ARTICLES
https://doi.org/10.1038/s41564-021-00996-8
nature microbiology

Check for updates

Prophage integration into CRISPR loci enables evasion of antiviral immunity in *Streptococcus pyogenes*

Andrew Varble[1]✉, Edmondo Campisi[2], Chad W. Euler[2,3], Pascal Maguin[1], Albina Kozlova[1], Jessica Fyodorova[1], Jakob T. Rostøl[1], Vincent A. Fischetti[2] and Luciano A. Marraffini[1,4]✉

CRISPR loci are composed of short DNA repeats separated by sequences, known as spacers, that match the genomes of invaders such as phages and plasmids. Spacers are transcribed and processed to generate RNA guides used by CRISPR-associated nucleases to recognize and destroy the complementary nucleic acids of invaders. To counteract this defence, phages can produce small proteins that inhibit these nucleases, termed anti-CRISPRs (Acrs). Here we demonstrate that the ΦAP1.1 temperate phage utilizes an alternative approach to antagonize the type II-A CRISPR response in *Streptococcus pyogenes*. Immediately after infection, this phage expresses a small anti-CRISPR protein, AcrIIA23, that prevents Cas9 function, allowing ΦAP1.1 to integrate into the direct repeats of the CRISPR locus, neutralizing immunity. However, *acrIIA23* is not transcribed during lysogeny and phage integration/excision cycles can result in the deletion and/or transduction of spacers, enabling a complex modulation of the type II-A CRISPR immune response. A bioinformatic search identified prophages integrated not only in the CRISPR repeats, but also the cas genes, of diverse bacterial species, suggesting that prophage disruption of the CRISPR–*cas* locus is a recurrent mechanism to counteract immunity.

Clustered, regularly interspaced, short, palindromic repeat (CRISPR) loci and CRISPR-associated (Cas) proteins provide prokaryotic organisms with adaptive immunity against invading nucleic acids, including viruses[1] and plasmids[2]. CRISPR loci are composed of short direct repeats (DRs) separated by similarly short sequences, known as spacers, that match the genomes of the invader. Usually, the terminal repeat (DRt) is not identical to the rest of the DRs and marks the end of the CRISPR array[3]. In the first phase of CRISPR immunity, new spacer sequences are acquired from the infecting DNA and integrated between two repeats in the CRISPR array[1]. In the second phase, spacers are transcribed and processed to generate RNA guides used by Cas nucleases to recognize and destroy the complementary nucleic acids of invaders (also known as protospacers)[2,4,5]. Depending on the content of CRISPR-associated (*cas*) genes, CRISPR loci are classified into six different types and multiple subtypes[6]. The type II-A CRISPR–Cas system of *Streptococcus pyogenes* has been extensively studied, not only for its role in phage defence[7,8], but also for the biotechnological applications of its signature RNA-guided nuclease, Cas9[9–11]. During phage infection, a small number of cells acquire a 30-nucleotide spacer sequence from the viral genome, which is integrated between the 36-nucleotide DRs[12–14]. DRs and spacers are transcribed into a single pre-CRISPR-RNA (pre-crRNA), which is subsequently processed into mature crRNA guides after interaction of the DR and the 5' end of the trans-activating crRNA (tracrRNA, a small RNA cofactor of Cas9)[15]. The double-stranded RNA formed through this interaction is cleaved by RNase III[15] to create the mature crRNA used by the Cas9 nuclease to find complementary targets in invading DNA[10]. Target recognition requires the presence of a short motif flanking the protospacer (protospacer adjacent motif, PAM)[16], and leads to cleavage and destruction of the infecting genome[5].

Phages have evolved anti-CRISPR (Acr) proteins to prevent the function of different type II Cas9 nucleases[17–22]. Acrs are typically present in temperate phages and are robustly transcribed in both the lytic and lysogenic cycles immediately after infection[23,24]. However, high levels of transcription of the *acr* operon interfere with the proper transcription of other viral genes required for the later stages of the infection cycle and must be reduced shortly after infection. This is accomplished by the anti-CRISPR-associated (aca) repressor, a protein containing a helix-turn-helix DNA-binding domain that regulates the transcription of the operon[24,25]. Interestingly, Aca repression is not complete, allowing prophages to produce low levels of Acr that can inhibit CRISPR immunity in the lysogen[23].

S. pyogenes (also known as the group A streptococcus), an important human pathogen, is the causative agent of a variety of diseases such as scarlet fever, pharyngitis, impetigo, cellulitis, necrotizing fasciitis and toxic shock syndrome[26]. A large number of the virulence genes responsible for *S. pyogenes* pathogenicity are harboured within prophages integrated throughout the chromosome[27]. Integration typically occurs after the injection and circularization of the viral genome, during the lysogenic cycle of temperate phages[28]. It usually requires recombination between specific sequences in the phage (*attP*) and the bacterial (*attB*) DNA, which generates two new sites at the left (*attL*) and right (*attR*) borders of the inserted prophage. Changes in the environment causing bacterial stress can induce the lytic cycle of the prophage, resulting in the excision and replication of the viral genome, followed by the formation and release of new viral particles that can infect other hosts[28]. In *S. pyogenes*, this process contributes to the spread of the virulence genes

[1]Laboratory of Bacteriology, The Rockefeller University, New York, NY, USA. [2]Laboratory of Bacterial Pathogenesis and Immunology, The Rockefeller University, New York, NY, USA. [3]Department of Medical Laboratory Sciences, Hunter College, CUNY, New York, NY, USA. [4]Howard Hughes Medical Institute, The Rockefeller University, New York, NY, USA. ✉e-mail: avarble@rockefeller.edu; marraffini@rockefeller.edu

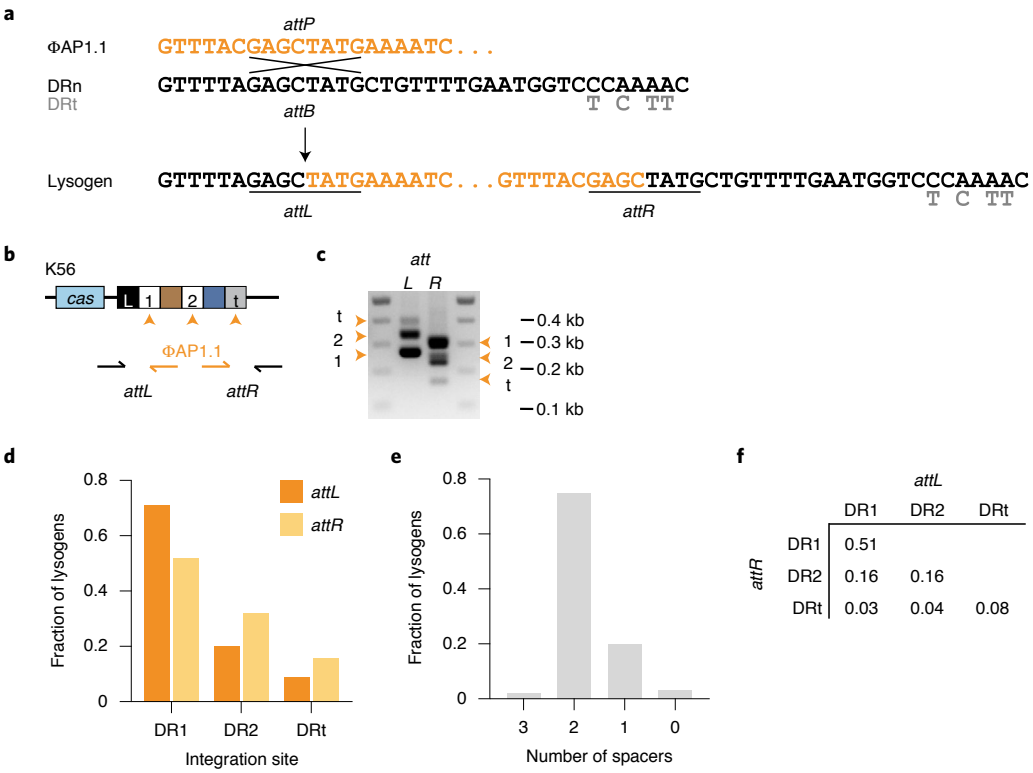


Fig. 1 | ΦAP1.1 lysogenization into the CRISPR locus. a, Recombination between the *attP* site of ΦAP1.1 (orange) and the *attB* site present in the direct repeat (DR, black) of *S. pyogenes* type II-A CRISPR loci. The mutations of the terminal repeat (DRt) are shown in grey. Recombination results in prophage integration and the generation of *attL* and *attR* sites that will recombine during prophage induction and excision. **b**, Type II-A CRISPR–*cas* locus of *S. pyogenes* K56, showing DR1, DR2 and DRt as white or grey boxes; spacers as coloured boxes. Orange arrowheads indicate the three possible integration sites of ΦAP1.1 in the different repeats. Arrows indicate the primers that anneal on the host (black) or phage (orange) genome to amplify the resulting *attL* or *attR* sites. **c**, Agarose gel electrophoresis of the *attL* or *attR* PCR products obtained after the amplification shown in **b**. Orange arrowheads indicate the inferred integration sites. Representative of three independent replicates. **d**, Frequency of lysogens ($n=100$) carrying *attL* or *attR* sites in each of the DRs of the K56 CRISPR locus. **e**, Frequency of lysogens ($n=100$) carrying different numbers of spacers. **f**, Same as **d** but showing the combined distribution of *attL* or *attR* sites.

carried by the temperate phages[27]. A bioinformatic study reported two closely related *S. pyogenes* strains (SSI-1 and MGAS315) harbouring prophages inserted into the repeats of their type II-A CRISPR array[29] – an observation that suggests a previously unidentified mechanism used by phages to inhibit CRISPR. Here we studied a member of this temperate phage family, ΦAP1.1, to determine whether and how type II-A CRISPR–Cas immunity is affected by prophage integration into the CRISPR repeats.

