CRISPR–Cas (clustered, regularly interspaced short palindromic repeats coupled with CRISPR-associated proteins) is a bacterial immunity system that protects against invading phages or plasmids\(^9\)–\(^6\). It comprises multiple \(cas\) genes, as well as an array of short sequences (‘spacers’) that are mostly derived from exogenous DNA and are interleaved by short DNA repeats. The CRISPR–Cas mode of action is divided into three main stages: adaptation (or ‘acquisition’), expression and interference. In the adaptation stage, a new spacer is acquired from the foreign DNA and integrated into the CRISPR array. During the interference stage, the effector complex identifies foreign nucleic acid via base pairing with the crRNA and targets it for degradation.

Numerous recent studies have characterized the molecular mechanisms governing the expression and interference stages of CRISPR activity, but the molecular details of the primary adaptation stage are still elusive. It was shown that the Cas1 and Cas2 proteins are necessary for primary spacer acquisition\(^7\), and that they form a single active complex\(^8\). Several systems to study spacer acquisition in the model bacterium \(E.\ coli\) have been established\(^7\)–\(^13\). Some of these systems only express \(cas\) genes on its genome, the effectors contributing DNA molecule is not targeted for degradation\(^7\)–\(^8\),\(^11\)–\(^13\). Strikingly, despite the lack of selection against spacer acquisition from the self chromosome, the vast majority of spacers acquired in such interference-free systems are derived from plasmid DNA\(^7\),\(^8\),\(^11\), suggesting intrinsic preference for the Cas1–Cas2 complex to acquire spacers from the exogenous DNA. The mechanism by which the Cas1–Cas2 complex preferentially recognizes the foreign DNA as a source for acquisition of new spacers, while avoiding taking spacers from the self chromosome, remains a major unresolved question.

CRISPR–Cas is an adaptive defence system in bacteria and archaea that provides acquired immunity against phages and plasmids\(^1\)–\(^6\). It comprises multiple \(cas\) genes, as well as an array of short sequences (‘spacers’) that are mostly derived from exogenous DNA and are interleaved by short DNA repeats. The CRISPR–Cas mode of action is divided into three main stages: adaptation (or ‘acquisition’), expression and interference. In the adaptation stage, a new spacer is acquired from the foreign DNA and integrated into the CRISPR array. In the expression stage, the repeat-spacer array is transcribed and further processed into short CRISPR RNAs (crRNAs). These mature crRNAs, in turn, bind to Cas proteins and form the effector protein–RNA complex. During the interference stage, the effector complex identifies foreign nucleic acid via base pairing with the crRNA and targets it for degradation.

Numerous recent studies have characterized the molecular mechanisms governing the expression and interference stages of CRISPR activity, but the molecular details of the primary adaptation stage are still elusive. It was shown that the Cas1 and Cas2 proteins are necessary for primary spacer acquisition\(^7\), and that they form a single active complex\(^8\). Several systems to study spacer acquisition in the model bacterium \(E.\ coli\) have been established\(^7\)–\(^13\). Some of these systems only express \(cas\) genes on its genome, the effector contributing DNA molecule is not targeted for degradation\(^7\)–\(^8\),\(^11\)–\(^13\). Strikingly, despite the lack of selection against spacer acquisition from the self chromosome, the vast majority of spacers acquired in such interference-free systems are derived from plasmid DNA\(^7\),\(^8\),\(^11\), suggesting intrinsic preference for the Cas1–Cas2 complex to acquire spacers from the exogenous DNA. The mechanism by which the Cas1–Cas2 complex preferentially recognizes the foreign DNA as a source for acquisition of new spacers, while avoiding taking spacers from the self chromosome, remains a major unresolved question.

Preference for exogenous DNA

We set out to understand the mechanism governing the self/non-self discrimination of the DNA source for spacer acquisition during the adaptation stage. For this, we used a previously described experimental system that monitors spacer acquisition in vivo in the \(E.\ coli\) type I-E CRISPR system\(^9\),\(^10\),\(^14\),\(^15\). In this system, \(cas\) genes are carried on a plasmid (pCas1+2) and their expression is regulated by an arabinose-inducible T7 RNA polymerase (Extended Data Fig. 1). We have previously shown that expression of Cas1–Cas2 in this system leads to spacer acquisition: that is, expansion of the chromosomal encoded CRISPR I array in \(E.\ coli\) BL21-AI\(^1\). Since this strain of \(E.\ coli\) harbours a CRISPR array but lacks any \(cas\) genes on its genome, this system is interference-free, and thus does not allow ‘primed’ CRISPR adaptation\(^9\),\(^10\),\(^14\),\(^15\).

After overnight growth of an \(E.\ coli\) BL21-AI culture carrying pCas1+2, we amplified the leader-proximal end of the CRISPR I array using a forward primer on the leader and a reverse primer matching spacer 2 of the native array. The amplification product, containing both native and expanded arrays, was sequenced using low-coverage Illumina technology (MiSeq) to accurately quantify the fraction of arrays that acquired a new spacer in each experiment. In parallel, high-coverage Illumina sequencing (HiSeq) was performed on gel-separated expanded arrays, to characterize the source, location and frequency of newly acquired spacers in high resolution (Extended Data Fig. 1). Overall, over 38 million newly acquired spacers were sequenced in this study (Extended Data Tables 1–3).

In cultures overexpressing Cas1–Cas2 for 16 h, 36.92% (±1.2) of the sequenced arrays contained a new spacer. Conversely, in cultures where Cas1–Cas2 were not induced, 2.61% (±0.5) of the arrays contained a new spacer after 16 h of incubation, indicating that the leakage of Cas1–Cas2 transcription (as measured by RNA sequencing; Supplementary Table 1) still resulted in spacer acquisition in a...
significant fraction of the cells (Extended Data Table 1a). Examining the origin of new spacers showed strong preference for spacer acquisition from the plasmid, with only 22.86% (±0.46) and 1.8% (±0.03) of the spacers derived from the self chromosome in the induced and non-induced cultures, respectively (Extended Data Table 1b). Considering the size of the plasmid (4.7 kilobases (kb)) and its estimated copy number of 20–40, this represents 100- to 1,000-fold enrichment for acquisition of spacers from the plasmid, compared with what is expected by the DNA content in the cell. These results also show that lower expression of Cas1–Cas2 leads to higher specificity for exogenous DNA. Therefore, most of the analyses henceforth are based on spacers acquired in conditions in which Cas1–Cas2 are expressed but not overexpressed.

**Replication-dependent adaptation**

Although only a small minority of spacers was derived from the *E. coli* chromosome, the extensive number of sequenced spacers allowed us to examine chromosome-scale patterns of spacer acquisition. Remarkably, strong biases in spacer acquisition were observed, defining several protospacer hotspots (Fig. 1a). As the protospacer adjacent motif (PAM) density on the chromosome scale is largely uniform (Fig. 1b and Extended Data Fig. 2), these protospacer hotspots could not be explained by excessive localization of PAM sequences in specific areas of the genome. We further investigated each of the hotspots in search of a mechanism that would explain the observed biases.

