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

Joshua W. Modell1, Wenyan Jiang1 & Luciano A. Marraffini1

Clustered regularly interspaced short palindromic repeats (CRISPR)–Cas systems provide protection against viral1 and plasmid2 infection by capturing short DNA sequences from these invaders and integrating them into the CRISPR locus of the prokaryotic host1. These sequences, known as spacers, are transcribed into short CRISPR RNA guides3–5 that specify the cleavage site of Cas nucleases in the genome of the invader6–8. It is not known when spacer sequences are acquired during viral infection. Here, to investigate this, we tracked spacer acquisition in Staphylococcus aureus cells harbouring a type II CRISPR–Cas9 system after infection with the staphylococcal bacteriophage φ12. We found that new spacers were acquired immediately after 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 a successful CRISPR immune response.

The acquisition of new spacer sequences from foreign DNA elements, also known as adaptation, is a hallmark of the CRISPR–Cas immune response1. There are six CRISPR–Cas types9, but the molecular mechanisms underlying this process have mostly been studied in the type I CRISPR–Cas system of Escherichia coli. Recent work in this bacterium examined the mechanism of CRISPR–Cas autoimmunity, which involves the acquisition of spacer sequences from the host genome and resident plasmids10,11. In this scenario, new chromosomal spacers are acquired primarily from the terminus of the host genome, owing to the presence of double-stranded DNA breaks (DSBs) that occur during replisome stalling. Adaptation is limited by chromosomal chi sequences and depends strongly on RecBCD, a complex required for the repair of genomic breaks in Gram-negative bacteria12 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 complex13. However, how and when new spacers are acquired from viral invaders during the CRISPR–Cas immune response remains unclear.

Here we studied spacer acquisition in the Gram-positive bacterium S. aureus RN4220, which lacks an endogenous CRISPR system, harbouring a plasmid engineered to carry the Streptococcus pyogenes type II-A CRISPR–Cas locus12 (Extended Data Fig. 1a). Cas9 is the CRISPR RNA (crRNA)-guided nuclease of this system14,15, which requires the presence of a protospacer adjacent motif (PAM) immediately downstream of the target with the sequence NGG15,16. 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 genome15—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 hcas9, which enhances spacer acquisition by two orders of magnitude17. 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 were calculated (full data for all next generation sequencing experiments in this manuscript are provided in Supplementary Data File 1; a summary of these data in table format is provided in Supplementary Data File 2). 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′-GAAGGG-3′)18 upstream from the dif site on each DNA strand (Extended Data Fig. 1b–e). In addition, the introduction of an I-sceI site19 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 use free DNA ends resulting from DSBs as the source of new spacers for CRISPR adaptation.

Are 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 free DNA ends generated? Free DNA ends can be found during 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 packaging20. To address this fundamental question, we analysed the incorporation of new spacers shortly after infection with the lambda-like φ12γ bacteriophage (Extended Data Fig. 2a, b), a lytic derivative of the temperate dsDNA staphylococcal phage φ12 (ref. 21). We first estimated the onset of φ12γ lysis at approximately 45 min (Extended Data Fig. 2c); therefore, we chose to analyse spacer acquisition 30 min after infection, when most phages are still going through the first lytic cycle. In doing so, 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, in experiments with a multiplicity of infection (MOI) of either 10 (Fig. 1b) or 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.

1Laboratory of Bacteriology, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA.
of the linear phage genome after DNA injection\textsuperscript{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 hcas9 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 paralogue of RecBCD\textsuperscript{11}, because S. aureus hosts harbouring a mutation in the nuclease active site of the AddA subunit (addA\textsuperscript{a}) showed a significant reduction in spacer acquisition with a small peak clustering close to the cos site but not reaching the first chi site (Fig. 1b and Extended Data Fig. 3b, c). In addition, the introduction of two extra chi sites, approximately 2.5 kb upstream of the φ12–3 cos site, created a new limit of total spacer reads of viral origin (n = 3). Green and red, spacer reads obtained in wild-type and addA\textsuperscript{a} hosts, respectively.

