DNA motifs determining the efficiency of adaptation into the *Escherichia coli* CRISPR array

Ido Yosef[a], Dror Shitrit[a], Moran G. Goren[a], David Burstein[b], Tal Pupko[b], and Udi Qimron[a,1]

[a]Department of Clinical Microbiology and Immunology, Sackler Faculty of Medicine, and [b]Department of Cell Research and Immunology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel

Edited by Jennifer A. Doudna, University of California, Berkeley, CA, and approved July 17, 2013 (received for review January 10, 2013)

Clustered regularly interspaced short palindromic repeats (CRISPR) and their associated proteins constitute a recently identified prokaryotic defense system against invading nucleic acids. DNA segments, termed protospacers, are integrated into the CRISPR array in a process called adaptation. Here, we establish a PCR-based assay that enables evaluating the adaptation efficiency of specific spacers into the type I-E *Escherichia coli* CRISPR array. Using this assay, we provide direct evidence that the protospacer adjacent motif along with the first base of the protospacer (5′-AAG) partially affect the efficiency of spacer acquisition. Remarkably, we identified a unique dinucleotide, 5′-AA, positioned at the 3′ end of the spacer, that enhances efficiency of the spacer's acquisition. Insertion of this dinucleotide increased acquisition efficiency of two different spacers. DNA sequencing of newly adapted CRISPR arrays revealed that the position of the newly identified motif with respect to the 5′-AAG is important for affecting acquisition efficiency. Analysis of approximately 1 million spacers showed that this motif is overrepresented in frequently acquired spacers compared with those acquired rarely. Our results represent an example of a short nonprotospacer adjacent motif sequence that affects acquisition efficiency and suggest that other as yet unknown motifs affect acquisition efficiency in other CRISPR systems as well.

defense mechanism | phage–host interaction | acquisition step

Clustered regularly interspaced short palindromic repeats (CRISPR) and their associated proteins (Cas) comprise an important prokaryotic defense system against horizontally transferred DNA (1–3) and RNA (4). This system shows remarkable analogies to the mammalian immune system (5, 6) and to eukaryotic RNA-interference mechanisms (7, 8). Three major types and 10 subtypes of CRISPR/Cas systems (9) have been found across ~90% of archaeal genomes and ~50% of bacterial genomes. All types consist of a CRISPR array—short repeated sequences called "repeats" flanking short sequences called "spacers." The array is usually preceded by a leader, AT-rich DNA sequence that drives CRISPR array expression and is important for acquiring new spacers into the array (10, 11). A cluster of *CRISPR-associated* (*cas*) genes encoding proteins that process the transcript, interfere with foreign nucleic acids, and acquire new spacers usually lies adjacent to the CRISPR array (12–14). RNA transcribed from the CRISPR array (crRNA) is processed by Cas proteins into RNA-based spacers flanked by partial repeats. These crRNAs specifically direct Cas interfering proteins to target nucleic acids matching the spacers. The spacers are acquired from these targeted sequences, termed "protospacers." Spacer acquisition into the CRISPR array consequently results in guiding the system to cleave DNA molecules harboring the corresponding protospacers. This feature renders the system competent in adaptively and specifically targeting invaders.

Spacer acquisition into a CRISPR array was first reported for *Streptococcus thermophilus* (2). It was shown that *S. thermophilus* that survives a phage challenge expands its CRISPR array with spacers identical to the protospacers of the challenging phage. Recently, we and others have demonstrated spacer acquisition by an *Escherichia coli* type I-E CRISPR array (11, 15–17). Those studies showed that (*i*) Cas1 and Cas2 are both essential for the acquisition step; (*ii*) a small portion of the leader sequence adjacent to the array is essential for the acquisition step; (*iii*) the first repeat is duplicated upon acquisition of a new spacer; (*iv*) a single repeat is necessary and sufficient for the acquisition of a new spacer.

As suggested from DNA sequence analyses, and later shown experimentally in different CRISPR/Cas subtypes, short, 2- to 5-bp sequences near the protospacer, termed protospacer adjacent motifs (PAMs), were found to be crucial for efficient recognition by the effector Cas proteins during the interference step; however, it was not clear whether these elements are also important for spacer acquisition (18, 19). Sequence analysis of newly obtained spacers from *E. coli* showed that a significant number (~75%) of the corresponding protospacers initiated with a G and the majority (50%) also had a 5′-AW (W = A or T) sequence upstream of the protospacer (11). The fact that the data were obtained without selection for functional spacers, due to the lack of interference proteins, indicated that this 5′-AWG motif not only determines the interference capability, but also enhances adaptation of the protospacers adjacent to it. Nevertheless, direct experimental evidence comparing the efficiency of adaptation of a protospacer that has or lacks this motif has not been provided.

Additional motifs affecting acquisition efficiency, other than the PAM, could theoretically be identified by high-throughput analysis of consensus motifs in acquired spacers. A recent study used high-throughput sequencing to obtain ~200,000 spacers derived from plasmids (20). Analysis of the obtained spacers showed that there is a clear preference for acquisition of certain plasmid protospacers over others, all having PAMs. Although the authors speculated that there are non-PAM motifs that account for this bias, thorough analysis of the acquired spacers did not reveal such a consensus sequence. In an earlier study, in *S. thermophilus*, ~500,000 spacers were analyzed. In this case too, a strong bias in protospacer selection was detected, nevertheless, a motif accounting for this bias was not identified (21).

In this study, we initially used a simpler, low-throughput technique, to identify motifs that determine acquisition efficiency and, consequently, validated them by using analyses of spacers obtained by high-throughput sequencing. We established two independent assays, based on PCR and DNA sequencing, to detect acquisition efficiency into the array. Using these assays, we provided direct evidence of a role for the 5′-AAG motif in determining acquisition efficiency. We further identified a dinucleotide motif, affecting acquisition of a pair of spacers, present at a 30-bp interval downstream of the 5′-AAG motif. We

Author contributions: I.Y., D.S., M.G.G., and U.Q. designed research; I.Y., D.S., and M.G.G. performed research; I.Y., D.S., M.G.G., D.B., T.P., and U.Q. analyzed data; and U.Q. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed: ehudq@post.tau.ac.il.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1300108110/-/DCSupplemental.

term this motif “acquisition affecting motif” (AAM). The higher occurrence of AAM in highly acquired spacers was then established in analysis of ∼1 million spacers from the *E. coli* chromosome, demonstrating that it plays a general role in determining acquisition efficiency.

Results and Discussion

Establishing an Assay for the Detection of Acquisition Efficiency of Specific Spacers. To gain insight into the genetic elements affecting the efficiency of acquisition, a robust assay was required. A PCR-based assay that monitors expansion of the CRISPR array upon acquisition can be used to determine the efficiency of total spacer acquisition but not the efficiency of a specific spacer’s acquisition (11). We therefore designed another simple PCR-based assay to detect acquisition efficiency of a specific spacer. In the modified assay, a culture of *E. coli* cells harboring plasmid pCas1+2, expressing the genes that are essential for adaptation, was propagated (11). The total culture, harboring numerous bacterial chromosomes with CRISPR arrays encoding newly adapted spacers from pCas1+2 and from the chromosome, was then used as the template in a PCR assay. We hypothesized that acquisition of random spacers could be detected by PCR using primers that match the protospacers’ sequences. For each PCR, we therefore used a primer complementary to the pCas1+2 plasmid on one strand and another primer complementary to the CRISPR array on the other (Fig. 1*A*). Thus, amplification could occur only on CRISPR arrays that had acquired a plasmid-derived protospacer that included the primer sequence (hereafter termed “matching spacer” and “matching primer,” respectively). We speculated that the intensity of the PCR product would correspond to the acquisition efficiency of the matching spacer, because the number of copies of initial template DNA depends on this efficiency. The assay could detect a single matching template among ∼1×10^5 different templates, as determined by PCR carried out on serial dilutions of the matching template mixed with nonmatching DNAs (Fig. S1).