Results

The temperate phage ΦAP1.1 integrates into the repeat sequences of the *S. pyogenes* type II-A CRISPR array. In agreement with previous analyses[29,30], we found 12 different *S. pyogenes* strains harbouring almost identical prophages inserted into the type II-A CRISPR array (Extended Data Fig. 1a). In most cases, these genomes contained a single degenerate terminal repeat (DRt) that was interrupted by the viral genome. To understand whether and how temperate phages integrate into the type II-A CRISPR locus, we isolated the phage ΦAP1.1 from strain AP1[31] (Extended Data Fig. 1b), a *cos* phage harbouring an *attP* sequence with 8 bp identical to nucleotides 7–14 of the DR, and DRt, which is its *attB* site (Fig. 1a). To facilitate the isolation and quantification of lysogens, we added a spectinomycin resistance cassette to ΦAP1.1 (Extended Data Fig. 1b). This phage was then used to infect *S. pyogenes* K56[32], a strain containing a CRISPR array of 2 spacers, 2 DRs and one DRt (Fig. 1b). We extracted genomic DNA from bacteria collected 24 h after infection and performed PCR with primers that specifically amplify the *attL* and *attR*. We obtained three PCR products separated by approximately 66 bp, indicative of phage integration into all the repeats, with a preference for DR1, DR2 and then DRt, according to the intensity of the DNA band (Fig. 1c). A similar analysis of 100 individual spectinomycin-resistant colonies (presumably lysogens) showed that indeed, ΦAP1.1 integrated into every direct repeat, but with a marked preference for DR1 (Fig. 1d). Prophage integration typically conserved the original repeat-spacer content, with 75/100 lysogens maintaining the two original spacers (Fig. 1e) where the *attL* and *attR* are inserted into the same repeat (Fig. 1f). A fraction of the insertion events led to CRISPR rearrangements that resulted in either the deletion of one or both spacers (23/100, Fig. 1e) where the *attL* and *attR* were located across different repeats (Fig. 1f), or in spacer duplication events (2/100, Fig. 1e and Extended Data Fig. 1c). We also deleted the integrase gene of this phage, generating ΦAP1.1$^{\Delta int}$, which failed to produce lysogenic integration into the CRISPR array (Extended Data Fig. 1d,e). Finally, after analysis of 110 spectinomycin-resistant colonies, we were unable to detect prophage integration into other *attB*-like sites of the genome (Extended Data Fig. 2). These data demonstrate that although ΦAP1.1-like phages are typically detected within a single DRt, they are able to integrate into any repeat of a CRISPR array using a site-specific recombinase, with a preference for the first DR.

ΦAP1.1 integration disrupts spacer transcription and crRNA processing. To determine whether lysogenization could be possible in the presence of a spacer that targets the infecting phage, we used

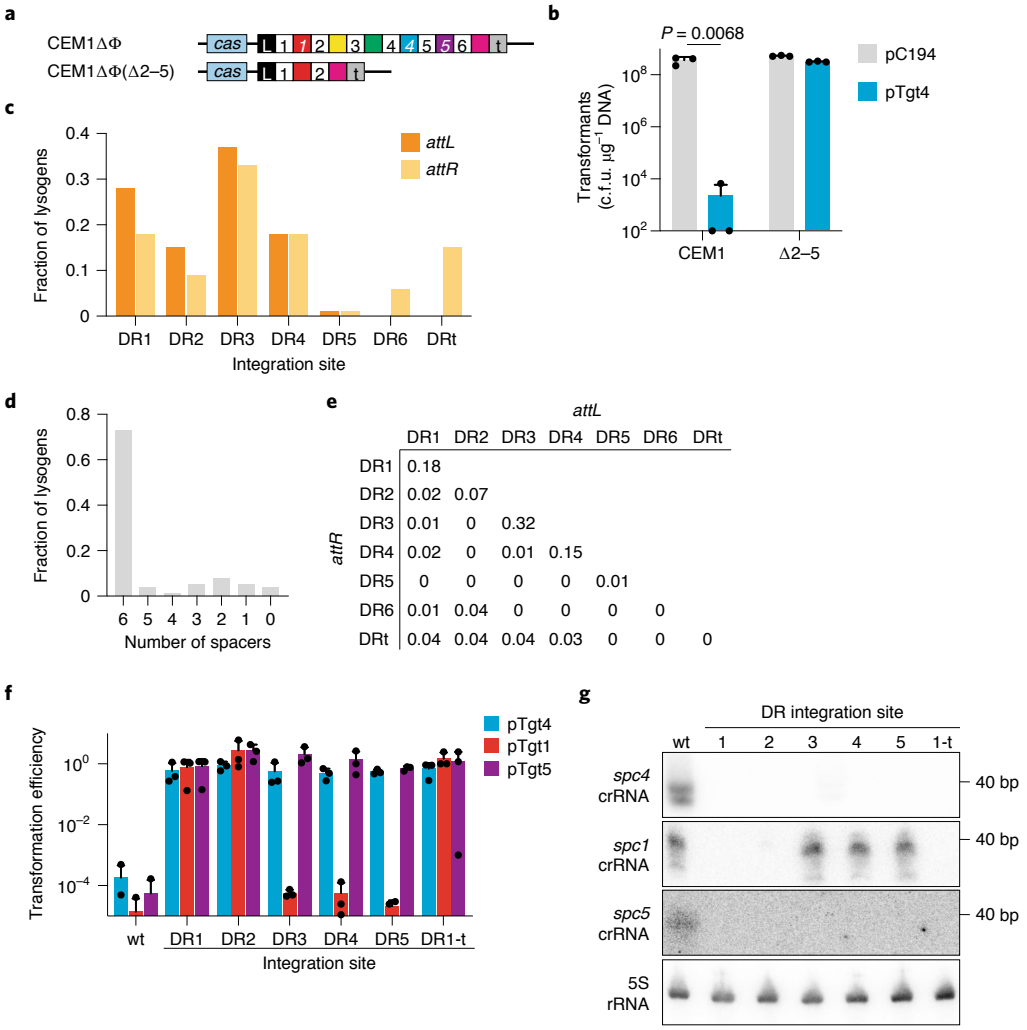


Fig. 2 | Disruption of CRISPR function by ΦAP1.1 lysogenization. a, *S. pyogenes* CEM1ΔΦ type II-A CRISPR–*cas* locus (wild-type, wt) and the mutant lacking spacer-repeat units 2, 3, 4 and 5 (Δ2–5). **b**, Transformation of the pC194 plasmid or a modified version harbouring a target for *spc4* from ΦAP1.1, pTgt4, after electroporation of wt or (Δ2–5) *S. pyogenes* CEM1ΔΦ competent cells. Mean + s.d. of three biological replicates are reported. Two-tailed unpaired *t*-test was used to calculate *P* value. **c**, Frequency of lysogens ($n = 100$) carrying *attL* or *attR* sites in each of the DRs of the CEM1ΔΦ CRISPR locus. **d**, Frequency of lysogens ($n = 100$) carrying different numbers of spacers. **e**, Same as **c** but showing the combined distribution of *attL* or *attR* sites. **f**, Transformation efficiency (relative to a pC194 vector control) of plasmids harbouring targets for *spc4*, *spc1* or *spc5* (pTgt4, pTgt1 or pTgt5, respectively) after electroporation of wt or lysogenic *S. pyogenes* CEM1ΔΦ competent cells. The ΦAP1.1 integration site for each lysogen is indicated. Mean + s.d. of three biological replicates are reported. **g**, Northern blot analysis of *spc4*, *spc1* and *spc5* crRNAs, as well as 5S rRNA produced by the strains transformed in **f**.