The results shown in Fig. 1b suggest that new spacers could be acquired soon after the 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\textsuperscript{23} 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 φ12 mutants impaired at different stages of the lytic cycle (Extended Data Fig. 6a–d). polA mutant phages lack DNA polymerase A and therefore cannot replicate the injected genome (Fig. 2a). Additionally, they cannot continue the lytic cycle after injection and produce neither intracellular (Fig. 2b) nor extracellular viral particles (Fig. 2c). These mutants provide a measure of the level of spacer acquisition that occurs after 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\textsuperscript{23}. After DNA injection they can replicate their genome (Fig. 2a) but cannot form either functional intracellular (Fig. 2b) or extracellular viral particles (Fig. 2c). Thus, spacer acquisition in the terS mutants can occur during DNA injection and replication but not before DNA packaging (Extended Data Fig. 6a, b). Finally, phages with a deletion encompassing the genes encoding the holin and lysin enzymes (hol/ly) are impaired at host lysis\textsuperscript{14}, but 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 packaging, but not during a second infectious cycle (Extended Data Fig. 6a, b). First, we compared adaptation levels between terS and hol/ly mutants 0 and 60 min after infection, when DNA packaging is mostly completed in the hol/ly mutant (Fig. 2b).
We did not detect significant differences in spacer incorporation between these two infections, either quantitatively (Fig. 2d) or qualitatively (Extended Data Fig. 6f), 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 0 and 30 min after infection, when DNA levels in the ho/ly mutant are approximately 30-fold higher than in the polA mutant (Fig. 2a).

Both mutants allowed similar spacer incorporation into the CRISPR array, both quantitatively (Fig. 2e) and qualitatively (Extended Data Fig. 6e), 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. This suggests that adaptation rates should align with the MOI, which directly determines how many phage dsDNA ends are injected. We found that adaptation rates were linearly correlated with the MOI (Extended Data Fig. 6g). Additionally, the region of the phage genome that is first injected into the host would be expected to be the source of most of the new spacers. To test this prediction, we inverted the cos site (116 base pairs) of φ12γ-3, generating φ12γ-3cos-flip (Fig. 2f), a phage with an inverted order of entry. In contrast to the pattern of CRISPR adaptation obtained after exposure to 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 occurring downstream of the flipped cos site at MOIs of both 10 (Fig. 2g) and 1 (Extended Data Fig. 3a). Acquisition peaks in φ12γ-3cos-flip were lower but covered a larger area than those in φ12γ-3; this is probably because there are no chi sites in the reverse strand of the φ12 genome to limit and concentrate adaptation activity. Accordingly, spacer acquisition from φ12γ-3cos-flip in the addA′ mutant displayed a major peak within 5 kb of the cos site, a mirror image pattern of that observed for φ12γ-3 (Extended Data Fig. 3e). Together, the data in Fig. 2 demonstrate that the acquisition of spacer sequences by the CRISPR–Cas adaptation machinery occurs primarily from the viral dsDNA ends that are first introduced during the injection of the φ12γ-3 genome.