Spacer acquisition was more pronounced at areas closer to the chromosomal origin of replication (oriC), with a clear gradient of reduced protospacer density as a function of the distance from oriC (Fig. 1a). In replicating cells, the DNA next to oriC is replicated first, hence the culture inevitably contains more copies of the origin-proximal DNA16. Indeed, upon sequencing of total genomic DNA extracted from the gin-proximal DNA16. Indeed, upon sequencing of total genomic DNA extracted from the *E. coli* BL21 genome (4.7 kilobases (kb)) and its estimated copy number of 20–40, this represents 100- to 1,000-fold enrichment for acquisition of spacers from the plasmid, compared with what is expected by the DNA content in the cell. These results also show that lower expression of Cas1–Cas2 leads to higher specificity for exogenous DNA. Therefore, most of the analyses henceforth are based on spacers acquired in conditions in which Cas1–Cas2 are expressed but not overexpressed.

Figure 1 | Chromosome-scale hotspots for spacer acquisition.

**a,** Distribution of protospacers across the *E. coli* BL21-AI genome. Protospacers were deduced from aligning new spacers, acquired into the CRISPR1 array after 16 h growth with no arabinose, to the bacterial genome. Only unique protospacers are presented, to avoid possible biases stemming from PCR amplification of the CRISPR array. Pooled protospacers from two replicates are presented. **b,** Protospacer density across a circular representation of the *E. coli* genome, normalized to the DNA content of the culture. Dark brown, normalized protospacer numbers; orange, PAM density. **c,** Protospacer distribution at the Ter region. Protospacer density is shown in 1-kb windows. **d,** Protospacer density in an *E. coli* BL21-AI in which the native 23 base pair (bp)-long TerB site was engineered into the pheA locus.

The most striking protospacer hotspot was observed around the chromosomal replication terminus (Ter), in two major peaks showing approximately 7- to 20-fold higher protospacer density than the surrounding area (Fig. 1b, c). The Ter macrodomain is the area where the two replication forks coming from opposite directions on the chromosome meet, leading to chromosome decatenation17. This chromosomal macrodomain contains unidirectional fork stalling sites called Ter sites (primarily TerA and TerC), which, during replication, stall the early-arriving replication fork until the late fork arrives from the other side17. We found that the primary fork-stalling sites TerA and TerC were the exact boundaries of the spacer acquisition hotspots (Fig. 1c). Moreover, the protospacer hotspots next to Ter sites were asymmetric relative to the fork direction of progression, with strong protospacer enrichment observed upstream of each fork stalling site and a relatively low, background protospacer density downstream of the stalled fork (Fig. 1b, c). Engineering of a native Ter site into the pheA locus on the bacterial chromosome generated a new localized protospacer hotspot, strongly supporting the idea that hotspots for spacer acquisition directly correlate with replication fork stalling sites (Fig. 1d).

The correlation between spacer acquisition biases and the replication fork stalling sites may suggest that CRISPR adaptation is promoted by active replication of the protospacer-containing DNA. We conducted a series of experiments to test this hypothesis. First, we used the replication-inhibitor quinolone nalidixic acid on *E. coli* BL21-AI cells during induction of Cas1–Cas2. As a control, we applied the RNA polymerase inhibitor rifampicin, which blocks transcription in *E. coli* but allows DNA replication (this antibiotic does not interfere with transcription of Cas1–Cas2 by the T7 RNA polymerase). Application of nalidixic acid resulted in an almost complete acquisition bias can largely be expected based on the average DNA content in the culture and, accordingly, normalizing protospacer density to DNA content eliminated most of the oriC-centred protospacer gradient (Fig. 1b).
elimination of spacer acquisition (164-fold reduction), but only an approximately twofold reduction in spacer acquisition rates was observed in the rifampicin-treated cells (Fig. 2a and Extended Data Table 1c), providing support to the hypothesis that spacer acquisition depends on DNA replication.

To substantiate these observations further, we examined the acquisition rates in E. coli K-12 cells carrying the temperature-sensitive allele dnaC2 (ref. 18). In these cells, initiation of DNA replication is blocked at 39 °C but is permitted at 30 °C. These cells were transformed with a pBAD-Cas1+2 vector, in which the Cas1–Cas2 operon is directly controlled by an arabinose-inducible promoter. Since these cells encode the full set of cas genes, the casC gene was also knocked out to avoid CRISPR interference or priming. As a control, we used an isogenic K-12 strain encoding the wild-type dnaC gene. After overnight growth in the replication-permissive temperature, the two strains showed similar rates of spacer acquisition. However, when the temperature-sensitive dnaC2 cells were grown at 39 °C, acquisition was almost completely abolished, with less than 0.1% of the sequenced arrays found to be expanded (Fig. 2b and Extended Data Table 2a). These results further strengthen the hypothesis that Cas1–Cas2-mediated spacer acquisition in the E. coli type I-E CRISPR system requires active replication of the protospacer-containing DNA.

We next asked whether spacer acquisition preferences correlate with the position of the replication fork. For this, we transferred a culture of the temperature-sensitive dnaC2 cells to 39 °C for 70 min. Since in this temperature replication re-initiation is inhibited, after 70 min there are no more progressing forks in these cells. We then induced Cas1–Cas2 expression for 30 min, and transferred the culture to 30 °C, resulting in synchronized initiation of replication. At these conditions, it takes the replication forks on average about 60 min to complete a full DNA replication cycle in dnaC2 cells19. In accordance, we sequenced the newly acquired spacers at 20, 40, 60, 90 and 120 min following synchronous replication initiation. Strikingly, the fraction of spacers derived from the Ter region gradually increased with the progression of the replication cycle, reaching 31% after 60 min (compared with only 6.4% at the 20 min time point; Fig. 2c, Extended Data Fig. 3 and Extended Data Table 2b). The pattern repeated itself in the second cycle of replication (90 and 120 min; Fig. 2c). These results demonstrate temporal correlation between the predicted position of stalled replication forks and the preference to acquire spacers from that position.

Combined, the above results support a model where the Cas1–Cas2 complex has preference for acquiring spacers from the area of a stalled replication fork during DNA replication. This model is intriguing, as it largely explains the observed preference for spacer acquisition from high copy number plasmids. During DNA replication in a cell, the chromosome occupies two replication forks travelling from the oriC to the Ter, where their stalling will promote spacer acquisition. At the same replication cycle, each copy of the plasmid will occupy a

mark the first properly oriented Chi site upstream relative to the fork stalling site. a, The CRISPR region in E. coli BL21-AI. b, The CRISPR region in E. coli K-12. c, The TerC region and d, the TerA region in E. coli BL21-AI. In c, the Chi site drawn at −2,260 kb represents a cluster of three consecutive Chi sites found in the same 1 kb window.
travelling fork, which will also be stalled during the termination of plasmid replication (in a Ter-independent manner\(^2\)). Therefore, the vast majority of stalled forks in a replicating cell localize to the multiple plasmid copies, and, if spacer acquisition is promoted by fork-stalling, the probability to acquire spacers from the plasmid is much higher. The model is in line with previous observations in *Sulfolobus*, showing that spacer acquisition from an infective virus does not occur unless the viral DNA is being replicated\(^2\).