S. pyogenes CEM1ΔΦ[33], an SF370[34] derivative lacking all prophages (Fig. 2a) but harbouring a spacer (*spc4*) that matches ΦAP1.1 (although with a single mismatch at nucleotide position +8 that does not impact targeting; Extended Data Fig. 1b,f). We transformed this strain and a mutant lacking spacers 2–5 (CEM1ΔΦ(Δ2–5)) with a plasmid harbouring the phage target (pTgt4) and confirmed that this sequence is targeted by the type II-A CRISPR–Cas system (Fig. 2b). We then infected strain CEM1ΔΦ with ΦAP1.1 and analysed the location of *attL* and *attR* by sequencing the PCR products of 100 lysogens. Interestingly, integration of the *attL* was only observed in the first 5 DRs, but not in DR6 or DRt, of the CRISPR array of strain CEM1ΔΦ (Fig. 2c). Other parameters were similar to those obtained for the non-targeting strain K56, with a low frequency of spacer deletions (Fig. 2d,e). The presence of *spc4* in most lysogens suggested that the immunity mediated by this spacer was neutralized by the insertion of ΦAP1.1 into the CRISPR array. Indeed, targeting of the pTgt4 plasmid was abolished in lysogens that had the prophage integrated into DRs 1–5 (maintaining the entire spacer repertoire), as well as a lysogen where all spacers were eliminated due to integration into both DR1 and DRt (DR1-t, Fig. 2f).

We also measured type II-A immunity mediated by *spc1* and *spc5* through transformation of additional plasmids harbouring targets for each of these spacers (Fig. 2f). Immunity by *spc1* was abolished by the presence of prophages integrated into DR1 and DR2, that is, immediately upstream and downstream of the targeting spacer. For *spc5*, all prophages were integrated in upstream DRs and eliminated efficient targeting. Taken together, these data suggest that all spacers downstream of the prophage integration site, as well as the one immediately upstream, lose their ability to provide immunity, most likely due to interruption of the transcription of the CRISPR array or crRNA processing[15] (Extended Data Fig. 1g,h). We confirmed this hypothesis after northern blot analysis of *spc4*, *spc1* and *spc5* crRNAs (Fig. 2g). Furthermore, we corroborated this pattern of integration by infecting *S. pyogenes* strain NCTC13743 (accession number NZ_LS483384), in which the type II-A CRISPR array contains three spacers that target ΦAP1.1: *spc2*, *spc6* and *spc9* (Extended

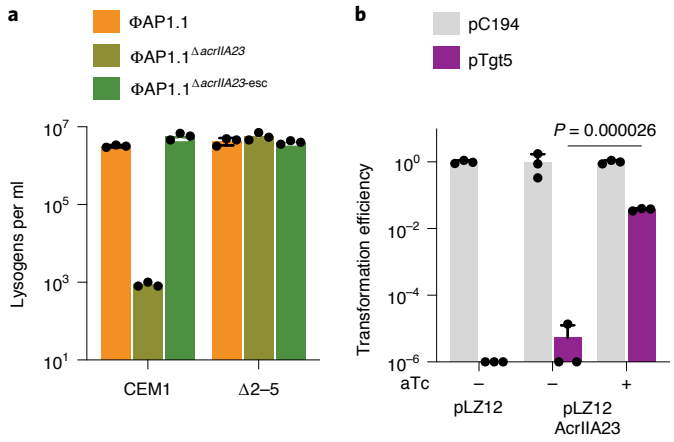


Fig. 3 | ΦAP1.1 carries a type II-A CRISPR–Cas inhibitor, AcrIIA23. **a**, Lysogenization rates of ΦAP1.1, ΦAP1.1$^{\Delta acrIIA23}$ or ΦAP1.1$^{\Delta acrIIA23\text{-esc}}$ after infection of wt or (Δ2–5) *S. pyogenes* CEM1ΔΦ cells. Mean + s.d. of three biological replicates are reported. **b**, Transformation efficiency (relative to a pC194 vector control) of the pC194 plasmid or a modified version harbouring a target for *spc5* (pTgt5) after electroporation of *S. pyogenes* CEM1ΔΦ competent cells carrying the pLZ12 vector or a version expressing AcrIIA23 under the control of a tetracycline-inducible promoter, in the presence or absence of aTc. Mean + s.d. of three biological replicates are reported. Two-tailed unpaired *t*-test was used to calculate *P* value.

Data Figs. 1b, 3a–c). We found that a majority of phages integrated either into DR2 or DR3, neutralizing the immunity provided by all three targeting spacers (Extended Data Fig. 3d), or downstream of DR3, in which case deletions encompassing the upstream targeting spacers were also detected. These results demonstrate that lysogenization of ΦAP1.1 into repeats neutralizes immunity, either by preventing the biogenesis of the targeting crRNA or through deletion of the targeting spacer DNA.

ΦAP1.1 harbours an inhibitor of the *S. pyogenes* type II-A CRISPR–Cas system. Even if integration into the CRISPR repeats inhibits immunity, targeting of the ΦAP1.1 DNA immediately after infection by Cas9 molecules loaded with *spc4* crRNA should decrease its frequency of lysogenization. Surprisingly however, the number of lysogens after infection of CEM1ΔΦ and CEM1ΔΦ(Δ2–5) streptococci were almost identical (Fig. 3a). Similarly, only about a 10-fold reduction in lysogens was observed when the triple-targeting *S. pyogenes* strain NCTC13743 was infected with ΦAP1.1 compared with its non-targeting mutant derivative NCTC13743(Δ1–10) (Extended Data Fig. 4a). These results suggest the presence of a mechanism that avoids Cas9 targeting before integration occurs.

ΦAP1.1 contains a four-gene locus with a similar arrangement as an anti-CRISPR (*acr*) region previously identified by our lab in the listeriophage ΦLS46[35], that is, it occurs at the end of the phage genome and it is transcribed in the opposite direction of the lytic genes (Extended Data Fig. 1b). Due to difficulty in detecting plaque-forming units after treatment of *S. pyogenes* with ΦAP1.1, we used a *Staphylococcus aureus* heterologous system to determine whether the *genes* of this operon are capable of inhibiting the type II-A CRISPR–Cas response against phage infection (Extended Data Fig. 4b,c). ORF2 inhibited Cas9, but did not display amino-acid homology to previously identified Acrs, and therefore we named it AcrIIA23, according to the current nomenclature[36]. To test whether AcrIIA23 is active in *S. pyogenes*, we transformed strain CEM1ΔΦ with the streptococcal plasmid pLZ12[37] or a version containing AcrIIA23 under the control of an anhydrous tetracycline (aTc) inducible promoter, pLZ12-AcrIIA23. To measure type II-A immunity, each of these strains was then transformed with a second plasmid containing the *spc5* target sequence, pTgt5, or an empty vector control, pC194, in the presence or absence of aTc (Fig. 3c). We found that in the absence of AcrIIA23 expression, there was low transformation efficiency for pTgt5 and that this was reverted to high levels by the addition of aTc. Altogether, these results demonstrate that ΦAP1.1 harbours a previously unidentified inhibitor of the *S. pyogenes* CRISPR–Cas immune response.

AcrIIA23 is required for ΦAP1.1 lysogenization under targeting conditions. To investigate the importance of AcrIIA23 for ΦAP1.1 lysogenization under targeting conditions, we generated a phage with a deletion of the inhibitor gene (which extended into ORF1 as well, since deletion of ORF2 only was not possible), ΦAP1.1$^{\Delta acrIIA23}$. Infection with this mutant phage showed that, in contrast to the wild-type ΦAP1.1, integration into the CRISPR array of strain CEM1ΔΦ, but not in that of CEM1ΔΦ(Δ2–5), was markedly impaired (Fig. 3a). A similar trend in lysogenization was detected when strain NCTC13743 was infected with the ΦAP1.1$^{\Delta acrIIA23}$ phage (Extended Data Fig. 4a). These results suggest that Cas9 cleavage of the phage genome in the absence of the inhibitor prevents lysogenization. To corroborate this, we isolated ΦAP1.1$^{\Delta acrIIA23\text{-esc}}$, a virus able to escape Cas9 targeting[10,38] due to the presence of a mutation in the PAM of the *spc4* target (Extended Data Fig. 1e), and found equal levels of lysogen formation in the presence or absence of a targeting spacer (Fig. 3a). Importantly, in contrast to wild-type ΦAP1.1 (Fig. 2c), this mutant phage also integrates into DR6 and DRt (Extended Data Fig. 5a,b). These two integrations do not affect *spc4* crRNA generation (Extended Data Fig. 5c) or immunity against pTgt4 (Extended Data Fig. 5d). On the basis of these results, we conclude that, through the inhibition of type II-A CRISPR–Cas targeting, AcrIIA23 protects the phage during the early stages of infection, allowing ΦAP1.1 integration into the CRISPR array.

acrIIA23 transcription is turned off during ΦAP1.1 lysogenization. Despite carrying an active inhibitor of the type II-A CRISPR–Cas immune response, the ΦAP1.1 prophage does not prevent plasmid targeting when the crRNA is still produced by the host (Fig. 2f). This is a striking contrast to other Acrs found in prophages, which are active during both the lytic and lysogenic stages of infection[23,24]. Using next-generation RNA sequencing (RNA-seq), we detected expression of a transcript spanning ORF1 and the inhibitor (as well as another containing ORF3 and ORF4) as early as 5 min after infection, continuing for the duration of the lytic infection (Fig. 4a). In contrast, when we performed RNA-seq of a lysogen where ΦAP1.1 was integrated into DR3, we found that only the ORF3-ORF4 mRNA is produced (Fig. 4b). These results demonstrate that ΦAP1.1 harbours a conditional inhibitor that is transcribed, and thus active, during the lytic but not during the lysogenic cycles of this temperate phage.