This pattern of spacer acquisition presents a potential advantage to the immunized host population upon subsequent phage encounters, as the cos-adjacent spacers will direct the Cas9 nuclease to cleave the next invading virus shortly after its injection. To test this idea, we generated a library of spacers that were evenly distributed throughout the φ12γ-3 genome (Fig. 3a and Extended Data Fig. 7a–f). Cells harbouring these spacers were infected with φ12γ-3 or φ12γ-3cos-flip phages in conditions in which the incorporation of additional spacers is inhibited (Extended Data Fig. 7c, d). Staphylococci that survived infection overnight were collected and their spacer content was assessed by next-generation sequencing to determine whether the selective pressure of the phage caused enrichment for spacers targeting any particular region of φ12γ-3. Spacers targeting the region immediately upstream of the cos site—the region that is injected first—were enriched after infection with wild-type φ12γ-3 (Fig. 3a and Extended Data Fig. 7g). The opposite pattern was observed following φ12γ-3cos-flip infection, with the enrichment of spacers targeting the region that is downstream of the flipped cos site (Fig. 3a and Extended Data Fig. 7g). This pattern of selection, which was opposite for spacers targeting different sides of the cos site, is more evident when spacer enrichment is plotted as the ratio of φ12γ-3/φ12γ-3cos-flip RPMs (Fig. 3b and Extended Data Fig. 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 programmed with spacers targeting viral sequences immediately (within 1 kb) upstream or downstream of the cos site (Fig. 3c). During infection with wild-type φ12γ-3, the three upstream spacers target the first region to be injected; during infection with φ12γ-3cos-flip, the order of targeting is inverted (Fig. 3c). In host populations in which all bacteria carry one of these six spacer sequences, CRISPR–Cas immunity allows cells to survive upon infection with either of the two phages at a low MOI (Extended Data Fig. 8). However, the incorporation of new spacers during CRISPR immunization is a very rare event.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 (ref. 25), resulting in extremely high MOIs following the lysis of non-CRISPR cells. We then infected the mixed cultures with either φ12γ-3 or φ12γ-3cos-flip. The population harbouring the three upstream targeting spacers was able to recover and survive infection with both φ12γ-3 and φ12γ-3cos-flip, but recovery was delayed with the latter (Fig. 3d). By contrast, the culture carrying the three downstream targeting spacers recovered rapidly and survived infection with φ12γ-3cos-flip but
not ϕ12:3 (Fig. 3e). These data confirm that the target location can strongly influence the efficiency of Cas9 targeting, especially immediately after CRISPR adaptation, when only a small fraction of cells acquire a new spacer. Together, 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). In addition, 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 the 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 ϕac staphylococcal phage ϕNM4-4, which lacks a defined genomic entry site, also shows a mode of spacer incorporation consistent with that seen in CRISPR adaptation during phage DNA injection (Extended Data Fig. 10c, d), suggesting that our findings apply to both cos and ϕac dsDNA phages. We propose that, although not strictly required for spacer acquisition (Extended Data Fig. 3e), DNA degradation by the AddAB helicase–nuclease 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 complex12 to incorporate new viral sequences into the CRISPR array (Extended Data Fig. 9). We did not detect significant spacer acquisition in the opposite direction of the cos site (from the cos site DNA end that enters the cell last). This suggests that spacer acquisition occurs during DNA ejection. Notably, many dsDNA phages have evolved to protect their DNA ends from this degradation by producing proteins that cap the injected dsDNA end or inhibit degradation by RecBCD or AddAB26. It is plausible 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 protection for the phage against both RecBCD or AddAB degradation and CRISPR adaptation. Recent studies have shown that the injection of the phage lambda genome takes on average 5 min and can last for up to 20 min; this provides a window of opportunity for spacer acquisition before DNA injection is complete. It also allows cleavage of the viral DNA before it is completely ejected, which we have found to be important for an effective type II CRISPR–Cas immune response. Finally, CRISPR–Cas systems also acquire spacers from conjugative plasmids2, most of which transfer to neighbouring cells as ssDNA. Since ssDNA is not a good substrate for RecBCD28, it is unlikely that spacer acquisition happens during plasmid transfer; rather, it is likely to occur during plasmid replication, as reported previously19. In conclusion, our study highlights the close 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 prophagocytic predators.

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

Received 13 December 2016; accepted 22 February 2017. Published online 29 March 2017.