**Involvement of the DNA repair machinery**

Another hotspot for spacer acquisition was observed just upstream of the CRISPR I array in the *E. coli* BL21-AI genome (Fig. 3a). This CRISPR-associated protoscaler hotspot clearly depends on CRISPR activity, because no hotspot was observed near the *E. coli* BL21-AI CRISPR II array, which lacks a leader sequence and is hence inactive\(^7\) (Fig. 3a). Indeed, in *E. coli* K-12, where both arrays are known to be active, spacer acquisition assays showed a protoscaler peak upstream of each of the two arrays (Fig. 3b). The protoscaler peaks at the CRISPR region resembled the peaks seen at the Ter sites, in the sense that they were asymmetric with respect to the replication fork direction, implying that activity at the CRISPR array forms a replication fork stalling site. Presumably the DNA nicking that occurs after the leader during insertion of a new spacer\(^13\) stalls the replication fork, thus generating a fork-dependent hotspot for spacer acquisition. Frequent stalling of the fork at the CRISPR would mean that the fork coming from the other direction will often be stalled for a longer time at the respective Ter site, TerC, waiting for the fork coming from the CRISPR direction to arrive (Extended Data Fig. 4). This may be one of the factors explaining why the TerC site is a much more pronounced protoscaler hotspot than the TerA site (Fig. 1b, c). Another factor that can contribute to the observed TerC/TerA bias may be that the clockwise replichore in *E. coli* (oriC to TerA) is longer than the anti-clockwise one (oriC to TerC), leading the forks to naturally stall at TerC more often than at TerA.

All of the spacer acquisition hotspots described above were defined by distinct peaks of high protoscaler density, with peak widths ranging between 10 and 50 kb (Fig. 3). On one end, these peaks were bounded by a fork stalling site, but the mechanism defining the boundary at the other end of the peaks was not clear. Strikingly, when searching for sequence motifs that preferentially appear at the other end of the peaks, we found that all protoscaler peaks were immediately flanked by the octamer motif GCTGGTGG, which is the canonical sequence of the Chi site (Fig. 3a–d). Chi sites interact with the double-strand break repair helicase/nuclease complex RecBCD and regulate the repair activity\(^2\). When a double-stranded DNA (dsDNA) break occurs, RecBCD localizes to the exposed end, and then unwinds...
and degrades the DNA until reaching a Chi site. Upon recognition of the Chi site, RecBCD generally ceases to degrade the DNA, and instead yields a single-stranded DNA that is bound by RecA and invades a homologous duplex DNA, which forms a template for completion of the missing DNA. Chi sites work in an asymmetric manner, meaning that the GCTGGTGG motif will only interact with RecBCD coming from the right-end of the DNA molecule (downstream of the site), whereas the reverse complement of Chi will only interact with RecBCD complexes coming from the left-end of the DNA. RecBCD indiscriminately degrades linear DNA, including phage DNA, and it was therefore suggested that this complex is one of the lines of defence against phages. Since Chi sites occur every ~5 kb in the E. coli genome, which is about 14 times more frequent than expected by chance, these sites were suggested as markers of viral infection by phages. The proper orientations (Fig. 4b), directly linking dsDNA breaks to spacer acquisition hotspots. Fourth, co-immunoprecipitation assays suggested that Cas1 interacts with RecB and RecC (although these interactions were not verified using purified proteins), supporting a model where the Cas1–Cas2 complex is directly fed from RecBCD DNA degradation products. Finally, Cas1 was shown to efficiently bind single-stranded DNA (ssDNA), which is amply generated during RecBCD DNA processing activity.

Our results show that protospacer hotspots are defined between sites of stalled forks and Chi sites (Fig. 3). Stalled replication forks are known to be major hotspots for dsDNA breaks and, it was demonstrated that the vast majority of dsDNA breaks in bacteria occur during DNA replication. These data therefore may imply that Cas1–Cas2 acquires spacers from degradation intermediates of RecBCD activity during the processing of dsDNA breaks that frequently occur at stalled replication forks.

Several lines of evidence support this hypothesis. First, the orientation of the Chi sites at the protospacer peaks was always consistent with the dsDNA break occurring at the fork direction rather than the other side, and the first properly oriented Chi site upstream of the stalled fork was always the site of peak boundary (Fig. 3a–d). Second, even outside the strong protospacer hotspots, there was a significant asymmetry in protospacer density upstream and downstream of Chi sites (Fig. 4a). The effect of this asymmetry was seen up to a distance of about 5–10 kb from the Chi site, consistent with an average distance of ~5 kb between Chi sites in the E. coli genome. Third, inducing a single, site-specific dsDNA break in the chromosome using the homing endonuclease I-SceI resulted in a clear protospacer hotspot that peaked at the site of the dsDNA break and was confined by Chi sites in the proper orientations (Fig. 4b), directly linking dsDNA breaks to spacer acquisition hotspots. Fourth, co-immunoprecipitation assays suggested that Cas1 interacts with RecB and RecC (although these interactions were not verified using purified proteins), supporting a model where the Cas1–Cas2 complex is directly fed from RecBCD DNA degradation products. Finally, Cas1 was shown to efficiently bind single-stranded DNA (ssDNA), which is amply generated during RecBCD DNA processing activity.

To test whether spacer acquisition indeed depends on the activity of the RecBCD complex, we used E. coli strains in which recB, recC or recD were deleted. Deep-sequencing-based quantification of spacer acquisition rates in these mutants showed reduced acquisition in all of these deletion strains (Fig. 4e and Extended Data Table 3a). Moreover, analysis of chromosomal protospacers in these mutants showed loss of spacer acquisition asymmetry near Chi sites (Fig. 4c), resulting in broader protospacer hotspots on the self chromosome (Fig. 4d). In accordance, the fraction of spacers derived from the self chromosome was ~10-fold higher in the recB, recC and recD deletion strains compared with the wild-type strain (Fig. 4f and Extended Data Table 3a). These results show that CRISPR adaptation is partly dependent on the activity of the RecBCD dsDNA break repair complex, and that this activity is responsible for some of the self/non-self discrimination properties of the CRISPR adaptation process. Consistent with these results, expression of a RecBCD inhibitor protein, the product of gene 5.9 of the T7 bacteriophage, showed reduced spacer acquisition compared with exogenous expression of a control protein (Extended Data Fig. 5).

It is noteworthy that in recB and recC deletions, the RecBCD complex is entirely non-functional, whereas the recD deletion produces a complex, RecBC, that is fully functional for DNA unwinding but entirely lacks nuclease activity. Our observation that the recD deletion mutant has poor spacer acquisition activity suggests that the nuclease activity of the RecBCD enzyme is important for spacer acquisition and implies that the degradation products generated by RecBCD during DNA processing between a dsDNA break and a Chi site may be the source of new spacers.

The involvement of Chi sites, as points where spacer acquisition activity is terminated, provides another axis for the avoidance of self
DNA in CRISPR adaptation. Since the pCas plasmid is completely devoid of Chi sites, its DNA will be fully degraded by RecBCD following any dsDNA break, providing plenty of potential substrate for Cas1–Cas2. In contrast, the high density of Chi sites on the bacterial chromosome serves for the relative avoidance of Cas1–Cas2 to acquire spacers from the chromosome, because RecBCD will only degrade the chromosomal DNA until reaching the nearest Chi site (Fig. 5a, b). Indeed, the ~10 fold higher acquisition frequency from the self chromosome seen in the recB, recC and recD deletion strains conforms with the natural 14-fold enrichment of Chi sites on the chromosome. To examine further whether Chi sites limit spacer acquisition, we performed spacer acquisition experiments with a plasmid that was engineered to contain a cluster of four consecutive Chi sites. As expected, an increased preference for chromosomal DNA in spacer acquisition was measured for the Chi-containing plasmid (Fig. 4g, Extended Data Table 3b and Extended Data Fig. 6).