Aberrant ΦAP1.1 excision during prophage induction can transduce immunity. In most of the strains sequenced to date, ΦAP1.1-like phages are located in a spacer-less CRISPR locus, flanked by a partial DR and a DRt (Extended Data Fig. 1a). In our lysogenization experiments, we observed that integration into a full CRISPR array can lead to spacer deletions (Figs. 1e, 2d and Extended Data Fig. 3d). We hypothesized that the cycling between integration and 'unproductive' excision, that is, the rare cases when the phage does not complete its lytic cycle and does not kill the host upon induction[39], could eventually lead to the elimination of all spacers and the pattern typically observed in sequenced *S. pyogenes* genomes. To investigate this, we grew two CEM1ΔΦ(ΦAP1.1) lysogens with prophage insertions in DR3 and DR5 for 24 h in liquid media containing mitomycin C (MMC) to trigger rounds of induction and integration. We then collected bacteria and amplified the

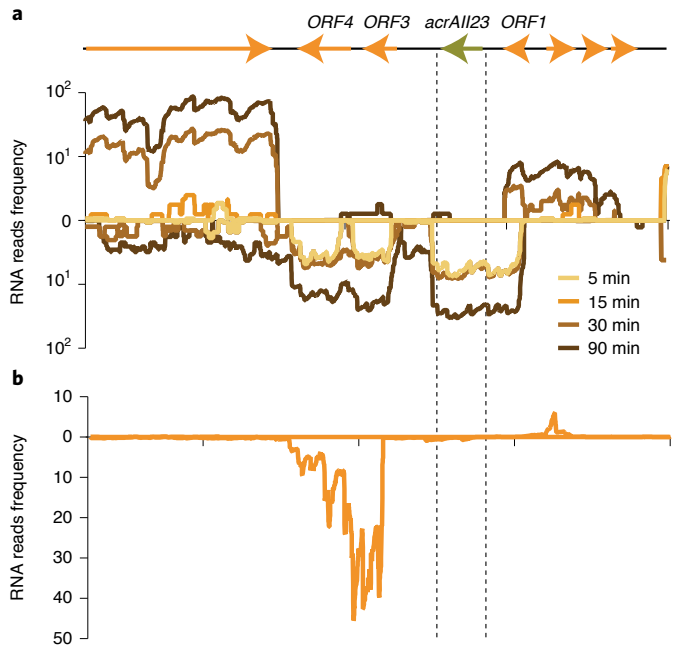


Fig. 4 | AcrIIA23 is not transcribed from ΦAP1.1 prophages. a, RNAseq of *S. pyogenes* CEM1ΔΦ cells at different times after infection with ΦAP1.1. The fraction of reads (relative to the total number of RNA-sequencing reads) that map to the *acr* region of the phage are plotted. **b**, Same as **a** but using RNA from a culture of *S. pyogenes* CEM1ΔΦ (ΦAP1.1::DR3) lysogen. The experiment was performed once.

attL and *attR* sites to detect changes in the number of spacers that flank each integration site. For the prophages originally inserted into DR3, we did not detect changes in the *attL* site, but the *attR* site appeared in PCR products of smaller sizes that indicated a location more proximal to the terminal repeat, either due to the deletion of spacer-repeat units, or re-insertion into repeats downstream of DR3 (Fig. 5a). Importantly, these new PCR products were diminished in the absence of MMC treatment, indicating that phage excision may lead to spacer deletion or rearrangement. A similar result was obtained with the prophage originally inserted in DR5 (Extended Data Fig. 6a).

The possibility that spacers are deleted during ΦAP1.1 induction raises the question of whether these spacers can become part of the viral genome during prophage excision. In principle, every DR and DRt contains both *attR* and *attL* sites (Fig. 1a) and therefore can be used by the phage integrase during excision. Amplification of the *attP* site of the phages induced with MMC produced a DNA fragment consistent with the size of an additional spacer-repeat unit, which was not seen when DNA from the ΦAP1.1 stock was used as template (Fig. 5b). PCR amplification of the *attP* using *spc1*- or *spc2*-specific primers indicated the presence of spacer-repeat units in the viral DNA (Extended Data Fig. 6b–e). To obtain a more accurate representation of the population of spacer-carrying phages, we subjected the *attP* amplicons with a larger size than expected (Fig. 5b) to next-generation sequencing, to obtain the relative fraction of reads of each spacer (Fig. 5c). We observed a 'proximity' effect, whereby the spacer sequences most commonly found in excised viral genomes were the closest to the DR in which the prophage was originally integrated, that is, *spc1* was the most commonly acquired spacer by phages previously residing in DR1 and DR2, *spc1* and *spc2* for DR3, and *spc3* and *spc4* for DR4 and DR5. However, this trend was not perfectly uniform and the results therefore suggest that there are certain hotspots where *attP* formation is preferred. This was also reflected in the number of repeat-spacer units that were excised along with the prophage. Whereas prophages integrated into DR1 and DR2 typically contained a single unit, at least half of the spacer-contaning reads of the DR3–5 lysogens contained two or more spacers (Fig. 5d).

The presence of repeat-spacer units in viral particles suggests that these could be added to the CRISPR array of new hosts following lysogenization (Extended Data Fig. 6f). To test this, we used MMC to induce CEM1ΔΦ (ΦAP1.1) lysogens in which the prophage was integrated into DR3, DR4, DR5 or a non-lysogen as a control. Viral particles were filtered from the culture supernatants and used to infect strain K56. PCR analysis of the surviving lysogens, using primers that anneal outside of the K56 CRISPR array in combination with primers that anneal to either the top or bottom strand of *spc1* or *spc3* sequences from the CEM1ΔΦ CRISPR locus, resulted in many different products for each lysogen population, with different numbers and mixtures of repeat-spacer units (Fig. 5e). This was most likely a consequence of the presence of several repeat sequences in both phage and bacterial genomes, each being a suitable *attP* and *attB* site, respectively, thus multiplying the possibilities for integration (Extended Data Fig. 6f). From this data, we conclude that all of the three ΦAP1.1 populations were able to transduce *spc1* and *spc3* from the CEM1ΔΦ strain into the CRISPR locus of the K56 strain. Sanger sequencing of PCR products obtained after transduction with a ΦAP1.1 prophage previously integrated in the DR3 of CEM1ΔΦ confirmed this conclusion (Fig. 5e).

Diverse instances of prophage integration into CRISPR–*cas* loci. To determine whether phage integration into CRISPR–*cas* loci is a unique feature of *S. pyogenes*, we performed a bioinformatic search. We identified prophages integrated into both *cas* genes and CRISPR arrays primarily in various type I subtypes (Fig. 6), which harboured both site-specific integrases and non-specific, Mu-type-like integrases[40,41]. We also found two *Streptococcus* species with phages encoding site-specific recombinases integrated into a type II-A CRISPR repeat. One prophage closely resembled ΦAP1.1 in *S. pyogenes* and was inserted into CRISPR repeats downstream of the *cas* genes (*S. dysgalactiae*); the other was inserted into an isolated type II-A CRISPR array distant from the *cas* gene operon (*S. canis*). Compared to most of the *S. pyogenes* cases, however, the CRISPR array was not reduced to a single repeat. Finally, a subset of the CRISPR–Cas systems contained spacers targeting the interrupting prophage (Extended Data Fig. 7), suggesting that integration disabled immunity. We also looked for the distribution of AcrIIA23 across phage genomes (Supplementary Table 1). We found it in multiple *S. pyogenes* prophages, and approximately half of these are integrated into the type II-A CRISPR locus. Most of the other AcrIIA23 homologues were located in prophages that integrate upstream of the *glgP* gene. Interestingly, in all cases regardless of the prophage integration site, AcrIIA23 is part of the same four-gene operon, suggesting the presence of a regulation mechanism similar to the one found in ΦAP1.1.