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. Brouns, S. J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321, 960–964 (2008).
4. Carte, J., Wang, R., Li, H., Terns, R. M. & Terns, M. P. Cas9 is an endonuclease that generates guide RNAs for invader defense in prokaryotes. Genes Dev. 28, 3489–3496 (2008).
5. Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor Rnase III. Nature 471, 602–607 (2011).
6. Garneau, J. E. et al. The CRISPR/Cas bacterial immune system cleaves DNA by a homing endonuclease mechanism. Nature 468, 67–71 (2010).
7. Jore, M. et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat. Struct. Mol. Biol. 18, 529–536 (2011).
8. Samai, P. et al. Co-transcriptional DNA and RNA cleavage during type III CRISPR–Cas immunity. Cell 161, 1164–1174 (2015).
9. Makarova, K. S. et al. An updated evolutionary classification of CRISPR–Cas systems. Nat. Rev. Microbiol. 13, 722–736 (2015).
10. Levy, A. et al. CRISPR adaptation biases explain preference for acquisition of foreign DNA. Nature 520, 47–50 (2015).
11. Wigley, D. B. Viral DNA repair: recent insights into the mechanism of RecBCD, AddAB and AdnAB. Nat. Rev. Microbiol. 11, 9–13 (2013).
12. Heler, R. et al. Cas9 specifies functional viral targets during CRISPR–Cas adaptation. Nature 519, 189–202 (2015).
13. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816–821 (2012).
14. Sapraunskas, R. et al. The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic Acids Res. 39, 9275–9282 (2011).
15. Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the target of the prokaryotic CRISPR defence system. Microbiology 155, 733–740 (2009).
16. Oates, K. A. et al. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat. Commun. 3, 945 (2012).
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 65, 168–175 (2017).
18. Halpern, D. et al. Identification of DNA motifs implicated in maintenance of bacterial core genomes by predictive modeling. PLoS Genet. 3, 1614–1621 (2007).
19. Colleux, 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 44, 521–533 (1986).
20. Wegrzyn, G., Licznerska, K. & Wegrzyn, A. Phage λ—new insights into regulatory circuits. Adv. Virus Res. 62, 155–178 (2012).
21. Novick, R. Properties of a cosmetic high-frequency transducing phage in Staphylococcus aureus. Virology 33, 155–166 (1967).
22. Hershey, A. D., Burgi, E. & Ingraham, L. Cohesion of DNA molecules isolated from phage lambda. Proc. Natl Acad. Sci. USA 49, 748–755 (1963).
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 111, 6101–6102 (2014).
24. Frischti, V. A. Bacteriophage lysins as effective antibacterials. Curr. Opin. Microbiol. 11, 393–400 (2008).
25. 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).
26. Bobay, L. M., Touchon, M. & Rocha, E. P. Manipulating or superseding host recombination functions: a dilemma that shapes phage evolvability. PLoS Genet. 9, e1003825 (2013).
27. Van Valen, D. et al. A single-molecule Hershey-Chase experiment. Curr. Biol. 22, 1339–1343 (2012).
28. Taylor, A. F. & Smith, G. R. Substrate specificity of the DNA unwinding activity of the RecBCD enzyme of Escherichia coli. J. Mol. Biol. 185, 431–443 (1985).

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

Acknowledgements We thank R. Heler for plasmid pPH163; J. McGinn for plasmid pJM62; J. Penades for providing e12 as well as assistance in working with it; B. Dujon for permission to use the l-Scl endonuclease; and The Rockefeller University Genomics Resource Center core facility for performing next-generation sequencing. J.W.M. is a Fellow of The Jane Coffin Childs Memorial Fund for Medical Research. L.A.M. is supported by the Rita Allen Scholars Program, an Irma T. Hirschl Award, a Siminsher Foundation Award, a Burroughs Wellcome Fund PATH award, an NIH Director’s New Innovator Award (1DP2AI104556-01) and an HHMI-Simons Faculty Scholar Award.

Author Contributions J.W.M. and L.A.M. conceived the study and designed experiments. J.W.M. and W.J. designed the spacer library construction method. W.J. performed the CRISPR immunization simulation assay. All other work was executed by J.W.M. and L.A.M. and J.W.M. wrote the paper with the help of W.J.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to L.A.M. (marraffini@rockefeller.edu).