To show that the assay actually detects specific acquisition events, we tested acquisition of 11 different spacers, using 11 matching primers, on cultures harboring pCas1+2. These primers, all annealing to plasmid sequences, were selected based on similar melting temperatures to minimize PCR variations. As a control, we used the same primers to test acquisition on cultures harboring pCas1^{D221A}+2, a plasmid encoding a point mutation that renders Cas1 nonfunctional (11, 22). In all instances, PCR amplification was evident in cultures harboring the functional Cas1 but not in those with the altered Cas1, indicating that the assay specifically detects acquisition events (Fig. 1*B*). The intensity of the PCR products amplified by the different primers varied significantly, suggesting that the matching spacers were acquired at different efficiencies (Fig. 1*B*). For example, the PCR amplification product using primer P6 was more than 13 times more intense than that obtained by using primer P11, as measured by imaging software (ImageJ). We designated primer P6—the primer showing the highest amplification level—as primer “HE” for “high efficiency” and primer P11, showing the lowest amplification level, as primer “LE” for “low efficiency.” The DNA sequences of primers HE and LE, along with 12 bp upstream and downstream of these primers are shown in Fig. 1*B*.

The different intensities of the DNA bands resulting from these two primers suggested that the acquisition efficiency of the respective matching spacers differed, the priming efficiency of each primer differed, or both. To estimate the contribution of each of these two possibilities to the observed difference in amplification intensities, we compared the priming efficiencies of both primers. To this end, the HE and LE primers were used to amplify an identical amount of template (plasmid pCas1+2) with an identical primer annealing to the reverse strand. PCR under these conditions showed that the band obtained by HE is approximately two times more intense than that obtained by LE, indicating that the priming efficiency of HE is approximately twice that of LE (Fig. S2). The fact that the differences observed in the acquisition efficiency assay exceeded 13-fold and that the priming efficiency could not account for the entire difference indicated that the efficiency of acquisition of the HE-matching spacers was significantly higher than that of the LE-matching spacer.

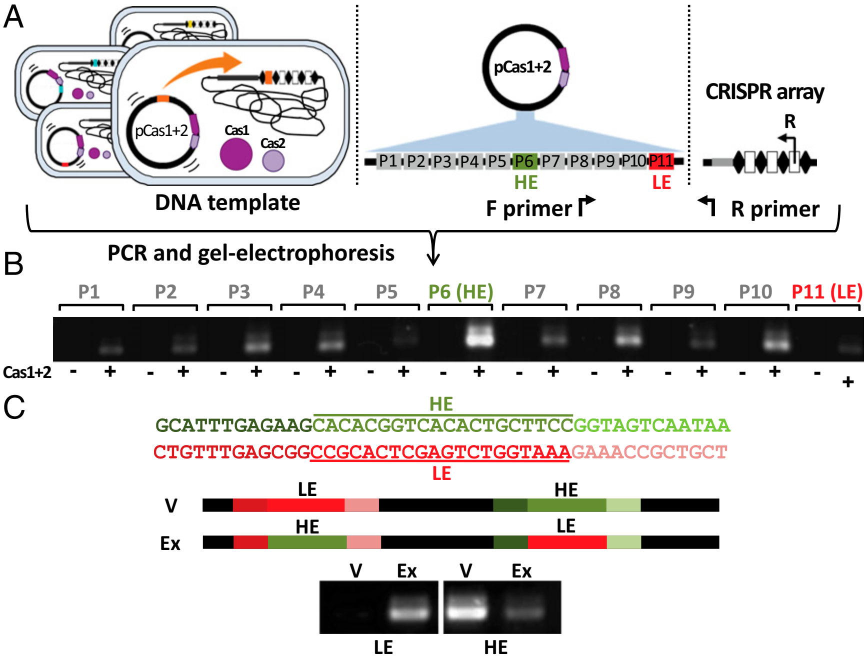

Fig. 1. Assay for quantification of specific spacer-acquisition efficiency. (*A*) Schematic representation of the assay. *E. coli* bacterial cells express Cas1 and Cas2 and, thus, acquire different spacers (derived from the chromosome and the resident plasmid) into their CRISPR array. DNA from these cells, having extended arrays, is used as a template in a PCR. One primer in this PCR (R) is homologous to the CRISPR array in the chromosome, and another primer (F) is homologous to a specific sequence in the plasmid DNA. Amplification is observed only when a plasmid segment matching the primer sequence is acquired by the array. The intensity of the amplification should thus be proportional to the amount of each acquired spacer. (*B*) Eleven primers annealing to pCas1+2 and to the control plasmid pCas1^{D221A}+2 were tested in the assay. −, bacterial culture harboring pCas1^{D221A}+2 expressing a nonfunctional Cas1; +, bacterial culture harboring pCas1+2 expressing functional Cas1. Expected product size is ∼320 bp. HE and LE primers are P6 and P11, respectively. HE and LE primers and their upstream and downstream regions are shown. Color codes: green, HE primer; dark green, 12 bp upstream of HE primer; light green, 12 bp downstream of HE primer; red, LE primer; dark red, 12 bp upstream of LE primer; pink, 12 bp downstream of LE primer. (*C*) Sequences of primers LE and HE were exchanged on pCas1+2 as depicted. PCR efficiency was determined on cultures harboring exchanged (Ex) or unmodified (V) plasmids. LE or HE primers were used as the F primer as indicated below the gel image. Gel images represent two experiments yielding similar results.

To unequivocally show that the acquisition efficiency of the matching spacers was the major factor determining the intensity of the bands, we measured the intensities of the PCR products after exchanging the positions of LE and HE on the plasmid: If the priming efficiency was the main reason for the differences, then the differences in intensity would remain. However, if acquisition efficiency was the dominant factor affecting intensity, then these intensities should reverse upon the exchange. The exchange resulted in reversal of the amplification intensity, i.e., the LE primer resulted in higher amplification than the HE primer (Fig. 1*C*). Despite this reversal, the intensity of the PCR product of LE on cultures harboring the exchanged vector was approximately twofold fainter than that obtained by using the HE primer on cultures harboring the original vector; this result was expected, in accordance with the approximately twofold decrease in LE priming efficiency compared with HE. These results indicated that the assay monitored differences in spacer-acquisition efficiencies and not merely priming efficiencies of each primer. Moreover, these results suggested that the elements that were left intact following the exchange of primers, i.e., the elements upstream and/or downstream of the primer, determine the acquisition efficiencies of the spacers.

DNA Sequences both Upstream and Downstream of the Primers Determine Acquisition Efficiency of the Matching Spacers. The above-described assay allowed us to map the location and determine the sequence of DNA elements affecting the efficiency of spacer acquisition. We exchanged 12 nt upstream and downstream of the sequence matching primer HE on the plasmid with the 12 nt found upstream and downstream, respectively, of the sequence matching primer LE (Fig. 2*A*). We then measured the efficiency of acquisition of the spacer matching HE in cultures harboring the modified or unmodified plasmids. As a loading control, we monitored the acquisition efficiency of the spacer matching primer LE, for which the adjacent upstream and downstream sequences remained identical in both plasmids. Exchange of 12 nt upstream of the HE primer resulted in a decrease in acquisition efficiency of the matching spacer. Exchange of 12 nt downstream of primer HE also resulted in a decrease in acquisition efficiency, indicating that both the upstream and downstream regions encode elements that determine acquisition efficiency. As expected, simultaneous replacement of the upstream and downstream DNA sequences of primer HE with those of primer LE resulted in an even greater decrease in acquisition of the matching spacer (Fig. 2*A*). Taken together, these results suggested that both the upstream and downstream regions of primer HE determine the acquisition efficiency of the spacer matching it. To show that the exchanged regions could also enhance acquisition efficiency of the spacer matching the LE primer, we carried out the reverse experiment, i.e., we exchanged 12 nt upstream and downstream of primer LE with the corresponding nucleotides of primer HE. Exchanging only the upstream region increased acquisition efficiency of the spacer matching primer LE. Exchanging only the downstream region also increased the acquisition efficiency, albeit to a lesser extent. Exchanging both the upstream and downstream 12-bp sequences increased the acquisition efficiency most dramatically (Fig. 2*B*). In a separate experiment, we showed that scrambling the sequence of the HE primer does not change its acquisition efficiency (Fig. S3). Taken together, these results indicated that the upstream and downstream sequences of primer LE each encode at least one motif that determines acquisition efficiency. Moreover, they showed that the upstream motif is slightly more dominant in determining acquisition efficiency, because its exchange resulted in a more pronounced effect.