In conclusion, these results converge to a single, unifying model that explains the preference of the CRISPR adaptation machinery to acquire spacers from foreign DNA, as well as the observed biases in spacer acquisition patterns (Fig. 5). Under this model, Cas1–Cas2 takes the DNA substrate for spacer acquisition from degradation products of RecBCD activity during the processing of dsDNA breaks. Since the vast majority of dsDNA breaks in the cell occur during DNA replication\(^2\), with stalled replication forks being major hotspots for such breaks\(^2\)\(^4\)\(^5\), high-copy-number plasmids are much more prone to spacer acquisition owing to the higher number of forks on plasmids (Fig. 5c). The high-density presence of Chi sites on the bacterial chromosome further protects it from extensive spacer acquisition (Fig. 5b). Moreover, as most phages enter the cell as a linear DNA, and since RecBCD would bind any exposed linear DNA and process it until the nearest Chi site\(^2\), unprotected phage DNA will be a target for spacer acquisition immediately upon entry to the cell, providing an additional preference for spacer acquisition specifically from phage DNA (Fig. 5d). If entry to the cell were successful, the extensive replication activity of the phage DNA would provide another anchor for spacer acquisition from phage.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 1 October 2014; accepted 9 February 2015.**

**Published online 13 April 2015.**

1. Terns, M. P. & Terns, R. M. CRISPR-based adaptive immune systems. *Curr. Opin. Microbiol.* **14**, 321–327 (2011).
2. Westra, E. R. et al. The CRISPRs, they are a-changin’: how prokaryotes generate adaptive immunity. *Annu. Rev. Genet.* **46**, 311–339 (2012).
3. Wiedenhoff, B., Sternberg, S. H. & Doudna, J. A. RNA-guided genetic silencing and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Acknowledgements** We thank N. Barkai, A. Tanay, E. Mick, S. Doron, A. Stern, T. Dagan and M. Shamir for discussion. We also thank the Skarstad group for providing the MG1655dnaC2 strain, the Michel group for providing the JJC1819 strain and D. Dar for assistance in Illumina sequencing. R.S. was supported, in part, by the Israel Science Foundation (grant 268/14) and the Israeli Ministry of Health (grant 9988-3). DNA motifs determining the efficiency of adaptation into the Escherichia coli chromosome. *Nucleic Acids Res.* **42**, 7884–7893 (2014).

**Supplementary Information** is available in the online version of the paper.

**Author Contributions** M.G., U.Q., O.A., M.M., G.A. and R.E. performed the experiments; A.L., M.G., U.Q. and M. Shamir for discussion. We also thank the Skarstad group for providing the MG1655dnaC2 strain, the Michel group for providing the JJC1819 strain and D. Dar for assistance in Illumina sequencing. R.S. was supported, in part, by the Israel Science Foundation (personal grant 1303/12 and I-CORE grant 1796), the European Research Council Starting Grant programme (grant 260432), Human Frontier Science Programme (grant RGP0011/2013), the Absis-Frenkel foundation, the Pasteur-Weizmann Council, the Minerva Foundation, and by a Deutsch-Israelische Projektkooperation grant from the Deutsche Forschungsgemeinschaft. U.Q. was supported, in part, by the European Research Council Starting Grant programme (grant 336079), the Israeli Science Foundation (grant 268/14) and the Israeli Ministry of Health (grant 9988-3). A.L. is grateful to the Azrieli Foundation for the award of an Azrieli Fellowship.

**Author Information** RNA sequencing data are available in the National Center for Biotechnology Information Sequence Read Archive database under accession numbers SRX862155–SRX862158 in study SRP053013. Raw data of spacer sequences are accessible at http://www.weizmann.ac.il/molgen/Sorek/files/CRISPR_adaptation_2015/crispr_adaptation_2015_data.html. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to U.Q. (ehuda@post.tau.ac.il) or R.S. (rotem.sorek@weizmann.ac.il).
METHODS
No statistical methods were used to predetermine sample size.

Reagents, strains and plasmids. Luria–Bertani (LB) medium (10 g l−1 tryptone, 5 g l−1 yeast extract, 5 g l−1 NaCl) and agar were from Acumedia. Antibiotics and transposon cassette (pCas1–Cas2) were from Calbiochem. Isopropyl-β-D-thiogalactoside (IPTG) was from Bio-Lab. Calcium chloride (CaCl2), sodium citrate (Na-citrate), restriction enzymes, T4 Polynuclease Kinase and Phusion high fidelity DNA polymerase were from New England Biolabs. Rapid ligation kit was from Roche. Taq DNA polymerase was from LAMDA Biotech. NucleoSpin Gel and PCR Clean-up kit was from Macherey-Nagel. The bacterial strains, plasmids and oligonucleotides used in this study are listed in Supplementary Table 2.

Plasmid construction. Plasmids were constructed using standard molecular biology techniques. DNA segments were amplified by PCR. Standard digestion of the PCR products and vector by restriction enzymes was done according to the manufacturer’s instructions.

pBAD plasmid encoding Cas1 and Cas2 was constructed by amplifying cas1 and cas2 from pWUR399 plasmid32 using oligonucleotides IY68F and IY68R (Supplementary Table 2). The amplified DNA and pBAD18 (ref. 30) vector were both digested with SacI and Sall and ligated to yield pBAD-Cas1+2. The DNA insert was sequenced to exclude mutations introduced during cloning. pWUR plasmid encoding Cas1 and Cas2 under lac promoter was constructed by amplifying the lac promoter from pCA24N plasmid29 using oligonucleotides IY56F and OA12F (Supplementary Table 2). The amplified products were ligated to yield pCas1+2-IPTG and sequenced to exclude mutations introduced during cloning. The construct was sequenced to ensure the presence of an intact I-SceI site.

pWURV2 plasmid was constructed by amplicating the pCas1 plasmid32 by amplifying the pCas1+2-IPTG and sequencing to exclude mutations introduced during cloning. pBAD33-gp5.9 plasmid was obtained by ligating pCas1+2-IPTG and pBAD33-gp5.9. The plasmid was confirmed using restriction enzymes. The construct was sequenced to ensure the presence of the I-SceI restriction site.

pWUR plasmid encoding Cas1 and Cas2 under pBAD promoter was constructed by amplifying the pBAD promoter from pCA24N plasmid using oligonucleotides IY68F and IY68R. The amplified DNA and pBAD18 (ref. 30) vector were both digested with SacI and Sall and ligated to yield pBAD33-gp5.9.

Strain construction using recombination-based genetic engineering (recombineering). All deletion mutants used for recombineering were obtained from the Keio collection31. BL21-AI was transformed with pCas1–Cas2 plasmid32 and the transformants were selected on LB agar plates containing 30 μg ml−1 ampicillin and 10 μg ml−1 chloramphenicol. E. coli BL21-AI strain was sequence-verified for the presence of an intact I-SceI site.

PCR preparation for CRISPR. Different CRISPR libraries were prepared by PCR amplification from donor strain BL21-AI or K-12 harboring pCas1–Cas2 plasmid32. Colonies were spread on LB plates containing 30 μg ml−1 ampicillin and 10 μg ml−1 chloramphenicol and incubated overnight at 37 °C or 30 °C (for K-12 strain). A single colony was picked from each plate and used as template in a PCR amplifying CRISPR array I for BL21-AI or array II for K-12. Primers MG7R/ OAAR and MG7R/MG34F were used to detect array expansion for BL21-AI and K-12, respectively (Supplementary Table 2). Only colonies that did not undergo array expansion were used in the assays described below.

