevant primers and DNA oligonucleotides can be found in Supplementary Table S3. All relevant strains and plasmids can be found in Supplementary Table S4.

**Construction of Yfp-CAR.** A yfp gene (sYFP2, Genbank KM0183003) was cloned after the CRISPR-array from pCSIR-T, which corresponds to parts of CRISPR-I. The yfp gene was cloned out of frame (+2) of the start codon (ATG) with the CRISPR-I array between the ATG and the yfp. This also places a stop codon (located in the leader) in frame with the start codon, before the yfp. The partial array includes 69 bp of the leader, two repeats, one spacer and one partial spacer. The sequence is identical to CRISPR-I in MG1655 except for a point mutation in the yfp leader and repeat. This also places a stop codon (located in the leader) (ATG) with the CRISPR-I array between the ATG and the yfp translational start codon and a stop codon (located in the leader) (ATG) with the CRISPR-I array between the ATG and the yfp. The partial array includes 69 bp of the leader, two repeats, one spacer and one partial spacer. The sequence is identical to CRISPR-I in MG1655 except for a point mutation in the yfp leader and repeat. This also places a stop codon (located in the leader) (ATG) with the CRISPR-I array between the ATG and the yfp.

For time course experiment, overnight cultures were made in LB with 50 mg/L streptomycin, 1% glucose and 0.1 mM IPTG and 0.2% arabinose to induce Cas protein expression. Assay was performed at 37 °C for 24 h. To allow for expression and maturation of Yfp from late integration events, the cultures were subsequently diluted 1:100 in LB with 0.2% glucose, to repress further Cas protein expression, and grown for 5 h at 37 °C. After 5 h, samples were taken for flow cytometry for single-cell detection of integration as well as for PCR for detection of integration in the whole population (see below).

For time course experiment, overnight cultures were made in LB with 50 mg/L streptomycin, 1% glucose and 0.1 M potassium phosphate buffer in order to prevent leaky expression of Cas1 and Cas2 which may cause premature spacer integration. Overnight cultures were diluted as before and, at indicated time points, samples were taken and OD600 was measured. The samples were centrifuged for 2 minutes to pellet cells and cells were resuspended in LB with 0.2% glucose to approximately the same OD600 (0.025) for all time points. Regrowth was done for 1.5 h at 37 °C. After regrowth, cells were pelleted by centrifugation for 2 min, and resuspended in ice cold, sterile filtered phosphate-buffered saline (PBS). Samples were kept on ice for the duration of the time course and integration was detected as Yfp + cells by flow cytometry as described below.

**Spacer integration detection by PCR.** For PCR, 500 µl of culture was pelleted and cells were resuspended in 100 µL sterile water. The sampled volume was adjusted according to OD600 to sample approximately the same number of cells for all assayed cultures. For spacer integration assays, OD600 were 3.5–4.5 after 24h of growth. Cells were lysed by boiling 10 min at 95 °C. Cell debris was pelleted and 5 µl of culture supernatant was used for a 50 µl PCR reaction using DreamTaq Green DNA polymerase (Thermo Fisher Scientific) and primers amplifying the CRISPR array (for Yfp-CAR: LA112 and LA007, for Yfp-CAR: LA112 and LA168; Supplementary DNA oligonucleotides can be found in Supplementary Table S3. All relevant strains and plasmids can be found in Supplementary Table S4.
Yfp positive events for three (Yfp-CAR CR-II) or six (Yfp-CAR CR-I) biological replicates were averaged, error bars indicate standard deviation. Fluorescent colonies were visualized in ChemiDoc™ MP with Image lab v. 4.0 software (Bio-Rad) using a blue laser (488 nm; bandpass filter 525/50 channel B1). 100,000 events were recorded for each sample. Data were acquired with the MACSQuantify™ Software (Miltenyi Biotec) and processed with FlowJo Software (FlowJo, LLC).

Events were gated for bacterial cells using side scatter measurements; these were subsequently classified as Yfp positive or negative by bifurcating the Yfp channel. The Yfp signal was always gated in such a way that there were <0.01% positive events in the negative control (strain MLS904 or MLS989 with pCas1 D221A+2). The number of Yfp positive events for three (Yfp-CAR CR-II) or six (Yfp-CAR CR-I) biological replicates were averaged, error bars indicate standard deviation.

