ARTICLES
https://doi.org/10.1038/s41564-017-0097-z

nature microbiology

# Spontaneous CRISPR loci generation in vivo by non-canonical spacer integration

Jeff Nivala[1,2,4], Seth L. Shipman[1,2,3] and George M. Church[1,2*]

**The adaptation phase of CRISPR–Cas immunity depends on the precise integration of short segments of foreign DNA (spacers) into a specific genomic location within the CRISPR locus by the Cas1–Cas2 integration complex. Although off-target spacer integration outside of canonical CRISPR arrays has been described in vitro, no evidence of non-specific integration activity has been found in vivo. Here, we show that non-canonical off-target integrations can occur within bacterial chromosomes at locations that resemble the native CRISPR locus by characterizing hundreds of off-target integration locations within *Escherichia coli*. Considering whether such promiscuous Cas1–Cas2 activity could have an evolutionary role through the genesis of neo-CRISPR loci, we combed existing CRISPR databases and available genomes for evidence of off-target integration activity. This search uncovered several putative instances of naturally occurring off-target spacer integration events within the genomes of *Yersinia pestis* and *Sulfolobus islandicus*. These results are important in understanding alternative routes to CRISPR array genesis and evolution, as well as in the use of spacer acquisition in technological applications.**

The spacer acquisition process of the *Escherichia coli* type I-E clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated protein (Cas) system has been well characterized[1–8]. As is typical in all known types, spacer integration by this system requires two Cas proteins, Cas1 and Cas2, which form a heteromeric integration complex[7,9–11]. On the protospacer, the Cas1–Cas2 complex recognizes a 3′-TTC-5′ protospacer adjacent motif (PAM) on the bottom strand that largely determines the efficiency and directionality of protospacer integration into the CRISPR array[12–15]. The array is minimally composed of 60 nucleotides (nt) of the leader region, and a single 28-nt repeat[7]. Within the array, Cas1–Cas2 recognizes a conserved inverted repeat motif within the interior of the repeat. A non-Cas protein, integration host factor (IHF), binds to a conserved sequence within the leader and helps direct integration into the 5′ leader proximal end of the array[6]. In vitro, spacer integration events occur outside of canonical CRISPR arrays ('off-target' sites) with relatively high frequency, albeit with a lower occurrence in the presence of IHF[6,9]. Even so, this is surprising considering that specific integration into the 5′ end of the array is essential for robust immunity[16] and overall genomic integrity[17].

## Results

Recently, we demonstrated that electroporation of synthetic oligo protospacers into *E. coli* BL21 overexpressing Cas1–Cas2 led to the acquisition of these oligo sequences into the genomic CRISPR1 locus[15]. We reasoned that this method of defined spacer acquisition (DSA) would make the discovery of off-target spacer integrations easier to detect within the genome because, in contrast to previous paradigms, the spacer sequence is known a priori. We performed DSA using a previously characterized oligo protospacer that is integrated with high efficiency[9,15] (psAA33; Supplementary Table 1). The psAA33 sequence matches a 35-nt segment of the M13 bacteriophage genome, and includes a canonical 5′-AAG PAM. Following electroporation of this oligo into cultures of cells overexpressing Cas1–Cas2, cells were diluted into fresh lysogeny broth (LB) and allowed to recover. After culture outgrowth overnight, the total DNA content of the cells (genomic and plasmid) was extracted and subjected to whole-genome shotgun sequencing at a depth of ~350× genomic coverage on an Illumina MiSeq (Fig. 1a). Reads were mapped to the BL21 reference genome. Analysis conditions were set to allow for alignments with >50 nt insertions in each read (relative to the reference sequence), as canonical spacer integration into the CRISPR array results in a 61-nt expansion (33-nt spacer + 28-nt repeat duplication). After mapping, 32 reads aligning to the first position of the genomic CRISPR1 array showed array expansions resulting from the integration of the psAA33 sequence (20 reads) or endogenous genome/plasmid-derived spacers (12 reads). We also found a total of nine reads outside the genomic arrays ('off-target') that similarly contained all the hallmarks of spacer integration events. Each of these reads contained a 27-nt or 28-nt region of the genome duplicated on both sides of a 33-nt insertion. In 7 instances, the 33-nt insertion contained the psAA33 sequence. In each of these, the inserted bases excluded the 5′-AA bases of the PAM (Fig. 1b and Table 1), which is consistent with Cas1–Cas2-mediated PAM processing and integration of the oligo spacer that occurs in 'on-target' integrations[14,15]. The other two off-target instances were 33-nt spacer insertions whose sequences occur at other regions of the BL21 genome, and appear to be off-target integration events of genome-derived spacers. As *E. coli* BL21 lacks the Cas proteins needed for target interference (the Cascade proteins), the occurrence of genome-derived spacers would not lead to autoimmunity in this context. These initial results showed that, in these conditions, off-target integrations by Cas1–Cas2 occur with a frequency of ~1 off-target integration for every 4 on-target integrations into the CRISPR array, and that both oligo and genome-derived protospacers can be integrated off-target (Fig. 1c). Although we found over-represented nucleotides in the genomic sites surrounding these off-target integrations (Supplementary Fig. 1)

[1]Department of Genetics, Harvard Medical School, Boston, MA, USA. [2]Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, MA, USA. [3]Department of Stem Cell and Regenerative Biology, Center for Brain Science, and Harvard Stem Cell Institute, Harvard University, Cambridge, MA, USA. Present address: [4]Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA.
*e-mail: gchurch@genetics.med.harvard.edu