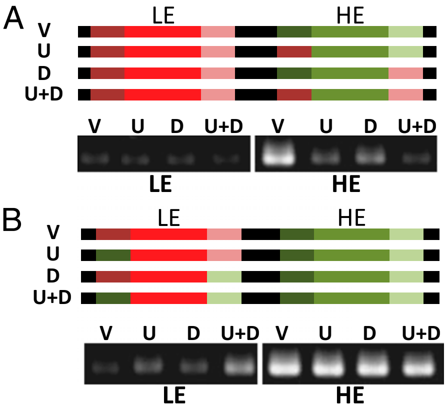

Fig. 2. Motifs affecting spacer-acquisition efficiency are present both upstream and downstream of the tested primers. Upstream and downstream nucleotides (12 each) of the HE (*A*) and LE (*B*) sequences were exchanged as depicted. Acquisition efficiency assays of cultures harboring the indicated plasmids were carried out as described in Fig. 1*A*. The LE or HE primer was used as the F primer as indicated below the gel image. V, unmodified pCas1+2; U, exchange of 12 bp upstream of the primer; D, exchange of 12 bp downstream of the primer; U+D, exchange of both upstream and downstream regions of the primers. Gel images represent three experiments yielding similar results. Green, HE primer; dark green, 12 bp upstream of HE primer; light green, 12 bp downstream of HE primer; red, LE primer; dark red, 12 bp upstream of LE primer; pink, 12 bp downstream of LE primer.

Identifying the Exact Motif Location and Validating Its Function. Spacers initiating with base G and having an upstream 5′-AW sequence are overrepresented, most likely due to their higher acquisition efficiency (11). As expected, a 5′-AAG motif was found immediately upstream of the HE primer, but not the LE primer. It was thus thought that 5′-AAG is the likely upstream motif responsible for enhanced acquisition of the spacer matching HE. However, direct evidence for this role was not provided. More importantly, the exact position or sequence of the DNA element that we identified downstream of the primer was unknown. To systematically locate and directly prove the functionality of both motifs, we exchanged 2–9 nt upstream of the HE primer with their counterparts upstream of the LE primer and monitored the decrease in acquisition efficiency. Note that six consecutive nucleotides at positions −4 to −9 upstream of both primers were identical and, thus, their exchange was redundant (Fig. 1*B*). Exchange of 3 bp upstream of the HE primer reduced the acquisition efficiency of the matching spacer to the same level as exchange of 12 bp, suggesting that these 3 bp, namely 5′-AAG, constitute the upstream motif that affects acquisition efficiency (Fig. 3*A*). In the region downstream of the primer, acquisition efficiency decreased only slightly following exchange of 3, 6, and 9 bp, but exchange of 12 bp decreased the acquisition efficiency dramatically (Fig. 3*B*). This result suggested that the major downstream motif is located between nucleotides 10 and 12 downstream of the end of the primer. To further assess the relative contribution of each of the identified nucleotides to acquisition efficiency, we point mutated each nucleotides at positions −3 and −2 upstream of primer HE, as well as positions 10, 11, and 12 downstream of primer HE, to the corresponding nucleotides flanking the LE primer. The 5′-G at position −1, which was similar in both HE and LE primers, was changed in the HE primer to a 5′-C. We then carried out the spacer-acquisition efficiency assay for each of the point-mutated plasmids. As can be seen in Fig. 3*C*, replacement of the A at position −3 with a C, replacement of the A at position −2 with a G, and replacement of the G at position −1 with a C, all resulted in a decrease in acquisition efficiency, with position −1 showing the largest decrease. This result provided direct evidence that the 5′-AAG motif determines acquisition efficiency and that the G in

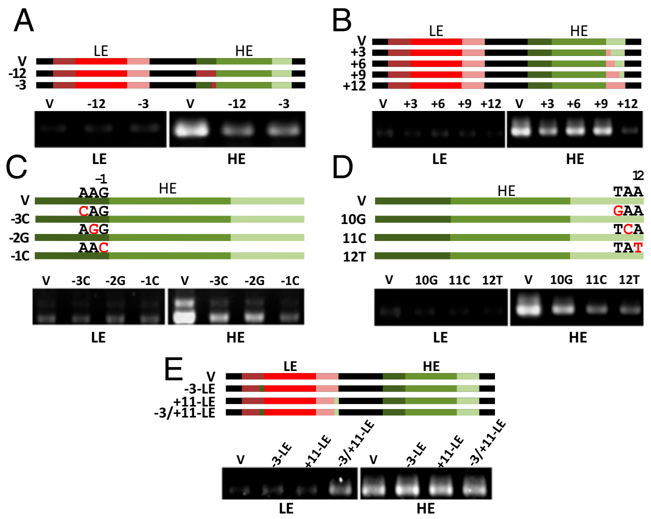

Fig. 3. Mapping of motifs affecting spacer-acquisition efficiency. (*A*) Acquisition efficiency assays following exchange of 12, 3, or 0 nt upstream of the HE primer, as depicted, were carried out as described in Fig. 1*A*. PCR was carried out by using the LE and HE primers as indicated below the gel images. V, unmodified pCas1+2; −12, exchange of 12 bp upstream of the sequence matching primer HE; −3, exchange of 3 bp upstream of the sequence matching primer HE. (*B*) Acquisition efficiency assays following exchange of 12, 9, 6, 3, or 0 nt downstream of the HE primer, as depicted. PCR was carried out by using the LE and HE primers as indicated below the gel images. V, unmodified pCas1+2; +3, +6, +9, +12, exchange of the indicated number of nucleotides downstream of the HE sequence. (*C*) Acquisition efficiency assays following point mutations of the indicated base pairs upstream of the sequence matching the HE primer. Numbers above the sequence indicate position number with respect to the primer start. (*D*) Acquisition efficiency assays following point mutations of the indicated bp downstream of the sequence matching the HE primer. Numbers above the sequence indicate position number with respect to the primer end. (*E*) Validation of the effect of the identified motifs on spacer-acquisition efficiency. Acquisition efficiency assays following exchange of 5′-AA at position −3 of the plasmid sequence matching the LE primer (-3-LE), at position +11 of the plasmid sequence matching the LE primer (+11-LE), or of both (-3/+11-LE), as depicted. PCR was carried out by using the LE and HE primers as indicated. Gel images represent three experiments yielding similar results. Green, HE primer; dark green, 12 bp upstream of HE primer; light green, 12 bp downstream of HE primer; red, LE primer; dark red, 12 bp upstream of LE primer; pink, 12 bp downstream of LE primer.

this motif is the most dominant nucleotide affecting acquisition. Importantly, in the primer's downstream motif, we found that the three nucleotides at positions 10, 11, and 12 affect acquisition efficiency. Exchange of nt 10 from T to G resulted in a small decrease in acquisition efficiency, whereas replacement of A at positions 11 and 12 to C and T, respectively, resulted in a more substantial decrease (Fig. 3*D*). We therefore concluded that the major element affecting acquisition efficiency in the region downstream of the primer is the dinucleotide at positions 11 and 12, with a minor contribution of the nucleotide at position 10.

To test whether the downstream dinucleotide is also responsible for the observed increase in the acquisition efficiency of the LE matching spacer, we constructed unique plasmids having the identified upstream and downstream motifs. We introduced these motifs at positions −3 and −2, and positions 11 and 12 relative to the 5′ and 3′ ends of primer LE, respectively. We then measured the acquisition efficiency of the spacer matching the LE primer in cultures harboring the modified vs. unmodified plasmid. Insertion of either the upstream or downstream motifs alone slightly increased acquisition efficiency, whereas insertion of both motifs dramatically increased acquisition efficiency of this spacer to levels similar to those observed by simultaneous exchange of 12 bp upstream and downstream of this primer (Fig. 3*E*). These experiments thus validated that the 5′-AAG found at position −3 of the primer, and 5′-AA found at position 11 downstream of the primer, significantly affect acquisition efficiency of the tested spacers. We term the 5′-AA motif AAM.

Sequence Analysis of Acquired Spacers Confirms That the Interval Between the Motifs Determine Acquisition Efficiency. To determine the borders of the acquired spacer matching the HE primer, we extracted the amplified DNA product, ligated it to a plasmid vector, and sequenced 20 products (Fig. 4*A*). The frequency of the obtained spacers was plotted against the distance from the 3′ end of the primer. In 95% of the cases, position 12 downstream of the primer was the last nucleotide of the sequenced spacer (Fig. 4*B*). The sequencing results thus indicated that the predominant acquired spacer initiates at position −1 with a 5′-G preceded by a 5′-AA motif and ends with the 5′-AA dinucleotide at positions 11 and 12 (Fig. 4*B*). We expected that in the LE-matching spacer, no discrete borders would be observed due to the lack of motifs that determine acquisition of a single predominant spacer. Indeed, sequencing of 20 LE-matching spacers revealed that there is no single predominant spacer acquired from this region, but that the acquired spacers are scattered near the end of the primer (Fig. 4*B*). We later verified these data by analyzing 22,493 spacers from these two regions, as shown in Fig. S4. These results also validated that the HE matching spacer was acquired more frequently then the LE matching spacer: HE-matching spacers were sampled 21,552 times, whereas LE-matching spacers were sampled only 941 times.