Reviewer Information Nature thanks R. Barrangou, J. Doudna and the other anonymous reviewer(s) for their contribution to the peer review of this work.
METHODS

Bacterial strains and growth conditions. Cultivation of S. aureus RN4220 and E. coli cells were carried out in brain–heart infusion (BHI) medium or LB liquid medium (BD), respectively, at 37 °C, 30 °C for maintenance of pMAY-based plasmids. S. aureus medium was supplemented with chloramphenicol at 10 μg ml⁻¹, erythromycin at 10 μg ml⁻¹, kanamycin at 25 μg ml⁻¹ or spectinomycin at 250 μg ml⁻¹. E. coli medium was supplemented with chloramphenicol at 25 μg ml⁻¹, ampicillin at 100 μg ml⁻¹, kanamycin at 50 μg ml⁻¹ or spectinomycin at 50 μg ml⁻¹ for plasmid maintenance. Media were supplemented with 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) to express genes from the inducible promoters PpWJ3 and PpWJ4 and with 2 μg ml⁻¹ 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 P1 buffer supplemented with 107 μg ml⁻¹ lysozymaph (AMBI Products) and incubated at 37 °C for 10 min followed by the standard QIAprep protocol. We used 150 ng (log phase) or 50 ng (overnight) of plasmid as input for the enrichment PCR CR of the CRISPR locus using Phusion DNA Polymerase (Thermo) with the following primer mix: three parts JW8 or JW419 and one part each JW3, JW4 and JW5 (Supplementary 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 Aligner30. Perfectly aligned spacers and subjected to Illumina high-throughput sequencing with either the MiSeq, HiSeq or NextSeq platforms.

Strain construction. The addA+ (D1159A) nuclease domain mutation was introduced into the S. aureus RN4220 genome (generating strain JW418), using the pMAY allelic replacement system31 with plasmid pJW233. An l-tet and polA site immediately adjacent to a kanamycin-resistance gene (for selection purposes) was introduced into the S. aureus RN4220 genome at the attR locus (generating strain JW263), using the pMAD allelic replacement system31 with plasmid pJW480. Mutations in φ12 were introduced into strains of S. aureus RN4220 harbouring the φ12 prophage, using the pJW244/pJW327 allelic replacement system developed in our laboratory by W.J. A pJW244-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 and grown at 37 °C. pJW327-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 either pJW244 or pJW327 were re-struck onto a second chloramphenicol plate and grown at 37 °C to remove non-transformed, background cells and their genotype was confirmed by PCR using primers that amplify the integration site. Bona fide co-integrants were grown overnight in plain medium at 28 °C. Logarithmic-phase cultures were treated with the pJW326 counter-selection phagemid, producing a Ca99 nuclease programmed to target and destroy the cat gene of the integrated plasmid for 1 h 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 sensitivity to chloramphenicol).

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 pJW326 and to alleviate this counter-selection plasmid. Cultures were plated on plain BHI agar at 37 °C overnight, 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 pJW217, respectively.

Propagation of φ12 mutants. The φ12 mutant lysogenic strains JW924, JW833 and JW563, harbouring the complementing plasmids pJW474, pJW383 and pJW278, respectively, were induced from logarithmic phase cultures by the simultaneous addition of 2 μg ml⁻¹ mitomycin C and 1 mM IPTG for 3 h. Supernatants were filtered through 0.45-μm syringe filters (Acrodisc), concentrated with Amicon Ultra centrifugal filter units (EMD Millipore) and resuspended in BHI. In order to generate phage stocks that were free of mitomycin C-induced DNA damage, logarithmic phase cultures of S. aureus RN4220 harbouring the complementing plasmids pJW474, pJW383 and pJW278 were infected with the PBS phage stock obtained from strains JW924, JW833 and JW563, respectively, at an MOI of 1 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 sensitivity to chloramphenicol). The phage stock was propagated on S. aureus RN4220 and used to infect culture media containing mitomycin C (1 μg ml⁻¹). After 3 h, final phage stocks were prepared by filtration of the supernatant.