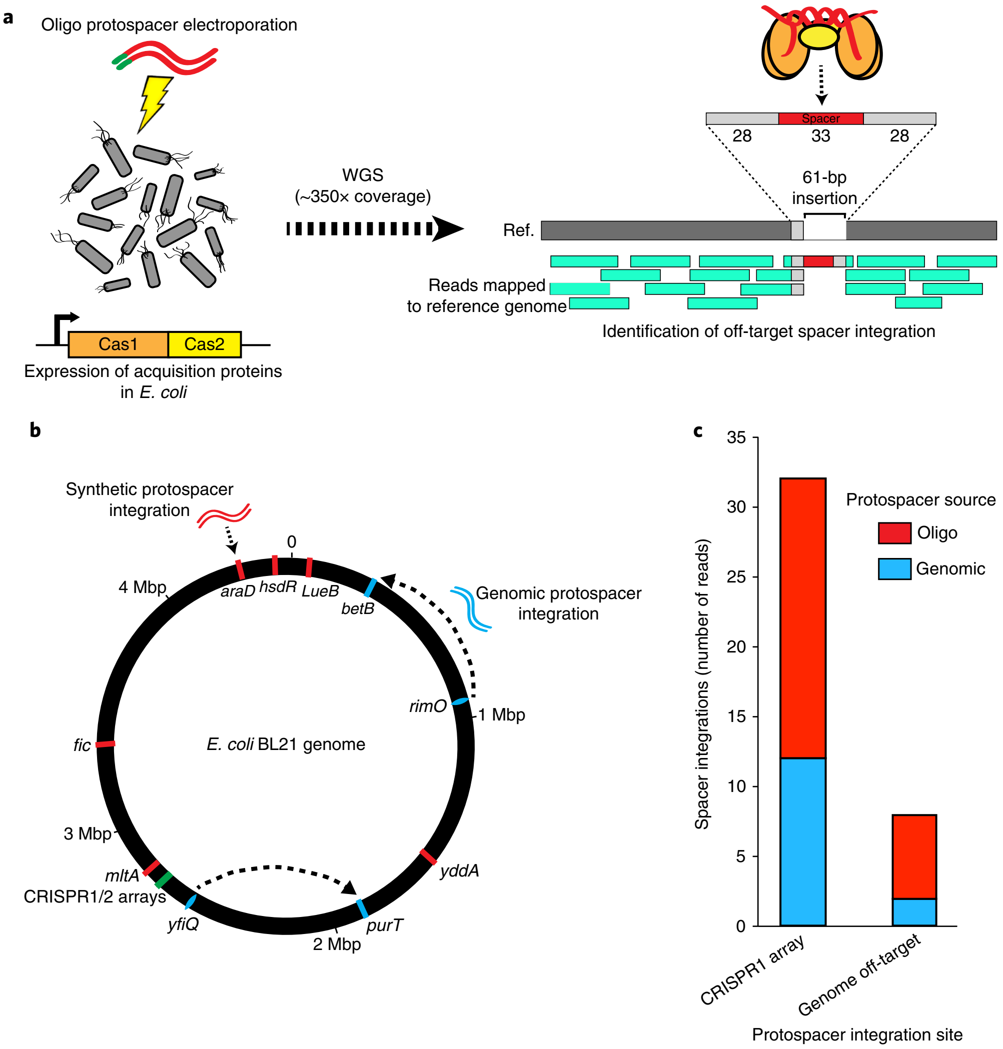

**Fig. 1 | Whole-genome deep sequencing reveals off-target spacer integration events within the *E. coli* genome. a**, Schematic of the experimental workflow. A culture of *E. coli* BL21 expressing Cas1 and Cas2 is electroporated with a 35-bp oligo protospacer that includes a 5′-AAG PAM. Following electroporation and outgrowth, the total DNA content of the cells is isolated, fragmented and shotgun sequenced on an Illumina high-throughput sequencing machine. Reads are mapped back to the BL21 reference genome. Spacer integration events are identified as an ~61-bp insertion, which includes the spacer sequence (33 bp) and the duplicated target site (~28-bp repeat). **b**, Eight of the off-target integration sites discovered within the genome, shown in the diagram labelled as the gene in which they were inserted. The origin of the dashed arrows indicates the site of the genome-derived spacer and point towards the site of integration. Note that off-target integration events within the *lacI* gene are not shown because they cannot be unambiguously mapped to the genome or plasmid. **c**, Comparison between the number of on-target integrations into the first position of the CRISPR1 array and off-target integrations elsewhere in the genome outside of the CRISPR1 array. Data represent the results from a single WGS experiment.

that partially agreed with previous work characterizing essential array sequence motifs[18,19], the small sample size made us hesitant to draw firm conclusions. Thus, we sought to characterize many more off-target integration events.

To radically expand the number of off-target sites that we could identify without having to continually sequence the genome to extreme depths, we developed a method to target our sequencing to spacer integration sites, which we term Spacer-seq (Fig. 2a). After prepping whole-genome libraries, the Spacer-seq approach utilizes an additional round of PCR with a specific primer that matches the defined spacer sequence to amplify only fragments of the genome that contain a new integration. Applying Spacer-seq to the genomic fragment library previously presented in Fig. 1 (as well as three additional biological replicates), we specifically enriched and sequenced only the genomic fragments that contained the psAA33 oligo protospacer sequence, and discovered an additional 695 unique off-target spacer integration sites (Fig. 2b, Supplementary Fig. 2 and Supplementary Table 2). To eliminate the potential for analysing fragments that did not contain bona fide spacer integrations, we performed the spacer-enrichment PCR step with primers that excluded the terminal 10 base pairs (bp) of the 3′ psAA33 sequence (Supplementary Table 1). This allowed us to filter out fragments that were amplified by mispriming on regions of endogenous DNA, as they would not contain the 10-bp spacer-specific sequence that was excluded in the primer. Of the Spacer-seq reads that passed this filter, ~86% of the integration sites mapped to a CRISPR locus, whereas the remaining reads aligned to off-target sites in the genome (~13%) or plasmid (~0.4%) (Fig. 2c). Normalizing for total DNA content within the cell, off-target integrations displayed no preference between inserting into genomic DNA and inserting into plasmid DNA (Fig. 2d), and were typically found within the protein-coding regions of non-essential genes (~94%). In addition, to investigate how threshold effects may influence the frequency of off-targeting, we replicated the experiment with decreasing levels