To further demonstrate that the AAM plays a significant role in determining the efficiency of this specific spacer acquisition, we sequenced CRISPR arrays that had adapted a spacer, as described above. We predicted that if the motifs 5′-AAG and AAM enhance the efficiency of adaptation of a spacer located in the 30-bp interval between them, then a protospacer encoding a 5′-AAG and having an AAM 30 nt downstream should be acquired at a higher efficiency than a protospacer having a 5′-AAG but lacking a downstream AAM. To test this hypothesis, we constructed three plasmids—all having two 5′-AAG motifs at 3-bp intervals from each other. On one plasmid, we inserted an AAM 30 bp downstream of the upper 5′-AAG; in the other, we inserted an AAM 30 bp from the lower 5′-AAG motif, and in another we did not insert AAM at any of these positions but rather inserted 5′-CT, the dinucleotide present in positions 32–33 of the LE matching spacer, at both positions (Fig. 4*C*). To determine the predominant acquired protospacer, we cloned and sequenced the amplified DNA harboring the acquired spacers. Approximately 30 spacers were sequenced for each plasmid. The frequency of the obtained spacers was again plotted against the distance from the 3′ end of the primer. As can be seen in Fig. 4*C*, 83% of the spacers acquired from the plasmid having the upper AAM had the expected upper border, whereas only 3% had the lower border. In the plasmid having the lower AAM, the trend reversed to 36 and 50% of the upper and lower borders, respectively. In the plasmid lacking the AAM, the frequencies were 47 and 27%, for the upper and lower 5′-AAG, respectively. These results demonstrate unequivocally that the AAM affects the acquisition efficiency of the LE-matching spacer. However, we expected that the ratios would be mirrored, such that in the lower pair of motifs, the frequency of spacers would be ∼80% from the lower border, whereas the frequency from the upper border would be ∼3%. The fact that the results did not completely reverse the ratios suggests that the acquisition machinery "scans" the DNA from the 5′ to 3′ end. In ∼35% of the cases, upon encountering a 5′-AAG motif, a protospacer is cleaved, regardless of the downstream motif, and the machinery runs off the DNA. In the remaining ∼65% of the cases it resumes scanning.

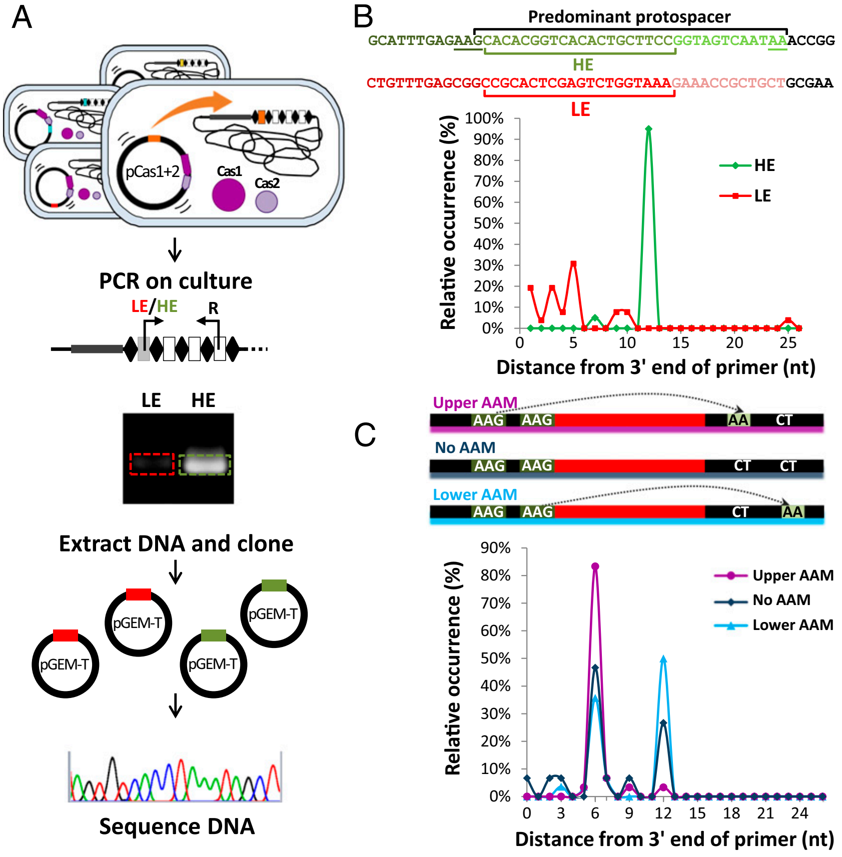

Fig. 4. DNA sequencing to characterize acquired spacers. (*A*) Schematic representation of the assay. *E. coli* bacterial culture expressing Cas1 and Cas2 and, thus, acquiring different plasmid-derived spacers in their CRISPR arrays, is used as a DNA template in a PCR. One primer is homologous to the CRISPR array in the chromosome, and another is homologous to either the LE or HE sequence in the plasmid DNA. The amplified PCR product is then extracted from the gel, ligated to a plasmid, and the cloned fragment is DNA sequenced. (*B*) Identifying the predominantly acquired spacer. Acquired spacers (20 HE-matching spacers and 26 LE-matching spacers) were sequenced as described above. The distance of the border of the acquired spacer with the repeat in the CRISPR array from the primer end is plotted on the *x* axis against the relative occurrence of spacers acquired at this position. The predominant acquired spacer matching HE is identified by using these results, and the motifs affecting acquisition efficiency are underlined. (*C*) Monitoring spacer acquisition efficiency following shifts in the dinucleotide position 30 nt downstream of a matching 5′-AAG. Spacers acquired from cultures harboring plasmids carrying the depicted DNA constructs were sequenced as described above. Dashed arrows bridge between the 5′-AA and their corresponding 5′-AAG located 30 nt upstream. The distance of the border of the acquired spacer with the repeat in the CRISPR array from the primer end is plotted on the *x* axis against relative occurrence of spacers acquired at this position.

However, when a 5′-AAG with a matching AAM downstream is encountered, the frequency of acquisition reaches more than 80%. Such a scanning model has been proposed to explain the observed "priming" mechanism, in which acquisition is facilitated from a particular DNA strand if a spacer from that strand is present in the array (15, 16). Although a recent study negated the scanning model in a different experimental settings (20), it is possible that it operates on certain DNA segments, under certain conditions. Our data thus corroborate the scanning model, but further studies are required to prove it. Another possible explanation, which we cannot exclude, might be that there are still more elements that slightly affect acquisition efficiency of the LE-matching spacer. Nevertheless, our overall results show that the 5′-AAG and the AAM 30 nt downstream of it are the most dominant factors affecting acquisition of this tested spacer.