Discussion

Here we describe the interaction between the temperate phage ΦAP1.1 and the type II-A CRISPR–Cas system of its host, *S. pyogenes*. Notably, this phage uses the CRISPR DR sequence as the attachment site for integration and produces an anti-CRISPR, AcrIIA23, that is only expressed during the lytic portion of its replication cycle. This particular mode of insertion eliminates the immunity of spacers located downstream and immediately upstream of the *attL*, but does not affect the function of spacers further upstream (Fig. 2f). Therefore, ΦAP1.1 integration can selectively turn off the immunity provided by different sets of spacers – a mode of action different from previously described anti-CRISPRs that broadly turn off CRISPR function[42]. This could be exploited

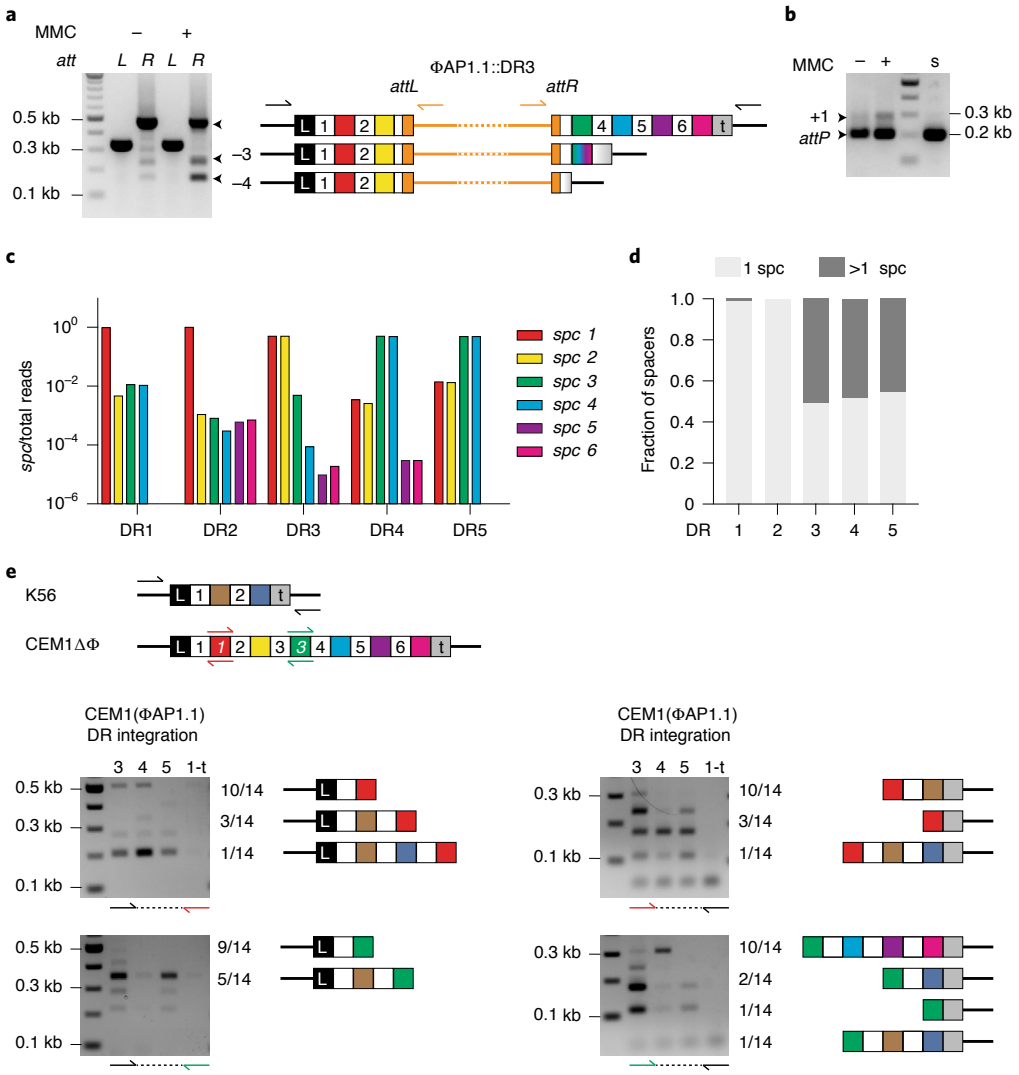
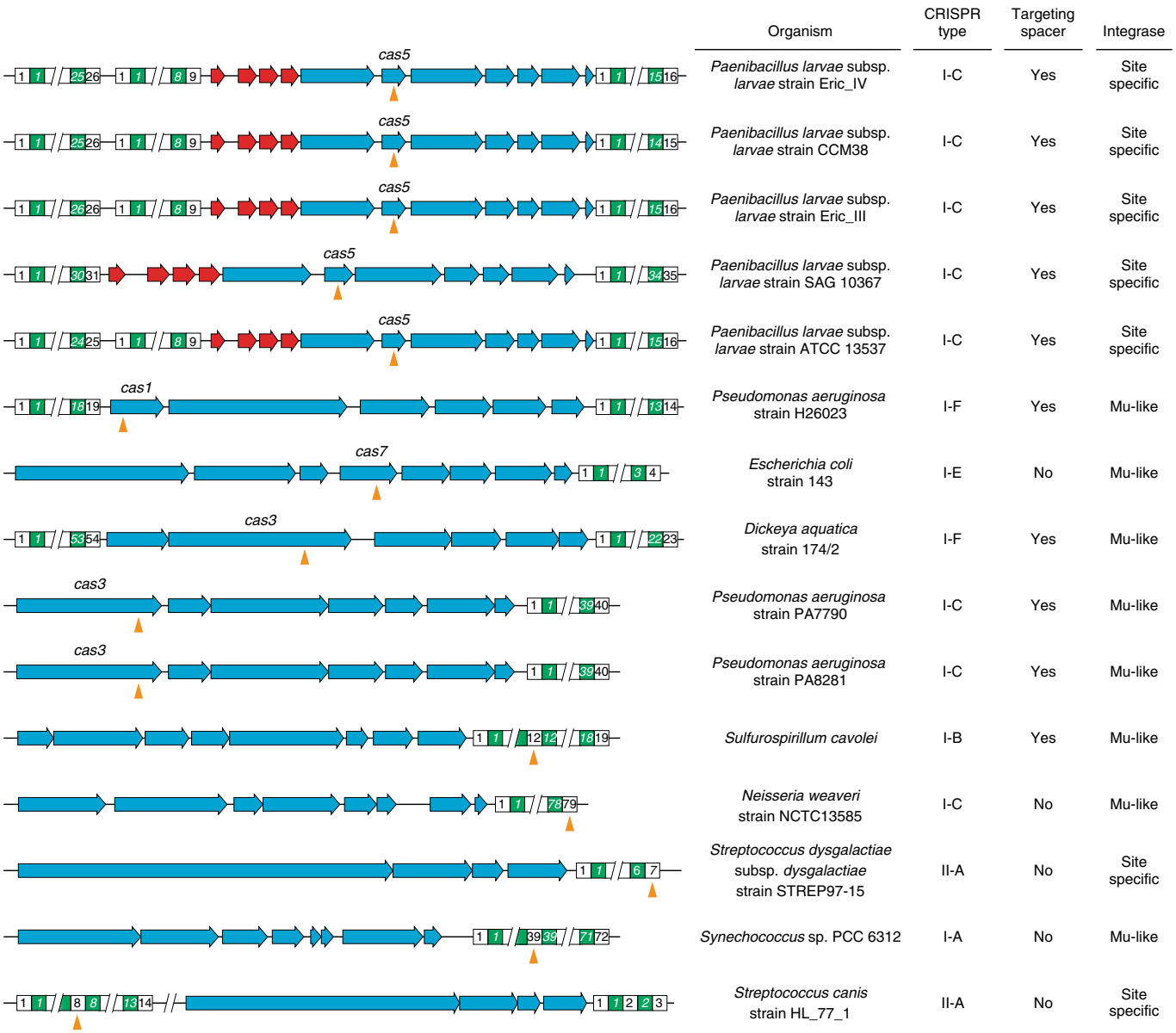


Fig. 5 | ΦAP1.1 mediates the transduction of CRISPR spacers. a, Agarose gel electrophoresis of the *attL* or *attR* PCR products obtained after treatment of *S. pyogenes* CEM1ΔΦ (ΦAP1.1::DR3) lysogens with MMC. Primers used and the expected products are shown in the diagrams to the right. Black arrowheads indicate PCR products that lack a number of the expected spacer-repeat units. Representative of 3 independent replicates. **b**, Same as **a** but after amplification of the ΦAP1.1 *attP* site. 's' indicates a control amplification of a phage stock that has never been lysogenized. Representative of 3 independent replicates. **c**, DNAseq of the *attP* amplification products obtained via PCR after induction of ΦAP1.1 prophages integrated in the indicated DR. The fraction of reads (relative to the total *attP* reads with an additional spacer) corresponding to each spacer sequence associated with the *attP* site is shown. **d**, Distribution of the number of spacers associated with the *attP* in **c**. **e**, Agarose gel electrophoresis of PCR products obtained from the amplification of the CRISPR locus of *S. pyogenes* K56 lysogens generated after infection with lysates obtained from *S. pyogenes* CEM1ΔΦ lysogens where ΦAP1.1 was integrated into DR3, DR4, DR5 or DR1-t (lacking all spacers). Top diagram shows the K56-, *spc1*- or *spc3*-specific primers indicated as black, red or green arrows, respectively. Diagrams next to gels show the spacer composition of 14 different PCR products of the CRISPR locus that were sequenced, these products being obtained after transduction of *S. pyogenes* CEM1ΔΦ (ΦAP1.1::DR3) lysates.

