HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2017 September 29.

Published in final edited form as:

Nature. 2017 April 06; 544(7648): 101–104. doi:10.1038/nature21719.

CRISPR-Cas systems exploit viral DNA injection to establish and maintain adaptive immunity

Joshua W. Modell, **Wenyan Jiang**, and **Luciano A. Marraffini**

Laboratory of Bacteriology, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA

Abstract

CRISPR-Cas systems provide protection against viral[1] and plasmid[2] infection by capturing short DNA sequences from these invaders and integrating them into the CRISPR locus of the prokaryotic host[1]. These sequences, known as spacers, are transcribed into short RNA guides[3–5] that specify the cleavage site of Cas nucleases in the genome of the invader[6–8]. When spacer sequences are acquired during viral infection is not known. To investigate this, we followed spacer acquisition in *Staphylococcus aureus* cells harboring a type II CRISPR-Cas9 system after infection with the staphylococcal bacteriophage ϕ12. We found that new spacers are acquired immediately following infection preferentially from the *cos* site, the viral free DNA end that is first injected into the cell. Analysis of spacer acquisition after infection with mutant phages demonstrated that most spacers are acquired during DNA injection, but not during other stages of the viral cycle that produce free DNA ends, such as DNA replication or packaging. Finally, we showed that spacers acquired from early-injected genomic regions, which direct Cas9 cleavage of the viral DNA immediately after infection, provide better immunity than spacers acquired from late-injected regions. Our results reveal that CRISPR-Cas systems exploit the phage life cycle to generate a pattern of spacer acquisition that ensures the success of the CRISPR immune response.

The acquisition of new spacer sequences from foreign DNA elements is hallmark of the CRISPR-Cas immune response[1]. The molecular mechanisms of this process, also known as CRISPR adaptation, have mostly been studied in the type I CRISPR-Cas system of *Escherichia coli*, one of the six different CRISPR-Cas types[9]. Recent work in this bacterium examined the mechanism of CRISPR-Cas autoimmunity; i.e. the acquisition of spacer sequences from the host genome and resident plasmids[10]. In this scenario, new spacers are

Correspondence to marraffini@rockefeller.edu.

Supplementary information

Extended Data Figures S1–10.
Extended Data Table 1.
Supplementary Data Files 1–2.
Supplementary sequences file.

Author Contributions. JWM and LAM conceived the study and designed experiments. JWM and WJ designed the spacer library construction method. WJ performed the CRISPR immunization simulation assay. All other work was executed by JWM. LAM and JWM wrote the paper with the help of WJ.

The authors have no conflicting financial interests.

acquired primarily from the chromosomal terminus, due to the presence of double-stranded DNA breaks (DSBs) that occur during replisome stalling. Adaptation is limited by chromosomal *chi* sequences and strongly depends on RecBCD, a complex required for the repair of genomic breaks in Gram-negative bacteria[11] that degrades DNA starting at the DSB and stopping at the *chi* site. It is believed that this degradation generates the DNA substrates used as new spacers that are incorporated into the CRISPR array by the Cas1-Cas2 integrase complex[10]. However, it remains unclear how and when new spacers are acquired from phages and other common prokaryotic invaders during the CRISPR-Cas immune response.

Here we studied spacer acquisition in the Gram-positive bacterium *S. aureus* RN4220, which lacks an endogenous CRISPR system, harboring a plasmid engineered to carry the *Streptococcus pyogenes* type II-A CRISPR-Cas locus (Extended Data Fig. 1a), an experimental system that was previously developed in our lab[12]. Cas9 is the crRNA-guided nuclease of this system[13,14], which requires the presence of a protospacer adjacent motif (PAM) immediately downstream of the target with the sequence NGG[13,15]. To avoid the potentially confusing effects of primed adaptation, a form of more frequent spacer acquisition that relies on the presence of pre-existing spacers with partial matches to the invading genome[16], we removed all spacers from the type II-A CRISPR locus and left only a single repeat sequence. We also used a *cas9* allele, hyper-*cas9* or h*cas9*, that enhances spacer acquisition by two orders of magnitude[17]. Using this system and a PCR amplification technique that enriches for newly incorporated spacers (Extended Data Fig. 1a), we were able to detect rare spacer acquisition events occurring minutes after phage infection. We first looked at the autoimmunity pattern of chromosomal spacer acquisition during exponential growth. The PCR products of adapted CRISPR loci were subjected to next generation sequencing and the reads per million (RPM) for each new spacer sequence was calculated (full data for all next generation sequencing experiments in this manuscript is provided in Supplementary Data File 1; a summary of this data in table format is provided in s). Plotting the RPM values for spacers of chromosomal origin along the *S. aureus* genome revealed a strong adaptation hotspot surrounding the *dif* site, which marks the terminus of the circular bacterial chromosome (Fig. 1a and Extended Data Fig. 1b–e). The peak is limited by the first staphylococcal *chi* sequence (5′-GAAGCGG-3′)[18] upstream from the *dif* site on each DNA strand (Extended Data Fig. 1b–e). In addition, the introduction of an I-*sceI* site[19] resulted in an additional adaptation hotspot when the I-SceI endonuclease was expressed, again limited by *chi* sites (Fig. 1a and Extended Data Fig. 1f–g). These results demonstrate that both type I and type II CRISPR-Cas systems can utilize DSBs as the source of new spacers for CRISPR adaptation.

Are DSBs and/or free DNA ends also used for spacer acquisition during the CRISPR-Cas immune response against viral infection? If so, when during the life cycle of the invading virus are these DSBs and free DNA ends generated? Free DNA ends can be found in several stages of the infectious cycle of lambda-like dsDNA (double-stranded DNA) bacteriophages, for example following DNA injection, through accidental DNA breaks during theta replication, following the transition to rolling circle replication and throughout DNA packaging[20]. To address this fundamental question, we analyzed the incorporation of new spacers shortly after infection with the lambda-like ϕ12γ3 bacteriophage (Extended Data

Fig. 2a–b), a lytic derivative of the temperate dsDNA staphylococcal phage ϕ12[21]. We first estimated the onset of ϕ12γ3 lysis at approximately 45 minutes (Extended Data Fig. 2c); therefore we chose to analyze spacer acquisition at 30 minutes post-infection, when most phages are still going through the first lytic cycle. In this way we avoided the selection and counter-selection of spacer sequences that produce more or less efficient crRNAs, respectively, that occurs during multiple rounds of infection, and which could skew the original distribution of newly acquired spacers. In contrast to the acquisition of chromosomal spacer sequences (where the CRISPR adaptation hotspot was flanked by two symmetrically opposed *chi* sites), the majority of newly acquired viral spacers mapped to a 13 kb region (the ϕ12 genome encompasses 45 kb) flanked by the *cos* site and its first upstream *chi* site, both in experiments with a multiplicity of infection (MOI) of 10 (Fig. 1b) and 1 (Extended Data Fig. 3a). The *cos* site marks the viral free DNA ends that are generated by staggered cleavage of the phage DNA during packaging, which serve as cohesive ends for re-circularization of the linear phage genome following DNA injection[22] (Extended Data Fig. 2d–e). The *cos*-adjacent pattern of spacer acquisition was not introduced by the enrichment primers or the use of the h*cas9* allele (Extended Data Fig. 4), was observed on both DNA strands upstream from the *cos* site (Extended Data Fig. 5a–c) and was not simply a reflection of the PAM sequence distribution on the phage genome (Extended Data Fig. 5d–e). Efficient spacer acquisition depended on the AddAB repair machinery, the Gram-positive functional paralog of RecBCD[11], since *S. aureus* hosts harboring a mutation in the nuclease active site of the AddA subunit (*addA*^{n}) led to a significant reduction of spacer acquisition, with a small peak clustered closely to the *cos* site and not reaching the first *chi* site (Fig. 1b and Extended Data Fig. 3b–c). In addition, the introduction of two extra *chi* sites ~2.5 kb upstream of the ϕ12γ3 *cos* site created a new limit for the spacer acquisition hotspot, with a peak occurring immediately adjacent to the *cos* site (Extended Data Fig. 3d). Altogether these results demonstrate that CRISPR adaptation against ϕ12γ3 exploits the presence of a natural and invariable free DNA end at the *cos* site of the virus to generate new spacers.

The results of figure 1b suggest that new spacers could be acquired soon after injection of the viral DNA. It is also possible that spacer acquisition occurs during *cos* site cleavage at the initiation of genome packaging by terminase enzymes[23] and/or, similarly to the case of chromosomal spacer acquisition near the genome terminus, DSBs generated during viral DNA replication could be substrates for CRISPR adaptation (Extended Data Fig. 6a). To distinguish between these possibilities we measured spacer acquisition in ϕ12γ3 mutants impaired at different stages of the lytic cycle (Extended Data Fig. 6a–d). *polA* mutant phages lack DNA polymerase A and cannot replicate the injected genome (Fig. 2a), cannot continue the lytic cycle after injection and produce neither intracellular (Fig. 2b) nor extracellular viral particles (Fig. 2c). Therefore these mutants provide a measure of the level of spacer acquisition that occurs following injection and before replication (Extended Data Fig. 6a–b). *terS* mutants lack the small subunit of the terminase complex and are incapable of *cos* site cleavage by this complex during DNA packaging[23]. After DNA injection they can replicate their genome (Fig. 2a) but cannot form functional intracellular (Fig. 2b) nor extracellular viral particles (Fig. 2c). Thus spacer acquisition in the *terS* mutants can occur during DNA injection and/or replication but before DNA packaging (Extended Data Fig. 6a–

b). Finally, phages with a deletion encompassing the genes encoding the holin and lysin enzymes (*ho/ly*) are impaired at host lysis[24], but they can replicate (Fig. 2a) and package their DNA, producing intracellular (Fig. 2b) but not extracellular (Fig. 2c) viral particles. In this mutant, spacer acquisition can occur throughout DNA injection, replication and/or packaging, but not during a second infectious cycle (Extended Data Fig. 6a–b). First we compared adaptation levels between *terS* and *ho/ly* mutants at 0 and 60 minutes post-infection, when DNA packaging is mostly completed in the *ho/ly* mutant (Fig. 2b). We did not detect significant differences in spacer incorporation between these two infections, both quantitatively (Fig. 2d) and qualitatively (Extended Data Fig. 6e), demonstrating that the free DNA ends generated at the *cos* site during phage DNA packaging do not contribute substantially to the generation of new spacers. To measure how much CRISPR adaptation occurs during phage DNA replication we compared the levels of spacer acquisition between *polA* and *ho/ly* mutants at 0 and 30 minutes post-infection, when DNA levels in the *ho/ly* mutant are ~30-fold higher than in the *polA* mutant (Fig. 2c). Both mutants allowed similar spacer incorporation into the CRISPR array both quantitatively (Fig. 2e) and qualitatively (Extended Data Fig. 6f), demonstrating that replication is not a major stage in the viral cycle for the acquisition of new spacers. These results suggest that the great majority of new spacers are acquired early in the phage life-cycle, during or shortly after injection of the viral dsDNA ends. In addition, because the phage mutants cannot propagate, these results show that the pattern of spacer acquisition is not a result of selection generated by multiple infection cycles. One prediction of this conclusion is that adaptation rates should scale with the MOI, which directly determines how many phage dsDNA ends are injected. Indeed, we found that adaptation rates are linearly correlated with MOI (Extended Data Fig 6g). A second prediction is that the region of the phage genome that is first injected into the host would be the source of most of the new spacers. To test this we inverted the *cos* site (116 base pairs) of ϕ12γ3, generating ϕ12γ3$^{cos\text{-flip}}$ (Fig. 2f), a phage with an inverted order of entry. In contrast to the pattern of CRISPR adaptation obtained following challenge with wild type phages, during infection with the mutant phage a minority of spacers were acquired from the region upstream of the *cos* site, with the highest spacer density mapping downstream of the flipped *cos* site at both MOI of 10 (Fig. 2g) and 1 (Extended Data Fig. 3a). Acquisition peaks in ϕ12γ3$^{cos\text{-flip}}$ were lower but covered a larger area compared to those in ϕ12γ3, likely because there are no *chi* sites in the reverse strand of the ϕ12 genome to limit and concentrate adaptation activity. In addition, spacer acquisition from ϕ12γ3$^{cos\text{-flip}}$ in the *addA*^{n} mutant displayed a major peak within 5 kb of the *cos* site, a mirror image pattern of that observed for ϕ12γ3 (Extended Data Fig. 3e). Altogether, the data in figure 2 demonstrate that the acquisition of spacer sequences by the CRISPR-Cas adaptation machinery occurs primarily from the viral dsDNA ends first introduced during the injection of the ϕ12γ3 genome.

This pattern of spacer acquisition presents a potential advantage to the immunized host population upon subsequent phage encounters, since the *cos*-adjacent spacers will direct the Cas9 nuclease to cleave the next invading virus shortly after its injection. To test this we generated a library of spacers evenly distributed throughout the ϕ12γ3 genome (Figs. 3a and Extended Data Figures 7a–f). Cells harboring this library were infected with ϕ12γ3 or ϕ12γ3$^{cos\text{-flip}}$ phages in conditions where the incorporation of additional spacers is inhibited

(see Extended Data Figure 7c–d). Staphylococci surviving infection overnight were collected and their spacer content was assessed by next-generation sequencing to determine if the selective pressure of the phage enriched for spacers targeting any particular region of ϕ12γ3. Indeed, spacers targeting the region immediately upstream of the *cos* site were enriched after infection with wild type ϕ12γ3 (Fig. 3a and Extended Data Figure 7g); i.e. the region that is injected first. The opposite pattern was observed following ϕ12γ3$^{cos\text{-flip}}$ infection, with the enrichment of spacers targeting the region downstream of the flipped *cos* site (Fig. 3a and Extended Data Figure 7g). This pattern of selection, opposite for spacers targeting different sides of the *cos* site, is more evident when spacer enrichment is plotted as the ratio of ϕ12γ3/ϕ12γ3$^{cos\text{-flip}}$ RPMs (Fig. 3b and Extended Data Figure 7h). These results suggest that during the targeting phase of CRISPR-Cas immunity, Cas9 cleavage of early-injected phage sequences provides better phage clearance than the cleavage of late-injected sequences. To corroborate this, we used a type II CRISPR-Cas system lacking the adaptation-specific genes *cas1*, *cas2* and *csn2* (strictly required for spacer acquisition[12]), and expressing Cas9 programmed with spacers targeting viral sequences immediately (within 1 kb) upstream or downstream from the *cos* site (Fig. 3c). During infection with wild type ϕ12γ3 the three upstream spacers target the first region to be injected; during infection with ϕ12γ3$^{cos\text{-flip}}$ the order of targeting is inverted (Fig. 3c). In host populations where all bacteria carry one of these six spacer sequences, CRISPR-Cas immunity allows cell survival upon infection with either of the two phages at low MOI (Extended Data Fig. 8). However, the incorporation of new spacers during CRISPR immunization is a very rare event[1,12]. To mimic this situation, we mixed bacteria programmed with either the three upstream or downstream spacers with staphylococci lacking a CRISPR-Cas plasmid at a ratio of 1:10,000[25], resulting in extremely high MOIs following the lysis of non-CRISPR cells. We then infected the mixed cultures with either ϕ12γ3 or ϕ12γ3$^{cos\text{-flip}}$. The population harboring the three upstream targeting spacers was able to recover and survive ϕ12γ3 but not ϕ12γ3$^{cos\text{-flip}}$ infection (Fig. 3d). In contrast, the culture carrying the three downstream targeting spacers recovered and survived ϕ12γ3$^{cos\text{-flip}}$ but not ϕ12γ3 infection (Fig. 3e). These data confirm that target location can strongly influence the efficiency of Cas9 targeting, especially immediately following CRISPR adaptation, when only a small fraction of cells acquire a new spacer. Altogether, our results demonstrate that spacer acquisition from early-injected viral sequences is an effective strategy for establishing a robust CRISPR-Cas immune response.