Analysis of Approximately a Million Spacers Shows that the AAM Enhances Adaptation Efficiency. If indeed, the AAM enhances adaptation efficiency of spacers, then it should be overrepresented in frequently acquired spacers compared with rarely acquired spacers. To test this hypothesis, we sequenced 2.67 million spacers acquired from the *E. coli* chromosome and pCas1+2 plasmid. Approximately a million of these spacers were uniquely mapped to the chromosome and the remaining ~1.2 million were mapped to the plasmid (Dataset S1). We contrasted the 5% spacers that were most frequently acquired versus the spacers that were acquired only once, as described in *SI Materials and Methods*. Remarkably, both the 5′-AAG at positions −2 to +1 and the 5′-AA at positions 32–33 of the protospacers from the *E. coli* chromosome were the most significantly overrepresented motifs in the group of frequently acquired spacers compared with the group of rarely acquired spacers ($P < 1 \times 10^{-12}$, Fisher exact test after correction for multiple testing; Fig. 5 and Dataset S2). Interestingly, a G at position 33 was significantly underrepresented ($P < 1 \times 10^{-54}$; Fig. 5 and Dataset S2). Notably, analysis of spacers from the plasmid did not reveal significant overrepresentation of the AAM in the frequently acquired spacer group (Fig. S5). This result corroborates a previous high-throughput spacer analysis from an *E. coli* plasmid (20). We speculate that this motif was not identified in spacers derived from the plasmid due to the low number of protospacers having a 5′-AAG in the plasmid. The total number of 5′-AAG on both strands of the pCas1+2 plasmid is 138, and 102 in the plasmid analyzed in ref. 20, whereas such protospacers occur in the chromosome 125,223 times. The analyses are further disturbed by the fact that other dinucleotides at this position may have similar effects as 5′-AA, as shown for one of the tested protospacers (Fig. S6*A*). It should also be emphasized that although the effect of AAM was shown for two different spacers, this effect in another tested spacer was subtle (Fig. S6*B*). This result suggests that the magnitude of this effect is probably context dependent and that there are still other unknown factors that determine spacer-acquisition efficiency. Nevertheless, the analysis of the current dataset, showing statistically significant evidence for this motif, along with the experimental demonstration that this motif enhances acquisition efficiency of two spacers, indicates that this motif is important for spacer selection by Cas1 and Cas2.

Prospects. To summarize, we established two unique assays to monitor acquisition efficiency of any particular spacer. Using these assays, we provided direct evidence that the 5′-AAG motif determines the efficiency of acquisition of a spacer into the CRISPR array. Remarkably, we identified that the last dinucleotide of the tested spacer, the AAM, also affects acquisition

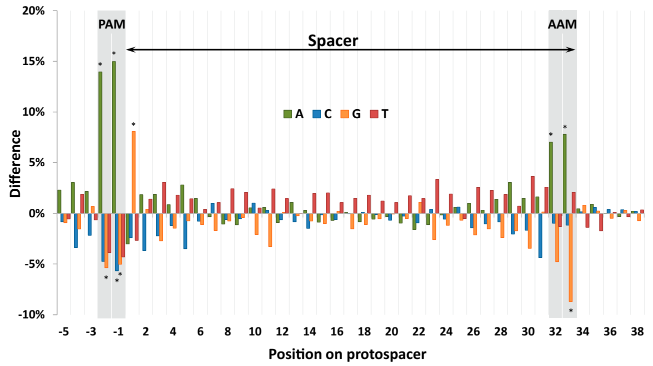

Fig. 5. Comparison of the occurrence of each base in spacers acquired from the *E. coli* chromosome. Each bar represents the difference between the occurrence of a specific base at the indicated position of the spacer in the frequently acquired spacers and its occurrence in the rarely acquired spacers. Asterisks mark the most significant differences (P value $< 1 \times 10^{-12}$). The graph was generated based on high-throughput sequencing data presented in Datasets S1 and S2.

efficiency. This study thus defines a unique element important for the acquisition step of the fascinating CRISPR/Cas defense system other than the PAM. It must be emphasized, however, that the AAM does not affect acquisition efficiencies of all tested spacers (e.g., Fig. S6*B*) and that there are probably additional factors determining its effect.

The fact that the motifs were identified exactly at or near the cleavage sites suggests that they may play a role in determining the orientation of spacer insertion. The significant underrepresentation of G at the end of highly acquired spacers suggests that this G may serve as a negative signal for insertion of the spacer in the leader-proximal end. In contrast, the overrepresentation of G as the initiating base of the spacer may serve as a positive signal for insertion in that orientation. Further biochemical studies are needed to address this issue.

Finally, bioinformatics studies revealed that in some cases, the PAM is located upstream of the protospacer (e.g., CRISPR/Cas type I; ref. 23), in other cases it is located downstream of the protospacer (e.g., CRISPR/Cas type II; ref. 23), and it can be completely absent (e.g., CRISPR/Cas type III; refs. 24 and 25). In light of our findings of an additional, unique sequence that determines acquisition efficiency, we propose that there are more such unidentified motifs located upstream and downstream of the protospacers and we believe that our experimental approach could be applied to identify these motifs.

Materials and Methods

Reagents, Strains, Plasmids, and Oligonucleotides. LB medium (10 g/L tryptone, 5 g/L yeast extract, and 5 g/L NaCl) was from Acumedia, agar was from Difco, and antibiotics, isopropyl-β-D-thiogalactopyranoside (IPTG), and L-arabinose were from Sigma-Aldrich. Taq 2× Master Mix for PCR was from Lamda Biotech. Restriction enzymes were from New England Biolabs. Rapid ligation kit was from Roche. The bacterial strains, plasmids, and oligonucleotides used in this study are listed in Tables S1 and S2. Construction of plasmids is detailed in *SI Materials and Methods*.

Acquisition Efficiency Assays. Overnight cultures of *E. coli* BL21-AI harboring pCas1+2 plasmid or derivatives thereof were diluted 1:600 and aerated at 37 °C in LB medium containing 50 μg/mL streptomycin with 0.2% (wt/vol) L-arabinose + 0.1 mM IPTG for 16 h. A sample of the culture was used as the template in a PCR amplifying CRISPR array 1 by using one primer homologous to the CRISPR array (MG7F) and another primer as indicated. The reaction contained 5 μL of 2× PCR master mix, 0.25 μL of 10 μM primer MG7F, 2.5 μL of 1 μM of the indicated primer, 0.5 μL of bacterial culture, and 1.75 μL of double-distilled water. The PCR started with 3 min at 95 °C followed by 30 cycles of 20 s at 95 °C, 20 s at 60 °C, and 20 s at 72 °C. The final extension step at 72 °C was carried out for 5 min. Samples of 4 μL each were loaded on a 1.5% (wt/vol) agarose gel and electrophoresed for 20 min at 120 V. For DNA sequencing of the amplified products, the amplified band was excised from the gel and ligated into a T-A compatible vector (pGEM-T; Promega). Colonies transformed with the ligation mixture and having the desired insert were picked, and the insert was PCR amplified and DNA sequenced.

ACKNOWLEDGMENTS. We thank Camille Vainstein for professional language editing. This work was supported by Israel Science Foundation Grant 611/10 and Marie Curie International Reintegration Grant PIRG-GA-2009-256340.

1. Marraffini LA, Sontheimer EJ (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322(5909):1843–1845.
2. Barrangou R, et al. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315(5819):1709–1712.
3. Brouns SJ, et al. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321(5891):960–964.
4. Hale CR, et al. (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139(5):945–956.
5. Goren M, Yosef I, Edgar R, Qimron U (2012) The bacterial CRISPR/Cas system as analog of the mammalian adaptive immune system. *RNA Biol* 9(5):549–554.
6. Abedon ST (2012) Bacterial 'immunity' against bacteriophages. *Bacteriophage* 2(1): 50–54.
7. Bhaya D, Davison M, Barrangou R (2011) CRISPR-Cas systems in bacteria and archaea: Versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet* 45: 273–297.
8. Wiedenheft B, Sternberg SH, Doudna JA (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482(7385):331–338.
9. Makarova KS, et al. (2011) Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 9(6):467–477.
10. Pougach K, et al. (2010) Transcription, processing and function of CRISPR cassettes in Escherichia coli. *Mol Microbiol* 77(6):1367–1379.
11. Yosef I, Goren MG, Qimron U (2012) Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res* 40(12):5569–5576.
12. Deveau H, Garneau JE, Moineau S (2010) CRISPR/Cas system and its role in phage-bacteria interactions. *Annu Rev Microbiol* 64:475–493.
13. Sorek R, Kunin V, Hugenholtz P (2008) CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* 6(3): 181–186.
14. Marraffini LA, Sontheimer EJ (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet* 11(3):181–190.
15. Swarts DC, Mosterd C, van Passel MW, Brouns SJ (2012) CRISPR interference directs strand specific spacer acquisition. *PLoS ONE* 7(4):e35888.
16. Datsenko KA, et al. (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat Commun* 3:945.
17. Goren MG, Yosef I, Auster O, Qimron U (2012) Experimental definition of a clustered regularly interspaced short palindromic duplicon in Escherichia coli. *J Mol Biol* 423(1): 14–16.
18. Deveau H, et al. (2008) Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J Bacteriol* 190(4):1390–1400.
19. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151(Pt 8):2551–2561.
20. Savitskaya E, Semenova E, Dedkov V, Metlitskaya A, Severinov K (2013) High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in E. coli. *RNA Biol* 10(5):716–725.
21. Paez-Espino D, et al. (2013) Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nat Commun* 4:1430.
22. Babu M, et al. (2011) A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. *Mol Microbiol* 79(2):484–502.
23. Mojica FJ, Díez-Villaseñor C, García-Martínez J, Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155(Pt 3):733–740.
24. Marraffini LA, Sontheimer EJ (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* 463(7280):568–571.
25. Hale CR, et al. (2012) Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Mol Cell* 45(3):292–302.