by ΦAP1.1 (and related phages) to (1) neutralize spacers that target its genome, but at the same time maintain the immunity provided by other spacers against competitor phages and (2) allow the acquisition of new, functional spacers in the first position of the CRISPR array. Furthermore, due to the presence of multiple *attL* and *attR* sites – one in every upstream and downstream DR, respectively – induction of the ΦAP1.1 prophage can result in the incorporation of repeat-spacer units into the viral genome and their subsequent transduction (Fig. 5e). Such spacer transduction events would increase the range of viral targets in the CRISPR system in an adaptation-independent manner, and benefit both the residing prophage and its host. Both the gain- and loss-of-function of spacers that result from integration of ΦAP1.1 into the CRISPR array could play an important role in the restriction or tolerance of prophages targeted by type II-A immunity, including those carrying virulence genes that play fundamental roles in the different manifestations of group A streptococcus disease[43]. In most of the *S. pyogenes* strains sequenced so far, ΦAP1.1-related prophages are integrated into a single DRt element (Extended Data Fig. 1a). We suggest that after the initial insertion into a full CRISPR array, subsequent rounds of excision and integration over evolutionary timescales ultimately erase all repeat-spacer units. Although a rare event that requires the excision without lysis, this process is termed 'active lysogeny' and is known to occur in some *S. pyogenes* and *Listeria monocytogenes* prophages[39]. We hypothesize that the lack of spacers in most ΦAP1.1 lysogens decreases barriers against viral infection, facilitating



Organism	CRISPR type	Targeting spacer	Integrase
Paenibacillus larvae subsp. *larvae* strain Eric_IV	I-C	Yes	Site specific
Paenibacillus larvae subsp. *larvae* strain CCM38	I-C	Yes	Site specific
Paenibacillus larvae subsp. *larvae* strain Eric_III	I-C	Yes	Site specific
Paenibacillus larvae subsp. *larvae* strain SAG 10367	I-C	Yes	Site specific
Paenibacillus larvae subsp. *larvae* strain ATCC 13537	I-C	Yes	Site specific
Pseudomonas aeruginosa strain H26023	I-F	Yes	Mu-like
Escherichia coli strain 143	I-E	No	Mu-like
Dickeya aquatica strain 174/2	I-F	Yes	Mu-like
Pseudomonas aeruginosa strain PA7790	I-C	Yes	Mu-like
Pseudomonas aeruginosa strain PA8281	I-C	Yes	Mu-like
Sulfurospirillum cavolei	I-B	Yes	Mu-like
Neisseria weaveri strain NCTC13585	I-C	No	Mu-like
Streptococcus dysgalactiae subsp. *dysgalactiae* strain STREP97-15	II-A	No	Site specific
Synechococcus sp. PCC 6312	I-A	No	Mu-like
Streptococcus canis strain HL_77_1	II-A	No	Site specific

Fig. 6 | Prophage integration into other CRISPR–*cas* loci. Schematic of phages integrated into either CRISPR or *cas* loci. Blue arrows, *cas* genes; red arrows, non-*cas* genes; white squares, CRISPR repeats; green squares, spacers; numbers indicate the quantity of repeats or spacers in the CRISPR array. Prophage insertions are marked by an orange arrowhead, with the name of the *cas* gene that contains the integration site indicated. The bacterial strain, CRISPR type, the presence of spacers targeting the inserted prophage and the type of integrase are shown to the right of the locus representation.

the incorporation of multiple prophages, and therefore allowing lateral gene flow that plays a critical role in *S. pyogenes* survival and colonization of the human host[44].

Following robust transcription at the onset of phage infection, *acr* genes are repressed by anti-CRISPR associated (Aca) transcription factors to prevent the deleterious effects of the *acr* high transcription on the downstream phage genes[24,45]. The *acr* region of ΦAP1.1 contains two genes, *orf3* and *orf4*, which are transcribed during lysogeny. Since *aca* genes are invariably found in the vicinity of *acr* genes[24,46,47], we hypothesize that *orf3* or *orf4* produces the Aca that regulates transcription of *acrIIA23*, providing the tight repression that is observed in the ΦAP1.1 prophage. Regardless of the mechanism that prevents *acrIIA23* transcription in ΦAP1.1 lysogens, the absence of Cas9 inhibition after integration only selects prophage insertions that prevent the generation of the targeting crRNA. This resembles the strategy used by listeriophages harbouring AcrIIA12, which binds to Cas9 and triggers its degradation[48]. While this mode of inhibition functions during lysogeny, it is not efficient enough to protect the phage during lytic infection, and therefore other, more rapid inhibitors are required during this stage.

Interestingly, we found cases of prophage integration within other CRISPR–*cas* loci, mostly encoding type I systems. Some of these prophages carry Mu-like integrases, which lead to lysogenization in random places of the host genome[40,41]. Therefore, these temperate phages do not seem to have developed a specific mechanism to target CRISPR loci. On the other hand, many of the other prophages appear to encode site-specific integrases and it can therefore be hypothesized that these have evolved to 'weaponize' lysogenization, that is, disrupt the expression of the CRISPR–*cas* locus and its immunity. Our study presents a unique dynamic in the parasite–host conflict, where the phage can modulate CRISPR immunity in streptococci and possibly other bacteria through lysogenization.

Methods

Bacterial strains and growth conditions. Culture of *S. pyogenes* and *S. aureus* was carried out in brain–heart infusion (BHI) medium at 37 °C. For *S. pyogenes*, liquid experiments were carried out in 10 ml medium in 15 ml conical tubes unless otherwise noted. *S. pyogenes* media was supplemented with 5 μg ml^{-1} chloramphenicol or 50 μg ml^{-1} spectinomycin for plasmid maintenance and/or lysogen selection. For *S. aureus*, liquid experiments were carried out in 3 ml medium in 15 ml conical tubes. *S. aureus* medium was supplemented with 5 μg ml^{-1} chloramphenicol or 10 μg ml^{-1} erythromycin for plasmid maintenance and/or strain selection. All strains used in this study are listed in Supplementary Table 2.

Induction of lysogens. To isolate phage, overnight cultures of the indicated lysogens were diluted 1:50 into 10 ml BHI and grown at 37 °C without agitation. At OD_{600} = 0.3, cells were induced with 0.5 μg ml^{-1} mitomycin C. At 4 h post induction, cells were spun down and the supernatant was filtered through 0.45 μm syringe filters (Acrodisc).

Strain construction. To make *S. aureus* strains with the integrated SF370 CRISPR locus (JAV17 and JAV16, Supplementary Table 2), pCL55-integration plasmids were constructed containing the SF370 *cas* genes with the wildtype CRISPR array (pAV121) or with a CRISPR array containing a ΦNM4γ4-targeting spacer (pAV122) (5'-TGATACATTAACATTTAGTAAATCATTACG-3'). These vectors were transformed in RN4220, recovered at 30 °C for 2 h, and plated on 5 μg ml^{-1} chloramphenicol. Integration was confirmed using primers oGG50 and oGG96. Primers used in this study are detailed in Supplementary Table 3.

Deletions of spacers in *S. pyogenes* strains CEM and NCTC 13743 were isolated by curing ΦAP1.1-spec lysogens. Prophages were cured by transforming lysogens as described in the transformation assay section, with a plasmid containing the SF370 type II-A *cas9* with a CRISPR array targeting the ΦAP1.1-spec, pAV315 (5'-TATATTTGGTGTGAAAGAAAGAAGCACATG-3'). Following transformation, clones were isolated and CRISPR arrays were screened using primers AV1171 and AV1015 to detect spacer deletions. Positive clones were then passaged without selection for two 1:100 dilutions and patched on plain and chloramphenicol plates to isolate clones that had lost the targeting plasmid.