Our study showed that type II CRISPR-Cas systems use free DNA ends as an entry point for the acquisition of new spacers. During infection by dsDNA phages, the linear genomic end is injected first. We demonstrated that this initial step in viral infection is exploited by the CRISPR-Cas immune response both to acquire new anti-viral spacers and to direct Cas nucleases to the invading target rapidly after infection begins, immediately halting the lytic cycle (Extended Data Fig. 9). At the same time the preference for free DNA ends could provide a mechanism to prioritize the acquisition of spacers from the invading virus, which produces free DNA ends as an obligate life-cycle step, over acquisition from chromosomal DNA (to avoid CRISPR “autoimmunity”). Our studies used the staphylococcal phage ϕ12, whose DNA packaging mechanism ensures that every viral particle injects the *cos* site first (Extended Data Fig. 10a–b). Analysis of the pattern of spacer acquisition after infection with

the *pac* staphylococcal phage ϕNM4γ4, which lacks a defined genomic entry site, also shows a mode of spacer incorporation consistent with CRISPR adaptation during phage DNA injection (Extended Data Fig. 10c–d), suggesting that our findings apply to both *cos* and *pac* dsDNA phages. We propose that, although not strictly required for spacer acquisition (Extended Data Fig. 3e), DNA degradation by AddAB of the injected viral genome (which starts at the *cos* site and ends at the first *chi* site) generates additional free DNA ends that are used as substrates for the Cas1-Cas2-Csn2-Cas9 spacer acquisition complex[12] to incorporate new viral sequences in the CRISPR array (Extended Data Fig. 9). We did not detect significant spacer acquisition in the opposite direction of the *cos* site; i.e. from the *cos* site DNA end that enters the cell last, a result that suggests that spacer acquisition occurs during DNA ejection. Interestingly, many dsDNA phages have evolved to protect their ends from this degradation by producing proteins that cap the injected dsDNA end or that inhibit RecBCD/AddAB[26]. It is tempting to speculate that the low frequency of spacer acquisition is related to these protection mechanisms, which would limit the generation of substrates for the Cas1-Cas2-Csn2-Cas9 complex. In this case, these mechanisms would provide a double protection for the phage, from RecBCD/AddAB degradation and CRISPR adaptation. Recent studies have shown that the injection of the phage lambda genome takes on average five minutes and can last up to twenty minutes[27]; this provides a window of opportunity for spacer acquisition before DNA injection is complete. It also provides a window for the cleavage of the viral DNA before it is completely ejected, something that we found to be important for an effective type II CRISPR-Cas immune response. Finally, CRISPR-Cas systems also acquire spacers from conjugative plasmids[2] most of which transfer to neighboring cells as ssDNA. Since ssDNA is not a good substrate for RecBCD[28], it is improbable that spacer acquisition happens during plasmid transfer. Most likely it occurs during plasmid replication, as reported previously[10]. In conclusion, our study highlights the intimate relationship between both stages of CRISPR-Cas immunity, immunization and targeting, and reveals how these immune systems have exploited the biology of phage infection to provide robust protection against prokaryotic predators.

Methods

Bacterial strains and growth conditions

Cultivation of *S. aureus* RN4220[29] and *E. coli* strains were carried out in brain-heart infusion (BHI) or LB liquid media (BD) respectively at 37°C, 30°C for maintenance of pE194-derived temperature sensitive plasmids or 28°C for maintenance of pIMAY-based plasmids. *S. aureus* media were supplemented with chloramphenicol at 10 μg/ml, erythromycin at 10 μg/ml, kanamycin at 25 μg/ml or spectinomycin at 250 μg/ml. *E. coli* media were supplemented with chloramphenicol at 25 μg/ml, ampicillin at 100 μg/ml, kanamycin at 50 μg/ml or spectinomycin at 50 μg/ml for plasmid maintenance. Media was supplemented with 1 mM IPTG to express genes from the inducible promoters P_{spac} and $P_{spac(h)}$ and with 2 μg/ml mitomycin C (A&G Scientific) to induce prophage excision and propagation.

PCR amplification of expanded CRISPR loci

CRISPR plasmids were harvested from *S. aureus* cells with a modified QIAprep Spin Miniprep Kit protocol: Bacterial cell pellets were resuspended in 250 μl of P1 buffer supplemented with 107 μg/ml lysostaphin (AMBI Products) and incubated at 37°C for 10 minutes followed by the standard QIAprep protocol. 150 ng (log phase) or 50 ng (overnight) of plasmid were used as input for the enrichment PCR of the CRISPR locus using Phusion DNA Polymerase (Thermo) with the following primer mix: 3 parts JW8 and 1 part each JW3, JW4 and JW5 (Extended Data Table 1). Variants of the primer JW8 with 3–8 bp bar codes at the 5′ end were used to distinguish experiments from each other during multiplexed high-throughput sequencing. PCRs with conventional primers were performed similarly using primers JW1131 and L401 with 5′ bar codes for multiplexing. Amplicons were gel or PCR purified and subjected to Illumina high-throughput sequencing with either the MiSeq, HiSeq or NextSeq platforms.

Deep sequencing data analysis

Spacers were aligned to chromosomal, plasmid and phage genomes using the Burrows-Wheeler Aligner[30]. Perfectly aligned spacers were assigned to the position in the reference genome at the 5′ end of the alignment, and the number of spacers occurring at each genomic position was aggregated. Two layers of normalization were next applied. First, spacer counts were multiplied by a factor which accounted for the base at that spacer's 3′-end. The enrichment PCR leverages the clamping of the reverse primers' 3-end (T, A or G) with the 3′ ends of newly acquired spacers[12]. A reverse primer ending in C is omitted from the primer mix in order to avoid priming against the single repeat in non-adapted plasmids. Because of the increased G-C bond strength relative to A-T, spacers ending in C are overrepresented compared to spacers ending in A or T. Spacers ending G, as well as non-adapted loci, are observed at low frequencies despite the absence of a perfectly matching reverse primer, likely because of 3′ trimming during primer synthesis or PCR. We therefore examined the observed frequencies of spacers ending in A, T, C or G following enrichment PCR and multiplied each by a factor that normalized the frequencies to the expected values of 22.8%, 41.9%, 18.1% and 17.2%. These values reflect the observed frequencies for phage- or chromosomally-derived spacers following amplification with non-enrichment primers (Extended Data Fig. S6a–b). When using conventional primers, the spacer 3′ base normalization step was skipped. Following the further aggregation of read counts into 1kb (phage) or 10kb (*S. aureus* chromosome) bins, we multiplied each bin by a factor accounting for the number of GG sequences within that bin, and on a given DNA strand where applicable. This is based on the observation that adaptation occurs primarily at sites immediately upstream from a 5′-NGG-3′ sequence, known as the PAM[12]. Finally, reads per million (RPM) were calculated either as RPM_{tot}, RPM_{spc}, $RPM_{\phi 12}$ or RPM_{chr}. RPM_{tot} corresponds to phage reads within a bin per million total CRISPR loci (including adapted and unadapted sites), while RPM_{spc} corresponds to phage reads per million total spacers (from chromosomal, plasmid and phage sequences). These are measures of absolute spacer acquisition rates where RPM_{tot} is used for comparisons across different host cell genetic backgrounds and RPM_{spc} for comparisons within the same genetic background. $RPM_{\phi 12}$ refers to phage reads within a bin per million total phage reads. RPM_{chr} refers to

chromosomal reads within a bin per million total chromosomal reads. $RPM_{\phi 12}$ and RPM_{chr} highlight the distribution of acquired spacers within a given genome.

Phage construction

We used a variation of a method previously described[31]. ϕ12γ3 was isolated as a spontaneous escaper plaque following an infection of ϕ12 on a soft agar lawn of *S. aureus* RN4220 cells carrying plasmid pJW479, which encodes a type III CRISPR-Cas system targeting the ϕ12 cI-like repressor. PCR of the cI locus confirmed a 1347 bp deletion spanning the complete cI gene and the first 616 base pairs of the upstream gene (RNaseT family protein) as indicated in Extended Data Fig. 2a. ϕ12γ3$^{cos\text{-flip}}$ was created by propagating ϕ12γ3 in liquid culture in cells harboring plasmid pJW481 which contains the flipped cos site and surrounding upstream and downstream homology regions for recombination with the phage genome. Recombinant ϕ12γ3$^{cos\text{-flip}}$ plaques were recovered on a soft agar lawn of RN4220 cells carrying plasmid pJW482, which encodes a type III CRISPR-Cas system targeting a sequence present only in ϕ12γ3 and not in ϕ12γ3$^{cos\text{-flip}}$. ϕ12γ3$^{chi\text{-extra}}$ was created in a similar fashion to ϕ12γ3$^{cos\text{-flip}}$ with pJW484 and pJW483 substituted for the recombination and targeting plasmids respectively.

Strain construction

The *addA*^{n} (D1159A) nuclease domain mutation was introduced into the *S. aureus* RN4220 genome (generating strain JW418) using the pIMAY allelic replacement system[32] with plasmid pJW233. An I-*sceI* recognition site immediately adjacent to a kanamycin-resistance gene (for selection purposes) was introduced into the *S. aureus* RN4220 genome at the *srtA* locus (generating strain JW263) using the pMAD allelic replacement system[33] with plasmid pJW480. Mutations in ϕ12 were introduced into strains of *S. aureus* RN4220 harboring the ϕ12 prophage using the pWJ244/pWJ327 allelic replacement system developed in our lab by Wenyan Jiang. A pWJ244-derived suicide vector containing the *ho/ly* mutation was introduced by electroporation into *S. aureus* RN4220::ϕ12 lysogens. Co-integrants (containing the plasmid integrated into the chromosome via recombination) were isolated on BHI agar plates supplemented with chloramphenicol grown at 37°C. pWJ327-derived temperature sensitive plasmids containing the *polA* and *terS* mutations were introduced by electroporation into *S. aureus* RN4220::ϕ12 lysogens and colonies were grown on plates supplemented with chloramphenicol at 28°C. Co-integrants were isolated by re-streaking colonies from 28°C plates directly onto 37°C plates. Co-integrants of pWJ244 or pWJ327 were re-struck onto a second chloramphenicol plate and grown at 37°C to remove non-transformed, background cells, and their genotype confirmed by PCR using primers that amplify the integration site. Bona-fide co-integrants were grown overnight in plain media at 28°C. Logarithmic-phase cultures were treated with the pWJ326 counter-selection phagemid, producing a Cas9 nuclease programmed to target and destroy the *cat* gene of the integrated plasmid, for 1 hour and plated on BHI agar supplemented with erythromycin at 30°C. The selected mutants were confirmed by PCR and Sanger sequencing and by checking their growth on chloramphenicol plates at 37°C (to ensure chloramphenicol-sensitivity). Mutants were re-struck at 30°C on BHI agar supplemented with erythromycin, and colonies were inoculated overnight in plain BHI at 37°C to inhibit replication of pWJ326 and cure this counter-selection plasmid. Overnight cultures were plated on plain

BHI agar at 37°C, and colonies were replica-plated on both plain BHI agar at 37°C and BHI agar supplemented with erythromycin at 30°C to check for plasmid loss. With this method, strains JW924, JW833 and JW563, containing the *polA*, *terS* and *ho/ly* mutations, were created using plasmids pJW430, pJW410 and pJW271 respectively.

Propagation of ϕ12 mutants

For the experiments in Fig. 2c–g, infections were performed at an approximate MOI of 10. Subsequently, the precise MOI was determined for each experiment, and the qPCR data and adaptation bar graphs were normalized to this precise MOI. Supplementary Data File 2 shows the exact MOI for each experiment.

pWJ326 phagemid propagation

S. aureus ST24 cells[34] harboring plasmid pWJ326 were inoculated at 30°C in BHI supplemented with erythromycin. Spontaneous induction of the prophage 80α results in the exclusive phage packaging of pWJ326, which harbors the sequence for the 80α packaging (*pac*) site, within the 80α *terS* gene. pWJ326 phagemids were isolated by concentrating the supernatant of overnight *S. aureus* ST24/pWJ326 cultures using Amicon Ultra centrifugal filter units (EMD Millipore) and resuspending the retentate in 1/10 the original volume of PBS.

Propagation of ϕ12 mutants

The ϕ12 mutant lysogenic strains JW924, JW833 and JW563, harboring the complementing plasmids pJW474, pJW383 and pJW278 respectively, were induced from logarithmic phase cultures by the simultaneous addition of 2 μg/ml of mitomycin C and 1 mM IPTG for 3 hours. Supernatants were filtered through 0.45 μm syringe filters (Acrodisc), concentrated with Amicon Ultra centrifugal filter units (EMD Millipore) and resuspended in PBS. In order to generate phage stocks that were free of mitomycin C-induced DNA damage, logarithmic phase cultures of *S. aureus* RN4220 harboring the complementing plasmids pJW474, pJW383 and pJW278 were infected with the PBS phage stocks obtained from strains JW924, JW833 and JW563, respectively, at an MOI of 1 and treated with 1 mM IPTG. After 3 hours, final phage stocks were prepared by filtration of the supernatant.

qPCR

Total DNA was extracted from *S. aureus* cells using a modified phenol/chloroform/isoamyl alcohol (Fisher) protocol as described[35]. qPCR was performed using Fast SYBR Green Master Mix (Life Technologies) and 7900HT Fast Real-Time PCR System (Applied Biosystems) with primer pairs JW703/JW704 (ϕ12) and W915/W916 (*S. aureus* RN4220 *rho* control). The sequences of the primers used in this study are listed in Extended Data Table 1.

Plaque formation assay

Plaque forming units (pfu) were counted on a layer of *S. aureus* cells suspended in 50% heart infusion agar (HIA) plated on a BHI agar base. Samples for measuring extracellular pfu were obtained by filtering culture supernatants through a 0.45 μm syringe filter. Samples

for measuring intracellular pfu were obtained by spinning cells at 10,000 rpm for 1 minute, resuspending cell pellets in BHI supplemented with lysostaphin (167 μg/ml) and incubating the suspensions at 37°C for 5 minutes. Samples were stored on ice until all time points were taken. Finally, samples were briefly vortexed, spun at 10,000 rpm for 1 minute, and 10-fold dilutions were made from the supernatant.

Spacer library generation

Total phage DNA was extracted by a modified phenol/chloroform/isoamyl (PCI) alcohol extraction protocol. Phage were concentrated with Amicon Ultra centrifugal filter units to ~10^9 pfu/μl, and 100 μl of phage were digested with 10 μl (8 units) proteinase K (NEB) in the presence of 1X CutSmart buffer (NEB) in a total volume of 500 μl at 37°C for 3 hours. 500 μl of PCI were added and the DNA extraction protocol was followed as described without bead beating[35]. 50 μg phage DNA was sonicated in a 130 μl total volume in microTUBE AFA Fiber Pre-Slit Snap-Cap 6×16mm tubes (Covaris) using the Covaris S220 Focused-ultrasonicator to a fragment size of 150 bp. The sonicated DNA was dialyzed and 50 μl (~20 μg) was transformed by electroporation into 50 μl of electro-competent *S. aureus* RN4220 cells carrying the pRH223 and pRH240 plasmids (L6308). L6308 cells were pretreated with 0.25 μg/ml anhydrotetracycline hydrochloride for 8 minutes prior to being made electro-competent to induce *cas1*, *cas2* and *csn2*. Following electroporation, cells were outgrown in plain BHI at 37°C for 3 hours and $CaCl_2$ was added for 20 minutes before infection with live phages.