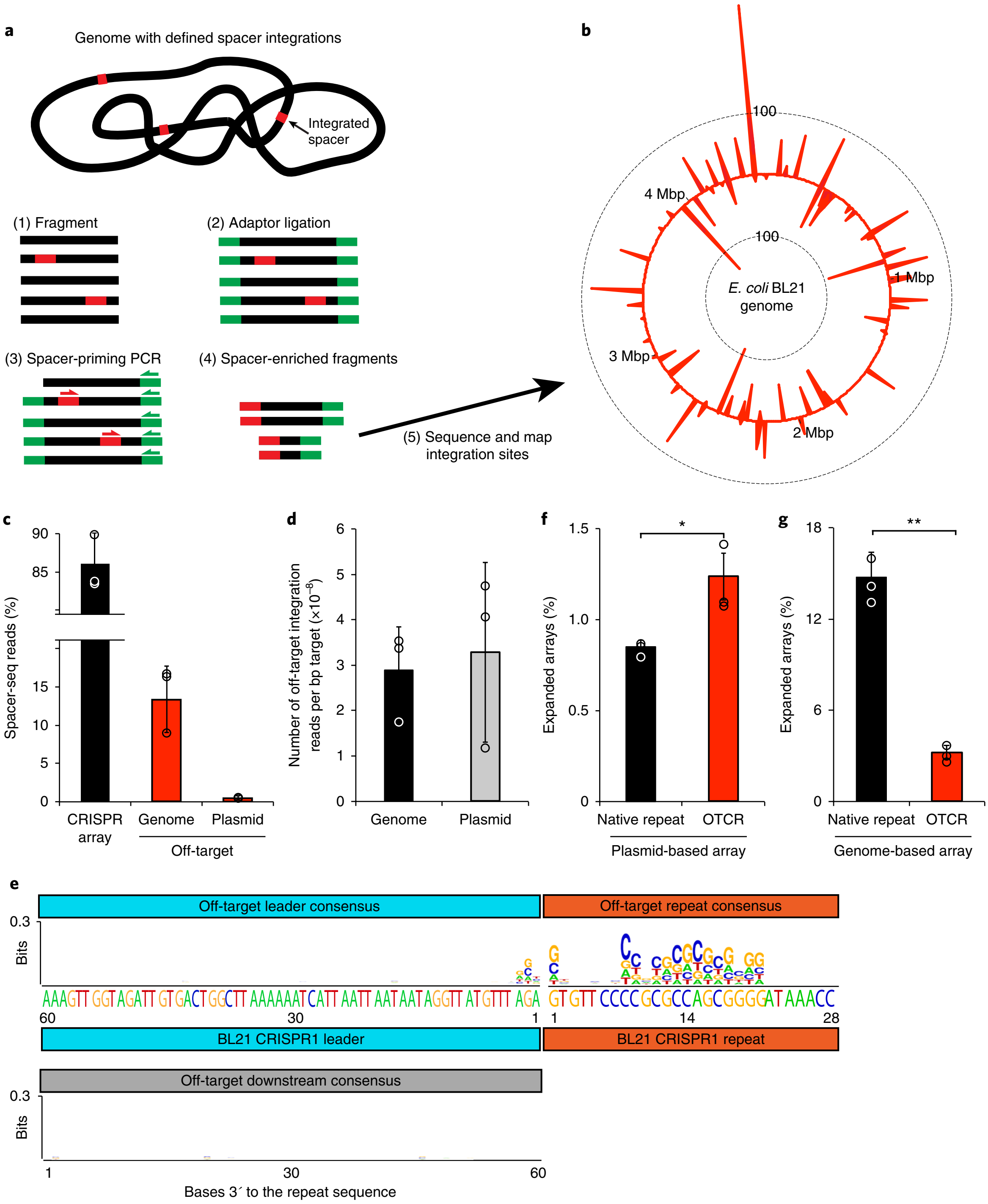

**Fig. 2 | Spacer-seq identifies hundreds of off-target spacer integration sites within the *E. coli* genome.** **a**, Schematic of the Spacer-seq workflow. Fragmentation of isolated genomic DNA containing DSA events (step 1). Ligation of adaptor sequences onto fragment ends (step 2). PCR amplification using the defined spacer sequence and adaptor sequence as primers (step 3) for specific enrichment of fragments containing spacer insertions (step 4). High-throughput sequencing of enriched fragments and mapping of reads to the reference genome (step 5). **b**, The diagram of the genome shows an example of a single Spacer-seq experiment with the number of reads mapped to the *E. coli* BL21 genome (binned per 10 kb). Dashed lines represent 100 reads. **c**, The percentage of Spacer-seq reads mapped to a CRISPR array, or to off-target sites within the genome or expression plasmid. Error bars represent mean ± s.d., $n = 3$ biological replicates. **d**, Comparison between the average number of off-target integration events mapped to the genome or plasmid, normalized by the total DNA content within the cell (assuming ~30 plasmids per cell). Error bars represent mean ± s.d., $n = 3$ biological replicates. **e**, WebLogo of the ~700 unique off-target integration sites identified by Spacer-seq, aligned to the BL21 CRISPR1 array leader and the repeat sequence. **f**, The percentage of expanded arrays after the DSA experiment. Plasmid containing the minimal version of the K12 CRISPR1 array (native repeat) is compared to a mutant version with repeat mutations C14G and A15C (OTCR). Error bars represent mean ± s.e.m. $n = 3$ biological replicates. $^*P = 0.04$ calculated with a two-sample unpaired *t*-test. **g**, The percentage of expanded arrays after the DSA experiment. The genomic CRISPR1 array (native repeat) is compared to a strain in which the entire CRISPR1 locus is replaced with a minimal array consisting of a 100-nt leader and a single mutant repeat (OTCR). Error bars represent mean ± s.e.m. $n = 3$ biological replicates. $^{**}P = 0.002$ calculated with a two-sample unpaired *t*-test. Open circles represent individual replicate data points.

**Table 1 | Off-target spacer integrations identified by WGS**

| Genomic integration site | Repeat 1 | Repeat 2 | Repeat size (bp) | Spacer | Protospacer origin |
|---|---|---|---|---|---|
| 84342 (*leuB*) | GTCAGTTCGCGCAC<br>ACACAGGATGTCG | GTCAGTTCGCGCA<br>CACACAGGATGTCG | 27 | psAA33 | Oligo |
| 297299 (*betB*) | GAGGAAGGCGCGC<br>GCGTACTGTGCGGCG | GAGGAAGGCGCG<br>CGCGTACTGTGCGGCG | 28 | (CC)GCATGTGGACGACGTCA<br>TCCCACTGATGGCAGA(GA) | Genome (rimO) |
| 1502611 (*yddA*) | ATTTATCGTCTACG<br>GGCAGGGGAAGTGC | ATTTATCGTCTACG<br>GGCAGGGCAAGTGC | 28 | (CT)GGCAGTGCGCCCTTATC<br>CGCATCAGCTGGAAGA(AT) | Genome (yfiQ) |
| 1846069 (*purT*) | TGTTGTCCCCTGC<br>GCTCGCGCAACGAAA | TGTTGTCCCCTGCG<br>CTCGCGCAACGAAA | 28 | psAA33 | Oligo |
| 2754345 (*mltA*) | ATCACCGCCTGCG<br>CGCAGCCACTCTGCC | ATCACCGCCAGCGC<br>GCAGCCACTCTTGC | 28 | psAA33 | Oligo |
| 3320059 (*fic*) | GCTTGGTCCGCTG<br>GTGCGCGGTTTACCG | GCTTGGTCCGCTGG<br>TGCGCGGTTTACCG | 28 | psAA33 | Oligo |
| 4299750 (*araD*) | TGATATCCCGTGC<br>ACGCGCGGATTAAGC | TGATATCCCGTGCAC<br>GCGCGGATTAAGC | 28 | psAA33 | Oligo |
| 4460195 (*hsdR*) | CTTTTTCCCGCAG<br>GCGCGAGGCGAAGCC | CTTTTTCCCGCAGGC<br>GCGAGGCGAAGCC | 28 | psAA33 | Oligo |
| 336969/plasmid (*lacI*) | ATCAGACCGTTTC<br>CCGCGTGGTGAACCA | ATCAGACCGTTTCCC<br>GCGTGGTGAACCA | 28 | psAA33 | Oligo |

Spacers derived from the genome show flanking genomic nucleotides in parentheses. Genomic integration site nucleotide numbering and gene annotations are referenced to the *E. coli* BL21 genome GenBank accession number CP010816.

of Cas1–Cas2 induction and oligo protospacer. We observed that, although decreasing the concentration of oligo protospacer by up to $10^{-2}$ or not inducing Cas1–Cas2 expression substantially lowers overall acquisition efficiency, no significant effect was similarly observed in the overall ratio of on-target to off-target integrations (Supplementary Fig. 3).

With the additional 695 off-target sites, we re-generated the off-target site sequence logos (Fig. 2e). When comparing the new logo with the original consensus sequence generated from the nine off-target sites identified by whole-genome sequencing (WGS) (Supplementary Fig. 1), the same palindromic motif within the repeat was present. However, the new logo also identified overrepresented bases near the putative leader–repeat junction (that is, the first base of the repeat and the first three bases of the leader)[20], and showed no conservation for nucleotides that were further upstream in the leader or in the 60 nt that were downstream of the repeat (Fig. 2e). This result is surprising considering that an IHF-binding motif located in the native leader sequence upstream of the first repeat was previously found to be essential for integrations into the canonical CRISPR array in vivo[6]. Thus, we performed DSA experiments and Spacer-seq on knockout strains that lacked the α-subunit or β -subunit of IHF (an obligate heterodimer) to determine what effect IHF has on off-target insertions. The IHF-knockout strains had substantially reduced integration efficiencies (~$10^3$-fold reduction) into the native CRISPR1 array (on-target), whereas the overall off-target integration rates only decreased ~10–20-fold, with ~95% of all spacer integrations going into off-target regions of the genome (Fig. 3a,b). We then compared the locations of the off-target sites found in the IHF-knockout strains with those of the wild-type (WT) strain, and observed similar distribution profiles (correlation coefficient of $r = 0.49 \pm 0.07$ for IHF knockouts versus WT, compared with $r = 0.43 \pm 0.09$ for WT versus WT experimental replicates), with the most frequent off-target sites being consistent across all strains (Fig. 3c,d). These results suggest that the presence of IHF increases the efficiency of both on-target and off-target integration activity, although off-target activity is less dependent on the presence of IHF overall. To better understand these results, we searched for potential IHF-binding motifs near the ten most prevalent off-target locations across all data sets and strains. We found that all of these sites had regions within 100 nt of the off-target repeat that shared at least 67% identity to the IHF-consensus-binding motif (Supplementary Fig. 4), which supports our results that IHF enhances the rates of both on-target and off-target integration events. Previous in vitro experiments have demonstrated that, even in the absence of IHF, efficient spacer integration into supercoiled plasmid-based CRISPR arrays can still occur[6]. Thus, we determined what effect the addition of a plasmid that included a canonical CRISPR array would have on the genomic contexts of the IHF-knockout strains during DSA. Spacer-seq results from these experiments demonstrated that the addition of a multicopy plasmid-based array reduces the frequency of off-target events into the genome from ~95% to ~35%, with ~60% of integrations now going into the plasmid-based array on-target, even in the absence of IHF (Supplementary Fig. 5). Meanwhile, the on-target integration frequency into the genomic array remained unchanged at ~5% of all Spacer-seq reads.