Phage construction. To create ΦAP1.1-spec, the K56 strain of *S. pyogenes* was transformed with pAV329, a plasmid containing spectinomycin resistance cassette and ~1 kb phage-homology arms, to insert the antibiotic resistance cassette immediately downstream of the phage integrase in the lysogeny region. This strain was diluted 1:50 into 10 ml BHI, supplemented with 5 mM $CaCl_2$, grown at 37 °C without agitation and infected with ΦAP1.1 at OD_{600} = 0.3. The resulting lysate was filtered through 0.45 μm syringe filters (Acrodisc) and used to infect K56. At 5 h post infection, cells were plated on 50 μg ml^{-1} spectinomycin. Insertion of the spectinomycin cassette into ΦAP1.1 was confirmed using primers AV1277 and AV1298.ΦAP1.1$^{\Delta int}$

To create ΦAP1.1$^{\Delta int}$, the K56 strain of *S. pyogenes* was transformed with pAV480, a plasmid removing the tyrosine site-specific integrase with ~1 kb phage-homology upstream and downstream. This strain was diluted 1:50 into 10 ml BHI and supplemented with 5 mM $CaCl_2$, grown at 37 °C without agitation and infected with ΦAP1.1-spec at OD_{600} = 0.3. The resulting lysate was filtered through 0.45 μm syringe filters (Acrodisc) and used to infect a CEM strain containing a plasmid expressing a type III-A CRISPR–Cas system from *S. epidermidis* with a spacer targeting the integrase (5'-GAACTATTTATCAGTCATATACATTTTTGTGTGGCCAAG-3', pAV483). Overnight cultures of CEMΔΦ(Δ2–5) (300 μl) were transformed with pAV483, mixed with 5 ml 50% BHI agar, supplemented with 5 mM $CaCl_2$ and infected. Escaper plaques were isolated and replaqued on this strain. Phage DNA was isolated and primers AV1201 and 1430 were used to amplify the targeted region.

ΦAP1.1$^{\Delta acrIIAr23}$ deletion was isolated by passaging ΦAP1.1-spec on a CEM strain containing a plasmid expressing a type III-A CRISPR–Cas system from *S. epidermidis* with a spacer targeting *acrIIA23* (5'-ATATTCCATTTTATTTCTCCTTATTACATAGATAA-3', pAV449). Overnight cultures of CEMΔΦ(Δ2–5) (300 μl) were transformed with pAV449, mixed with 5 ml 50% BHI agar, supplemented with 5 mM $CaCl_2$ and infected with ΦAP1.1-spec. Escaper plaques were isolated and replaqued on this strain. Phage DNA was isolated and primers AV1339 and 1340 were used to amplify the targeted region. A phage was isolated with a deletion spanning nucleotides 37,881 to 39,044 of the ΦAP1.1-spec genome, encompassing *acrIIA23* and *orf1*.

Isolation of ΦAP1.1$^{\Delta acrIIA23\text{-esc}}$ escaping from spacer 4 targeting of SF370 CRISPR array was carried out by passaging ΦAP1.1$^{\Delta acrIIA23}$ on a CEM strain containing a plasmid expressing a type II-A CRISPR spacer (pAV392). Spacer 4 from SF370 contains a mismatch at position 8 when targeting the ΦAP1.1, which was corrected in the spacer cloned into pAV392 to perfectly target ΦAP1.1. Overnight cultures of CEMΔΦ(Δ2–5) with pAV392 (300 μl) were mixed with 5 ml 50% BHI agar, supplemented with 5 mM $CaCl_2$ and infected with ΦAP1.1$^{\Delta acrIIA23}$. Escaper plaques were isolated and replaqued on this strain. Phage DNA was isolated and primers AV1127 and AV1128 were used to amplify the spacer 4 target region; a phage with a G to A mutation in the first PAM position was isolated for subsequent experiments.

Phage sequencing. ΦAP1.1-spec DNA was purified as previously described[49]. Phage DNA was sheared to 300 bp using the Covaris Ultrasonicator S220. Sheared DNA was prepared using the TruSeq Nano kit (Illumina) and sequenced using 150-nucleotide single-end read with an Illumina MiSeq. The phage genome was assembled using SPAdes[50] and the phage *cos* site was predicted using PhageTerm[51].

Quantification of lysogens. To quantify lysogens, the indicated strain was diluted 1:30 into BHI supplemented with 5 mM $CaCl_2$. At OD_{600} = 0.3, cultures were infected with 50 μl ΦAP1.1-spec. Cells were collected 90 min post infection and plated on BHI plates with 50 μg ml^{-1} spectinomycin. Individual colonies were restreaked and genomic DNA was isolated by resuspending colonies in colony lysis buffer (250 mM KCl, 5 mM $MgCl_2$, 50 mM Tris-HCl at pH 9.0, 0.5% Triton X-100) supplemented with 500U PlyC[52] and incubated at 37 °C for 20 min, then at 98 °C for 10 min. Samples were centrifuged and the supernatant was used for PCR amplification with primers H237 and AV1050 for the *attL* junction, AV1089 and AV1015 for the *attR* junction, and AV1425 and AV1441 for the phage genome.

Lysogen passaging. ΦAP1.1-spec lysogens integrated into the indicated DR of CEM1ΔΦ were inoculated from single colonies and serially passaged by making 1:100 dilutions after 24 h growth. Half were passaged in the presence of a sublethal concentration of mitomycin C (0.1 μg ml^{-1}).

Efficiency of plaquing assays. For *S. aureus*, phage titre assays were performed as previously described[53]. Briefly, *S. aureus* strains with either the wildtype integrated type II-A locus from SF370 (JAV17) or the locus with a ΦNM4γ4 targeting spacer (JAV16, 5'-TGATACATTAACATTTAGTAAATCATTACG-3') were transformed with either a Ctl vector (pJTR162), orf4 (pAV412), AcrIIA23 (pAV413) or orf1 (pAV414). Strains were outgrown for 2 h. At 30 min pre-collection, aTc was added at a concentration of 12.5 ng ml−1. Cultures (200 μl) were mixed with 5 ml 50% BHI agar and supplemented with 5 mM $CaCl_2$. Cells were plated on plain plates or plates supplemented with aTc. Serial dilutions of ΦNM4γ4 were then spotted on the plates.

Cell transformations. For *S. aureus*, transformations were carried out as previously described[53]. For *S. pyogenes*, overnight cultures were diluted 1:50 into 250 ml BHI and grown at 37 °C without agitation. At OD_{600} = 0.3–0.4, cultures were pelleted and washed three times with cold 10% glycerol. Plasmid DNA was prepared from *S. aureus* strain RN4220[54], and dialysed on 0.025 μm nitrocellulose filters (Millipore). DNA (100 ng) was added to 50 μl aliquots of electrocompetent cells and transformed in cold 1 mm cuvettes under the following settings: 2,500 V, 25 μF, 600 Ω, 1 mm. After electroporation, cells were resuspended in 950 μl BHI and recovered at 37 °C without agitation for 2 h.

Transformation assays. For *S. pyogenes*, cells were prepared as described in the cell transformation section. The plasmids used for targeting experiments were ctl (pC194), spacer 1 target (pAV344), spacer 4 target as it appears in ΦAP1.1 (pAV354), spacer 5 target (pAV346), and the three potential NCTC 13743 targets as they appear in ΦAP1.1 (pAV407). DNA (100 ng) was added to 50 μl aliquots of electrocompetent cells and transformed in cold 1 mm cuvettes under the following settings: 2,500 V, 25 μF, 600 Ω, 1 mm. After electroporation, cells were resuspended in 950 μl BHI and recovered at 37 °C without agitation for 2 h. Serial dilutions were then made and plated on the appropriate antibiotics. For cultures where AcrIIA23 was expressed, cells transformed with Ctl (pLZ12spec) and AcrIIA23 (pAV432) were grown as above, with aTc added 30 min pre-collection at 12.5 ng ml^{-1}. Cells were treated as above and plated on the appropriate antibiotics plus or minus aTc.

RNA isolation. Overnight cultures were diluted 1:50 into 100 ml BHI and grown at 37 °C without agitation. At OD_{600} = 0.3–0.4, cultures were pelleted. Cells were resuspended in 70 μl PBS supplemented with 500U PlyC[52] and incubated at 37 °C for 30 min with shaking. Sarkosyl (10 μl, 10%) was added and the sample was then vortexed. RNA was then purified and DNase-treated using the Direct-zol miniprep plus kit (Zymo Research) for RNA-seq or treated with TRIzol for northern blots.

Small RNA northern blots. Northern blot analysis was carried out as previously described[55]. Purified RNA (20 μg) was loaded. AV1149 was used to blot for spacer 1, AV1152 for spacer 4, AV1153 for spacer 5 and AV1156 for 5S.

RNA sequencing. For phage infection time courses, overnight cultures of *S. pyogenes* strain CEM were diluted 1:50 into 10 ml BHI and grown at 37 °C without agitation. At OD_{600} = 0.3, cells were infected with ΦAP1.1-spec, pelleted and frozen at indicated time points. RNA was then isolated as described above. For lysogen sequencing, an overnight culture of CEM with ΦAP1.1-spec integrated into DR 3 was diluted 1:50 into 10 ml BHI and grown at 37 °C without agitation. At OD_{600} = 0.3, RNA was collected as described above. Ribosomal RNA was depleted using Ribo-Zero Plus rRNA depletion kit (Illumina) and subsequently purified using the RNA Clean and Concentrate −5 kit (Zymo Research). RNA was then prepared for deep sequencing using the TruSeq Stranded mRNA Library preparation kit (Illumina) and run on the MiSeq System (Illumina). Reads were aligned to the ΦAP1.1-spec genome for the phage time course or the ΦAP1.1-spec

genome integrated into DR 3 of the CEM1ΔΦ CRISPR array using Bowtie 2[56]. Sequence alignment map (SAM) files were normalized by total read count and converted to Wiggle (WIG) files with a custom python script to plot coverage at each nucleotide. For the phage infection time course, reads were mapped from nucleotide 35,754 to 39,363 of the ΦAP1.1-spec genome. For lysogen sequencing, reads were mapped to this region plus the 286 bp of the 3' portion of the CRISPR array.