CRISPR immunization simulation assay

After overnight inoculations in TSB supplemented with spectinomycin, *S. aureus* RN4220 strains containing plasmids pJW465, pJW466 and pJW469 (“upstream targeting”) or plasmids pJW471, pJW476 and pJW477 (“downstream targeting”) were pooled in equal volumes. Each pool was mixed with *S. aureus* RN4220 cells at a ratio of 1:10^4 and 0.5 μl of this mix was added to 150 μl of TSB supplemented with calcium chloride in quadruplicate in a 96-well cell culture plate (Cellstar, 655180). Following 1 hour of shaking at 37°C in a microplate reader (TECAN Infinite® 200 PRO), ϕ12γ3 or ϕ12γ3$^{cos\text{-flip}}$ phages were added at an MOI of approximately 18. OD_{600} measurements were recorded every 10 minutes during further shaking at 37°C. Spectinomycin was added after 5 hours to maintain plasmid selection and to inhibit the growth of non-plasmid containing, phage-resistant, CRISPR-independent cells.

Plasmid construction

The sequence of the primers and oligonucleotides used in this study are listed in Extended Data Table 1. To make pRH163, pGG32[36] was amplified with the primers H103/H104 and a one-piece Gibson assembly was performed[37]. To make pJW479, oligonucleotides JW100 and JW101 were annealed and cloned into the BsaI site of pGG3-bsaI as described[38]. To make pJW480, pKOR[39] was amplified with JW264/JW265, RN4220 genomic DNA was amplified with JW406/JW407 and JW410/JW411, pKL55[36] was amplified with JW408/JW409, and a 4-piece Gibson assembly was performed. To make pJM76, pE194[40] was amplified with JM276/JM277, pJM62 was amplified with JM278/JM279, and a 2-piece Gibson assembly was performed. To make pJW259, pJM76 was amplified with JW116/

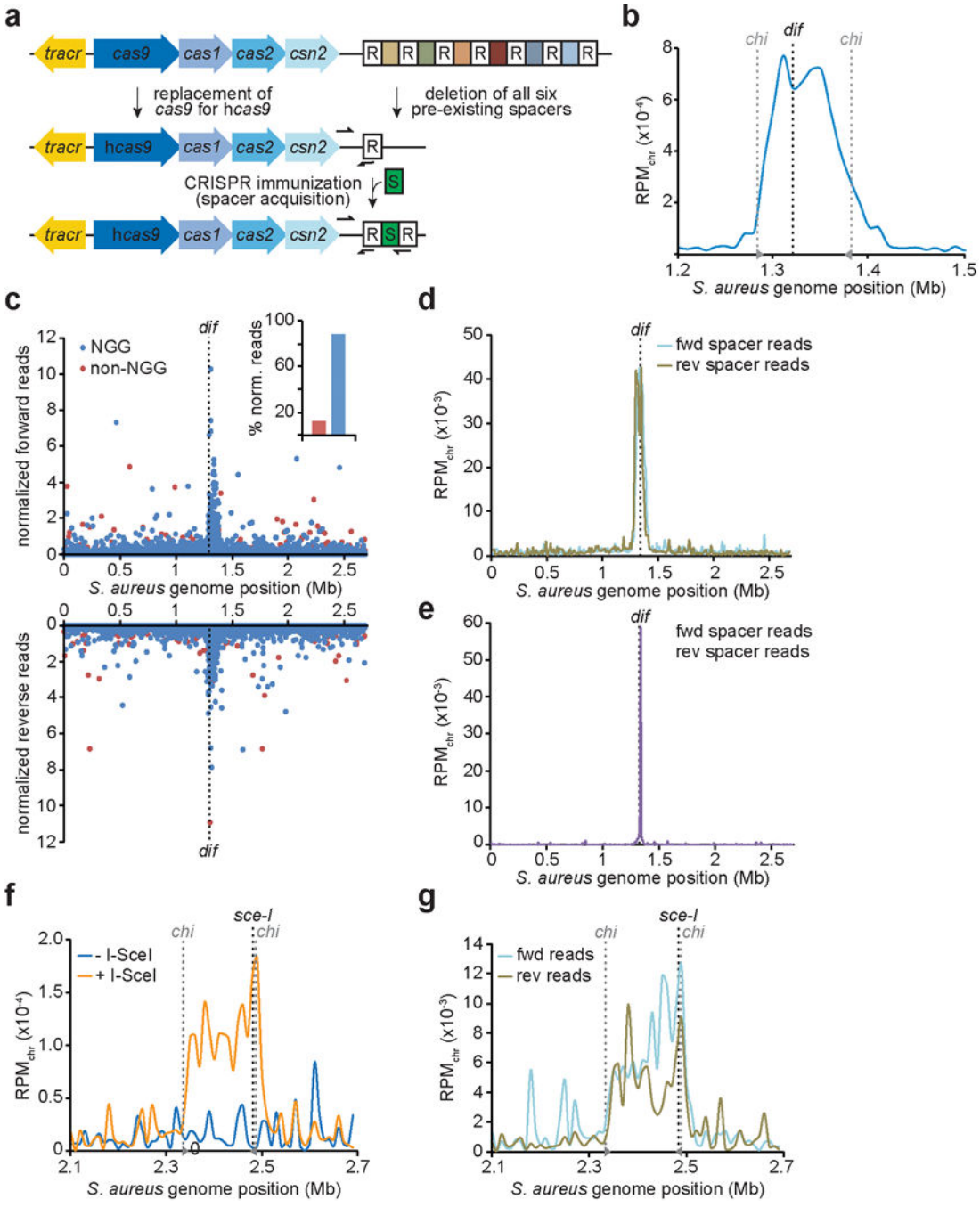
JW356, pRH163 was amplified with JW355/JW115, and a 2-piece Gibson assembly was performed. Plasmid pLM9-4B was constructed by digesting plasmid pLM9 with BamHI, re-ligating the resultant fragments and isolating a ~5kb plasmid with the *E. coli* replication and Amp^R genes removed. pJW197 is a pUC57-based plasmid containing the I-SceI endonuclease sequence which was codon-optimized for expression in *S. aureus* and synthesized by Genewiz LLC (see Supplementary sequences file). To make pJW215, pLM9-4B was amplified with JW412/JW413, pJW197 was amplified with JW460/JW461, and a 2-piece Gibson assembly was performed. To make pJW250, pJW215 with amplified with JW582/oGG104 and W145/JW584, and a 2-piece Gibson assembly was performed. To make pJW232, pIMAY[32] was amplified with JW319/JW320, RN4220 genomic DNA was amplified with JW532/JW533 and JW534/535, and a 3-piece Gibson assembly was performed. To make pJW233, pIMAY was amplified with JW319/JW320, RN4220 genomic DNA was amplified with JW536/JW537 and JW538/539, and a 3-piece Gibson assembly was performed. To make pJW127, pRH163 was amplified with JW285/JW286, and a 1-piece Gibson assembly was performed. To make pWJ322, pIMAY was amplified with W1245/W1246 and W1247/W1248, and a 2-piece Gibson assembly was performed. To make pWJ327, pWJ322 was amplified with W1255/W1256 and ϕNM1 genomic DNA[41] was amplified with W1251/W1252 and W1253/W1254, and a 3-piece Gibson assembly was performed. To make pJW430, pWJ327 was amplified with JW809/JW810, ϕ12 genomic DNA was amplified with JW903/JW904 and JW905/JW906, and a 3-piece Gibson assembly was performed. To make pJW410, pWJ327 was amplified with W1255/W1256, ϕ12 genomic DNA was amplified with JW771/JW883 and JW884/JW777, and a 3-piece Gibson assembly was performed. To make pJW271, pWJ244 was amplified with W1005/JW595, ϕ12 genomic DNA was amplified with JW639/JW653 and JW654/JW642, and a 3-piece Gibson assembly was performed. To make pWJ326, W1079 and W1080 were annealed and cloned into the BsaI site of pWJ295[38]. To make pJW474, pJW250 was amplified with JW673/JW674, a gBlocks® Gene Fragment (IDT) containing a codon-degenerated version of the ϕ12 *polA* gene (see Supplementary sequences file) was amplified with JW1003/JW1004, and a 2-piece Gibson assembly was performed. Codon scrambling was used to prevent recombination between the *polA*2STOP* mutant gene and the wild-type copy used for complementation. Synonymous mutations were introduced at each codon to the next most frequently used codon in *S. aureus* (http://www.kazusa.or.jp/codon/). If the frequency of the new codon was below 2 per 1000, the original codon was used. To make pJW383, pJW215 was amplified with JW673/JW674, ϕ12 genomic DNA was amplified with JW784/JW785, and a 2-piece Gibson assembly was performed. To make pJW278, pJW215 was amplified with JW673/JW674, ϕ12 genomic DNA was amplified with JW671/JW672, and a 2-piece Gibson assembly was performed. To make pJW462, pLZ12spec[42] was amplified with JW713/JW714, pDB184[36] was amplified with JW715/JW774, and a 2-piece Gibson assembly was performed. To make pJW465, pJW466, pJW469, pJW471, pJW476, and pJW477, oligonucleotide pairs JW1008/JW1009, JW1010/JW1011, JW1016/JW1017, JW1020/JW1021, JW790/JW791 and JW792/JW793 were annealed and cloned into the BsaI site of pJW462. To make pJW481, pT181[43] was amplified with AV108/AV109, ϕ12 genomic DNA was amplified with JW613/JW712, JW711/JW617 and JW614/JW615, and a 4-piece Gibson assembly was performed. To make pJW482, oligonucleotides JW618 and JW619 were annealed and cloned into the BsaI site of pGG-BsaI-R[8]. To make pJW484,

pC194[42] was amplified with JW155/JW156, ϕ12 genomic DNA was amplified with JW968/JW969 and JW970/JW971, and a 3-piece Gibson assembly was performed. To make pJW483, oligos JW966 and JW967 were annealed and cloned into the BsaI site of pDB114[38].

Data availability

The analyzed DNA-seq data that support the findings of this study have been provided in the Supplementary Data File 1. All other data are available from the corresponding author upon reasonable request.

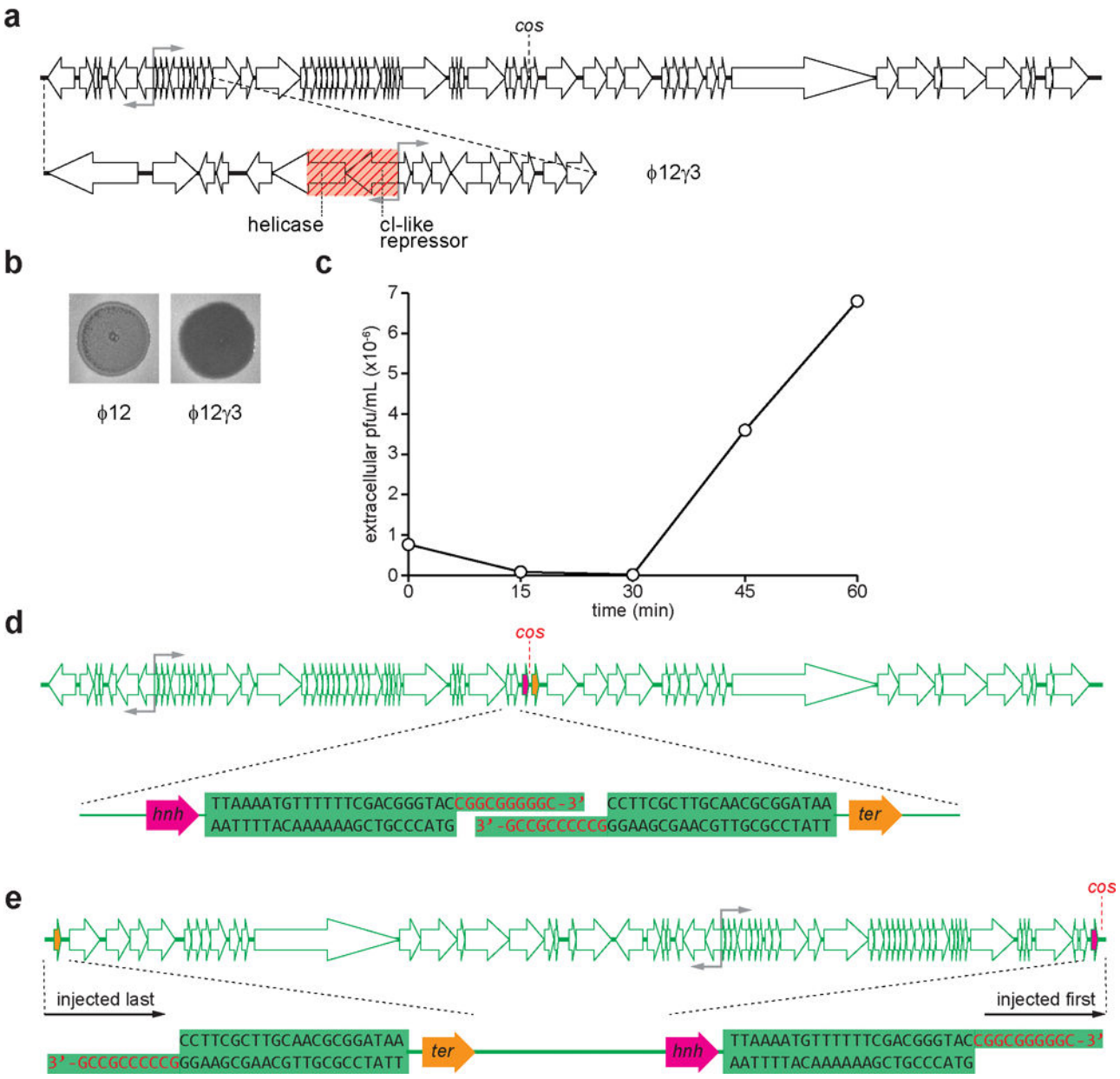
Extended Data



Extended Data Figure 1. Chromosomal spacer acquisition in the *S. pyogenes* type II-A CRISPR-Cas locus

a, Organization of the *S. pyogenes* type II CRISPR-Cas locus. Arrows indicate the annealing position of the primers used to enrich for PCR products containing expanded CRISPR loci. R, repeat; S, new spacer. **b**, Abundance (RPM_{chr}) of chromosomal sequences incorporated as spacers from wild type cells in Fig. 1a (zoom on the *dif* site, the chromosomal terminus). Data previously reported for spacer acquisition by the type I CRISPR-Cas system of *E. coli*