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 and data adaptation bar graphs were normalized to this value. Supplementary Data File 2 shows the exact MOI for each experiment.

qPCR. Total DNA was extracted from S. aureus cells using a modified phenol-chloroform-isoamyl alcohol (Fisher) protocol as described32. 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 JW278/JW279 (φ12 ρho/ly control). The sequences of the primers used in this study are listed in Supplementary 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 p.f.u. were obtained by filtering culture
supernatants through a 0.45-μm syringe filter. Samples for measuring intracellular p.f.u. were obtained by spinning cells at 10,000 r.p.m. for 1 min, resuspending cell pellets in BHI supplemented with lysostaphin (167 μg ml⁻¹) and incubating the suspensions at 37°C for 5 min. Samples were stored on ice until all time points were recorded. Finally, samples were briefly vortexed and spun at 10,000 r.p.m. for 1 min, and tenfold dilutions were made from the supernatant.

**Spacer library generation.** Total phage DNA was extracted by a modified phenol-chloroform-isoamyl alcohol extraction protocol. Phages were concentrated with Amicon Ultra centrifugal filter units to approximately 50 μl PCI was added and the DNA extraction protocol was followed as described without bead beating. 50 μg of phage DNA was sonicated in 130 μl total volume in microTUBE AFA Fibre Pre-Slit Snap-Cap 6 × 16 mm tubes (Covaris) using the Covaris S220 Focused-ultrasonicator to a fragment size of 150bp. The sonicated DNA was dialyzed and 30 μl (approximately 20 μg) was electroporated into 50 μl of electropotentient S. aureus RN4220 cells carrying the pRH223 and pRH240 plasmids (L6308). L6308 cells were pretreated with 0.25 μg ml⁻¹ ampicillin and 50 μg ml⁻¹ tetracycline hydrochloride for 8 min before being made electropotentient to induce cas1, cas2 and con2. Following electroporation, cells were outgrown in plain BHI at 37°C for 3h and CaCl₂ was added for 20 min before infection with live phages.

**CRISPR immunization simulation assay.** After overnight inoculations in tryptic soy broth (TSB) supplemented with spectinomycin, S. aureus RN4220 strains containing either 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 and 0.5 μl of this mix was added to 150 μl TSB supplemented with calcium chloride in quadruplicate, in a 96-well cell culture plate (Cellstar, 655180). Following 1h of shaking at 37°C in a microplate reader (TECAN Infinite 200 PRO), γ₁₂ genomic DNA was amplified with JW155/JW156, and a one-piece Gibson assembly was performed. To make pJW481, pT181 was amplified with AV108/AV109, 0.12 genomic DNA was amplified with JW613/JW712, JW711/JW617 and JW614/JW615, and a four-piece Gibson assembly was performed. To make pJW482, oligonucleotides JW618 and JW619 were annealed and cloned into the BsaI site of pJW462. To make pJW484, pJW485 was amplified with W155/W156, 0.12 genomic DNA was amplified with JW613/JW69 and JW970/JW971, and a three-piece Gibson assembly was performed. To make pJW483, oligos JW966 and JW967 were annealed and cloned into the BsaI site of pDH114.

**Data availability.** The analysed DNA sequencing data that support the findings of this study are provided in Supplementary Data File 1. All other data are available from the corresponding author upon reasonable request.
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 positions of the primers used to enrich for PCR products containing expanded CRISPR loci. R, repeat; S, new spacer. 