High-throughput sequencing of acquired spacers. Yfp-CARs were sequenced after a spacer acquisition assay using Single Molecule Real Time (SMRT) sequencing. To create the PCR products used for SMRT sequencing, cell lysate from a representative spacer integration assay of Yfp-CAR CR-I, Yfp-CAR CR-II was used as template. Cell lysate preparation was done as described above. PCR reactions were performed using Phusion Hot Start Flex DNA polymerase (New England Biolabs) for 35 rounds of amplification. The same primer pair, LA112 and LA168, was used for both Yfp-CARs to minimize potential amplification bias caused by primer differences. PCR product was purified using AMPure PB beads (Pacific BioSciences). Templates were prepared using maximum recommended amount of adapters and sequenced on a PacBio RS II sequencer (Pacific BioSciences) using one SMRT cell per sample. Circular consensus sequencing (CCS) algorithm was used to generate sequences, and information about newly acquired spacers was extracted.

Test of detection limit. To examine the detection limit of spacer integration by PCR and flow cytometry, we performed a dilution test using the Yfp-CAR CR-I pre-expanded Yfp positive control (MLS902) and the strain with unexpanded array (MLS904), both with the inactive Cas1 to prevent integration of new spacers. Overnight cultures were diluted 1:100 in LB with 0.2% glucose and grown for 5 h at 37°C. OD600 was adjusted to 4 for both cultures before serial dilution of the pre-expanded control in the strain with unexpanded CRISPR array. Samples were taken for analysis by flow cytometry and 500 µl was used for spacer integration PCR as described above. In the same way, three more dilution series were made using the Yfp-CAR CR-II positive control (MLS990) and unexpanded strain (MLS989), again with pCas1 D221A+2. Overnight cultures were back diluted as above and the fresh cultures were diluted to an OD600 of 4. The Yfp positive control strain (MLS990) was serially diluted in the unexpanded strain (MLS989) in three independent series and analyzed by flow cytometry. Gating of cells after flow cytometry was done as described above. The numbers of Yfp positive events from the four dilution series (one with Yfp-CAR CR-I and three with Yfp-CAR CR-II) were averaged and the standard deviation was calculated.

Fluorescence microscopy. After spacer acquisition assay and subsequent regrowth, samples were taken for microscopy. Cells were imaged using a Nikon Ti-Eclipse inverted microscope for phase contrast and Yfp fluorescence using a 20 ms capture time. Images were captured using a sCMOS camera (Andor, UK) with a Nikon CFI Plan Apo Lambda 60X Oil objective and all images were processed the same way using Adobe Photoshop CS6 software for overlays.

For real-time imaging (Supplementary Fig. S5) cells were placed on a 1% agarose pad supplemented with 1x LB directly after spacer acquisition assay, and images were acquired at three time points (0 h, 1 h, and 3 h) as described above.

References
1. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709–1712, doi:10.1126/science.1138140 (2007).
2. Bolotin, A., Quinquis, B., Sorek, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151, 2551–2561, doi:10.1099/mic.0.28048-0 (2005).
3. Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. Journal of molecular evolution 60, 174–182, doi:10.1007/s00239-004-0046-3 (2005).
4. 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 151, 653–663, doi:10.1099/mic.0.27437-0 (2005).
5. Brouns, S. J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 966–964, doi:10.1126/science.1159689 (2008).

6. Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNaSE III. *Nature* **471**, 602–607, doi:10.1038/nature99886 (2011).

7. Garneau, J. E. et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71, doi:10.1038/nature09523 (2010).

8. Westra, E. R. et al. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Molecular cell* **46**, 595–605, doi:10.1016/j.molcel.2012.03.018 (2012).

9. Yosef, I., Goren, M. G. & Qmuron, U. Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic acids research* **40**, 5569–5576, doi:10.1093/nar/gkx216 (2012).

10. Datsenko, K. A. et al. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature communications* **3**, 945, doi:10.1038/ncomms1937 (2012).

11. Staals, R. H. et al. Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. *Nature communications* **7**, 12853, doi:10.1038/ncomms12853 (2016).

12. Semenova, E. et al. Highly efficient primed spacer acquisition from targets destroyed by the Escherichia coli type I-E CRISPR-Cas interfering complex. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 7626–7631, doi:10.1073/pnas.1602639113 (2016).

13. Swarts, D. C., Mosterd, C., van Passel, M. W. & Brouns, S. J. CRISPR interference directs strand specific spacer acquisition. *Plos one* **7**, e35888, doi:10.1371/journal.pone.0035888 (2012).

14. Patterson, A. G. et al. Quorum Sensing Controls Adaptive Immunity through the Regulation of Multiple CRISPR-Cas Systems. *Molecular cell* **64**, 1102–1108, doi:10.1016/j.molcel.2016.11.012 (2016).

15. Hoyland-Krogshbo, N. M. et al. Quorum sensing controls the Pseudomonas aeruginosa CRISPR-Cas adaptive immune system. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 131–135, doi:10.1073/pnas.1617451113 (2017).

16. Levy, A. et al. CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **520**, 505–510, doi:10.1038/nature14302 (2015).