We next tested whether removing the only active CRISPR array in the BL21 genome (CRISPR1) might influence the distribution of off-target sites elsewhere in the genome. To answer that question, we created a CRISPR1-deletion strain by deleting the entire CRISPR1 array and leader region from the BL21-AI genome, and performed DSA followed by Spacer-seq. We then compared the locations of the off-target sites found in the CRISPR1-deletion strain to those of the WT and observed similar distribution profiles (correlation coefficient of $r = 0.53 \pm 0.12$ for the CRISPR1 deletion versus WT, compared with $r = 0.43 \pm 0.09$ for WT versus WT experimental replicates), with the most frequent off-target sites again being consistent across all strains (Fig. 3c,d).

Curiously, the inverted repeat of the off-target repeat consensus (Fig. 2e) is a perfect palindrome within bases C8 to G21 (CCNCGCGCGCGNGG), whereas the endogenous *E. coli* CRISPR repeats have two non-palindromic bases (nucleotides C14 and A15). The off-target perfect palindrome logo is similar to a logo generated from aligning all the repeats associated with the type I-E repeat class, as previously shown[21]. To test whether the perfect internal palindrome found in the off-target consensus is actually the most strongly preferred Cas1–Cas2 target site, we performed a DSA assay that compared the in vivo spacer acquisition efficiency of the endogenous repeat sequence to that of a mutant repeat representing the off-target consensus sequence (that is, containing the repeat mutations C14G and A15C), which we termed the

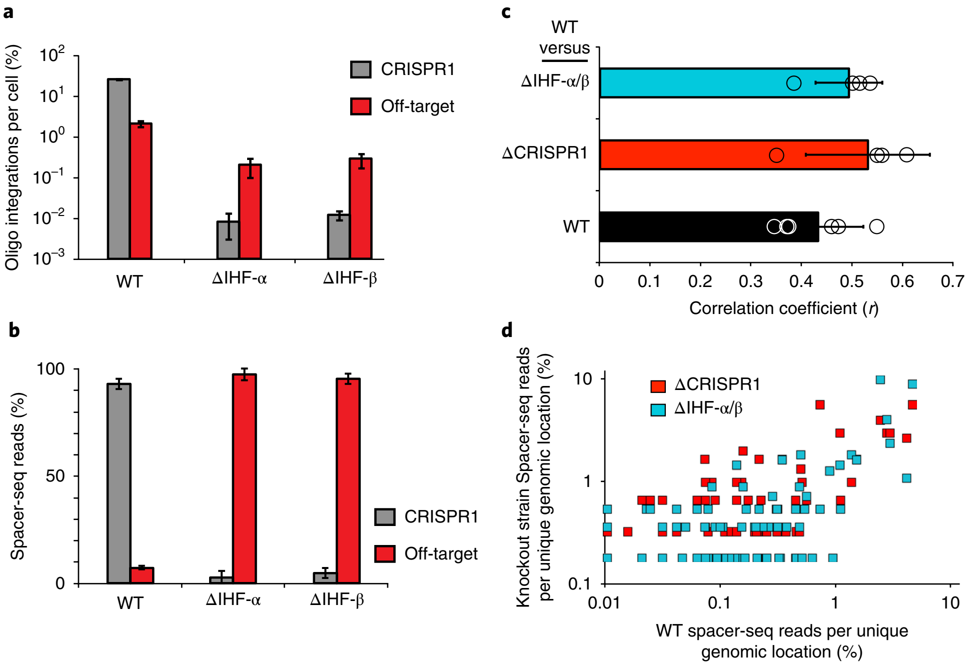

**Fig. 3 | Effects of genomic knockouts of IHF and the CRISPR1 locus on off-target spacer integration activity. a**, The percentage of oligo integrations into the CRISPR1 locus or the off-target sites normalized per cell (array) following DSA in the BL21-AI strain (WT), or the BL21-AI strain with either the IHF-α or the IHF-β subunits knocked out (ΔIHF-α and ΔIHF-β, respectively). Error bars represent mean ± s.d. **b**, The percentage of Spacer-seq reads aligned on-target to the CRISPR1 locus or to other regions in the genome (off-target) in the WT, ΔIHF-α, and ΔIHF-β strains. Error bars represent mean ± s.d. **c**, Pearson correlation coefficient (*r*) of the off-target site identities between the WT versus ΔIHF-α/β strain, WT versus CRISPR1 deletion (ΔCRISPR1) strain and WT versus WT replicates. Error bars represent mean ± s.d. Open circles represent individual replicate data points. **d**, The percentage of Spacer-seq reads aligned to unique off-target sites within the genome. The knockout strain percentages (*y* axis) of the ΔIHF-α/β strains and the ΔCRISPR1 strain are compared to those of WT (*x* axis). Each point represents a unique genomic site. For all panels, *n* = 4 (for WT) and *n* = 3 (for knockout strains) biological replicates.

'off-target consensus repeat' (OTCR). We found that the array containing the OTCR acquired nearly 50% more spacers than the native array (1.2 ± 0.1% and 0.85 ± 0.03% of all plasmid-based arrays were expanded, respectively) (Fig. 2f).

To see whether these results were specific to a plasmid-based array, we created a modified BL21-AI strain in which its native CRISPR1 locus was replaced with a minimal version of the array containing the OTCR. Engineering strains with enhanced spacer acquisition activity would also be useful in molecular recording applications[15,22]. To do this, we designed and integrated a synthetic CRISPR array containing the first 100 nt of the native CRISPR1 leader upstream of the OTCR sequence into the BL21-AI CRISPR1-deletion strain that we had previously constructed. However, in the DSA experiments quantifying oligo acquisition efficiencies, the OTCR strain actually displayed lower acquisition rates compared with those of the WT BL21-AI strain (Fig. 2g). This finding conflicts with the plasmid-based array results (Fig. 2f), suggesting that array activity is context dependent and that additional regions outside of the first repeat and leader might affect acquisition efficiency. For instance, in modifying the first repeat, we also deleted subsequent repeats. The presence of many repeats within an array may help to recruit Cas1–Cas2 localization to the CRISPR locus.

Canonical CRISPR leaders include promoter elements for the expression of CRISPR RNA (crRNA) transcripts, which are utilized by the Cas effector proteins for spacer-guided nuclease activity[23]. Most of the off-target spacer integrations we characterized occur within the protein-coding regions of non-essential genes, and therefore downstream of endogenous promoters. This is not surprising given the high density of genes in bacterial genomes. This observation suggests the possibility that these off-target integration products could be transcribed, dependent on the activity of proximal promoter elements. Thus, we asked whether we could detect the expression of off-target integration products within cellular transcripts. To do this, we performed Spacer-seq on complementary DNA (cDNA) derived from the total RNA isolated from cultures of BL21-AI cells following DSA. Sequencing results from these experiments confirmed the expression of off-target integration products, with the overall frequency of off-target reads within transcripts similar to the levels found in the genome (Supplementary Fig. 6a). These RNA Spacer-seq reads mapped to the most abundant cellular transcripts (Supplementary Fig. 6b), as further evidenced by enrichment for off-target sites within ribosomal operons (Supplementary Fig. 6a).