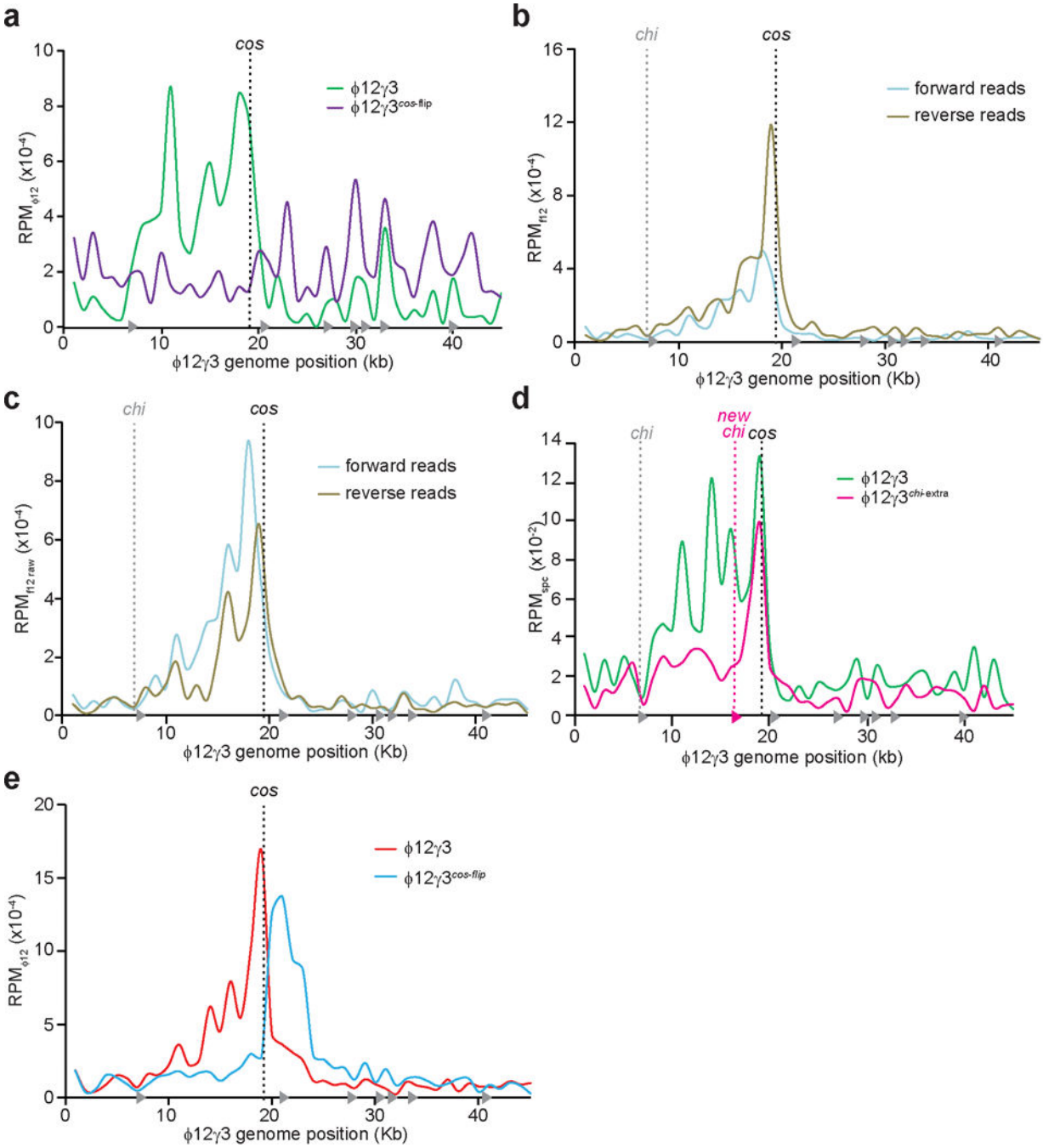
showed accumulation of new spacer sequences matching the two *ter* sites at the chromosomal terminus. *ter* sites are not characterized in *S. aureus*, but the absence of multiple peaks indicates that (i) DSB patterns at the terminus are more distributed in *S. aureus* or (ii) the nature of the breaks allows bidirectional adaptation in *S. aureus* but not in *E. coli*. **c**, Abundance (normalized reads) of individual spacers from the experiment in Fig. 1a, derived from the forward (top plot) or reverse (bottom plot) strand of the chromosome. Blue dots represent spacers derived from sites with a 5′-NGG-3′ PAM between immediately downstream from the 5′ end of the spacer. Red dots represent spacers with a non-NGG PAM. Insert, % of normalized spacers with or without 3′ PAMs. **d**, Abundance (RPM_{chr}) of chromosomal sequences from the experiment in Fig. 1a, derived from the forward (light blue) or reverse (olive) strand, incorporated as spacers into the CRISPR array. **e**, Abundance (RPM_{chr}) of chromosomal spacers in cells harboring wt *cas9* following a 30 minute infection with ϕ12γ3 at MOI 100. *dif*, chromosome terminus. **f**, Abundance (RPM_{chr}) of chromosomal sequences incorporated as spacers in cells with an *I-sceI* site with or without I-SceI expression (orange and blue respectively, zoom on the *I-sceI* recognition site). **g**, Abundance (RPM_{chr}) of chromosomal sequences, derived from the forward (light blue) or reverse (olive) strand, incorporated as spacers in *srtA:I-sceI* cells from Fig. 1b (zoom on the *I-sceI* recognition site). *dif*, chromosome terminus; *sce-I*, I-sceI recognition sequence; grey triangles, *chi* sites pointing in 5′-3′ direction, with the dotted lines marking the first *chi* sites upstream of the *dif* site.



Extended Data Figure 2. Generation of ϕ12γ3

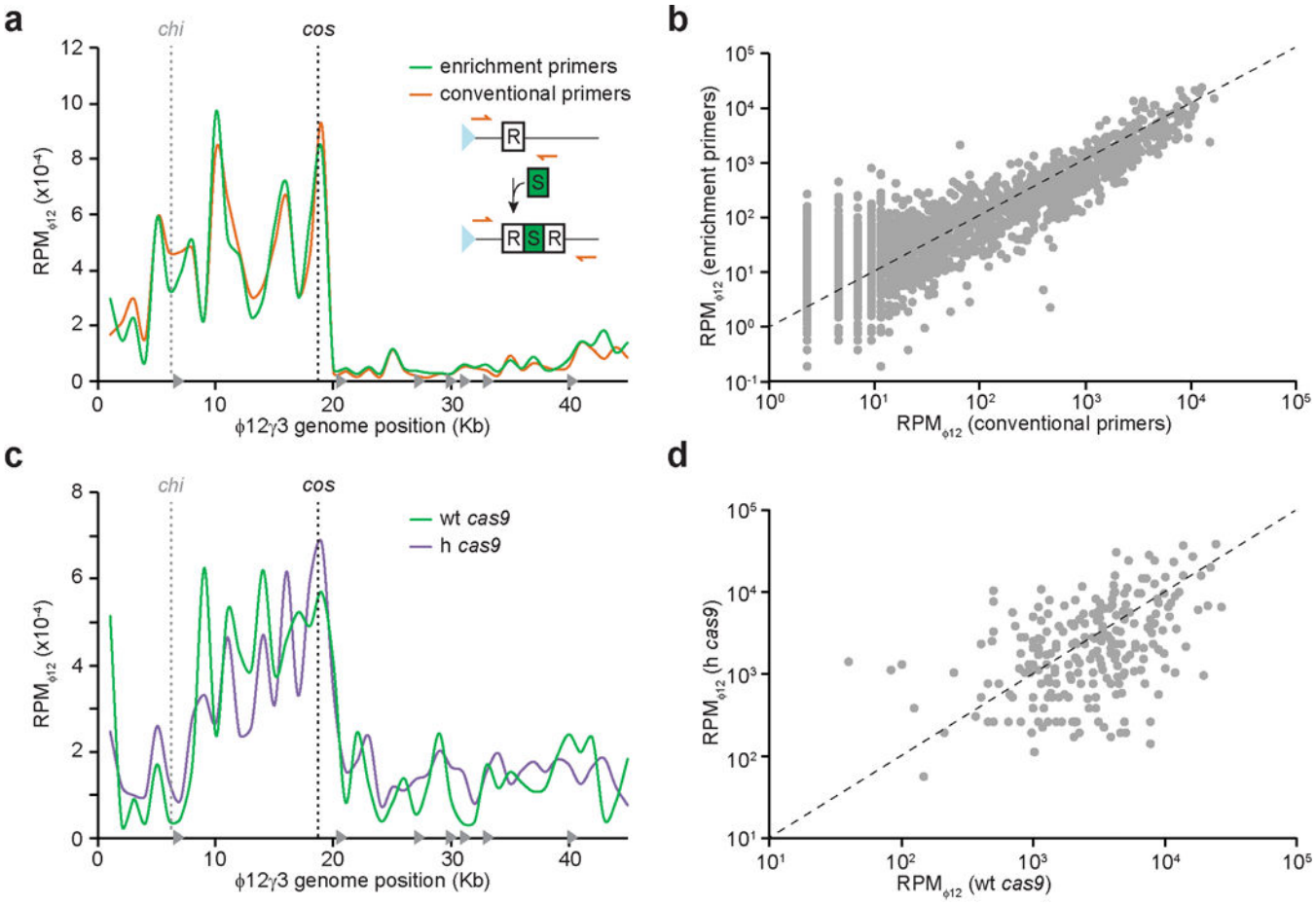
a, Genome organization of the staphylococcal temperate phage ϕ12. The grey arrows show the location of a bidirectional promoter controlling the expression of the lysogeny and lysis genes. The location of the *cos* site is also noted. ϕ12γ3 contains a deletion (highlighted in red) of the cI-like repressor gene (and part of a neighboring helicase gene) that prevents the establishment of lysogeny. **b**, ϕ12 has forms turbid plaques due to its ability to lysogenize and form colonies that are resistant to superinfection. In contrast ϕ12γ3 forms clear plaques due to its inability to lysogenize. **c**, Extracellular viral particles measured as plaque forming units (pfu) obtained from culture supernatants at 0, 15, 30, 45 and 60 minutes after infection with ϕ12γ3. After an initial period in which the phage particles in the media decrease due to phage adsorption, the pfu value increases following the burst of the infected cells. **d**, A

staggered cleavage by the terminase complex at the *cos* site generates free DNA ends with a 10 bp 3′ overhang during DNA packaging. **e**, A view of ϕ12 with the cleaved *cos* sites at either end of the genome. During injection, the *hnh* (magenta) proximal end enters the bacterial cell first while the *ter* (orange) proximal end remains temporarily in the phage capsid.



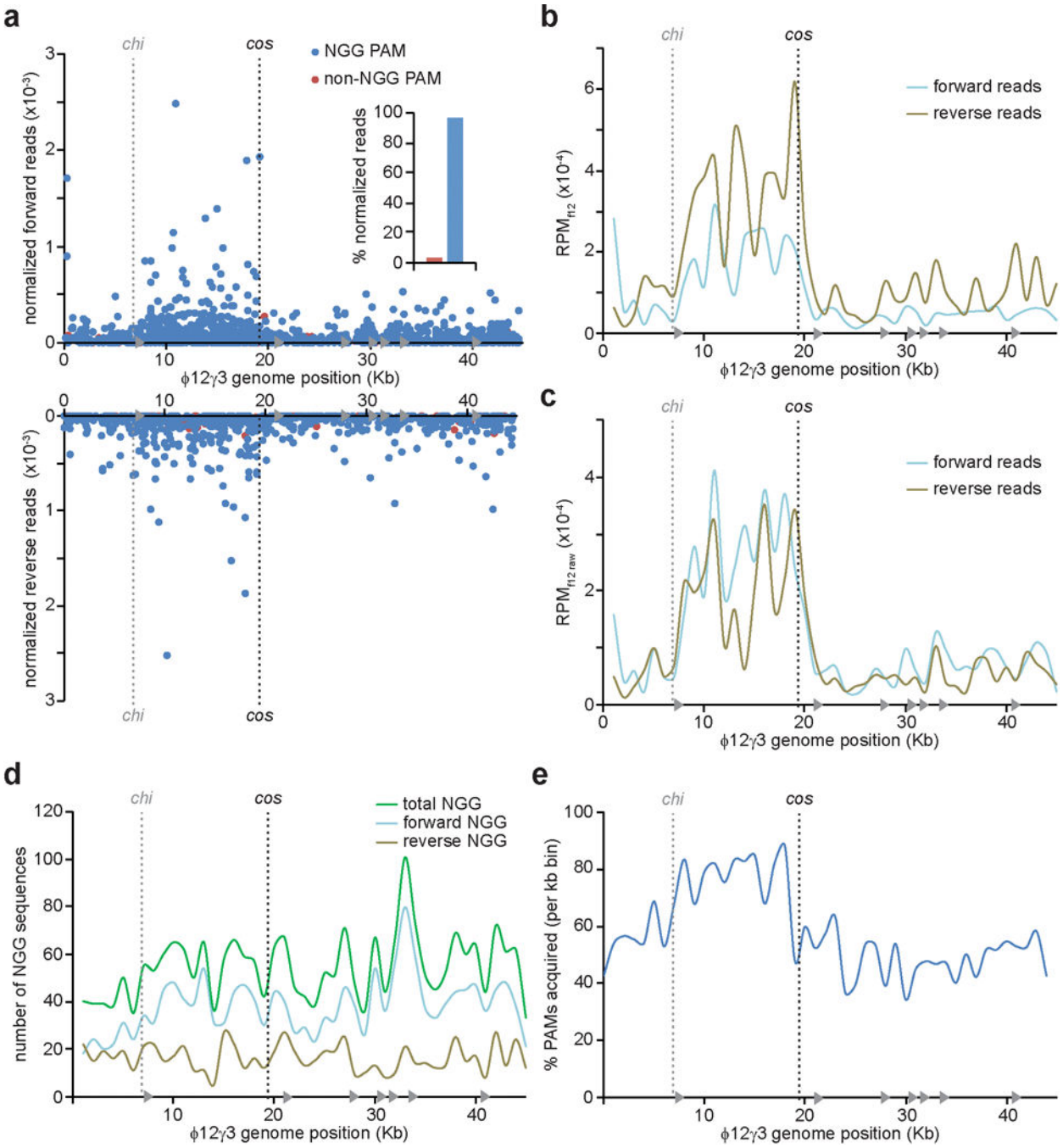
Extended Data Figure 3. Patterns of ϕ12γ3 spacer acquisition

a, Abundance ($RPM_{\phi12}$) of ϕ12γ3 (green) or ϕ12γ3$^{cos\text{-flip}}$ sequences incorporated as spacers into the CRISPR array 30 min after infection at MOI 1. *cos*, cohesive end; grey triangles, *chi* sites pointing in 5′-3′ direction. **b**, Abundance ($RPM_{\phi12}$) of ϕ12γ3 forward (light blue) or reverse (olive) strand sequences incorporated into the CRISPR array from the triplicate *addA*^{n} experiment in Fig. 1c. **c**, Same as panel **a**, but showing abundance in $RPM_{\phi12\ raw}$ which does not normalize reads to account for the GG content in each 1 kb bin. *cos*, cohesive end; *chi*, first *chi* site upstream of the *cos* site; grey triangles, *chi* sites pointing in 5′-3′ direction. **d**, To corroborate the involvement of *chi* sites in spacer acquisition obtained after infection with ϕ12γ3 (green), we introduced two forward facing *chi* sites (pink arrow and dotted line) in tandem into this phage about 2.5 kb upstream of the *cos* site generating ϕ12γ3$^{chi\text{-extra}}$. CRISPR adaptation against this phage showed comparable levels of spacer acquisition for sequences mapping to the region between the *cos* site and the new *chi* sites. However, adaptation in the region to the left of the new *chi* sites, normally a highly adapted region, was reduced significantly. Abundance (RPM_{spc}) of ϕ12γ3 (green) or ϕ12γ3$^{chi\text{-}extra}$ (pink) sequences incorporated as spacers into the CRISPR array 30 min after infection at MOI 10 is shown. **e**, Abundance ($RPM_{\phi12}$) of ϕ12γ3 (red) or ϕ12γ3$^{cos\text{-}flip}$ (blue) sequences incorporated as spacers following a 30 minute infection of *addA*^{n} cells at MOI 10. Despite the absence of *chi* sites pointing in the 3′-5′ direction, spacer acquisition from ϕ12γ3$^{cos\text{-}flip}$ is limited to the area immediately adjacent to the *cos* site, in contrast to that observed for wild-type cells in Fig. 2g. This suggests that *addA*^{n} degradation increases the number of free DNA ends used as adaptation substrates and that the *cos* site serves as the main entry point for the adaptation machinery.