b. Abundance (RPMchr) of chromosomal sequences incorporated as spacers from wild-type cells in Fig. 1a (close-up on the dif site, the chromosomal terminus). Data previously reported for spacer acquisition by the type I CRISPR–Cas system of *E. coli* showed the accumulation of new spacer sequences at two hotspots adjacent to the two major ter sites at the chromosomal terminus. ter sites are not characterized in *S. aureus*, but the absence of multiple peaks indicates that either DSB patterns at the terminus are more distributed in *S. aureus*, or 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) or reverse (bottom) strand of the chromosome. Blue dots represent spacers derived from sites with a 5′-NGG-3′ PAM immediately downstream of the 3′ end of the spacer. Red dots represent spacers with a non-NGG PAM. Insert, percentage of normalized spacers with or without 3′ PAMs. 

d. Abundance (RPMchr) 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 (RPMchr) of chromosomal spacers in cells harbouring wild-type cas9 following a 30-min infection with φ12 at an MOI of 100. 

f. Abundance (RPMchr) of chromosomal sequences incorporated as spacers in cells with an I-sceI site with or without I-SceI expression (orange and blue, respectively; close-up of the I-sceI recognition site). 

g. Abundance (RPMchr) of chromosomal sequences, derived from the forward (light blue) or reverse (olive) strand, incorporated as spacers in srtA I-sceI cells from Fig. 1b (close-up of the I-sceI recognition site). 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. © 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
Extended Data Figure 2 | Generation of $\phi 12\gamma 3$. a, Genome organization of the staphylococcal temperate phage $\phi 12$. 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. $\phi 12\gamma 3$ contains a deletion (red) of the cl-like repressor gene (and part of a neighbouring helicase gene) that prevents the establishment of lysogeny. 
b, $\phi 12$ forms turbid plaques owing to its ability to lysogenize and form colonies that are resistant to superinfection. By contrast, $\phi 12\gamma 3$ forms clear plaques owing to its inability to lysogenize. 
c, Extracellular viral particles measured as p.f.u. obtained from culture supernatants 0, 15, 30, 45 and 60 min after infection with $\phi 12\gamma 3$. After an initial period in which the phage particles in the medium decrease because of phage adsorption, the p.f.u. value increases after the burst of the infected cells. 
d, 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 $\phi 12$ with the cleaved cos sites at either end of the genome. During injection, the bnh (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 (RPMφ12γ3) of φ12γ3 (green) or φ12γ3cos-flip (purple) sequences incorporated as spacers into the CRISPR array 30 min after infection at an MOI of 1. Grey triangles, chi sites pointing in 5′–3′ direction. **b**, Abundance (RPMφ12γ3) of φ12γ3 forward (light blue) or reverse (olive) strand sequences incorporated into the CRISPR array from the triplicate addA experiment in Fig. 1c.  
**c**, As in b, but showing abundance in RPMφ12γ3raw, which does not normalize reads to account for the GG content in each 1-kb bin. chi, first chi site upstream of the cos site.  
**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 φ12γ3chi-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 substantially. Abundance (RPMφ12γ3) of φ12γ3 (green) or φ12γ3chi-extra (pink) sequences incorporated as spacers into the CRISPR array 30 min after infection at an MOI of 10 is shown. **e**, Abundance (RPMφ12γ3) of φ12γ3 (red) or φ12γ3cos-flip (blue) sequences incorporated as spacers following a 30-min infection of addA cells at an MOI of 10. Despite the absence of chi sites pointing in the 3′–5′ direction, spacer acquisition from φ12γ3cos-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 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 wild-type cas9. a, Abundance (RPMφ12γ3) of φ12γ3 sequences incorporated into the CRISPR array after overnight infection at an MOI of 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 data sets in a plotted with RPMφ12γ3 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 (RPMφ12γ3) of φ12γ3 sequences incorporated into the CRISPR array after a 30-min infection at an MOI of 100 of cells harbouring hcas9 (purple) or wild-type cas9 (green). d, Individual spacers common to both data sets in c were plotted with RPMφ12γ3 values for hcas9 on the y axis and wild-type cas9 on the x axis. 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) or reverse (bottom) strand of φ12γ3 following a 30-min infection at an MOI of 10. Blue dots represent spacers associated with NGG PAMs; red dots represent spacers with non-NGG flanking sequences. Insert, percentage of normalized spacers with or without canonical NGG PAMs.