After confirming that off-target integration products retain the potential to be expressed, we wondered whether some of these transcripts could function as crRNA in defence. If so, it would imply that off-target spacer acquisition activity has the potential to augment immunity by the incidental genesis and expression of 'neo'-CRISPR arrays (NCAs). To test this, we selected ten off-target integration sites that we discovered by Spacer-seq (sites within the *araD*, *cysI*, *fic*, *hsdR*, *mnmC*, *phnP*, *potG* and *yfic* genes, in addition to a site within an unnamed hypothetical protein ('hyp.')) that share considerable homology to the native CRISPR repeat and cloned them into expression plasmids along with a spacer that matches the M13 bacteriophage genome (Fig. 4a,b and Supplementary Table 1). These plasmids were introduced into a strain of *E. coli* that expresses the full set of type I-E Cas genes required for adaptation and defence (BW40114)[24]. First, we used these strains to see whether any of the cloned plasmid-based NCAs could function in direct interference through a plasmid interference assay[25] by attempting to transform an additional plasmid that also contained the M13 spacer target sequence into cultures expressing the NCA and Cas proteins. If the NCAs are functional, they should reduce the efficiency of this

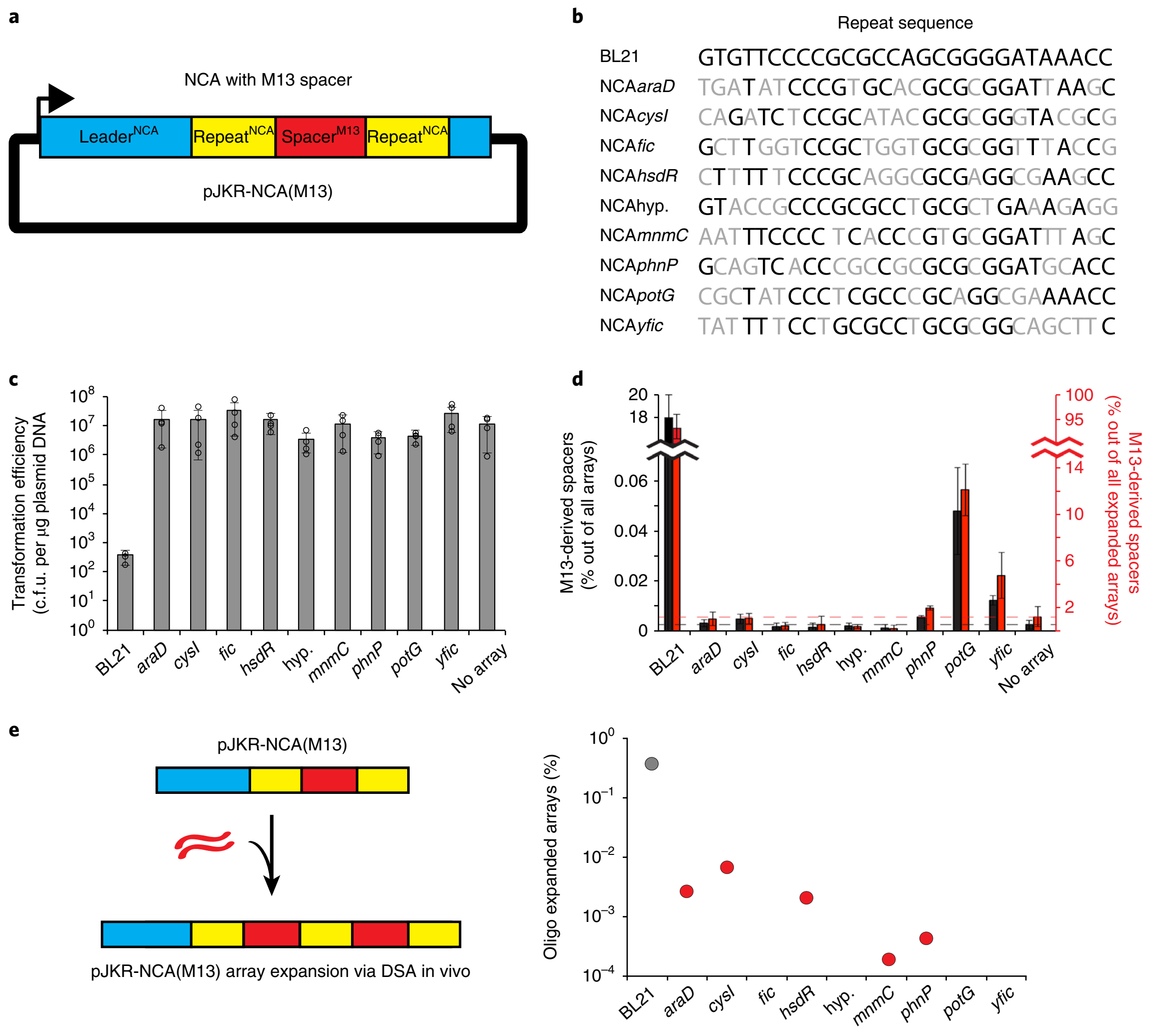

**Fig. 4 | Comparison of three different NCA sequences and their activity in target interference and primed acquisition. a**, Schematic of the plasmid-based NCAs that were used in the primed acquisition assays. The arrays contain an inducible promoter that drives the expression of 60 nt of the off-target leader (leader$^{NCA}$) along with a 33-nt spacer that matches the M13 bacteriophage genome (spacer$^{M13}$) that is flanked by the 28-nt off-target repeat sequences (repeat$^{NCA}$). **b**, Multiple sequence alignment of the NCA repeats aligned to the BL21 CRISPR repeat sequence. Residues conserved with the BL21 repeat are shown in black. **c**, Results of the plasmid interference assay with the strains harbouring the plasmids that encode either WT (BL21) or NCAs containing the M13 spacer, or a strain with no plasmid-based array. $n = 4$ biological replicates (circles). Error bars represent mean ± s.d. c.f.u., colony-forming unit. **d**, Results of the primed acquisition assay with the strains harbouring the plasmids that encode either WT (BL21) or NCAs containing strains. Black bars correspond to the percentage of total arrays containing new M13-derived spacers. The red bars correspond to the percentage of newly expanded arrays containing M13-derived spacers. $n = 3$ biological replicates. Error bars represent mean ± s.d. **e**, Comparison of plasmid-based NCA expansion frequencies following DSA. Expansion frequencies for each NCA were quantified by high-throughput sequencing of the plasmid-based arrays. Each point represents the percentage of expansions detected for each array. We did not detect any expansions for NCAs that do not display a point (*fic*, *potG* and *yfic*), indicating integration efficiencies of $<10^{-4}$%.

transformation relative to a negative-control plasmid that does not contain a matching protospacer target. The results of this interference assay are shown in Fig. 4c. Although a strain expressing the canonical BL21 array and M13 spacer reduced the transformation efficiency by more than four orders of magnitude compared to the negative control, none of the NCA strains demonstrated a significant effect on the transformation efficiency (Fig. 4c).

Although plasmid interference is the most direct test of a functional CRIPSR system, it is not the most sensitive. Recently, it was shown that a more sensitive in vivo test of crRNA function is a primed acquisition assay[25]. Briefly, 'priming' is the efficient acquisition of new spacers during a phage challenge of this system stimulated by a pre-existing spacer matching the phage genome that enhances the acquisition of additional phage spacers. Thus, the plasmid-based NCAs containing the M13 spacer should enhance the acquisition of phage-derived spacers during an M13 phage challenge if the NCAs express functional crRNA. The results of the primed acquisition assay are shown in Fig. 4d. Although the majority of the NCAs did not stimulate additional spacer acquisitions relative to a negative control that lacked a plasmid-based array, cells expressing NCA$^{potG}$ acquired ~16-fold more M13-derived spacers compared to background (0.048 ± 0.02% versus 0.003 ± 0.002%, respectively). In addition, the NCA$^{potG}$ strain had an increased bias for M13-derived spacers within its newly acquired spacer population compared to background (12.1 ± 2.2% versus 1.3 ± 0.9%). Although these frequencies are well below the rates observed for the native BL21 array strain, we have only tested a small fraction of the hundreds of possible NCA sequences, and therefore can envision additional off-target sequences with greater crRNA functionality. Even still, these results support a model in which an off-target integration event could lead to the expression of at least semi-functional crRNA.

A key feature of CRISPR–Cas immunity is the ability to store multiple spacers within a single locus. This is achieved through iterative integration events overtime into the same leader–repeat site, which is inherently preserved following integration and repeat duplication. To investigate whether NCA sites can also undergo multiple expansions beyond the original off-target event, we performed DSA on the strains containing the plasmid-based NCAs. Deep sequencing of the NCA loci following DSA revealed that five out of the nine NCAs could be expanded with an additional spacer, albeit at orders of magnitude less efficienct than the canonical array (Fig. 4e).