Extended Data Figure 4. Patterns of ϕ12γ3 spacer acquisition using conventional primers or wt *cas9*

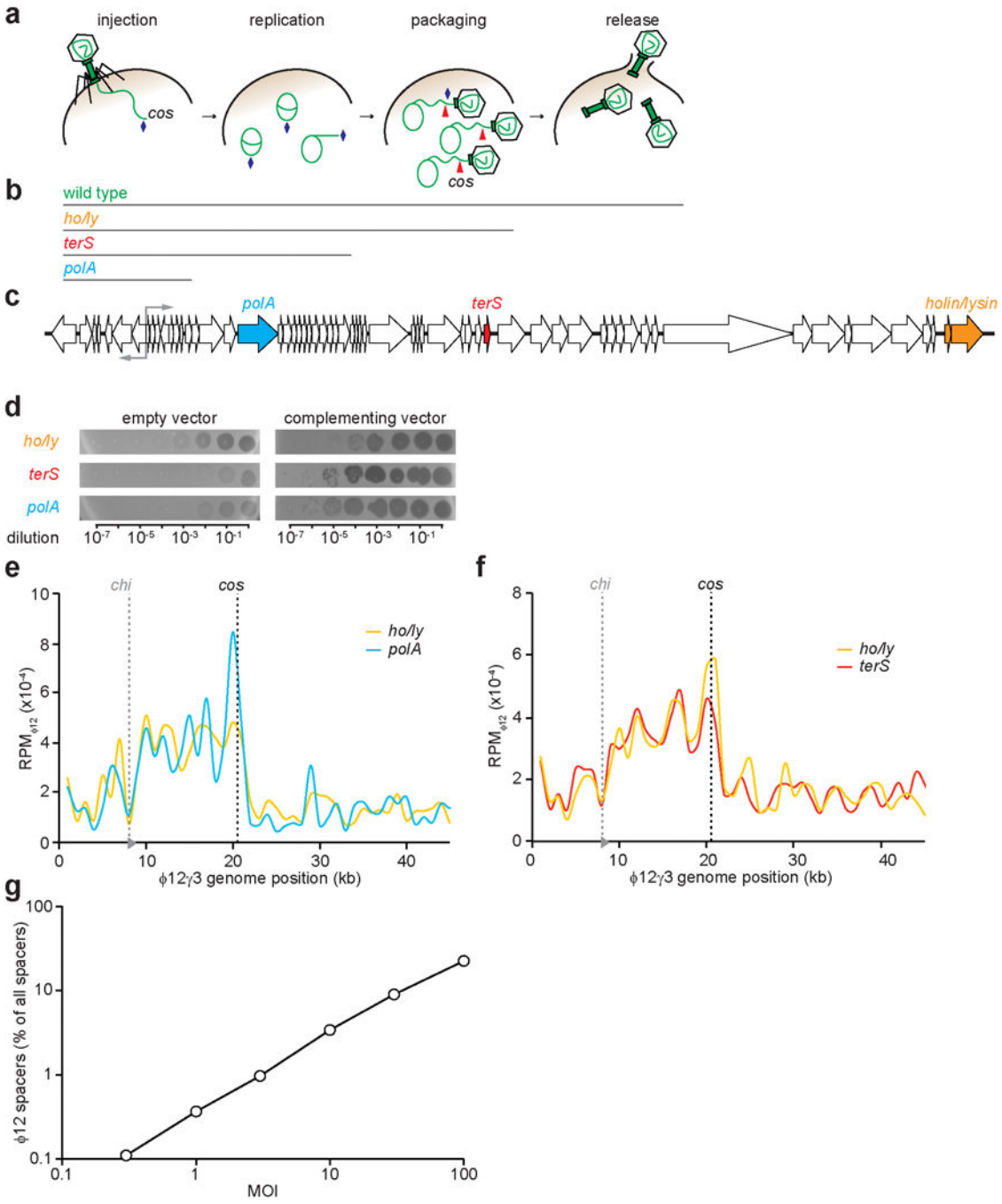
a, Abundance (RPM$_{\phi12}$) of ϕ12γ3 sequences incorporated into the CRISPR array after an overnight infection at MOI 10. CRISPR loci within a single sample were amplified using either enrichment primers (green) or conventional primers (orange). Insert: location of the conventional primers; outside the CRISPR repeats. **b**, Individual spacers common to both datasets in panel **a** were plotted with RPM$_{\phi12}$ values for enrichment primers on the y-axis and conventional primers on the x-axis. The diagonal dotted line indicates the identity line. **c**, Abundance (RPM$_{\phi12}$) of ϕ12γ3 sequences incorporated into the CRISPR array after a 30 minute infection at MOI 100 of cells harboring h *cas9* (purple) or wt *cas9* (green). **d**, Individual spacers common to both datasets in panel **c** were plotted with RPM$_{\phi12}$ values for h *cas9* on the y-axis and wt *cas9* on the x-axis. *cos*, cohesive end; *chi*, first *chi* site upstream of the *cos* site. The diagonal dotted line indicates the identity line.



Extended Data Figure 5. PAM preference and strand bias for ϕ12γ3 spacer acquisition

a, Abundance (normalized reads) of individual spacers from one of the wild-type replicates in Fig. 1c, derived from the forward (top plot) or reverse (bottom plot) strand of ϕ12γ3 following a 30 minute infection at MOI 10. Blue dots represent spacers associated with NGG PAMs while red dots represent spacers with non-NGG flanking sequences. Insert, % of normalized spacers with or without canonical NGG PAMs. **b**, Abundance ($RPM_{\phi 12}$) of ϕ12γ3 forward (light blue) or reverse (olive) strand sequences incorporated into the CRISPR array from the triplicate wild type experiment in Fig. 1b. **c**, Same as panel **b**, but showing

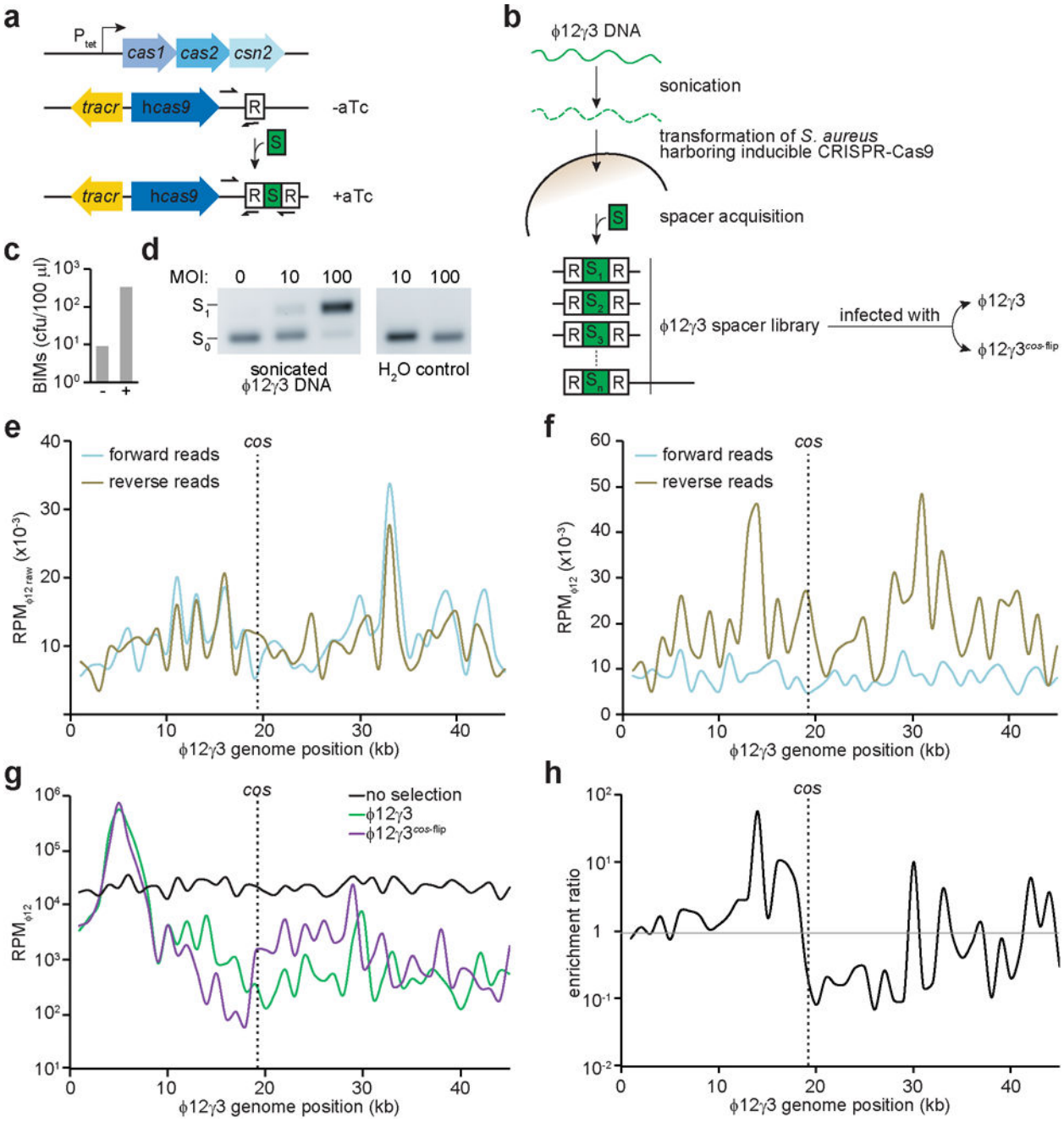
abundance in $RPM_{\phi 12\ raw}$ which does not normalize reads to account for the GG content in a given 1 kb bin. *cos*, cohesive end; *chi*, first *chi* site upstream of the *cos* site; grey triangles, *chi* sites pointing in 5′-3′ direction. **d**, The number of 5′-NGG-3′ PAM sites within 1kb bins on the forward (light blue), reverse (olive) or combined (green) strands of the ϕ12γ3 genome. **b**, From one of the wild type replicates in Fig. 1b, the percentage of PAMs within each 1kb bin that are represented by at least one spacer are plotted against the ϕ12γ3 genome. This pattern of "PAM acquisition" demonstrates that the spacer distribution pattern of ϕ12γ3 does not result from hyper-acquisition at a few sites. *cos*, cohesive end; *chi*, first *chi* site upstream of the *cos* site; grey triangles, *chi* sites pointing in 5′-3′ direction.



Extended Data Figure 6. Generation and spacer acquisition profile of ϕ12 mutant phages

a, Stages of the ϕ12 lytic cycle. Blue rhombuses indicate the location of free dsDNA ends generated in different stages that could be used for spacer acquisition. Red arrowheads indicate terminase cleavage of the *cos* site during viral DNA packaging. **b**, Mutations used in this study to halt different stages of the ϕ12 lytic cycle. **c**, Localization of the mutated genes in the ϕ12 genome. To eliminate *polA* (encoding DNA polymerase A) function we added two stop codons after the 110th codon. *terS* (encoding the small terminase subunit) function was eliminated through an in-frame deletion that left only the first and last 15 codons of the

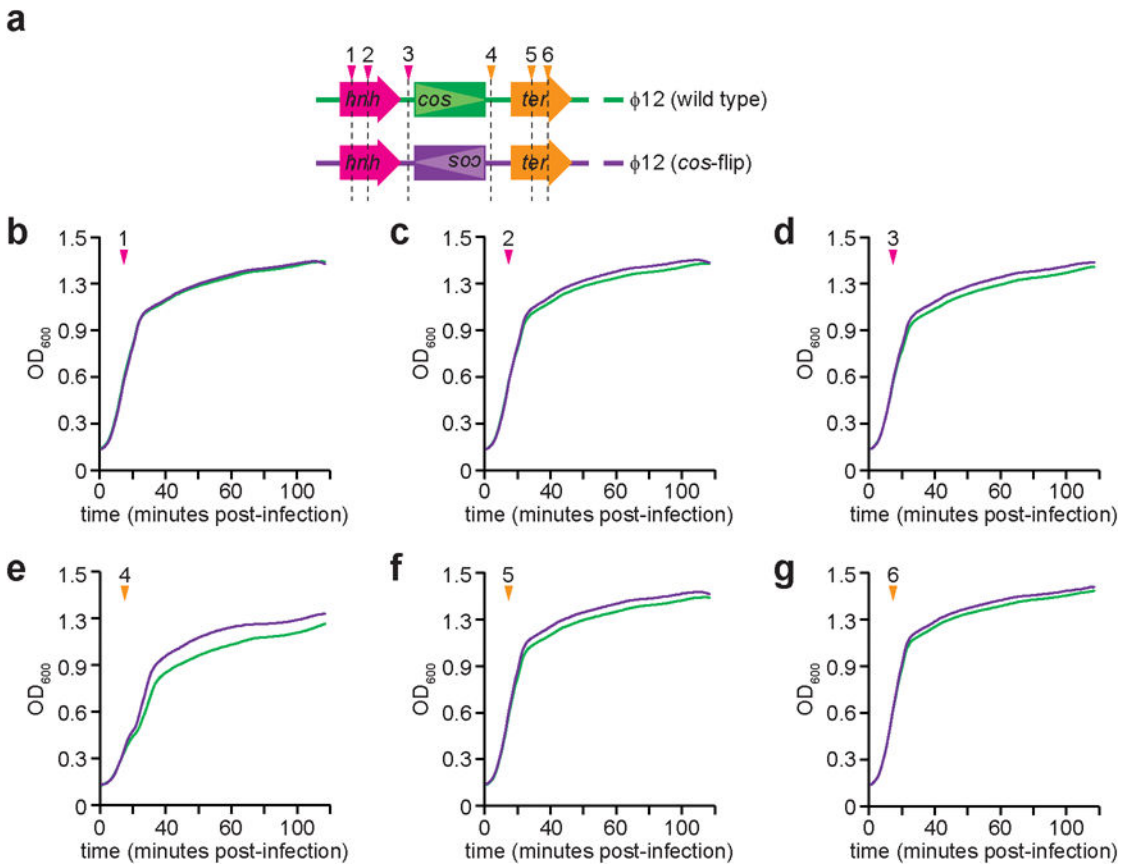
gene. The *holing/lysin* operon was disrupted through an in-frame deletion that left the first five codons of the *holin* gene fused to the last seven codons of the *lysin* gene. **d**, The mutations were generated in ϕ12 prophages integrated into the *S. aureus* RN4220 chromosome. The resulting lysogens harboring the mutant prophages, which were incapable of forming and/or release viral particles, were transformed with complementing plasmids carrying a wild type copy of the mutated gene. The transformed lysogens were induced and different dilutions of the resulting lysate were spotted on plates seeded with staphylococci with or without the complementing plasmid. In all cases the mutant phages were able to lyse the complemented cells but not bacteria carrying an empty plasmid. **e**, Abundance ($RPM_{\phi12}$) of ϕ12:*polA* (blue) or ϕ12:*ho/ly* (orange) sequences incorporated as spacers into the CRISPR array 30 min after infection at MOI 10. *cos*, cohesive end; grey triangles, *chi* sites pointing in the 5′-3′ direction, with the dotted line marking the first *chi* site upstream of the *cos* site. **f**, Abundance ($RPM_{\phi12}$) of ϕ12:*terS* (red) or ϕ12:*ho/ly* (orange) sequences incorporated as spacers into the CRISPR array 60 min after infection at MOI 10. *cos*, cohesive end; grey triangles, *chi* sites pointing in the 5′-3′ direction, with the dotted line marking the first *chi* site upstream of the *cos* site. **g**, The % of ϕ12 spacers (calculated as the ratio of ϕ12-specific spacer reads to the total spacer reads) was plotted as a function of the MOI.