Standard acquisition assay. A single colony of BL21-AI or BL21-ALPHA::terB, or BL21-AI/AREC::C/D strains harboring pCas1+2 plasmid, or BL21-AI strain harboring pWURV2 plasmid, or K-12/αcasC7T/RNP strain harboring pBAD-Cas1+2 plasmid, or BL21-AI ydhQ-1-SceI site strain harboring pCas1+2::IPTG and pBAD-1-SceI plasmids, or BL21-AI strain harboring pChi or pCtrl-Chi plasmids, and BL21-AI strain harboring pCas1+2 and pBAD33-gp5.9 plasmids were inoculated in LB medium containing 50 μg ml−1 streptomycin + 0.2% (w/v) glucose for BL21-AI strains carrying a single plasmid, or 100 μg ml−1 ampicillin + 0.2% (w/v) glucose for K12 strain, or 100 μg ml−1 ampicillin + 50 μg ml−1 streptomycin + 0.2% (w/v) glucose for BL21-AI ydhQ-1-SceI site/pCas1+2-IPTG/pBAD-1-SceI strain, or 200 μg ml−1 ampicillin + 55 μg ml−1 chloramphenicol for BL21-AI pCas1+2/ pBAD33-gp5.9. Cultures were aerated at 37 °C for 16 h. Each overnight culture was diluted 1:600 in LB medium containing appropriate antibiotics with or without 0.2% (w/v) L-arabinose + 0.1 mM IPTG for pCas1+2, pChi, and pCtrl-Chi harbouring strains, or 0.2% (w/v) L-arabinose for pBAD-Cas1+2 harbouring strains, or 0.02 mM IPTG and 0% L-arabinose for pCas1+2-IPTG and pBAD-1-SceI harbouring strains, or 0.4% (w/v) L-arabinose for pCas1+2 and pBAD33-gp5.9 harbouring strains. Cultures were aerated at 37 °C for an additional 16 h. DNA from these cultures was used as template (see DNA preparation for PCR) in PCRs using primers OA1F/IY130R (PCR1) and RE10RD/IY230R (PCR2) for amplifying BL21-AI CRISPR array I, or MG116F/MG34F (PCR1, see below) and RE10RD/MG115R (PCR2, see below) for amplifying K-12 array II.

Acquisition assay in the presence of antibiotics. A single colony of BL21-AI/ pCas1+2 was inoculated in LB medium containing 50 μg ml−1 streptomycin + 0.2% (w/v) glucose and aerated at 37 °C for 16 h. The overnight cultures were diluted 1:600 in LB medium containing 50 μg ml−1 streptomycin with or without 0.2% (w/v) L-arabinose + 0.1 mM IPTG and aerated at 37 °C. Once cultures reached 0.25, induction of Cas1–Cas2 was performed by adding 0.1 mM IPTG. Cultures were aerated for 16 h at 37 °C.

Acquisition assay in replication-deficient strain. A single colony of K-12/αcasC (control) or K-12/αcasCdnaC2 harbouring pBAD-Cas1+2 was inoculated in LB medium containing 100 μg ml−1 ampicillin + 0.2% (w/v) glucose and aerated at 30 °C, for 16 h. The overnight cultures were diluted 1:600 in LB medium containing 100 μg ml−1 ampicillin + 0.2% (w/v) L-arabinose and aerated at 30 °C or 39 °C for another 16 h. Cultures were then lysed and used as template in PCRs using primers MG116F/MG34F (PCR1) and RE10RD/MG115R (PCR2) for amplifying K-12 array II.

Synchronized acquisition assay. A single colony of K-12/αcasC (control) or K-12/αcasCdnaC2 harbouring pBAD-Cas1+2 was inoculated in LB medium containing 100 μg ml−1 ampicillin + 0.2% (w/v) glucose and aerated at 30 °C, for 16 h. The overnight cultures were diluted 1:600 in LB medium containing 100 μg ml−1 ampicillin + 0.2% (w/v) glucose and aerated at 30 °C until D600nm reached 0.25. Cultures were then split into six tubes and transferred to non-permissive temperature (39 °C). After 70 min, induction of Cas1–Cas2 sensitivity was amplified by PCR with oligonucleotides MG53F and MG53R that provided homology to a region downstream of the ydhQ gene. The Kan-sacB cassette was inserted into DY378 strain37 by recombineering (as described above). Colonies that were found to be resistant to kanamycin and sensitive to sucrose (that is, containing the Kan-sacB cassette) were picked and verified by PCR. The Kan-sacB cassette was transferred by P1 transduction from DY378 to BL21-AI. A second PCR was performed using oligonucleotides MG54F and MG54R that produced a short linear DNA containing the I-SceI restriction site with homology to upstream and downstream of the ydhQ stop codon. Recombineering of this DNA fragment to BL21-AI, ydhQ-Kan-sacB resulted in kanamycin-sensitive and sucrose-resistant colonies that replaced the Kan-sacB cassette with the I-SceI restriction site immediately after the ydhQ stop codon. DNA from the resulting strain was sequence-verified for the presence of an intact I-SceI site.

CRISPR array size determination before acquisition assay. All strains underwent a preliminary validation step aimed at eliminating acquisition before induction: E. coli BL21-AI or K-12 harbouring pCas1+2 or pBAD-Cas1+2 plasmids, respectively, were spread on LB with 50 μg ml−1 streptomycin or 100 μg ml−1 ampicillin + 0.2% (w/v) glucose plates and incubated overnight at 37 °C or 30 °C (for K-12/αcasCdnaC2). A single colony was picked from each plate and used as template in a PCR amplifying CRISPR array I for BL21-AI or array II for K-12. Primers MG7R/OAAR and MG7R/MG34F were used to detect array expansion for BL21-AI and K-12, respectively (Supplementary Table 2). Only colonies that did not undergo array expansion were used in the assays described below.
was performed: cells were centrifuged in a standard centrifuge (4,600g, 10 min), resuspended in LB medium containing 100 µg ml⁻¹ ampicillin + 0.2% (v/v) L-arabinose and aerated for an additional 30 min at 39 °C. Replication was then initiated by aerating the split cultures at 30 °C for 0, 20, 40, 60, 90 and 120 min. For replication arrest, cells were lysed and used as template in PCRs using primers MG116E/MG34F (PCR1) and RE10RD/MG115R (PCR2) for amplifying K-12 array.

DNA preparation for PCR. DNA was prepared from all cultures that underwent acquisition assays. One millilitre of each culture was centrifuged in a microcentrifuge for 1 min at 13,000g and resuspended in 100 µl LB medium. The concentrated culture underwent fast freeze in liquid nitrogen, was boiled at 95 °C for 10 min and placed on ice for 5 min. The lysate was then centrifuged in a microcentrifuge for 2 min at 13,000g; the supernatant was transferred to a new tube and served as template for PCR1 (see Preparation of DNA samples for deep sequencing).

Cultures preparation for RNA sequencing. A single colony of E. coli BL21-AI strain harbouring pCas1 + 2 plasmid was inoculated in LB medium containing 50 µg ml⁻¹ streptomycin + 0.2% (v/v) glucose and aerated at 37 °C for 16 h. Each overnight culture was diluted 1:600 in LB medium containing appropriated antibiotics with or without 0.2% (v/v) L-arabinose + 0.1 mM IPTG. Following growth, 15 ml of each culture was centrifuged in a standard centrifuge (4,600g, 10 min), the supernatant was discarded and the pellet underwent fast freeze in liquid nitrogen. Cell pellets were then thawed and incubated at 37 °C with 300 µl 2 mg ml⁻¹ lysozyme (Sigma-Aldrich catalogue number L6876-1G) in Tris 10 mM pH 8.0, and total nucleotides were extracted using the Tr-NucleoSpin Tissue kit. Total DNA samples were used for deep sequencing.