Supporting Information

Yosef et al. 10.1073/pnas.1300108110

SI Materials and Methods

Plasmid Construction. pCas1+2 plasmid encoding Cas1 and Cas2 (1) was used as a template for most plasmid constructions. Phosphorylated primers annealing to this plasmid in their 3′ end but having extended or modified 5′ ends as indicated in Table S1 were used to amplify linear fragments from this plasmid or other plasmids as indicated in Table S2. The linear DNAs were ligated and transformed into *Escherichia coli* NEB5α to yield the indicated plasmids. In some cases, site-directed mutations were introduced by using mutated primers and a PfuTurbo DNA polymerase (Stratagene). The original, nonmutated plasmid template was eliminated by using the methylation-dependent restriction enzyme DpnI according to the manufacturer instructions. The obtained amplified products were transformed into NEB5α to yield the indicated plasmids. All relevant amplified DNA segments were verified by DNA sequencing to have no mutations other than the desired mutations.

High-Throughput Sequencing and Analysis of Spacers. To generate a large number of spacers for high-throughput sequencing, overnight cultures of *E. coli* BL21-AI harboring pCas1+2 were induced to acquire spacers as described in *Materials and Methods*. These cultures were used as a template in a PCR by using primers OA1F+IY130R (Table S1) followed by amplification of the obtained products using primers IY230R7 and RE10RD (Table S1). The final PCR products were further manipulated by using Illumina's kit (catalog no. 15025064, barcode 3) according to the manufacturer's instructions. Sequencing was carried out by using Illumina's HiSEq. **2500**, in a rapid mode, single-read run.

Sequences containing the IY230R7 primer and a 33-nt-long spacer along with its flanking repeats were extracted from the Illumina HiSeq reads. Spacers from these reads were aligned to the *E. coli* BL21-AI genome (National Center for Biotechnology Information accession no. NC_012947.1), and to the pCas1+2 plasmid by using Novocraft's Novoalignment program (Novocraft Technologies). Of the 2.67 million spacers acquired, 2.19 millions were uniquely and fully aligned to 90,387 different locations on either the chromosome or the plasmid. The distinct spacers from the chromosome and the plasmid were separated to two groups: (*i*) "frequently acquired" group, consisting of 5% most frequently acquired spacers (in *E. coli* 4,290 spacers that were acquired more than 40 times, in the plasmid 268 spacers acquired 839 times or more); (*ii*) "rarely acquired" group, consisting of spacers that were acquired only once (in *E. coli* 22,294 spacers and 275 spacers in the plasmid). The protospacers, considered as the spacer along with 50 flanking nucleotides from each side, of the two groups were contrasted as follows: The number of each nucleotide in each position of the protospacers were compared between the two sets by using Fisher exact test. The *P* values were corrected for multiple testing by false detection rate control (2).

1. Yosef I, Goren MG, Qimron U (2012) Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res* 40(12):5569–5576.

2. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J Roy Stat Soc B Met* 57(1):289–300.

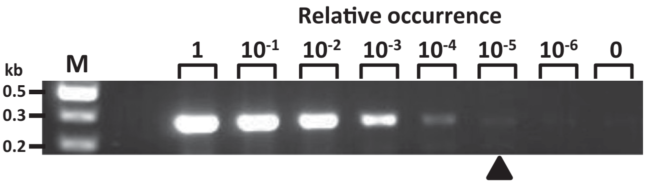

Fig. S1. Detection limit of the acquisition efficiency assay. Bacteria having a unique spacer in their CRISPR array were mixed with bacteria lacking this spacer. Numbers above each lane indicate the relative occurrence of bacteria having the unique spacer among bacteria lacking it (0 indicates none, and 1 indicates all). PCR using a primer annealing to this spacer, and a primer annealing to the CRISPR array was carried out on the different cell suspensions. The lowest ratio of cells in which an expanded band is still detected by PCR is indicated with an arrowhead. M, marker. Gel image represents three experiments yielding similar results.

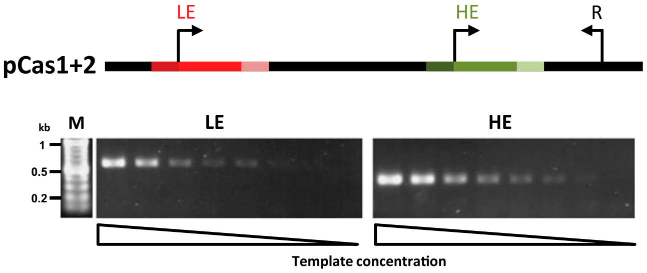

Fig. S2. Quantification of priming efficiency of low efficiency (LE) and high efficiency (HE) primers. A DNA template (pCas1+2) encoding both primer sequences on the same strand was used in a PCR to measure priming efficiency of each primer, as depicted. Equal amounts of template and reverse primer were used for both LE and HE PCRs. The highest template concentration used in the first lanes was 50 ng, and this concentration was serially diluted twofold with the last lane having no template. Gel image represents three experiments yielding similar results.

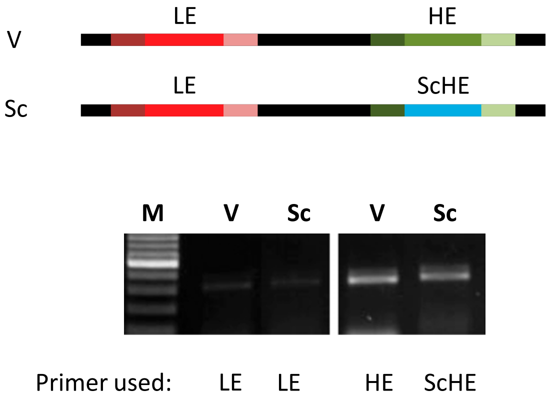

Fig. S3. Motifs in the primer sequence do not affect the efficiency of adaptation. The HE sequence was scrambled in a modified pCas1+2 plasmid, as depicted here and described in Tables S1 and S2. Acquisition efficiency assays were carried out on cultures harboring the indicated plasmid. Sc, exchange of the HE sequence with a scrambled sequence; ScHE, scrambled HE primer; V, unmodified pCas1+2.

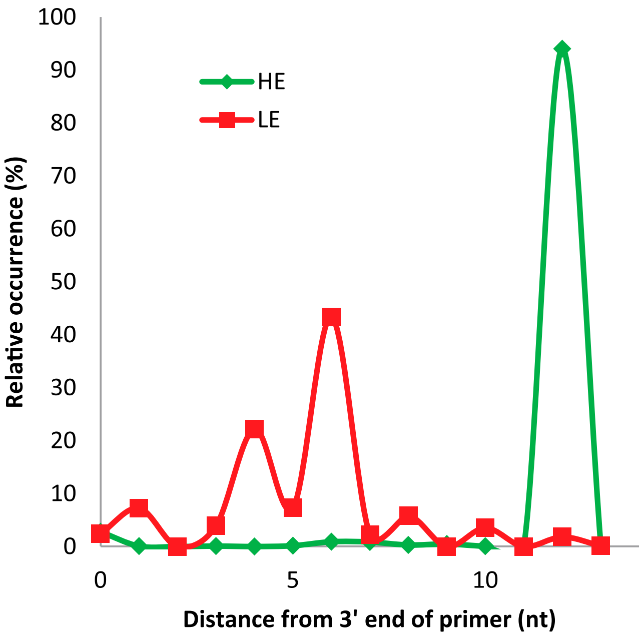

Fig. S4. Distribution of spacers matching the HE and LE primers. Sequence analysis of 22,493 spacers encoding the matching HE or LE primers was carried out by following acquisition assays described in *Materials and Methods*. The relative occurrence of acquired spacers out of the total spacers ending at the indicated distance from the 3′ end of the respective primer is shown for each position. The total number of spacers acquired from the HE-matching primer was 21,552, whereas those from the LE-matching primer was 941.