b, Abundance (RPM) 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, As in b, but showing abundance in RPMφ12γ3 raw, which does not normalize reads to account for the GG content in a given 1-kb bin. Grey triangles, chi sites pointing in 5′–3′ direction.

d, The number of 5′-NGG-3′ PAM sites within 1-kb bins on the forward (light blue), reverse (olive) or combined (green) strands of the φ12γ3 genome.

e, From one of the wild-type replicates in Fig. 1b, the percentage of PAMs within each 1-kb 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. Grey triangles, chi sites pointing in 5′–3′ direction.
Extended Data Figure 6 | See next page for caption.
Extended Data Figure 6 | Generation and spacer acquisition profile of φ12 mutant phages. a, Stages of the φ12 lytic cycle. Blue rhombi indicate the locations 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. The function of terS (encoding the small terminase subunit) was eliminated through an in-frame deletion that left only the first and last 15 codons of the gene. The ho/ly 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 harbouring the mutant prophages, which were incapable of forming and releasing 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 (RPMφ12) of φ12:polA (blue) or φ12:ho/ly (orange) sequences incorporated as spacers into the CRISPR array 30 min after infection at an MOI of 10. Grey triangle and dotted line mark the first 5′–3′ chi site upstream of the cos site. f, Abundance (RPMφ12) of φ12:terS (red) or φ12:ho/ly (orange) sequences incorporated as spacers into the CRISPR array 60 min after infection at an MOI of 10. Grey triangle and dotted line mark the first 5′–3′ chi site upstream of the cos site. g, The percentage of φ12-γ3 spacers (calculated as the ratio of φ12-specific spacer reads to the total spacer reads) plotted as a function of the MOI.
Extended Data Figure 7 | See next page for caption.
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 bp, which were introduced through electroporation (with a water control) into cells harbouring 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 were recovered for 3.3 h in ATC-free medium and treated with live φ12γ3 phage for 15 min. 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 colony-forming units (c.f.u.) per 100 μl, none of which had incorporated a new spacer (data not shown). This demonstrates that at least after 3.3 h without the aTc inducer, staphylococci harbouring 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 φ12γ3cos-flip. Conversely, cells transformed with sonicated φ12γ3 DNA yielded 327 c.f.u. per 100 μl, and all of those tested (n = 4) carried an expanded CRISPR array, indicating that substantial 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 φ12γ3cos-flip (not shown) at MOIs of 0, 10 or 100 for 24 h 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 did not enlarge the CRISPR array with viral spacers during this period. By contrast, strong PCR products corresponding to expanded CRISPR loci were observed following overnight phage infections, demonstrating the enrichment of adapted cells during CRISPR–Cas targeting, at MOIs of both 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. S1 and S0 corresponds to CRISPR loci with and without a new spacer sequence, respectively. E, The data from Fig. 3a with the abundance (RPMφ12raw) of spacers derived from forward (light blue) and reverse (olive) strands of φ12γ3. f, As in e, showing relative (to PAM content) abundance in RPMφ12. g, Same as Fig. 3a, showing the (RPMφ12) values for the full phage genome. h, Same as Fig. 3b, showing the φ12γ3/φ12γ3cos-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 φ12γ3cos-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 harbouring each of these CRISPR–Cas systems were infected at an MOI of 10 with either φ12γ3 (green) or φ12γ3cos-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 free DNA end injected (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 min after infection at an MOI of 10. This is similar to the results obtained at 30 min (Fig. 1c) but comparable to the infection conditions of d. The area highlighted with a yellow gradient shows the expected spacer acquisition pattern. 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 progressively 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 greater than 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 pac site. Therefore, the last sequence to be packaged into the phage capsid and 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 min 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 the observed ranges of pac phage processivity. Grey triangles, chi sites pointing in the 5′–3′ direction, with the dotted line marking the first chi site upstream of the cos site.