Extended Data Figure 7. Generation and testing of a library of spacers of ϕ12γ3 genomic DNA

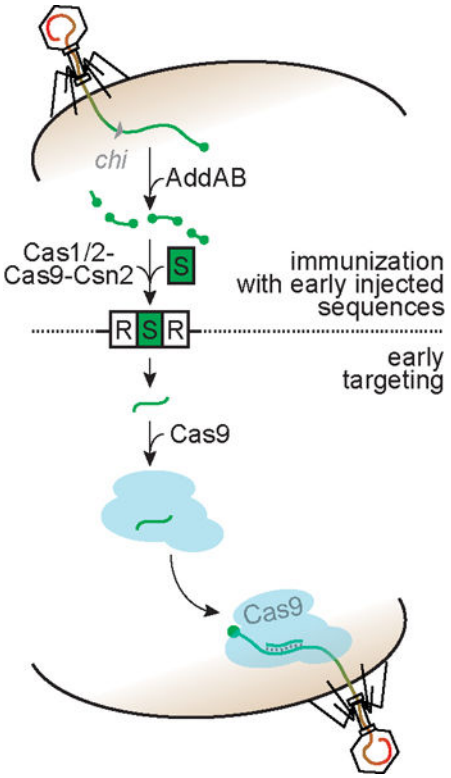
a, Engineering of the *S. pyogenes* type II-A CRISPR-Cas system to perform inducible spacer acquisition. The spacer acquisition genes cas1, cas2 and csn2 are under the control of a tetracycline-inducible promoter (P_{tet}) in one plasmid. Another plasmid contains *cas9* and the *tracrRNA* genes along with a single-repeat CRISPR array. Spacer acquisition can be detected only via PCR (arrows represent the primers used in the reaction) in the presence of the inducer anhydro-tetracycline (aTc). **b**, Procedure for constructing the library. ϕ12γ3 genomic DNA was sonicated to generate fragments of about 150 base pairs, which were introduced through electroporation (with a water control) into cells harboring the inducible

S. pyogenes CRISPR-Cas system in the presence of aTc. This generated a library of cells containing a type II CRISPR-Cas system programmed with different spacers from the ϕ12γ3 genome. **c**, After transformation, cells recovered for 3.3 hours in ATC-free media and were treated with live ϕ12γ3 phage for 15 minutes. The surviving bacteriophage-insensitive mutant (BIMs) colonies were counted by plating 100 μl of the infected culture. Bacteria transformed with a water control (−) yielded 9 cfu/100 μl, none of which had incorporated a new spacer (data not shown). This demonstrates that at least after 3.3 hours without the aTc inducer, staphylococci harboring the engineered CRISPR-Cas system cannot acquire new spacers during subsequent infection with live phage. This rules out new events of spacer acquisition during the infection of the library with ϕ12γ3 or ϕ12γ3$^{cos\text{-flip}}$. Conversely, cells transformed with sonicated ϕ12γ3 DNA yielded 327 cfu/100 μl, and all of those tested (n=4) carried an expanded CRISPR array, indicating that significant adaptation occurred during the electroporation of ϕ12γ3 fragments. **d**, Analysis of spacer selection from the library after infection. The library of spacers was treated with ϕ12γ3 or ϕ12γ3$^{cos\text{-flip}}$ (not shown) at MOIs of 0, 10 or 100 for 24 hours to determine whether selection during phage interference could influence the spacer distribution within the library. Spacer acquisition in the uninfected library (MOI =0) could not be detected by PCR, indicating that the majority of cells in the library do not enlarge the CRISPR array with viral spacers during this period. In contrast, strong PCR products corresponding to expanded CRISPR loci were observed following the overnight phage infections, demonstrating the enrichment of adapted cells during CRISPR-Cas targeting, both at MOI 10 and 100. Incorporation of new spacers could not be detected by PCR in the control cells transformed with water, providing further evidence that in this assay cells are only able to adapt during electroporation in the presence of aTc, but not after this treatment, in the absence of the inducer. S_1 and S_0, CRISPR loci with and without a new spacer sequence, respectively. **e**, The data from Fig. 3a with abundance ($RPM_{\phi12\ raw}$) of spacers derived from forward (light blue) and reverse (olive) strands of ϕ12γ3. **f**, Same as panel **e**, showing relative (to PAM content) abundance in $RPM_{\phi12}$. **g**, Same as Fig. 3a, showing the ($RPM_{\phi12}$) values for the full phage genome. **h**, Same as Fig. 3b, showing the ϕ12γ3/ϕ12γ3$^{cos\text{-flip}}$ enrichment ratio values for the full phage genome.



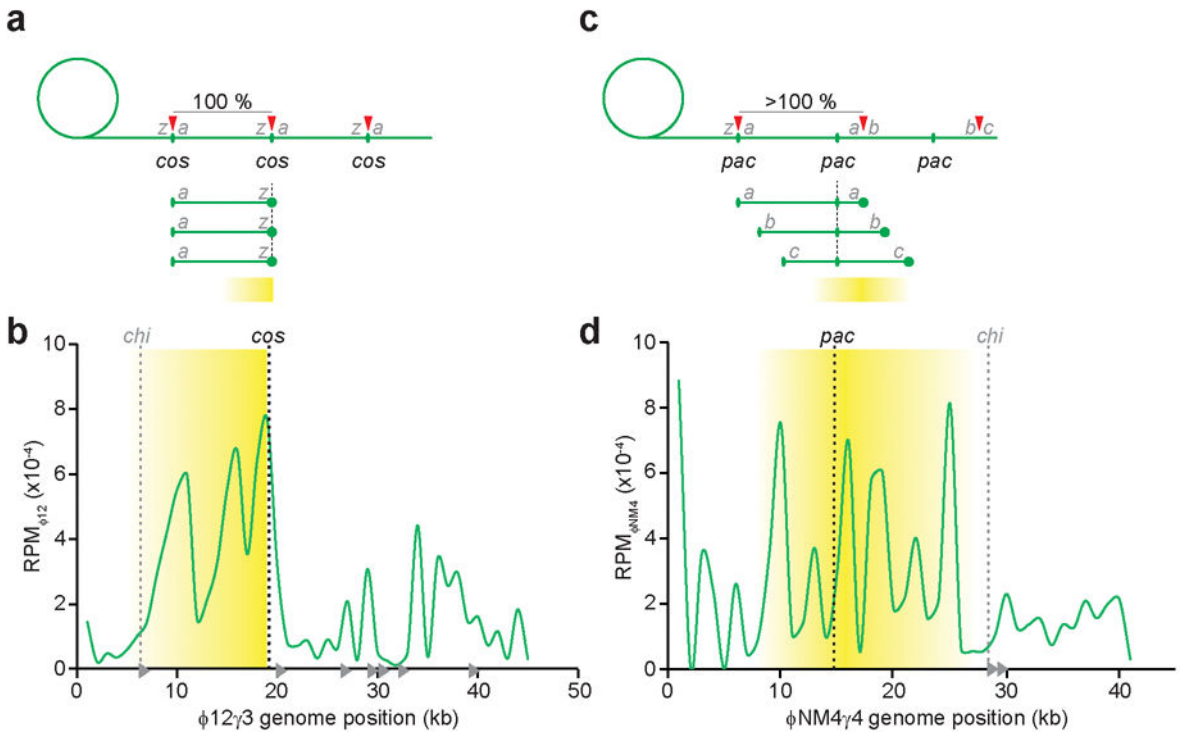
Extended Data Figure 8. Design and test of spacers targeting DNA sequences adjacent to the *cos* site

a, The *cos* site-proximal region of ϕ12γ3 or ϕ12γ3$^{cos\text{-flip}}$ was targeted by type II CRISPR-Cas systems programmed with a spacer matching the upstream (1,2,3) or downstream (4,5,6) region. **b–g**, Cells harboring each of these CRISPR-Cas systems were infected at an MOI of 10 with either ϕ12γ3 (green) or ϕ12γ3$^{cos\text{-flip}}$ (purple). CRISPR-mediated survival of the cultures was monitored by measuring their optical density at 600 nm (OD_{600}) over time.



Extended Data Figure 9. Coordination between the immunization and targeting phases of the type II CRISPR-Cas immune response

Immunization occurs shortly after the beginning of the infection through the acquisition of new viral spacer sequences preferentially from the first injected free DNA end (green circle). Degradation of this DNA end by the AddAB nuclease, limited by *chi* sites, generates additional free DNA end substrates for recognition by the Cas1-Cas2-Cas9-Csn2 spacer acquisition complex and subsequent integration into the CRISPR array by the Cas1-Cas2 integrase. During targeting, Cas9 nucleases loaded with the crRNA guides generated by the acquired spacers allow the majority of the cells of the immunized host population to target the first region to be injected (green; red, last injected region) by subsequent invading viruses, providing faster and more efficient immunity.



Extended Data Figure 10. Pattern of spacer acquisition during ϕNM4γ4 infection

a, Packaging of *cos* phages. In phage lambda, and presumably in ϕ12, the packaging of the viral genome invariably starts by cleavage of the terminase complex (red arrowhead) at the *cos* site of the phage concatamer generated by rolling circle replication, located between hypothetical genes *a* and *z*. The DNA to the left of the *cos* site (*z* gene, green circle) is the last to be packaged into the phage capsid and therefore always the first to be injected into a newly infected bacterial cell. The expected pattern of spacer acquisition starts from this dsDNA end and progressively decreases until the first *chi* site (yellow gradient box). **b**, Pattern of spacer acquisition for ϕ12γ3 10 minutes after infection at an MOI of 10. This is similar to the results obtained at 30 minutes (Fig. 1c) but comparable to the infection conditions of panel **d**. The area highlighted with a yellow gradient shows the expected spacer acquisition pattern. *cos*, cohesive end; grey triangles, *chi* sites pointing in the 5′-3′ direction, with the dotted line marking the first *chi* site upstream of the *cos* site. **c**, Packaging of *pac* phages. These phages employ a “headful” DNA packaging mechanism in which each genomic concatamer is cleaved first at the *pac* site with subsequent cleavages occurring processively but imprecisely, after packaging about 105 % genome lengths. The exact percentage is determined by how much DNA can be filled into the phage capsid and it is always >100 % to ensure duplicated sequences at each end of the injected genome for recombination and circularization after infection of the next host. *S. aureus pac* phages and pathogenicity islands (SaPIs) display a rightward packaging mechanism where the duplicated DNA is located downstream of the next *pac* site. Therefore the last sequence to be packaged into the phage capsid and therefore the first to be injected into a newly infected bacterial cell is variable for each infection (*a*, *b*, *c* gene, green circles) but lies immediately downstream of the *pac* site. The expected pattern of spacer acquisition starts from every different dsDNA end and progressively decreases leftward until the first *chi* site (yellow

gradient box). **d**, We determined the spacer acquisition pattern of the *pac* phage ϕNM4γ4 (a lytic derivative of ϕNM4, 10 minutes after infection at an MOI of 10. As expected for the injection of variable dsDNA ends downstream of the *pac* site, we detected a spacer acquisition hotspot in the 10–20 kb region to the right of this site (the expected pattern is highlighted in yellow). This is consistent with the rightward migration of *pac* phage injection points, with 10–20 kb corresponding to the packaging of about 5–10 viral genomes, well within observed ranges of *pac* phage processivity. *cos*, cohesive end; grey triangles, *chi* sites pointing in the 5′-3′ direction, with the dotted line marking the first *chi* site upstream of the *cos* site.

Extended Data Table 1

Oligonucleotides used in this study

Primer name	Sequence
JW8_pGG32 r1 upst F2	ggcttttcaagactgaagtctag
JW3_pGG32 r1 R_G	aaaacagcatagctctaaaacg
JW4_pGG32 r1 R_A	aaaacagcatagctctaaaaca
JW5_pGG32 r1 R_T	aaaacagcatagctctaaaact
H103 - I473F-F	ggaagtctgaagaaacatttaccccatgg
H104 - I473F-R	ccatggggtaaatgtttcttcagacttcc
JW400_phi12 type III spacer 1F	gaactggaacgtaaatttaaataatcgcttaaagtagat
JW401_phi12 type III spacer 1R	tttaatctactttaagcgattatttaaatttacgttcca
JW264_attP1 upstr R	cagaattcgtgactaagtcatagctg
JW265_attP2 dwnstr F	ccatccagctgatatcccctatag
JW406_srtA UHR F Gib pKOR1	agctatgacttagtcacgaattctgctagtcttgtaaatctaattcgtaaaatgc
JW407_srtA UHR R Gib KanR F+SceI target	aatacggtagggataacagggtaatcacctctatttaattgttcaggtgttg
JW408_KanR F+SceI target	attaccctgttatccctaccgtattaccgcctttgagtgag
JW409_KanR R	gcatagcgtgagctattaagctcg
JW410_srtA DHR F Gib KanR R	tcgagcttaatagctcacgctatgcgctttgcagaagaaaatgaatc
JW411_srtA DHR R Gib pKOR1	actataggggatatcagctggatggcctgttttataaattgattggcg
JM276_pJM76 gibson pE194 temp fwd 2	ggagaagattcgtaatttttatactgcaatcggatgcgat
JM277_pJM76 gibson pE194 temp rev 2	ccaaatcacagaatcatgttcatatttatcagagctcg
JM278_pJM76 gibson pJM62 temp fwd 2	gataaatatgaacatgattctgtgatttggatccttccagaagtc
JM279_pJM76 gibson pJM62 temp rev 2	cgattgcagtataaaaaattacgaatcttctcctgacgttttttaaatcttg
JW116 cas9 R for 2pGib	gttcatcgtagcgtttaatcattg
JW356_Sp CRISPR dwnstr F	cattgccgatgataacttgagaaagag
JW355_Sp CRISPR dwnstr R	ggtatccgactgctggtattaaccc
JW115 cas9 F for 2pGib	gcaaattggagatcaatatgctg
JW412_pLM9B R	gtacaagcttaattgttatccgctc
JW413_pLM9B F	agatctcgagagctcctaggc
JW460_Sa I-sceI F Gib Pspac	gagcggataacaattaagcttgtacttaggaggatgattatttatgaaaaac
JW461_pUC57-Amp R Gib Pspac	gctagcctaggagctctcgagatctcacaggaaacagctatgacc