Having demonstrated in our model system that off-target spacer integration by Cas1–Cas2 occurs in vivo at CRISPR repeat-like sequences within the *E. coli* genome, we asked whether we could find evidence for natural off-target activity in other species using existing genomic databases. To do this, we searched the literature for bacterial and archaeal species that have well-annotated phylogeny, published indications of active CRISPR–Cas systems and available whole-genome sequences. In addition, we combed the CRISPRdb[26] online database for related species and strains that had dissimilar numbers of CRISPR loci. Our investigation yielded support for off-target activity or 'neo-CRISPR genesis' (Supplementary Fig. 7) within related strains of two different microbial species with active CRISPR–Cas systems: *Yersinia pestis* and *Sulfolobus islandicus*. The first we describe are those of *Y. pestis*.

Owing to its potential as a human pathogen, *Y. pestis* phylogeny has been heavily studied, with many strains of the species whole-genome sequenced[27]. One of the modern *Y. pestis* strains, CO92, is typically used as the reference strain[28]. All but one strain of *Y. pestis* have three active CRISPR loci (YPa, YPb and YPc), and only one of these loci is proximal to a set of Cas genes (YPa)[27] (Figs. 4b and 5a). The exception to this is the Angola strain, which only has the YPa CRISPR–Cas locus, and is considered an ancient strain in the *Y. pestis* lineage[28]. In place of the other two loci are single degenerate repeats and accompanying leader regions, with both loci lying within hypothetical protein-coding regions. We postulate that arrays YPb and YPc are the result of off-target integration events that became fixed in strains following the divergence from the ancient Angola strain through the process of neo-CRISPR genesis.

The second example of native off-target spacer integration we found was in three closely related strains of the hyperthermophilic archaeal species *S. islandicus*: LAL14/1, HVE10/3 and REY15A[29] (Fig. 5c). Although all ten strains within this species possess multiple active CRISPR–Cas systems[30], only these three strains contain a region with a 37-nt spacer flanked by 24-nt repeats following the end of a hypothetical ABC-transporter-related protein (Fig. 5d). The other seven genomes only contain a single copy of the repeat. Intriguingly, the repeat is the same size as and shares sequence homology with the other two confirmed CRISPR array repeat sequence types found within the species (Fig. 5e). The spacer length is also typical for these CRISPR types. Furthermore, a BLASTn search of the spacer sequence uncovered a partial match with a known *S. islandicus* plasmid (pLD8501) that is not present in these strains[31] (Fig. 5f). This is important because canonical CRISPR spacers often share homology to known phages and plasmids. Taking all of these observations together, we speculate that this unique genomic feature is the result of an off-target spacer integration event following the divergence of this strain lineage from the rest of the *S. islandicus* species.

## Discussion

Spacer integration into the leader proximal end of CRISPR loci is an essential phenomenon of CRISPR–Cas systems. However, whether spacer integrations occur outside of canonical CRISPRs and the potential biological consequences of this were both previously unknown. We found, using DSA, WGS and Spacer-seq, that off-target spacer integrations can occur at many unique sites throughout the *E. coli* genome and carried plasmids. Off-target spacer integrations are potentially deleterious events that could affect genome integrity[17]. Conversely, the process of evolution itself is predicated on chance sampling of beneficial mutations through lapses in genetic fidelity. It has been previously shown that spacer acquisition is optimized for integration into the leader proximal end of the array to achieve a robust immune response, as spacers at the trailing end of the array are poorly expressed[16]. We found this to be true in the case of *E. coli* BL21 CRISPR1 array expression (Supplementary Fig. 8). Thus, off-target spacer integration activity, although probably deleterious in most instances, has the potential to boost crRNA expression levels and increase spacer diversity. To extend the relevance of our findings beyond our experimental model system, we also uncovered several examples of putative off-target spacer integration activity in previously sequenced genomes within the *Y. pestis* and *S. islandicus* lineages, and term this phenomenon 'neo-CRISPR genesis'. As the number of whole-genome-sequenced microbial species increases, particularly within clades of closely related strains, we suspect that further instances of neo-CRISPR genesis will come to light.

## Methods

**DSA.** The process of DSA using electroporated oligos has been described[14]. Briefly, liquid cultures of *E. coli* BL21-AI cells (Thermo) harbouring a plasmid expressing Cas1 and Cas2 under the control of a T7-lac promoter (pWUR 1 + 2, which was a generous gift from U. Qimron) were started from plates and grown overnight in LB. In the morning, cultures were diluted 1:30 in 3 ml fresh LB containing L-arabinose (Sigma-Aldrich) at a final concentration of 0.2% (w/w) and 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG; Sigma-Aldrich), unless otherwise noted, and grown for an additional 2 h. Cells were then pelleted, re-suspended and washed in water three times to remove residual media. Cells were then re-suspended in 50 µl (per 1 ml of the 3 ml culture) of water containing the psAA33 forward and reverse oligo strands each at a concentration of 3.1 µM and electroporated with a Bio-Rad Gene Pulser set to 1.8 kV, 25 uF and 200 Ω. Immediately following electroporation, cells were re-suspended in fresh LB and allowed to recover overnight. In the morning, the cultures were pelleted and frozen at −20 °C until DNA extraction.

**Whole-genome preparation, sequencing and analysis.** The total DNA content of the cell pellets were extracted and purified with a QIAmp DNA Mini Kit (Qiagen) following the manufacturer's protocol for bacterial cultures. The isolated DNA was then sheared to ~500-bp fragments using a Covaris S2 ultrasonicator. DNA fragments were then prepped for sequencing using NEBNext Ultra DNA Library Prep Kit for Illumina (NEB) and sequenced on an Illumina MiSeq machine (MiSeq Reagent Kit V2, paired-end 2 × 250 read lengths). Sequencing data were analysed using the Geneious assembler (Biomatters) by aligning reads to the BL21 reference genome (GenBank accession number CP010816) allowing for up to 70-nt insertions, and manually searching for reads containing the psAA33 sequence or insertions of ~61 nt.

**Spacer-seq and analysis.** Using the sheared and adaptor-ligated DNA fragments previously prepared for WGS as input, PCR was performed using a forward primer that contained the NEBNext Adaptor sequence (5′) and a portion of the psAA33 sequence (3′), and a reverse primer that matched the NEBNext Adaptor sequence. Enriched fragments were then indexed and sequenced on an Illumina MiSeq machine. Sequencing data were then analysed using custom-written software (Python). Briefly, primer sequences were removed from the reads and filtered for sequences that contained a match to the remaining psAA33 sequence that was not included in the primer. Sequences satisfying these criteria were then mapped to the BL21 reference genome.

**RNA Spacer-seq and analysis.** The total RNA content of the cell pellets was extracted and purified with a RNeasy Mini Kit (Qiagen) according to the manufacturer's protocol for bacterial cultures. The purified RNA was then used to produce cDNA using the ProtoScript II First Strand cDNA Synthesis Kit (NEB), and was then made double-stranded with the Second Strand cDNA Synthesis protocol according to NEB. The double-stranded cDNA was finally sheared, adaptor ligated and subjected to the same protocol as the genomic DNA Spacer-seq process and analysis. To compare RNA Spacer-seq reads to the total transcript abundance, a traditional RNA-sequencing was also performed on the isolated total RNA.