Total DNA purification. Overnight cultures of E. coli BL21-AI or K-12 lacAacC T7RNAP harbouring pCas1 + 2 plus pBAD-Cas1 + 2 plasmid, respectively, were diluted 1:600 and aerated for 16 h at 37 °C in LB medium containing 50 µg ml⁻¹ streptomycin or 100 µg ml⁻¹ ampicillin + 0.2% (v/v) glucose. These overnight cultures were then diluted 1:600 in LB medium containing 50 µg ml⁻¹ streptomycin or 100 µg ml⁻¹ ampicillin + 0.2% (v/v) L-arabinose + 0.1 mM IPTG or without inducers and aerated at 37 °C. Once cultures reached OD₆₀₀nm = 0.5–0.6, 3 ml were removed and used for total DNA purification using a Macherey-Nagel NucleoSpin Tissue kit. Total DNA samples were used for deep sequencing (MiSeq).

Preparation of spacer PCR products for deep sequencing. DNA from bacterial cultures that underwent various acquisition assays was amplified in two consecutive PCRs termed PCR1 and PCR2. In PCR1, the reaction contained 20 µl of Taq 2× Master Mix master mix, 1 µl of 10 µM forward and reverse primers (see Supplementary Table 2), 4 µl of bacterial lysate and 14 µl of double-distilled water. The PCR started with 3 min at 95 °C followed by 35 cycles of 20 s at 95 °C, 20 s at 55 °C and 20 s at 72 °C. The final extension step at 72 °C was performed for 5 min. Half of the PCR1 content (20 µl) was purified using the DNA clean-up kit and used for standard library preparation procedures followed by deep sequencing (MiSeq), while the other half (20 µl) was loaded on a 2% (w/v) agarose gel and electrophoresed for 60 min at 120 V. Following gel separation, the expanded band was excised from the gel and purified using the DNA clean-up kit. One nanogram from the extracted band served as a template for the PCR2 reaction aimed at amplifying the expanded CRISPR array products. PCR2 contained 10 µl of Taq 2× Master Mix master mix, 0.5 µl of 10 µM forward and reverse primers (Supplementary Table 2). 1 ng of the gel-extracted DNA from PCR1 and double-distilled water up to 20 µl. PCR2 program was identical to that of PCR1. The entire PCR2 content was loaded on a 2% (w/v) agarose gel, electrophoresed, excised and purified from the gel using the same conditions as in PCR1. Reads were mapped against the E. coli genome and pCAS plasmid using blastn (with parameters: -e 0.0001 -F F). For strain K-12, the Resf accession NC_000913.2 was used; for strain BL21-AI (for which the genome sequence is unavailable), the E. coli BL21-Gold(DE3)pLysS AG was used (Resf accession NC_012947.1).

New spacer insertions were called on the basis of sequence alignments of the resulting reads. For round 1 of the PCR (Extended Data Fig. 1), alignments supporting non-acquisition events were also recorded to quantify acquisition level. If the sequence read was fully mapped to the parental CRISPR locus in the leader-proximal side, a non-acquisition event was inferred. New acquisition events were inferred if the read alignment began by a substring that was mapped to the CRISPR locus (‘pre-acquisition’ mapping) followed by a spacer-length substring that mapped elsewhere on the genome or the plasmid. Uninformative alignments, resulting from sequencing of the leader-distal side of the PCR amplicon, were discarded. Spacer acquisition level for a sample was defined as the number of reads supporting acquisition events divided by the number of reads either supporting or rejecting spacer acquisition.

For round 2 of the PCR (enriching for expanded arrays only, Extended Data Fig. 1) we used only unambiguously mapped spacers (for example, spacers mapped to repetitive tRNA genes were discarded). If a spacer was mapped equally well both to the genome and the pCAS plasmid, only the plasmid spacer position was used.

For the plots of spacer distribution and hotspots (except for the plot in Extended Data Fig. 3), spacer positions were recorded only once (meaning that if there were multiple spacers hitting the exact same position, the position was considered only once). This procedure was done to avoid biases stemming from PCR amplification of the CRISPR array, as well as local biases stemming from differential PAM preferences.

Perl and R scripts were used for data analysis. Data visualization and statistical analysis used Microsoft Excel and R, including the R circular package (http://cran.r-project.org/web/packages/circular/circular.pdf) for Fig. 1 and Extended Data Fig. 4.
Extended Data Figure 1 | Graphic overview of the procedure for characterizing the frequency and sequence of newly acquired spacers. DNA from cultures of either E. coli K-12 (left) or E. coli BL21-AI (right) strains expressing Cas1–Cas2 from two different plasmids were used as templates for PCR. Round 1 was used to determine the frequency of spacer acquisition by comparing occurrences of expanded arrays to wild-type (WT) arrays. Round 2 amplified only the expanded arrays and, followed by deep sequencing, was used to determine the sequence, location and source of newly acquired spacers.
Extended Data Figure 2 | PAMs and DNA content along the *E. coli* BL21-AI genome. 

a. Distribution of PAM (AAG) sequences. Each data point represents the number of PAMs in a window of 10 kb.
b. DNA content of a culture growing in log phase. Genomic DNA was extracted from *E. coli* BL21-AI cells carrying the pCas plasmid, grown at log phase, and was sequenced using the Illumina technology. The resulting reads were mapped to the sequenced *E. coli* BL21(DE3) genome (GenBank accession number NC_012947). Areas where few or no reads map to the genome represent regions that are present in the reference BL21(DE3) genome but are missing from the genome of the sequenced strain (BL21-AI).
Extended Data Figure 3 | Distribution of newly acquired spacers on the genome during synchronized replication. *E. coli* K-12 ΔcasCΔbatC2 cells were transferred from 39 °C (replication restrictive temperature) to 30 °C (replication permissive). Cas1–Cas2 were induced in these cells 30 min before the transfer to 30 °C and during the growth in 30 °C. Newly acquired spacers were sequenced at the given time points: a, following 20 min; b, following 40 min; c, following 60 min from replication initiation. The positions of the newly acquired spacers in windows of 100 kb are shown, and their fraction out of the total new spacers in the sample.
Extended Data Figure 4 | A model explaining the preference for spacer acquisition near TerC compared with TerA in *E. coli* BL21-AI. The DNA manipulation at the CRISPR region forms a replication fork stalling site, and leads to extensive spacer acquisition upstream of the CRISPR. While the clockwise fork is stalled at the CRISPR, the anticlockwise fork reaches the Ter region and is stalled at the respective Ter site, TerC, leading to extensive spacer acquisition upstream of TerC. Another factor that can contribute to the observed TerC/TerA bias may be that the clockwise replichore in *E. coli* (oriC to TerA) is longer than the anticlockwise one (oriC to TerC), leading the forks to stall at TerC more often than at TerA.
Extended Data Figure 5 | The protein product of T7 gene 5.9 inhibits spacer acquisition activity. *E. coli* BL21-AI strains harbouring pBAD-Cas1 +2 and pBAD33-gp5.9 (lane 1) or pBAD33 vector control (lane 2) were grown overnight in the presence of inducers (0.4% l-arabinose). Gel shows PCR products amplified from the indicated cultures using primers annealing to the leader and to the fifth spacer of the CRISPR array. Results represent one of three independent experiments.
Extended Data Figure 6 | Distribution of protospacers across the plasmids. 
a, Distribution across pCtrl-Chi; b, distribution across pChi plasmids. Circular representation of the 4.7 kb plasmid is presented, with the inserted 4-Chi cluster present at the top of the circle. Black bars indicate the number of PAM-derived spacers sequenced from each position; green bars represent non-PAM spacers. Scale bar, 100,000 spacers. Pooled protospacers from two replicates are presented for each panel.
### Extended Data Table 1 | Spacer acquisition in normal and perturbed conditions