Primer name	Sequence
JW582_pLM9B Pspac(h) F	ctttatctacaaggtgtggcataatgtgtgtaattgtgagcggataacaattaagc
oGG104_pLM9-4B R	tataggtatgtggttttgtattgg
W145_pLM9-4B F	cttaacaatcccaaaacttgtcg
JW584_Pspac(h) R	acacacattatgccacaccttg
JW319_pIMAY MCS R	cttatcgataccgtcgacctcgag
JW320_pIMAY MCS F	cttgatatcgaattcctgcagcc
JW532_rexB(D945A) UHR F Gib pIMAY	ccccccctcgaggtcgacggtatcgataagtgaagaaaagaaatactttgaacagc
JW533_rexB(D945A) UHR R	tgcaatgatattaacaaaacttgtatcattctttg
JW534_rexB(D945A) DHR F	aagaatgatacaagttttgttaatatcattgcatataaatcctctgaaggtagtgcg
JW535_rexB(D945A) DHR R Gib pIMAY	tcccccgggctgcaggaattcgatatcaagcataatgctgttgaattaatttcaagc
JW536_rexA(D1159A) UHR F Gib pIMAY	ccccccctcgaggtcgacggtatcgataaggaaatggcatttccatctttagc
JW537_rexA(D1159A) UHR R	tgctacaaaataatgcacaccatctttaac
JW538_rexA(D1159A) DHR F	tttgttaaagatggtgtgcattattttgtagcatataaaaccgatgcatttaatcgtc
JW539_rexA(D1159A) DHR R Gib pIMAY	tcccccgggctgcaggaattcgatatcaagcattttcaggtaatggtagttcgtc
JW285_pRH163(C-7A) F	caaaaatagtatacgaggttttagagctatgctgttttg
JW286_pRH163(C-7A) R	taaaacctcgtatactatttttgtctaaaaaattttgtaatcgc
W1245_K864601 -pI MAY-f	gacaaaaatcaccttgcgctaatgctctgttacagctgttagattatgaaagccgatg
W1246_pIMAY-B1006-r	ccgccctgtcaggggcggggttttttttgagaattacaacttatatcgtatgg
W1247_B1006-pIMAY-f	aaaccccgcccctgacagggcggggttttttttgacaaataactctatcaatgatag
W1248_pIMAY-K864601-r	attagcgcaaggtgatttttgtcttcttgcgctaatttttcgctgcaaagccacgctac
W1251_TerS-up-f	ggtttttttctcaaataactctaatgtcagctgatgaagatggc
W1252_TerS-up-r	ctttcgtcatttcatttaccaccaactctcgcg
W1253_TerS-down-f	ggtggtaaatgaaatgacgaaagttaaattaaactttaac
W1254_TerS-down-r	cgctaatttttcgctgcaaagcttgcattaagaattgtaacccttg
W1255_pWJ322-f	gctttgcagcgaaaaaattagcg
W1256_pWJ322-r	tagagttatttgagaaaaaaaaccccgc
JW809_pWJ327 F	gttgtacaagggttacaattcttaatgc
JW810_pWJ327 R	ccatcttcatcagctgacattagag
JW903_f12 polA UHR F Gib pWJ327	aataactctaatgtcagctgatgaagatggggacggtatataaacatgaaagcaaaag
JW904_f12 polA STOP2 UHR R	taaaacttctccaactttatcaagcgaagcttattaaggtaagccaatacgcattgaattaac
JW905_f12 polA STOP2 DHR F	atggttaattcaatgcgtattggcttaccttaataagcttcgcttgataaagttggag
JW906_f12 polA DHR R Gib pWJ327	ttgcattaagaattgtaacccttgtacaacccttagttatgctttctaccggtac
W1255_pWJ327 F2 W1256_pWJ327 R2	gctttgcagcgaaaaaattagcg tagagttatttgagaaaaaaaaccccgc
JW771_phi12 dCos UHR F Gib pWJ322	gggcggggttttttttctcaaataactctagaggaaattatagaggagtctcaaggcc
JW883_TerS Dint78 R Gib F	tccaccttcttcttgaactatctttttttcttttcagctgtttttttgttaatttc
JW884_TerS Dint78 F Gib R	tatatgaaattaacaaaaaaacagctgaaaagaaaaaagatagttcaagaagaagg
JW777_dTerS DHR R Gib pWJ322	cttcttgcgctaatttttcgctgcaaagccttttagatccttcaatgcagggc
W1005_pWJ244 F	gtgaagacgaaagggcctcgtg
JW595_pWJ244 R2	gtgggatatttttaaaatatatatttatgttacag

Primer name	Sequence
JW639_phi12 Dholin-lysin UHR F Gib pWJ244	acataaatatatattttaaaaatatcccacctctcaagtttttagacctaaagcc
JW653_phi12 Dlysin UHR R Gib DHR	aagtaaatactaaatcgtgctaaacttacctacttttgcatccatttgtttgc
JW654_phi12 Dlysin DHR F Gib UHR	aaggagcaaacaaatggatgcaaaagtaggtaagtttagcacgatttagtatttacttag
JW642_phi12 Dholin-lysin DHR R Gib pWJ244	aggcgtatcacgaggccctttcgtcttcacgtgtgatattttgcaatctctttatcag
W1079_spc cat-f	aaacttttaccctgcatttattttcttagtgacag
W1080_spc cat-r	aaaactgtcactaagaaaataaatgcagggtaaaa
JW673_pLM9B for insert R	aaataatcatcctcctaagtacaagc
JW674_pLM9B for insert F	atgcaaatatgagccaaataaatatattc
JW1003_polA_decoded2 F	ttaagcttgtacttaggaggatgattatttatgaacattgacatagagacttattcttc
JW1004_polA_decoded2 R	agaatatatttatttggctcatatttgcatttagtctttcatatagaaaggacttgtg
JW784_TerS F Gib pLM9B	ttaagcttgtacttaggaggatgattatttatgaaaggggtctttatatgaaattaac
JW785_TerS R Gib pLM9B	agaatatatttatttggctcatatttgcatttaatagtcaccgaatccaccttc
JW671_phi12 holin F Gib pLM9B	ttaagcttgtacttaggaggatgattatttatggatgcaaaagtaataacaagatac
JW672_phi12 lysin R Gib pLM9B	agaatatatttatttggctcatatttgcatgtaaatactaaatcgtgctaaacttacc
JW713 pLZ12fwd	ggtcataacctgaaggaagatctggatcc
JW714 pLZ12rev	atctgtgccagttcgtaatgtctggtc
JW715_pRH163 F Gib pLZ12	ccagaccagacattacgaactggcacagatttagtaggtttagcaagatggcagcg
JW774_pDB184 R Gib pLZspec	aggatccagatcttccttcaggttatgaccccactttatccaattttcgtttgaac
JW1008_12vFlip 1F	aaacgcttatagcacatccagcaagtgcagggcag
JW1009_12vFlip 1R	aaaactgccctgcacttgctggatgtgctataagc
JW1010_12vFlip 2F	aaacgaactttgttaaagcggtagcgtatcatgcg
JW1011_12vFlip 2R	aaaacgcatgatacgctaccgctttaacaaagttc
JW1016_12vFlip 5F	aaacttgtaagtttaatccatgccctgcacttgcg
JW1017_12vFlip 5R	aaaacgcaagtgcagggcatggattaaacttacaa
JW1020_12vFlip 7F	aaacagttgcatcgtaatgaatttcatctctcaag
JW1021_12vFlip 7R	aaaacttgagagatgaaattcattacgatgcaact
JW790_phi12 tII target 2F	aaacatagtgatttgatgatagagcatacaaacag
JW791_phi12 tII target 2R	aaaactgtttgtatgctctatcatcaaatcactat
JW792_phi12 tII target 3F	aaacaattgatttcagttataaaaaactcaagagg
JW793_phi12 tII target 3R	aaaacctcttgagtttttataactgaaatcaatt
AV108_pT181_F	gcaaaaacaggtttaagcctcgc
AV109_pT181_R	aatgagtggcaaaatgctagcc
JW613_phi12 cos flip UHR R	ttaaatttttaatactctaattttcttaagattacttttg
JW712_JW613-long	taaacttttatgaaaggggggttaaatttttaatactctaattttcttaagattacttttg
JW711_JW616-long	agtaatcttaagaaaattagagtattaaaaatttaacccccctttcataaaagtttatccg
JW617_phi12 cos F Gib DHR	aaagtttatccgcgttgcaaaataaaaaaataatttatttttatagcccc
JW614_phi12 cos flip DHR F	ttgcaacgcggataaacttttatg
JW615_phi12 cos flip DHR R Gib pT181	ctgcgaggcttaaacctgtttttgccgtaatgaatttcatctctcaaaggtc
JW618_phi12 cos flip type III F	gaacggctataaaaataaattatttttttatttaaattt

Primer name	Sequence
JW619_phi12 cos flip type III R	gatcaaatttaaataaaaaaataatttatttttatagcc
JW155_pC194 Hindlll site F	agcttttaaaaagcaaatatgagcc
JW156_pC194 Hindlll site R	tctgtaggtttttaggcataaaactatatg
JW968_phi12 +chi4 UHR F Gib pC194	catatagttttatgcctaaaaacctacagaccaaagatgagattgaccaaatttg
JW969_phi12 +chi4 UHR R Gib DHR	ccccttttactccctccgcttcccgcttcaacactacttttttaaactttatagtgaatttg
JW970_phi12 +chi4 DHR F Gib UHR	ctataaagtttaaaaagtagtgttgaagcgggaagcggagggagtaaaggggtatagggg
JW971_phi12 +chi4 DHR R Gib pC194	tatttggctcatatttgctttttaaaagctggttacatagatatcagcctctgtg
JW966_phi12 +chi4 type II F	aaacaattcactataaagtttaaaaagtagtgttg
JW967_phi12 +chi4 type II R	aaaacaacactacttttttaaactttatagtgaatt
JW703_phi12 qPCR 2F	cgaattagctttaggctatcaag
JW704_phi12 qPCR 2R	gccaactatcaactaaaccttg
W915_rho-f	gtcaatgaccataacgcagaag
W916_rho-r	caatcggtgttactaaatccatg
JW1131_conv F	gcacttttagacaaaaatagtctacgag
L401_conv R	taaccctctttctcaagttatc

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Robert Heler for plasmid pRH163 and Jon McGinn for plasmid pJM62. We are grateful to Dr. Jose Penades for providing ϕ12 as well as assistance to work with it. We also thank Dr. Bernard Dujon for permission to use the I-SceI endonuclease. We would like to acknowledge The Rockefeller University Genomics Resource Center core facility for performing next-generation sequencing. JWM is a Fellow of The Jane Coffin Childs Memorial Fund for Medical Research. LAM is supported by the Rita Allen Scholars Program, an Irma T. Hirschl Award, a Sinsheimer Foundation Award, a Burroughs Wellcome Fund PATH award, an NIH Director's New Innovator Award (1DP2AI104556-01) and a HHMI-Simons Faculty Scholar Award.

References

1. Barrangou R, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 2007; 315:1709–1712. [PubMed: 17379808]
2. Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science. 2008; 322:1843–1845. [PubMed: 19095942]
3. Brouns SJ, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science. 2008; 321:960–964. [PubMed: 18703739]
4. Carte J, Wang R, Li H, Terns RM, Terns MP. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. Genes Dev. 2008; 22:3489–3496. [PubMed: 19141480]
5. Deltcheva E, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature. 2011; 471:602–607. [PubMed: 21455174]
6. Garneau JE, et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature. 2010; 468:67–71. [PubMed: 21048762]
7. Jore MM, et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol. 2011; 18:529–536. [PubMed: 21460843]

8. Samai P, et al. Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. Cell. 2015; 161:1164–1174. [PubMed: 25959775]
9. Makarova KS, et al. An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol. 2015; 13:722–736. [PubMed: 26411297]
10. Levy A, et al. CRISPR adaptation biases explain preference for acquisition of foreign DNA. Nature. 2015; 520:505–510. [PubMed: 25874675]
11. Wigley DB. Bacterial DNA repair: recent insights into the mechanism of RecBCD, AddAB and AdnAB. Nat Rev Microbiol. 2013; 11:9–13. [PubMed: 23202527]
12. Heler R, et al. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. Nature. 2015; 519:199–202. [PubMed: 25707807]
13. Jinek M, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012; 337:816–821. [PubMed: 22745249]
14. Sapranauskas R, et al. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. Nucleic Acids Res. 2011; 39:9275–9282. [PubMed: 21813460]
15. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology. 2009; 155:733–740. [PubMed: 19246744]
16. Datsenko KA, et al. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat Commun. 2012; 3:945. [PubMed: 22781758]
17. Heler R, et al. Mutations in Cas9 enhance the rate of acquisition of viral spacer sequences during the CRISPR-Cas immune response. Mol Cell. 2016; 65
18. Halpern D, et al. Identification of DNA motifs implicated in maintenance of bacterial core genomes by predictive modeling. PLoS Genet. 2007; 3:1614–1621. [PubMed: 17941709]
19. Colleaux L, et al. Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into *E coli* as a specific double strand endonuclease. Cell. 1986; 44:521–533. [PubMed: 3004738]
20. Wegrzyn G, Licznerska K, Wegrzyn A. Phage lambda–new insights into regulatory circuits. Adv Virus Res. 2012; 82:155–178. [PubMed: 22420854]
21. Novick R. Properties of a cryptic high-frequency transducing phage in *Staphylococcus aureus*. Virology. 1967; 33:155–166. [PubMed: 4227577]
22. Hershey AD, Burgi E, Ingraham L. Cohesion of DNA molecules isolated from phage lambda. Proc Natl Acad Sci U S A. 1963; 49:748–755. [PubMed: 16591099]
23. Quiles-Puchalt N, et al. Staphylococcal pathogenicity island DNA packaging system involving cos-site packaging and phage-encoded HNH endonucleases. Proc Natl Acad Sci USA. 2014; 111:6016–6021. [PubMed: 24711396]
24. Fischetti VA. Bacteriophage lysins as effective antibacterials. Curr Opin Microbiol. 2008; 11:393–400. [PubMed: 18824123]
25. McGinn J, Marraffini LA. CRISPR-Cas systems optimize their immune response by specifying the site of spacer integration. Mol Cell. 2016; 64:616–623. [PubMed: 27618488]
26. Bobay LM, Touchon M, Rocha EP. Manipulating or superseding host recombination functions: a dilemma that shapes phage evolvability. PLoS Genet. 2013; 9:e1003825. [PubMed: 24086157]
27. Van Valen D, et al. A single-molecule Hershey-Chase experiment. Curr Biol. 2012; 22:1339–1343. [PubMed: 22727695]
28. Taylor AF, Smith GR. Substrate specificity of the DNA unwinding activity of the RecBC enzyme of *Escherichia coli*. J Mol Biol. 1985; 185:431–443. [PubMed: 2997450]
29. Kreiswirth BN, et al. The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage. Nature. 1983; 305:709–712.
30. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754–1760. [PubMed: 19451168]
31. Martel B, Moineau S. CRISPR-Cas: an efficient tool for genome engineering of virulent bacteriophages. Nucleic Acids Res. 2014; 42:9504–9513. [PubMed: 25063295]