17. Jackson, S. A. et al. CRISPR-Cas: Adapting to change. *Science* **356**, doi:10.1126/science.aal5056 (2017).

18. Diz-Villasenor, C., Almendros, C., García-Martínez, J. & Mojica, F. J. Diversity of CRISPR loci in Escherichia coli. *Microbiology* **156**, 1373–1361, doi:10.1099/mic.0.036046-0 (2010).

19. Heber, R. et al. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* **519**, 199–202, doi:10.1038/nature14245 (2015).

20. Wei, Y. et al. CRISPR-Cas9 function and host genome sampling in Type II-A CRISPR-Cas adaptation. *Genes & development* **29**, 356–361, doi:10.1101/gad.257550.114 (2015).

21. Savitskaya, E., Semenova, E., Dedkova, Y., Meliltikova, A. & Severinov, K. High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in E. coli. *RNA Biol* **10**, 716–725, doi:10.4161/rna.24325 (2013).

22. Fineran, P. C. et al. Degenerate target sites mediate rapid primed CRISPR adaptation. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E1629–1638, doi:10.1073/pnas.1400771111 (2014).

23. Ivanic-Bace, I., Cass, S. D., Wearne, S. J. & Bolt, E. L. Different genome stability proteins underpin primed and naive adaptation in E. coli CRISPR-Cas immunity. *Nucleic acids research* **43**, 10821–10830, doi:10.1093/nar/gkt1213 (2015).

24. Vercoe, R. B. et al. Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLoS genetics* **9**, e1003454, doi:10.1371/journal.pgen.1003454 (2013).

25. Diz-Villasenor, C., Guzman, N. M., Almendros, C., García-Martínez, J. & Mojica, F. J. CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of Escherichia coli. *RNA biology* **10**, 792–802, doi:10.4161/rna.24023 (2013).

26. Kelly, J. R. et al. Measuring the activity of BioBrick promoters using an in vivo reference standard. *J Biol Eng* **3**, 4, doi:10.1186/1754-1611-3-4 (2009).

27. Nunez, J. K., Bai, L., Harrington, L. B., Hinder, T. L. & Doudna, J. A. CRISPR Immunological Memory Requires a Host Factor for Specificity. *Molecular cell* **62**, 824–833, doi:10.1016/j.molcel.2016.04.027 (2016).

28. Pul, U. et al. Identification and characterization of E. coli CRISPR-Cas promoters and their silencing by H-NS. *Molecular microbiology* **75**, 1495–1512, doi:10.1111/j.1365-2958.2010.07073.x (2010).

29. Ali Azam, T., Iwata, A., Nishimura, A., Ueda, S. & Ishihama, A. Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. *Journal of bacteriology* **192**, 158, doi:10.1128/JB.01875-14 (2014).

30. Chen, C. et al. Identification of E. coli CRISPR-cas promoters and their silencing by H-NS. *Molecular microbiology* **47**, 1335–1343, doi:10.1111/j.1365-2958.2013.07955.x (2013).

31. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci USA* **97**, 6640–6645, doi:10.1073/pnas.100114397 (2000).

32. Checroun, P. P. & Wackernagel, W. 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 (1995).

33. Pietz, F. et al. Ciprofloxacin selects for RNA polymerase mutations with pleiotropic antibiotic resistance effects. *J Antimicro Chemother* **72**, 75–84, doi:10.1093/jac/dkw364 (2017).

34. Gullberg, E., Albrecht, L. M., Karlsson, C., Sandgren, L. & Andersson, D. I. Selection of a multidrug resistance plasmid by sublethal levels of antibiotics and heavy metals. *mBio* **5**, e01918–01914, doi:10.1128/mBio.01918-14 (2014).

**Acknowledgements**

We thank Erik Gullberg for generously sharing strains. We also thank Martin Larsson, Tim Blokker and Jonas Kjellin for their contributions to the project. This work was supported by the Swedish Research Council (M.L., E.G.H.W.), the Wenner-Gren Foundations, the Royal Swedish Academy of Sciences and the Uppsala RNA Research Center (M.L.). We acknowledge the National Genomics Infrastructure (NGI) Uppsala Genome Center (UGC) and UPPMAX for providing assistance in massive parallel sequencing and computational infrastructure, and Moritz Buck at National Bioinformatics Infrastructure Sweden (NBIS) for assistance with sequence analysis. Work performed at NGI UGC has been funded by RFI/VR and Science for Life Laboratory, Sweden.

**Author Contributions**

L.A. and S.K. conceived the idea. L.A., M.H., E.G.H.W., S.K. and M.L. participated in experimental design. L.A., M.H. and S.K. performed the experiments. L.A., M.H., E.G.H.W., S.K. and M.L. participated in analyzing the results and writing the manuscript.

**Additional Information**

Supplementary information accompanies this paper at doi:10.1038/s41598-017-10876-z
Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017