#### a.

| Sample                              | rep | # of reads spanning the CRISPR array | # of reads supporting unmodified parental array | # of reads supporting acquisition of a new spacer | % expanded arrays |
|-------------------------------------|-----|--------------------------------------|-----------------------------------------------|-------------------------------------------------|------------------|
| BL21-AI, no ara                     | 1   | 25,718                               | 25,163                                        | 555                                             | 2.16%            |
| BL21-AI, no ara                     | 2   | 32,807                               | 31,800                                        | 1,007                                           | 3.07%            |
| BL21-AI, 0.2% ara                   | 1   | 28,188                               | 17,438                                        | 10,750                                          | 38.14%           |
| BL21-AI, 0.2% ara                   | 2   | 33,973                               | 21,843                                        | 12,130                                          | 35.70%           |
| BL21-AI, empty vector, no ara       | 1   | 12,021                               | 12,021                                        | 0                                               | 0%               |
| BL21-AI, empty vector, no ara       | 2   | 14,729                               | 14,729                                        | 0                                               | 0%               |
| BL21-AI, empty vector, 0.2% ara     | 1   | 28,251                               | 28,251                                        | 0                                               | 0%               |
| BL21-AI, empty vector, 0.2% ara     | 2   | 6,827                                | 6,827                                         | 0                                               | 0%               |

#### b.

| Sample                              | rep | # new spacers sequenced | # spacers from chromosome | # spacers from plasmid | % spacers from plasmid | % spacers from genome |
|-------------------------------------|-----|-------------------------|---------------------------|------------------------|------------------------|-----------------------|
| BL21-AI, no ara                     | 1   | 2,594,637               | 48,300                    | 2,546,337              | 98.14%                 | 1.86%                 |
| BL21-AI, no ara                     | 2   | 2,056,397               | 35,911                    | 2,020,486              | 98.25%                 | 1.75%                 |
| BL21-AI, 0.2% ara                   | 1   | 647,929                 | 151,181                   | 496,748                | 76.67%                 | 23.33%                |
| BL21-AI, 0.2% ara                   | 2   | 851,824                 | 190,791                   | 661,033                | 77.60%                 | 22.40%                |
| BL21-AI pheA::TerB                  | 1   | 2,937,147               | 46,015                    | 2,891,132              | 98.43%                 | 1.57%                 |
| BL21-AI pheA::TerB                  | 2   | 3,400,210               | 44,748                    | 3,355,462              | 98.68%                 | 1.32%                 |

#### c.

| Sample                              | rep | # of reads spanning the CRISPR array | # of reads supporting unmodified parental array | # of reads supporting acquisition of a new spacer | % expanded arrays |
|-------------------------------------|-----|--------------------------------------|-----------------------------------------------|-------------------------------------------------|------------------|
| BL21-AI + Nalidixic acid            | 1   | 71,941                               | 71,800                                        | 141                                             | 0.20%            |
| BL21-AI + Nalidixic acid            | 2   | 77,774                               | 77,714                                        | 60                                              | 0.08%            |
| BL21-AI + Rifampicin                | 1   | 36,976                               | 34,145                                        | 2,831                                           | 7.66%            |
| BL21-AI + Rifampicin                | 2   | 38,702                               | 28,147                                        | 10,555                                          | 27.27%           |

---

a. Adaptation experiments with *E. coli* BL21-AI cells. After overnight growth with or without induction of Cas1–Cas2 cloned on pWUR plasmid, the CRISPR array was amplified and sequenced to determine the fraction of arrays that acquired a new spacer. Results with BL21-AI with an empty pWUR vector (without Cas1–Cas2) are presented as a control. b. Identity of acquired spacers in *E. coli* BL21-AI cells. After overnight growth with or without induction of Cas1–Cas2, gel-separated expanded arrays were amplified and sequenced to study the identity of newly acquired spacers in high resolution. c. Effect of antibiotics on adaptation levels. The Cas1–Cas2 operon was induced in *E. coli* BL21-AI cells using 0.2% l-arabinose and 0.1 mM IPTG overnight, in the presence of either nalidixic acid (50 µg ml⁻¹) or rifampicin (100 µg ml⁻¹). After overnight induction, the CRISPR array was amplified and sequenced.
Extended Data Table 2 | Replication-dependent spacer acquisition

**a.**

| Sample                  | rep | # of reads spanning the CRISPR array | # of reads supporting unmodified parental array | # of reads supporting acquisition of a new spacer | % expanded arrays |
|-------------------------|-----|-------------------------------------|-----------------------------------------------|-----------------------------------------------|------------------|
| K-12 ΔcasC 30°C         | 1   | 98,884                              | 96,299                                        | 2,585                                         | 2.61%            |
| K-12 ΔcasC 30°C         | 2   | 117,522                             | 115,030                                       | 2,492                                         | 2.12%            |
| K-12 ΔcasC, dnaC2 30°C  | 1   | 152,827                             | 149,644                                       | 3,183                                         | 2.08%            |
| K-12 ΔcasC, dnaC2 30°C  | 2   | 100,125                             | 98,053                                        | 2,072                                         | 2.07%            |
| K-12 ΔcasC, 39°C        | 1   | 87,036                              | 83,688                                        | 3,348                                         | 3.85%            |
| K-12 ΔcasC, 39°C        | 2   | 86,580                              | 82,474                                        | 4,106                                         | 4.74%            |
| K-12 ΔcasC, dnaC2 39°C  | 1   | 66,618                              | 66,618                                        | 0                                             | 0.00%            |
| K-12 ΔcasC, dnaC2 39°C  | 2   | 60,325                              | 60,321                                        | 4                                             | 0.01%            |