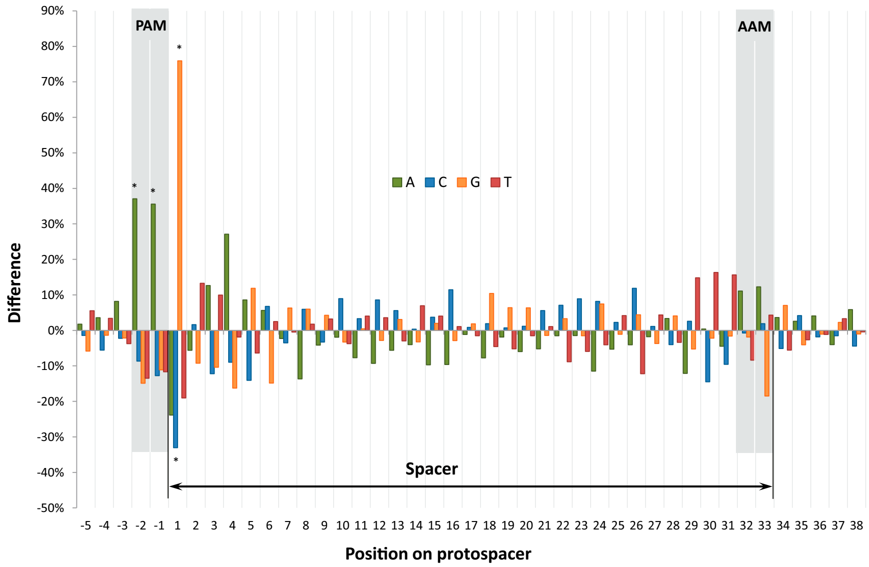

Fig. S5. Comparison of the occurrence of each base in spacers acquired from the pCas1+2 plasmid. Each bar represents the difference between the occurrence of a specific base at the indicated position of the spacer in the frequently acquired spacers and its occurrence in the rarely acquired spacers. Asterisks mark the most significant differences ($P < 1 \times 10^{-12}$). The graph was generated based on high-throughput sequencing data presented in Datasets S1 and S2.

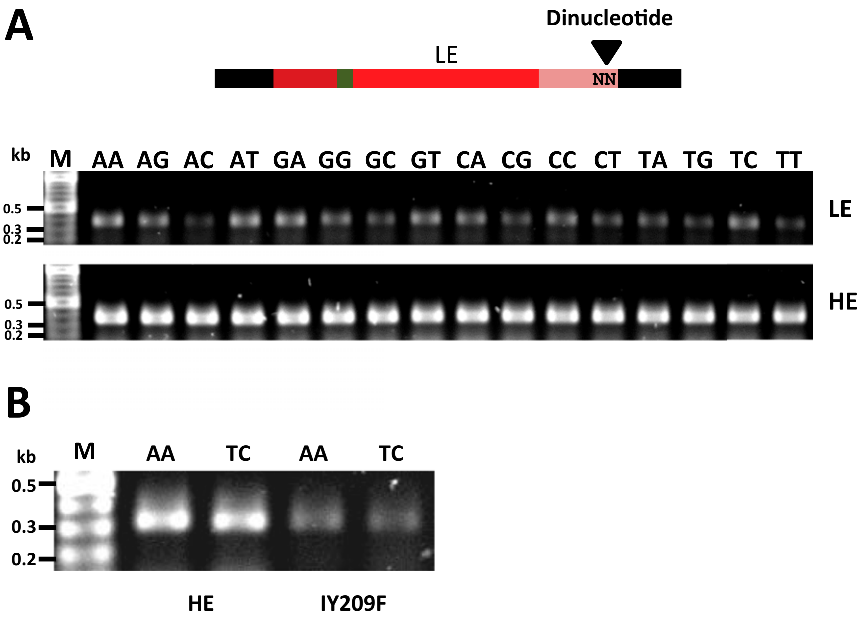

Fig. S6. (*A*) Relative contribution to acquisition efficiency conferred by each dinucleotide at the last positions of the LE-matching spacer. All possible dinucleotide combinations were inserted 30 nt downstream of a 5′-AAG that was inserted immediately upstream of the sequence matching the LE primer on pCas1+2 plasmid, as depicted. Acquisition efficiency assays were carried out on cultures harboring plasmids encoding the indicated dinucleotides by using the primer indicated to the right of each image, as described in Fig. 1*A*. (*B*) Effect of AAM on acquisition efficiency of an additional spacer. The 5′-AA dinucleotide was inserted 30 bp downstream 5′-AAG on the pCas1+2 plasmid, yielding plasmid pIY5006 (Table S2). Acquisition efficiency assays were carried out as described in Fig. 1*A*. The HE primer was used as a loading control, whereas primer IY209F was used to determine changes in acquisition efficiency due to the dinucleotide modification from 5′-TC to 5′-AA. Images represents two experiments yielding similar results.

Table S1. Bacteria, plasmids, and oligonucleotides used in this study

Bacteria/plasmids/oligonucleotides	Description/sequence	Source
Bacterial strains		
BL21-AI	F^- *ompT hsdSB*(rB–, mB–) *gal dcm araB*::*T7RNAP-tetA*, tet^r	Invitrogen
NEB5α	F^- ϕ*80lacZ*Δ*M15*Δ*(lacZYA-argF)* U169 *deoR recA1 endA1 hsdR17* (r_k^-, m_k^+) *gal* $^-$ *phoA supE44* λ^- *thi* $^-$*1 gyrA96 relA1*	New England Biolabs
Plasmids		
pCas1+2	pCDF-1b (Novagen) cloned with *cas1,2* under T7 promoter, str^r	1
pCas1^{D221A}+2	pCas1+2 encoding a nonfunctional Cas1 due to a D221A substitution	1
pGEM T-vector	PCR cloning vector	Promega
pWUR-HH-cas12	pCas1+2 having the LE sequence exchanged with the HE sequence	This study
Oligonucleotides 5′→3′		
DS5F	CAGGCATTTGAGAACCACACGGTCACACTGC	—
DS5R	GCAGTGTGACCGTGTGGTTCTCAAATGCCTG	—
DS6F	CAGGCATTTGAGAGGCACACGGTCACACTGC	—
DS6R	GCAGTGTGACCGTGTGCCTCTCAAATGCCTG	—
DS7F	CAGGCATTTGAGCAGCACACGGTCACACTGC	—
DS7R	GCAGTGTGACCGTGTGCTGCTCAAATGCCTG	—
DS10F	ACTCCCCGTTCAGCCCGACT	—
IY130R	CGTTTTTGGAATTTACAGCGAGG	—
IY137F (P1)	CCCTCGGCTTGAACGAATTG	—
IY138F (P2)	GGACTCGTCTACTAGCGCAG	—
IY139F (P3)	TGCTGCCACCGCTGAGCAAT	—
IY140F (P4)	GGGCCTCTAAACGGGTCTTG	—
IY141F (P5)	GCTGAAACCTCAGGCATTTG	—
IY142F (P6) (HE)	CACACGGTCACACTGCTTCC	—
IY142Fc (ScHE)	CCACACCCGCTCTATTGACG	—
IY143F (P7)	CAATAAACCGGTAAACCAGC	—
IY144F (P8)	AAGCGGCTATTTAACGACCC	—
IY145F (P9)	GGTCATCGTGGCCGGATCTT	—
IY146F (P10)	TGCTGCGAAATTTGAACGCC	—
IY147F (P11) (LE)	CCGCACTCGAGTCTGGTAAA	—
IY175R	AGTGCTTACGTTGTCCCGCA	—
IY147R	TTTACCAGACTCGAGTGCGG	—
IY142R	GGAAGCAGTGTGACCGTGTG	—
IY180F	GAAACCGCTGCTACCGGTAAACCAGCAATAGA	—
IY180R	CCGCTCAAACAGCTGAGGTTTCAGCAAAAAACCC	—
IY181F	GGTAGTCAATAAGCGAAATTTGAACGCCAGCA	—
IY181R	CTTCTCAAATGCGTAAAAAAGACACCAACCTTAAACC	—
IY187Fa	CCACACCCGCTCTATTGACGGGTAGTCAATAAACCGGTAA	—
IY187R	CTTCTCAAATGCCTGAGGTT	—
IY192F	GAAAGTCAATAAACCGGTAAACC	—
IY192R	CCGCTCAAATGCCTGAGGTTTCA	—
IY193F	GAAACCCAATAAACCGGTAAACCAGC	—
IY194F	GAAACCGCTTAAACCGGTAAACCAGCAAT	—
IY195Fd	GGTAGTCAAGAAACCGGTAAACCAGC	—
IY195Fe	GGTAGTCAATCAACCGGTAAACCAGC	—
IY195Ff	GGTAGTCAATATACCGGTAAACCAGC	—
IY196Fa	GTCTGGTAAAGAAACCGCTGCTGCGAAATTTGAACG	—
IY196Fc	GTCTGGTAAAGAAACCGCTGAAGCGAAATTTGAACG	—
IY196Ra	TCGAGTGCGGCCGCTCAAACAGGTAAAAAAGACACC	—
IY196Rb	TCGAGTGCGGCTTCTCAAACAGGTAAAAAAGACACC	—
IY201R	TCGAGTGCGGCTTCTCCTTCAGGTAAAAAAGACACC	—
IY201Fb	GTCTGGTAAAGAAACTGCTGAAGCGAAATTTGAACG	—
IY201Fc	GTCTGGTAAAGAAAAAGCTGCTGCGAAATTTGAACG	—
IY204F	ATGTCTAACAATTCGTAAAAGCCGAGGGGCCGCA	—
IY204R	TGCGGCCCCTCGGCTTTTACGAATTGTTAGACAT	—
IY209F	GTAGTCGGCAAATAATGTCT	—
IY214F	TGGTAAAGAAACCGCTGAGGCGAAATTTGAACGCCA	—
IY214R	TGGCGTTCAAATTTCGCCTCAGCGGTTTCTTTACCA	—
IY215F	TGGTAAAGAAACCGCTGACGCGAAATTTGAACGCCA	—