32. Monk IR, Shah IM, Xu M, Tan MW, Foster TJ. Transforming the untransformable: application of direct transformation to manipulate genetically *Staphylococcus aureus* and *Staphylococcus epidermidis*. MBio. 2012; 3
33. Arnaud M, Chastanet A, Debarbouille M. New vector for efficient allelic replacement in naturally nontransformable, low-GC-content, gram-positive bacteria. Appl Environ Microbiol. 2004; 70:6887–6891. [PubMed: 15528558]
34. Ubeda C, et al. Specificity of staphylococcal phage and SaPI DNA packaging as revealed by integrase and terminase mutations. Mol Microbiol. 2009; 72:98–108. [PubMed: 19347993]
35. Jiang W, Samai P, Marraffini LA. Degradation of phage transcripts by CRISPR-associated RNases enables type III CRISPR-Cas immunity. Cell. 2016; 164:710–721. [PubMed: 26853474]
36. Goldberg GW, Jiang W, Bikard D, Marraffini LA. Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. Nature. 2014; 514:633–637. [PubMed: 25174707]
37. Gibson DG, et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009; 6:343–345. [PubMed: 19363495]
38. Bikard D, et al. Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. Nat Biotechnol. 2014; 32:1146–1150. [PubMed: 25282355]
39. Bae T, Schneewind O. Allelic replacement in *Staphylococcus aureus* with inducible counter-selection. Plasmid. 2006; 55:58–63. [PubMed: 16051359]
40. Horinouchi S, Weisblum B. Nucleotide sequence and functional map of pE194, a plasmid that specifies inducible resistance to macrolide, lincosamide, and streptogramin type B antibodies. J Bacteriol. 1982; 150:804–814. [PubMed: 6279574]
41. Bae T, Baba T, Hiramatsu K, Schneewind O. Prophages of *Staphylococcus aureus* Newman and their contribution to virulence. Mol Microbiol. 2006; 62:1035–1047. [PubMed: 17078814]
42. Louwen R, et al. A novel link between *Campylobacter jejuni* bacteriophage defence, virulence and Guillain-Barre syndrome. Eur J Clin Microbiol Infect Dis. 2013; 32:207–226. [PubMed: 22945471]
43. Khan SA, Novick RP. Complete nucleotide sequence of pT181, a tetracycline-resistance plasmid from *Staphylococcus aureus*. Plasmid. 1983; 10:251–259. [PubMed: 6657777]

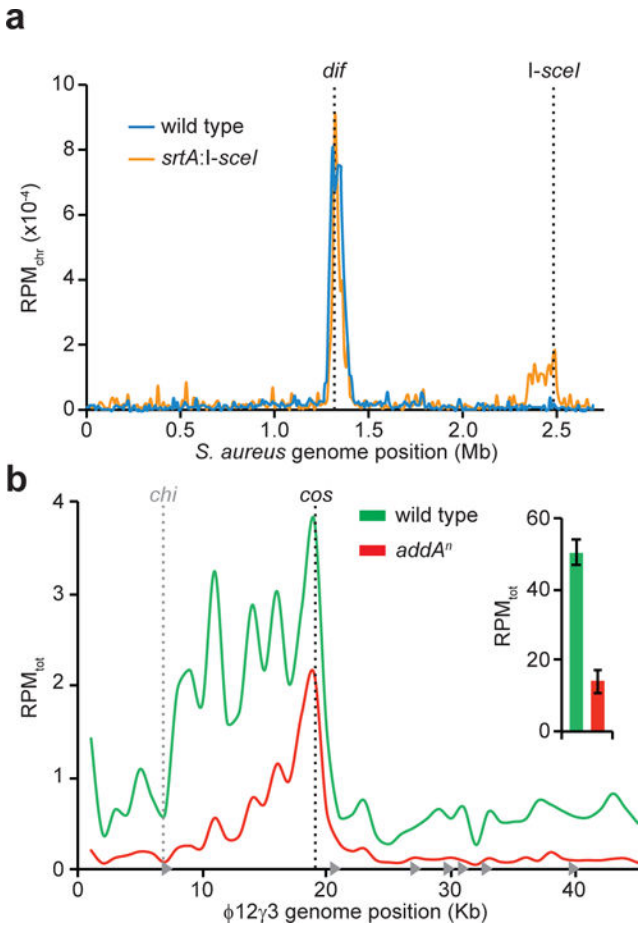


Figure 1. DsDNA ends at the viral *cos* site are hotspots for spacer acquisition during phage infection

a, Abundance (in reads per million, RPM_{chr}) of chromosomal sequences incorporated as spacers into the CRISPR array in wild-type cells (blue) or in cells with an insertion of a I-SceI cleavage site (orange). *dif*, chromosome terminus. **b**, Average (n=3) abundance (RPM_{tot}) of ϕ12γ3 sequences incorporated as spacers into the CRISPR array 30 minutes after infection at MOI 10. *cos*, cohesive end; grey triangles, *chi* sites pointing in the 5′-3′ direction. Insert, average abundance (RPM_{tot}) of total spacer reads of viral origin (n=3). Green and red, spacer reads obtained in wild type and $addA^n$ hosts, respectively.

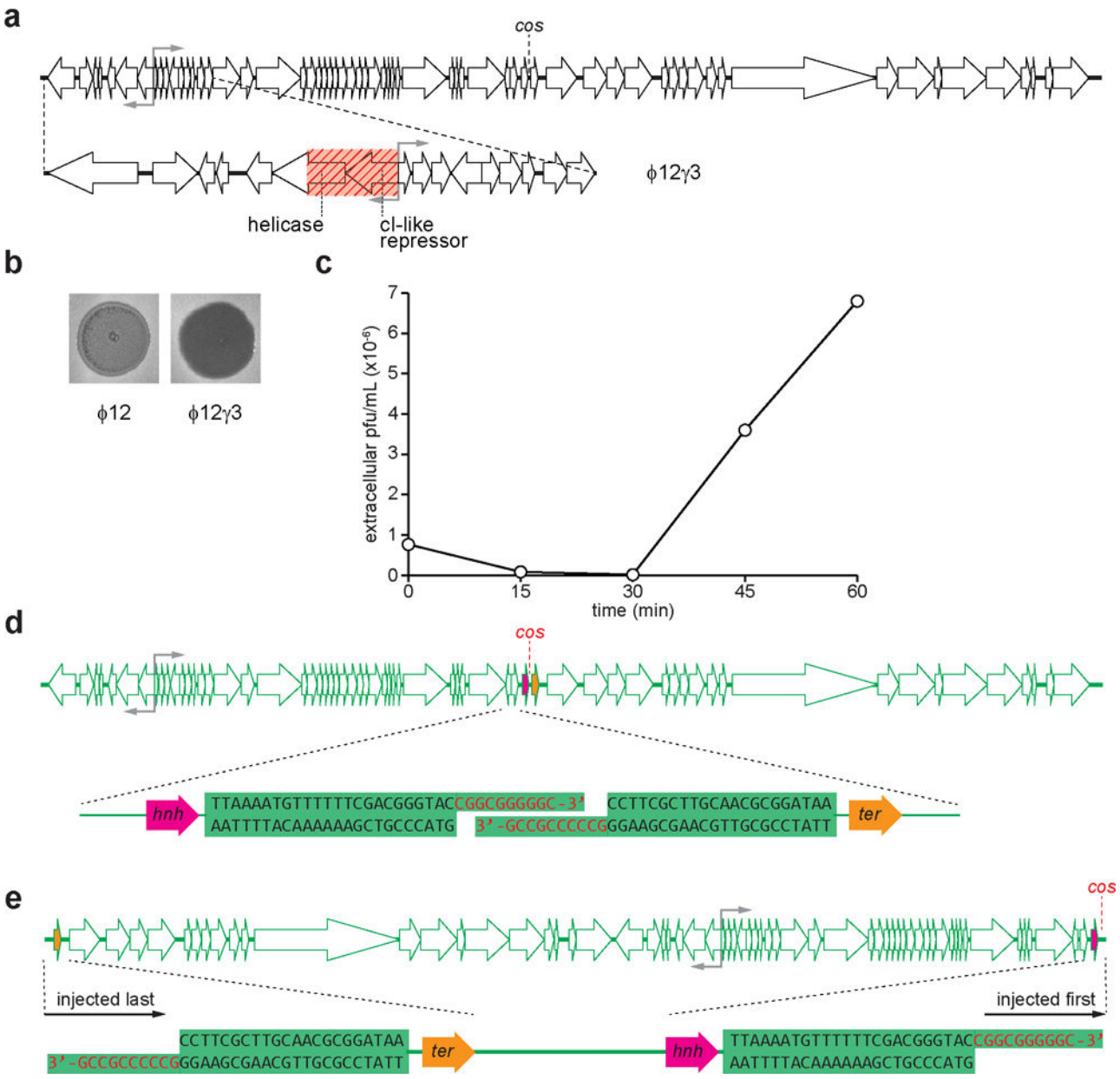


Figure 2. New viral spacers are acquired during DNA injection

a, qPCR amplification of ϕ12 DNA at 0, 30 and 60 minutes after infection (MOI ~ 10) with *polA* (blue), *terS* (red) and *ho/ly* (orange) mutants. Average (n=3) fold-change values relative to the *polA* 30-minute time point value are reported. **b**, Intracellular viral particles measured as plaque forming units (pfu) obtained by lysing hosts at 0, 30 and 60 minutes after infection with wild type ϕ12 or *polA* (blue), *terS* (red) and *ho/ly* (orange) mutants. **c**, Same as panel **b**, but measuring pfu from culture supernatants to quantify extracellular viral particles. **d**, Average ± SD (n=3) % of total spacer reads of viral origin at 0 and 60 minutes after infection with *terS* (red) or *ho/ly* (orange) ϕ12 mutants. **e**, Average ± SD (n=3) % of total spacer reads of viral origin at 0 and 30 minutes after infection with *polA* (blue) or *ho/ly* (orange) ϕ12 mutants. **f**, Construction of ϕ12γ3$^{cos\text{-flip}}$, in which the order of injection of the

viral genome is inverted. **g**, Abundance ($RPM_{\phi 12}$) of ϕ12γ3 (green) or ϕ12γ3$^{cos\text{-flip}}$ (purple) sequences incorporated as spacers into the CRISPR array 30 minutes after infection at MOI 10. *cos*, cohesive end; grey triangles, *chi* sites pointing in 5′-3′ direction.

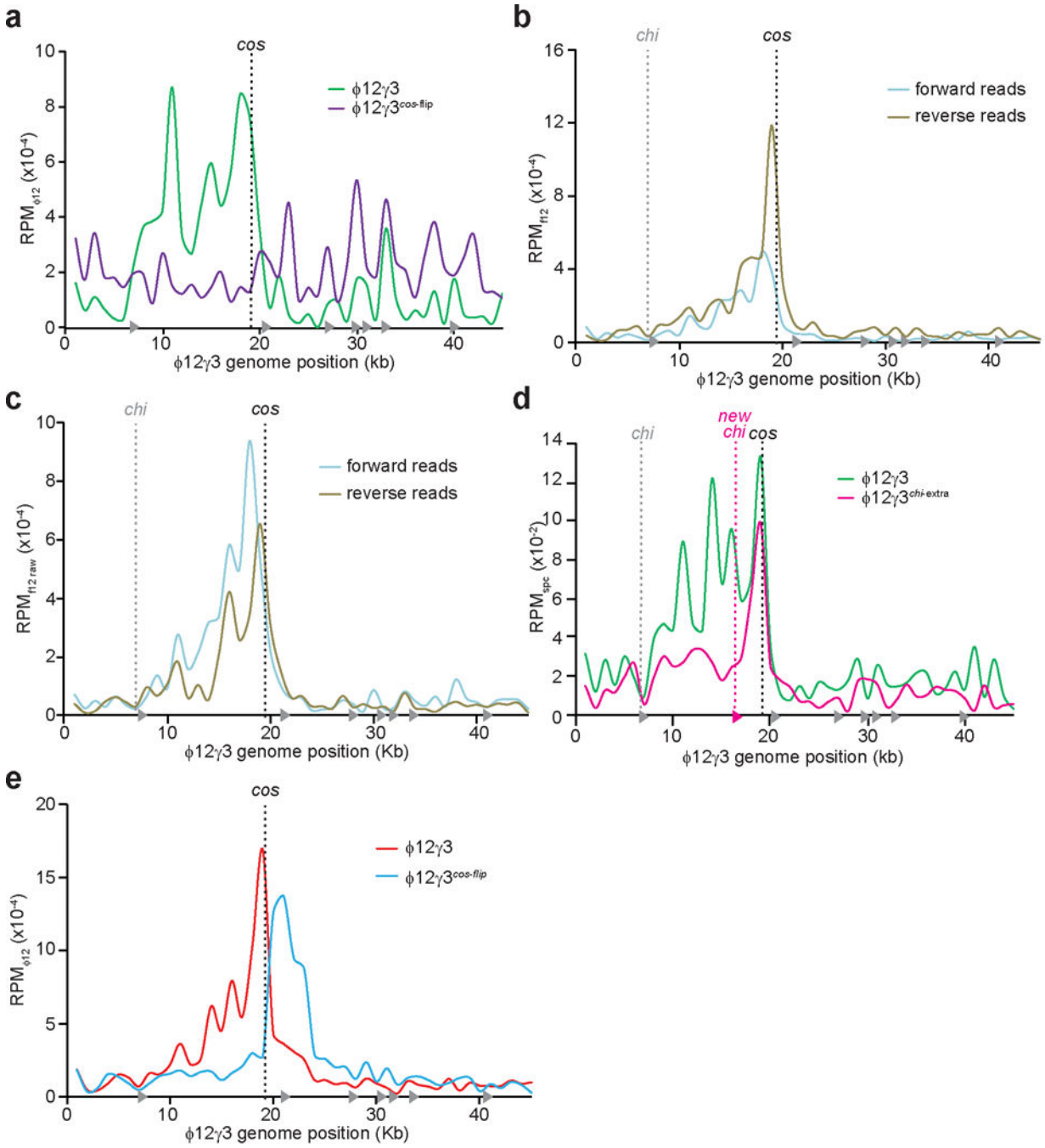


Figure 3. Targeting of early-injected phage genomic regions provides better immunity

a, Abundance ($RPM_{\phi 12}$) of ϕ12γ3 sequences incorporated as spacers into the CRISPR array after generation of an unbiased library of spacers (brown), and after overnight infection of cells harboring this library with ϕ12γ3 (green) or ϕ12γ3$^{cos\text{-flip}}$ (purple). *cos*, cohesive end. **b**, Ratio of ϕ12γ3/ϕ12γ3$^{cos\text{-flip}}$ RPM values obtained in panel **a**. **c**, Design of spacers targeting the regions immediately upstream (pink arrowhead) or downstream (orange arrowhead) of the ϕ12γ3 or ϕ12γ3$^{cos\text{-flip}}$ *cos* site. **d**, Average (n=4) optical density at 600 nm (OD_{600}) of cultures harboring CRISPR-Cas systems targeting the region upstream of the

cos site after infection with ϕ12γ3 (green) or ϕ12γ3$^{cos\text{-flip}}$ (purple) phages. **e**, same as panel **d** but using cells with CRISPR-Cas systems programmed to target the region downstream of the *cos* site.