**b.**

| Sample                  | rep | # of reads spanning the CRISPR array | # of reads supporting unmodified parental array | # of reads supporting acquisition of a new spacer | % expanded arrays | # spacers from chromosome | # spacers from Ter region | % spacers from Ter |
|-------------------------|-----|-------------------------------------|-----------------------------------------------|-----------------------------------------------|------------------|--------------------------|--------------------------|---------------------|
| dnaC2 0 min, 30°C       | 1   | 99,683                              | 99,669                                        | 14                                            | 0.014%           | 3,684                    | 508                      | 13.79%              |
| dnaC2 0 min, 30°C       | 2   | 107,825                             | 107,814                                       | 11                                            | 0.010%           | 456                      | 26                       | 5.70%               |
| dnaC2 20 min, 30°C      | 1   | 107,679                             | 107,671                                       | 8                                             | 0.007%           | 1,250                    | 128                      | 10.24%              |
| dnaC2 20 min, 30°C      | 2   | 113,040                             | 113,030                                       | 10                                            | 0.009%           | 1,402                    | 36                       | 2.57%               |
| dnaC2 40 min, 30°C      | 1   | 394,058                             | 394,018                                       | 40                                            | 0.010%           | 4,830                    | 930                      | 19.25%              |
| dnaC2 40 min, 30°C      | 2   | 100,975                             | 100,964                                       | 11                                            | 0.011%           | 10,541                   | 2,821                    | 26.76%              |
| dnaC2 60 min, 30°C      | 1   | 63,978                              | 63,967                                        | 11                                            | 0.017%           | 5,563                    | 1,604                    | 28.83%              |
| dnaC2 60 min, 30°C      | 2   | 108,605                             | 108,588                                       | 17                                            | 0.016%           | 6,551                    | 2,183                    | 33.32%              |
| dnaC2 90 min, 30°C      | 1   | 109,652                             | 109,636                                       | 16                                            | 0.015%           | 3,221                    | 348                      | 10.80%              |
| dnaC2 90 min, 30°C      | 2   | 206,652                             | 206,567                                       | 85                                            | 0.041%           | 2,827                    | 320                      | 11.32%              |
| dnaC2 120 min, 30°C     | 1   | 80,213                              | 80,192                                        | 21                                            | 0.026%           | 3,373                    | 848                      | 25.14%              |
| dnaC2 120 min, 30°C     | 2   | 121,583                             | 121,530                                       | 53                                            | 0.044%           | 3,135                    | 721                      | 23.00%              |

---

a. Adaptation experiment with dnaC2 temperature-sensitive cells. E. coli K12 cells were transformed with a pBAD-Cas1–Cas2 vector, in which the Cas1–Cas2 operon was directly controlled by an arabinose-inducible promoter. After overnight induction by 0.2% l-arabinose and 0.1 mM IPTG, the CRISPR array was amplified and sequenced. b. Time-course adaptation experiments with synchronously replicating dnaC2 temperature-sensitive cells. Temperature-sensitive K-12ΔcasD dnaC2 culture was transferred to 39 °C for 70 min. Cas1–Cas2 expression was then induced for 30 min using 0.2% l-arabinose and 0.1 mM IPTG, and the culture was transferred to 30 °C with continuous induction of Cas1–Cas2. The culture was sampled at successive time points after synchronous replication initiation, and the CRISPR array was amplified and sequenced to determine the fraction of cells that acquired a new spacer. In addition, gel-separated expanded arrays were amplified and sequenced, to study the localization of spacers derived from the chromosome.
Extended Data Table 3 | Involvement of the DNA repair machinery in spacer acquisition

**a.**

| Sample                              | rep | # of reads spanning the CRISPR array | # of reads supporting unmodified parental array | % of reads supporting acquisition of a new spacer | % expanded arrays | # new spacers sequenced | # spacers from chromosome | # spacers from plasmid | % spacers from plasmid | % spacers from genome |
|-------------------------------------|-----|--------------------------------------|-----------------------------------------------|-----------------------------------------------|------------------|-------------------------|--------------------------|------------------------|------------------------|------------------------|
| BL21-Al ΔrecB, no ara               | 1   | 35,060                               | 34,615                                        | 445                                           | 1.27%            | 663,470                 | 107,260                  | 556,210                | 83.83%                 | 16.17%                 |
| BL21-Al ΔrecB, no ara               | 2   | 36,116                               | 35,778                                        | 338                                           | 0.94%            | 441,290                 | 75,260                   | 366,030                | 82.95%                 | 17.05%                 |
| BL21-Al ΔrecC, no ara               | 1   | 116,840                              | 115,012                                       | 1,828                                         | 1.56%            | 704,870                 | 96,707                   | 608,163                | 86.28%                 | 13.72%                 |
| BL21-Al ΔrecC, no ara               | 2   | 132,549                              | 130,724                                       | 1,825                                         | 1.38%            | 507,844                 | 55,057                   | 452,787                | 89.16%                 | 10.84%                 |
| BL21-Al ΔrecD, no ara               | 1   | 85,877                               | 85,253                                        | 624                                           | 0.73%            | 2,938,455               | 353,353                  | 2,585,102              | 87.97%                 | 12.03%                 |
| BL21-Al ΔrecD, no ara               | 2   | 90,498                               | 89,802                                        | 696                                           | 0.77%            | 4,437,733               | 1,405,158                | 3,032,575              | 68.34%                 | 31.66%                 |
| BL21-Al ydhQ-I-SceI site/pCas12-IPTG/ pBAD-I-SceI, no ara | 1   | 87,419                               | 83,625                                        | 3,794                                         | 4.34%            | 221,721                 | 16,906                   | 204,815                | 92.38%                 | 7.62%                  |
| BL21-Al ydhQ-I-SceI site/pCas12- IPTG/ pBAD-I-SceI, no ara | 2   | 89,357                               | 86,745                                        | 2,612                                         | 2.92%            | 192,597                 | 15,995                   | 176,642                | 91.72%                 | 8.28%                  |

**b.**

| Sample                              | rep | # of reads spanning the CRISPR array | # of reads supporting unmodified parental array | % of reads supporting acquisition of a new spacer | % expanded arrays | # new spacers sequenced | # spacers from chromosome | # spacers from plasmid | % spacers from plasmid | % spacers from genome |
|-------------------------------------|-----|--------------------------------------|-----------------------------------------------|-----------------------------------------------|------------------|-------------------------|--------------------------|------------------------|------------------------|------------------------|
| BL21-Al pCtrl-Chi, no ara           | 1   | 4,221,820                            | 42,055                                        | 4,179,765                                     | 99%              | 4,179,765               | 4,179,765               | 0%                     | 1%                     | 1%                     |
| BL21-Al pCtrl-Chi, no ara           | 2   | 5,743,373                            | 50,345                                        | 5,693,028                                     | 99.12%           | 5,693,028               | 5,693,028               | 0.88%                  | 1%                     | 1%                     |
| BL21-Al pChi, no ara                | 1   | 2,726,923                            | 78,079                                        | 2,648,844                                     | 96.97%           | 2,648,844               | 2,648,844               | 3.03%                  | 1%                     | 1%                     |
| BL21-Al pChi, no ara                | 2   | 2,841,509                            | 87,106                                        | 2,754,403                                     | 96.74%           | 2,754,403               | 2,754,403               | 3.26%                  | 1%                     | 1%                     |

**a.** Adaptation experiments with *E. coli* BL21-Al ΔrecB, ΔrecC, ΔrecD and ydhQ-I-SceI cells. After overnight growth without induction of Cas1–Cas2, the CRISPR array was amplified and sequenced to determine the fraction of cells that acquired a new spacer. In addition, gel-separated expanded arrays were amplified and sequenced, to study the identity of newly acquired spacers in high resolution. **b.** Adaptation experiments with *E. coli* BL21-Al pCas1–2-Chi. After overnight growth without induction of Cas1–Cas2 from the pChi plasmid (which contained a cluster of four anticlockwise Chi sites on a 50 bp cassette inserted at position 1,300 of the pCas plasmid) gel-separated expanded arrays were amplified and sequenced, to differentiate between spacers acquired from self and plasmid DNA. As a control, a similar plasmid with a 50-bp Chi-less insertion at the same position in the pCas plasmid was used.

©2015 Macmillan Publishers Limited. All rights reserved