**Plasmid interference assay.** The NCAs containing the M13 spacer were synthesized as gBlocks (IDT) and cloned by Gibson assembly into the pJKR-

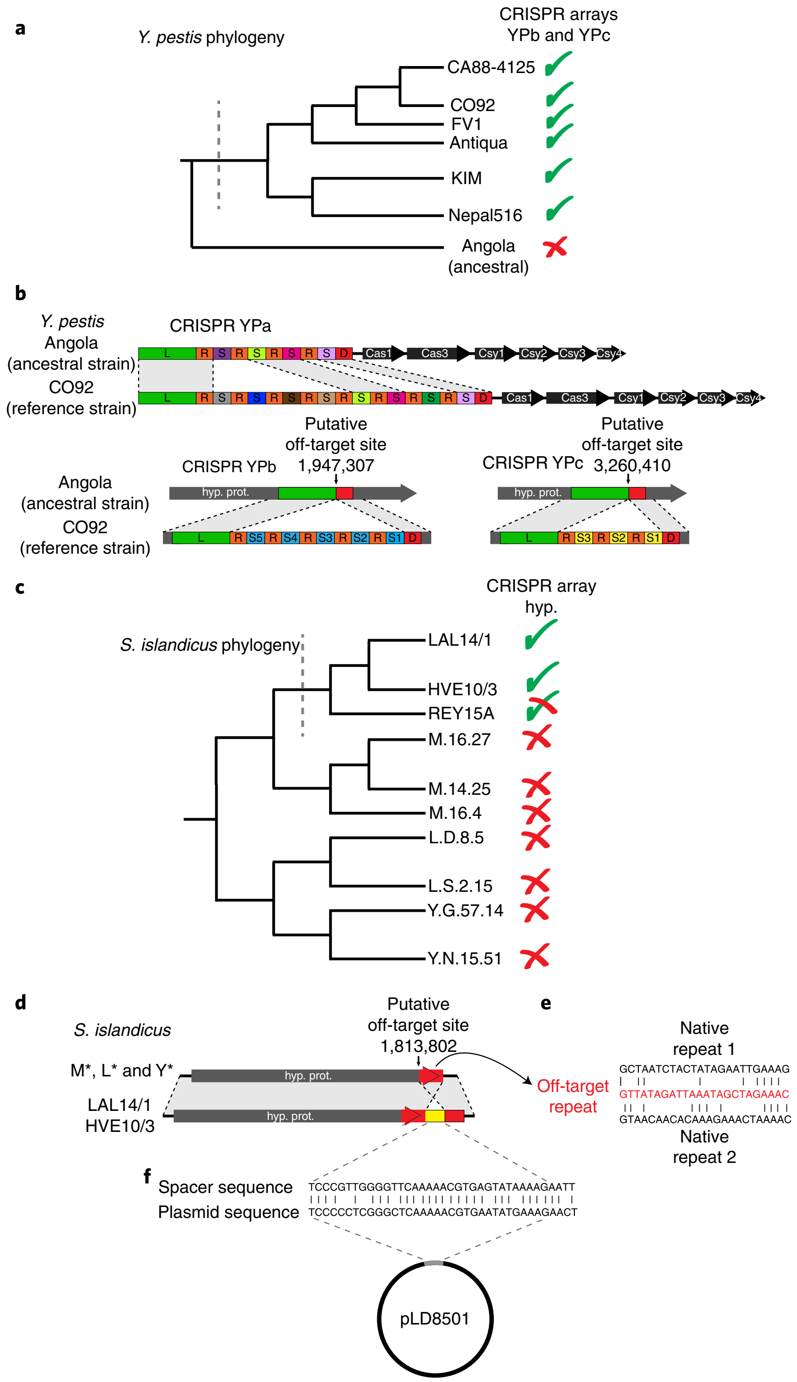

**Fig. 5 | Evidence for native off-target spacer integrations. a**, Diagram of *Y. pestis* phylogeny and the presence or absence of CRISPR arrays YPb and YPc, as denoted by a green check or red X, respectively. The dashed line demarks the branch between the absence or presence of the YPb and YPc arrays along the lineage. Figure adapted from ref. [27], PLoS. **b**, *Y. pestis* contains three canonical CRISPR arrays (YPa, YPb and YPc) and one set of type I-F Cas genes. Each array within the CO92 genome contains a leader (L), which shares 63% sequence identity across all three 200-nt leaders, and between 3 and 8 spacers (S) separated by 100% identical repeat sequences (R), with the exception of the terminal repeats, which are degenerate (D). The *Y. pestis* Angola strain, which is considered to be an ancestral strain of the species, contains only the Cas proximal array (YPa). At the Angola genomic locations homologous to the CO92 arrays, YPb and YPc, there are hypothetical protein-coding regions (hyp. prot.) that only contain the corresponding YPb and YPc array leader and terminal/degenerate repeat sequence. The putative spacer integration off-target site (between the pre-neo-CRISPR leader (green) and pre-neo-CRISPR 'degenerate' repeat (red)) within the ancestral Angola genome that eventually generated the YPb and YPc arrays of the descendant CO92 strain are demarcated. Grey regions within the dashed lines have 100% sequence homology. **c**, Diagram of *S. islandicus* phylogeny and the presence or absence of a putative off-target integration site within the genome at 1,813,802 (numbering based on the M.16.4 genome), as denoted by a green check or red X, respectively. The REY15A strain does not have a complete second repeat site. Figure adapted from ref. [29], Royal Society. **d**, Diagram comparing the genomic features of *S. islandicus* strains M*, L* and Y* with those of the LAL14/1 and HVE10/3 strains at the location of a putative off-target spacer integration event within the latter strains. The repeat and spacer regions are highlighted in red and yellow, respectively. **e**, The off-target repeat shares sequence homology with the other two canonical CRISPR repeat sequence types that are present within the species (the *S. islandicus* lineage contains three distinct CRISPR-Cas types: IA, IIIB-Cmr-α and IIIB-Cmr-β). **f**, Spacer sequence homology to a known *S. islandicus* plasmid (pLD8501).

H-tetR vector (replacing the *GFP* gene downstream of the pLtetO promoter). Sequence-verified plasmids were transformed into *E. coli* K12 BW40114. A plasmid containing the M13 spacer target site was constructed by cloning the 33-bp target sequence into the pFN19K plasmid via PCR. A plasmid interference assay has been previously described[24]. Briefly, overnight, cultures of strains containing the NCA plasmids were started from plates. In the morning, cultures were diluted in fresh LB containing the inducers arabinose, IPTG and anhydrotetracycline (Clontech), and grown for an additional 2 h. Cells were then washed three times in cold water, transformed with 50 ng pFN19K + M13 target plasmid and allowed to recover for ~1 h in LB at 37 °C before plating on LB plus Kan plates (absolute efficiency) and LB plus Carb plates (to normalize efficiencies).

**Primed acquisition assay and analysis.** A primed acquisition assay has been previously described[24]. Briefly, overnight, cultures of strains containing the NCA plasmids were started from plates. In the morning, cultures were diluted in fresh LB containing the inducers arabinose, IPTG and anhydrotetracycline, and grown for an additional 2 h. Cells were then diluted 1:10 into fresh LB (with inducers) and M13KE phage (NEB) at a concentration of $1\times10^9$ p.f.u. (plaque-forming units) per ml. Cultures were then grown overnight. In the morning, an aliquot of the sample was boiled and used as input for PCR that amplified the K12 CRISPR2 array locus. Amplicons were prepped with Illumina NEBNext Ultra DNA Library Prep Kit for Illumina (NEB) and sequenced on an Illumina MiSeq machine. Sequence data were analysed using custom-written software (Python). Briefly, the sequence of the first spacer within each array was extracted and blasted against a local database to quantify the number of spacers matching the M13 phage genome.

**Strain knockouts.** BL21-AI strains containing IHF-α and IHF-β knockouts were a generous gift of J. Doudna (Univ. California). The BL21-AI CRISPR1 array knockout strain and the OTCR strain were constructed by following the lambda red plus Cas9 gene-editing strategy[32].

**Life Sciences Reporting Summary.** Further information on experimental design is available in the Life Sciences Reporting Summary.