Table S1. Cont.

Bacteria/plasmids/oligonucleotides	Description/sequence	Source
IY215R	TGGCGTTCAAATTTCGCGTCAGCGGTTTCTTTACCA	—
IY216F	TGGTAAAGAAACCGCTGATGCGAAATTTGAACGCCA	—
IY216R	TGGCGTTCAAATTTCGCATCAGCGGTTTCTTTACCA	—
IY217F	TGGTAAAGAAACCGCTGGAGCGAAATTTGAACGCCA	—
IY217R	TGGCGTTCAAATTTCGCTCCAGCGGTTTCTTTACCA	—
IY218F	TGGTAAAGAAACCGCTGGGGCGAAATTTGAACGCCA	—
IY218R	TGGCGTTCAAATTTCGCCCCAGCGGTTTCTTTACCA	—
IY219F	TGGTAAAGAAACCGCTGGCGCGAAATTTGAACGCCA	—
IY219R	TGGCGTTCAAATTTCGCGCCAGCGGTTTCTTTACCA	—
IY220F	TGGTAAAGAAACCGCTGGTGCGAAATTTGAACGCCA	—
IY220R	TGGCGTTCAAATTTCGCACCAGCGGTTTCTTTACCA	—
IY221F	TGGTAAAGAAACCGCTGCAGCGAAATTTGAACGCCA	—
IY221R	TGGCGTTCAAATTTCGCTGCAGCGGTTTCTTTACCA	—
IY222F	TGGTAAAGAAACCGCTGCGGCGAAATTTGAACGCCA	—
IY222R	TGGCGTTCAAATTTCGCCGCAGCGGTTTCTTTACCA	—
IY223F	TGGTAAAGAAACCGCTGCCGCGAAATTTGAACGCCA	—
IY223R	TGGCGTTCAAATTTCGCGGCAGCGGTTTCTTTACCA	—
IY224F	TGGTAAAGAAACCGCTGTAGCGAAATTTGAACGCCA	—
IY224R	TGGCGTTCAAATTTCGCTACAGCGGTTTCTTTACCA	—
IY225F	TGGTAAAGAAACCGCTGTGGCGAAATTTGAACGCCA	—
IY225R	TGGCGTTCAAATTTCGCCACAGCGGTTTCTTTACCA	—
IY226F	TGGTAAAGAAACCGCTGTCGCGAAATTTGAACGCCA	—
IY226R	TGGCGTTCAAATTTCGCGACAGCGGTTTCTTTACCA	—
IY227F	TGGTAAAGAAACCGCTGTTGCGAAATTTGAACGCCA	—
IY227R	TGGCGTTCAAATTTCGCAACAGCGGTTTCTTTACCA	—
IY230R7	NNNNCCGTGAGCGATGATATTTGTGCT	—
MG7F	ATTTTGCGTTTCGTTCAGGT	—
MG56Fa	CACACGGTCACACTGCTTCCGAAACCGCTGCTGCGAAATT	—
MG56Fb	CCGCACTCGAGTCTGGTAAAGGTAGTCAATAAACCGGTAA	—
MG56R	CCGCTCAAACAGGTAAAAAAG	—
OA1F	ACATACTAGTTAATCAATGGATTAAGTACT	—
RE10RD	NNNNTGGATGTGTTGTTTGTGTG	—

[1] Yosef I, Goren MG, Qimron U (2012) Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res* 40(12):5569–5576.

Table S2. Primers and templates used for plasmid construction

Plasmid	Used in Fig.	Name in legend	Primers for PCR	DNA template
pWURHLcas12	1*C*	Ex	MG56Fb+IY187R	pWUR-HH-cas12
pIYULHDL12	2*A*	U+D	IY180F+IY142R	pIYULH12
pIYULH12	2*A* and 3*A*	U, −12	IY142F+IY180R	pCas1+2
pIYDLH12	2*A* and 3*B*	D, +12	IY180F+IY142R	pCas1+2
pIYUHL12	2*B*	U	IY147F+IY181R	pCas1+2
pIYDHL12	2*B*	D	IY181F+IY147R	pCas1+2
pIYUHLDH12	2*B*	U+D	IY147F+IY181R	pIYDHL12
pIYULH3	3*A*	−3	IY142F+IY192R	pCas1+2
pIYDLH3	3*B*	+3	IY192F+IY142R	pCas1+2
pIYDLH6	3*B*	+6	IY193F+IY142R	pCas1+2
pIYDLH9	3*B*	+9	IY194F+IY142R	pCas1+2
pCAS12DS7	3*C*	−3C	DS7F+DS7R	pCas1+2
pCAS12DS6	3*C*	−2G	DS6F+DS6R	pCas1+2
pCAS12DS5	3*C*	−1C	DS5F+DS5R	pCas1+2
pDHT10G	3*D*	10G	IY195Fd+IY142R	pCas1+2
pDHA11C	3*D*	11C	IY195Fe+IY142R	pCas1+2
pDHA12T	3*D*	12T	IY195Ff+IY142R	pCas1+2
pIY5000	3*E* and 5	−3-LE; CT	IY196Fa+IY196Rb	pCas1+2
pIY5010	3*E*	+11-LE	IY196Fc+IY196Ra	pCas1+2
pIY5003	3*E* and 5	−3/+11-LE; AA	IY196Fc+IY196Rb	pCas1+2
pIY5011	4*C*	Upper AAM	IY201Fb+IY201R	pCas1+2
pIY5012	4*C*	Lower AAM	IY201Fc+IY201R	pCas1+2
pIY5014	5	AG	IY214F +IY214R	pIY5003
pIY5015	5	AC	IY215F +IY215R	pIY5003
pIY5016	5	AT	IY216F +IY216R	pIY5003
pIY5017	5	GA	IY217F +IY217R	pIY5003
pIY5018	5	GG	IY218F +IY218R	pIY5003
pIY5019	5	GC	IY219F +IY219R	pIY5003
pIY5020	5	GT	IY220F +IY220R	pIY5003
pIY5021	5	CA	IY221F +IY221R	pIY5003
pIY5022	5	CG	IY222 +IY222R	pIY5003
pIY5023	5	CC	IY223F +IY223R	pIY5003
pIY5024	5	TA	IY224F +IY224R	pIY5003
pIY5025	5	TG	IY225F +IY225R	pIY5003
pIY5026	5	TC	IY226F +IY226R	pIY5003
pIY5027	5	TT	IY227F +IY227R	pIY5003
pSC142Fc	S3	Sc	IY187Fa+IY187R	pCas1+2
pIY5006	S6*B*	pIY5006	IY204F+IY204R	pCas1+2

Other Supporting Information Files

Datasets S1 and S2 (XLSX)