**Code availability.** The custom Python code used for analysis is available upon request.

**Data availability.** Spacer-seq Illumina sequencing data have been deposited to the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA; BioSample accession SAMN08134321). Additional data that support the findings of this study are available upon request.

Received: 27 October 2017; Accepted: 11 December 2017;
Published online: 29 January 2018

## References

1. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
2. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843–1845 (2008).
3. Wright, A. V., Nuñez, J. K. & Doudna, J. A. Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell* **164**, 29–44 (2016).
4. Sternberg, S. H., Richter, H., Charpentier, E. & Qimron, U. Adaptation in CRISPR–Cas systems. *Mol. Cell.* **61**, 797–808 (2016).
5. Brouns, S. J. J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
6. Nuñez, J. K., Bai, L., Harrington, L. B., Hinder, T. L. & Doudna, J. A. CRISPR immunological memory requires a host factor for specificity. *Mol. Cell.* **62**, 824–833 (2016).
7. Yosef, I., Goren, M. G. & Qimron, U. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* **40**, 5569–5576 (2012).
8. Levy, A. et al. CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **520**, 505–510 (2015).
9. Nuñez, J. K., Lee, A. S., Engelman, A. & Doudna, J. A. Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature* **519**, 193–198 (2015).
10. Nuñez, J. K. et al. Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nat. Struct. Mol. Biol.* **21**, 528–534 (2014).
11. Wright, A. V. et al. Structures of the CRISPR genome integration complex. *Science* **357**, 1113–1118 (2017).
12. Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A. & Severinov, K. High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in *E. coli*. *RNA Biol.* **10**, 716–725 (2013).
13. Shmakov, S. et al. Pervasive generation of oppositely oriented spacers during CRISPR adaptation. *Nucleic Acids Res.* **42**, 5907–5916 (2014).
14. Wang, J. et al. Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR–Cas systems. *Cell* **163**, 840–853 (2015).
15. Shipman, S. L., Nivala, J., Macklis, J. D. & Church, G. M. Molecular recordings by directed CRISPR spacer acquisition. *Science* **353**, aaf1175 (2016).
16. McGinn, J. & Marraffini, L. A. CRISPR–Cas systems optimize their immune response by specifying the site of spacer integration. *Mol. Cell.* **64**, 616–623 (2016).
17. Wright, A. V. & Doudna, J. A. Protecting genome integrity during CRISPR immune adaptation. *Nat. Struct. Mol. Biol.* **10**, 876–883 (2016).
18. Wang, R., Ming, L., Gong, L., Hu, S. & Xiang, H. DNA motifs determining the accuracy of repeat duplication during CRISPR adaptation in *Haloarcula hispanica*. *Nucleic Acids Res.* **44**, 4266–4277 (2016).
19. Goren, M. G. et al. Repeat size determination by two molecular rulers in the type I-E CRISPR array. *Cell. Rep.* **16**, 2811–2818 (2016).
20. Rollie, C., Schneider, S., Brinkmann, A. S., Bolt, E. L. & White, M. F. Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *eLife* **4**, e08716 (2015).
21. Kunin, V., Sorek, R. & Hugenholtz, P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* **8**, R61 (2007).
22. Shipman, S. L., Nivala, J., Macklis, J. D. & Church, G. M. CRISPR–Cas encoding of a digital movie into the genomes of a population of living bacter. *Nature* **547**, 345–349 (2017).
23. Marraffini, L. A. CRISPR–Cas immunity in prokaryotes. *Nature* **526**, 55–61 (2015).
24. Datsenko, K. A., Pougach, K., Tikhonov, A., Wanner, B. L., Severinov, K. & Semenova, E. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* **3**, 945 (2012).
25. Kuznedelov, K. et al. Altered stoichiometry *Escherichia coli* cascade complexes with shortened CRISPR RNA spacers are capable of interference and primed adaptation. *Nucleic Acids Res.* **44**, 10849–10861 (2016).
26. Grissa, I., Vergnaud, G. & Pourcel, C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinforma.* **8**, 172 (2007).
27. Barros, M. P. S. et al. Dynamics of CRISPR loci in microevolutionary process of *Yersinia pestis* strains. *PLoS ONE* **9**, e108353 (2014).
28. Eppinger, M. et al. Genome sequence of the deep-rooted *Yersina pestis* strain angola reveals new insights into the evolution and pangenome of the plague bacterium. *J. Bacteriol.* **192**, 1685–1699 (2010).
29. Jaubert, C. Genomics and genetics of *Sulfolobus islandicus* LAL14/1, a model hyperthermophilic archaeon. *Open. Biol.* **3**, 130010 (2013).
30. Gudbergsdottir, S. et al. Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol. Microbiol.* **79**, 35–49 (2011).
31. Reno, M. L., Held, N. L., Fields, C. J., Burke, P. V. & Whitaker, R. J. Biogeography of the *Sulfolobus islandicus* pan-genome. *Proc Natl. Acad. Sci. USA* **106**, 8065–8610 (2009).
32. Jiang, Y. et al. Multigene editing in *Escherichia coli* genome via the CRISPR–Cas9 system. *Appl. Environ. Microbiol.* **81**, 2506–2514 (2015).

## Acknowledgements

The authors thank Ekaterina Semonova and Konstantin Severinov (Rutgers) for generously providing strain BW40114, and J. Doudna and L. Harrington (UCB) for generously providing BL21-AI IHF-knockout strains. The project was supported by grants from the National Human Genome Research Institute (5R01MH103910), the National Human Genome Research Institute (5RM1HG008525) and the Simons Foundation Autism Research Initiative (368485) to G.M.C.

## Author contributions

J.N. and S.L.S. conceived the study. J.N. designed the work, performed the experiments, analysed the data and wrote the manuscript with input from S.L.S. and G.M.C. S.L.S. and G.M.C. discussed the results and commented on the manuscript.

## Competing interests

J.N., S.L.S. and G.M.C. are inventors on a provisional patent (62/490,901) filed by the President and Fellows of Harvard College that covers the work in this manuscript. A complete account of the financial interests of G.M.C. is listed at: http://arep.med.harvard.edu/gmc/tech.html.

## Additional information

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41564-017-0097-z.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Correspondence and requests for materials** should be addressed to G.M.C.

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

nature research

Corresponding author(s): George M. Church, Jeff Nivala

□ Initial submission □ Revised version ☒ Final submission

# Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

## Experimental design

### 1. Sample size

Describe how sample size was determined.

No sample size calculation was performed, but a sufficient number of replicates were performed to support statistical differences, if any, observed between samples.

### 2. Data exclusions

Describe any data exclusions.

No data was excluded.

### 3. Replication

Describe whether the experimental findings were reliably reproduced.

Experimental findings were reliably reproducible, and were replicated as described in Figure captions.

### 4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

No randomization was included in this study.

### 5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

No blinding was performed.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

### 6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

| n/a | Confirmed | |
|---|---|---|
| □ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
| □ | ☒ | A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| □ | ☒ | A statement indicating how many times each experiment was replicated |
| □ | ☒ | The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section) |
| ☒ | □ | A description of any assumptions or corrections, such as an adjustment for multiple comparisons |
| □ | ☒ | The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted |
| □ | ☒ | A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range) |
| □ | ☒ | Clearly defined error bars |

*See the web collection on statistics for biologists for further resources and guidance.*

## Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

Geneious sequence alignment software (v 5.4.4) and custom python scripts were used for analysis of sequencing data. Code is available upon request.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

## Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

Materials and sequences, such as strains and plasmids, are available upon request.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used in this study.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No eukaryotic cell lines were used in this study.

b. Describe the method of cell line authentication used.

No eukaryotic cell lines were used in this study.

c. Report whether the cell lines were tested for mycoplasma contamination.

No eukaryotic cell lines were used in this study.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

No eukaryotic cell lines were used in this study.

## Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

No animals were used in this study.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

No humans were used in this study.