Characterizing phage *attP*. To isolate phage carrying spacers from the CEM1ΔΦ CRISPR array, overnight cultures of the indicated lysogens were diluted 1:50 into 10 ml BHI and grown at 37 °C without agitation. At OD_{600} = 0.3, cells were induced with 0.5 μg ml^{-1} mitomycin C. At 4 h post induction, cells were spun down and the supernatant was filtered through 0.45 μm syringe filters (Acrodisc). Phage DNA was purified from these inductions as previously described[49]. From purified DNA, specific PCRs were carried out for spacer 1, 2 and 5. For spacers 1 and 2, AV1089 was used as the forward primer, while AV1149 and AV1151 were used as reverse primers for spacer 1 and 2, respectively. For spacer 5, H57 was the spacer 5'-specific forward primer, while AV1008 was the reverse primer. To deep sequence the phage *attP*, the phage junction was amplified from purified phage particles with AV1169 and AV1170, and products above the typical 75-nucleotide phage *attP* were gel purified and prepared for deep sequencing with the TruSeq Nano kit (Illumina). DNA was sequenced using 150-nucleotide single-end reads with an Illumina MiSeq. A custom python script was used to quantify spacers in the phage *attP*, with read length limiting the detection of spacers in the *attP* to two.

Detection of spacer transfer. Phage carrying spacers from the CEM1ΔΦ CRISPR array were isolated as described in the characterizing *attP* section. *S. pyogenes* strain K56 was diluted 1:30 into 10 ml BHI, supplemented with 5 mM $CaCl_2$ and grown at 37 °C without agitation. At OD_{600} = 0.3, cells were infected with phage induced from the indicated DR. Cells were collected 5 h post infection. Genomic DNA was isolated using the Wizard Genomic DNA purification kit (Promega). Cells were resuspended in 100 μl 50 mM EDTA with 500U PlyC and incubated at 37 °C with shaking for 30 min. Nuclei lysis solution (600 μl) was then added and DNA was purified following the manufacturer's instructions. PCRs were then carried out to detect transfer of spacers either 5' or 3' of the integrated lysogen in K56. For the amplification of 5' spacers, AV1171 was the forward primer, while AV1149 and AV1151 were used to specifically detect CEM1ΔΦ spacers 1 and 3, respectively. For the amplification of 3' spacers, AV1015 was the reverse primer, while H49 and H53 were used to specifically detect CEM1ΔΦ spacers 1 and 3, respectively.

Identification of prophage integration into CRISPR–*cas* loci. The database containing all identified *cas* loci and CRISPR arrays from CRISPRCasFinder[57] was downloaded and parsed with PostgreSQL to extract strain name and genomic coordinates of *cas* loci and CRISPR array. A custom python script was used to search for strains present in the PHASTER database[58] and return strains that had prophages within 2,000 bp of a *cas* locus or a CRISPR array. These strains were then manually checked for prophage integration into either the CRISPR array or *cas* locus. CRISPR arrays were checked for targeting spacers that matched the prophage with less than 4 mismatches. HHpred[59] was used to identify the type of recombinase encoded by the prophage.

Plasmid construction. Primers used in this study are detailed in Supplementary Table 3. Supplementary Table 4 contains the information of the DNA fragments used to construct plasmids in this study.Supplementary Table 5 contains the information of the primers and templates used to construct all the plasmids in this study.

Statistics and reproducibility. Statistical tests were carried out in GraphPad Prism 7.05.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Source data are provided with this paper. The data from this study are available from the corresponding authors upon request. The raw data for the RNA-seq experiments in Figs. 4a,b, 5c,d and the DNA reads of the ΦAP1.1-spec sequence can be found at the Sequence Read Archive (NIH), BioProject accession number PRJNA668016. The sequence of ΦAP1.1-spec is deposited under accession number MW168838. CRISPR–Cas databases were accessed from CRISPRCasFinder and the prophage database was accessed from PHASTER.

Code availability

Custom python scripts are deposited at https://github.com/Marraffini-Lab/Varble_etal_2021.git. Included are scripts that interact with PHASTER to determine CRISPR proximity to prophages (PHASTER_CRISPR.py), search deep sequencing data for spacer sequences in the phage ΦAP1.1 *attB* site (Phage_attb_II.py) and convert.sam files to.wig files for RNA sequencing analysis (SAM_WIG.py).

Received: 28 January 2021; Accepted: 15 October 2021;
Published online: 24 November 2021

References

1. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
2. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843–1845 (2008).
3. Jansen, R., Embden, J. D., Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**, 1565–1575 (2002).
4. Brouns, S. J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
5. Garneau, J. E. et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
6. Makarova, K. S. et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).
7. Heler, R. et al. Cas9 specifies functional viral targets during CRISPR–Cas adaptation. *Nature* **519**, 199–202 (2015).
8. Modell, J. W., Jiang, W. & Marraffini, L. A. CRISPR–Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. *Nature* **544**, 101–104 (2017).
9. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
10. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
11. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
12. 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).
13. Wright, A. V. & Doudna, J. A. Protecting genome integrity during CRISPR immune adaptation. *Nat. Struct. Mol. Biol.* **23**, 876–883 (2016).
14. Xiao, Y., Ng, S., Nam, K. H. & Ke, A. How type II CRISPR–Cas establish immunity through Cas1–Cas2-mediated spacer integration. *Nature* **550**, 137–141 (2017).
15. Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
16. Deveau, H. et al. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400 (2008).
17. Eitzinger, S. et al. Machine learning predicts new anti-CRISPR proteins. *Nucleic Acids Res.* **48**, 4698–4708 (2020).
18. Forsberg, K. J. et al. Functional metagenomics-guided discovery of potent Cas9 inhibitors in the human microbiome. *Elife* **8**, e46540 (2019).
19. Hynes, A. P. et al. An anti-CRISPR from a virulent streptococcal phage inhibits *Streptococcus pyogenes* Cas9. *Nat. Microbiol.* **2**, 1374–1380 (2017).
20. Mahendra, C. et al. Broad-spectrum anti-CRISPR proteins facilitate horizontal gene transfer. *Nat. Microbiol.* **5**, 620–629 (2020).
21. Rauch, B. J. et al. Inhibition of CRISPR–Cas9 with bacteriophage proteins. *Cell* **168**, 150–158.e10 (2017).
22. Uribe, R. V. et al. Discovery and characterization of Cas9 inhibitors disseminated across seven bacterial phyla. *Cell Host Microbe* **25**, 233–241.e5 (2019).
23. Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429–432 (2013).
24. Stanley, S. Y. et al. Anti-CRISPR-associated proteins are crucial repressors of Anti-CRISPR transcription. *Cell* **178**, 1452–1464.e13 (2019).
25. Birkholz, N., Fagerlund, R. D., Smith, L. M., Jackson, S. A. & Fineran, P. C. The autoregulator Aca2 mediates anti-CRISPR repression. *Nucleic Acids Res.* **47**, 9658–9665 (2019).
26. Ferretti, J. J., Stevens, D. L. & Fischetti, V. A. (eds) *Streptococcus pyogenes: Basic Biology to Clinical Manifestations* (University of Oklahoma Health Sciences Center, 2016).
27. Fischetti, V. A. In vivo acquisition of prophage in *Streptococcus pyogenes*. *Trends Microbiol.* **15**, 297–300 (2007).
28. Howard-Varona, C., Hargreaves, K. R., Abedon, S. T. & Sullivan, M. B. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *ISME J.* **11**, 1511–1520 (2017).
29. Nozawa, T. et al. CRISPR inhibition of prophage acquisition in *Streptococcus pyogenes*. *PLoS ONE* **6**, e19543 (2011).
30. Yamada, S. et al. Phylogenetic relationship of prophages is affected by CRISPR selection in Group A *Streptococcus*. *BMC Microbiol.* **19**, 24 (2019).

31. Fiebig, A. et al. Comparative genomics of *Streptococcus pyogenes* M1 isolates differing in virulence and propensity to cause systemic infection in mice. *Int. J. Med. Microbiol.* **305**, 532–543 (2015).
32. Kjems, E. Studies on streptococcal bacteriophages. 2. Adsorption, lysogenization, and one-step growth experiments. *Acta Pathol. Microbiol. Scand.* **42**, 56–66 (1958).
33. Euler, C. W. et al. Targeted curing of all lysogenic bacteriophage from *Streptococcus pyogenes* using a novel counter-selection technique. *PLoS ONE* **11**, e0146408 (2016).
34. Desiere, F., McShan, W. M., van Sinderen, D., Ferretti, J. J. & Brussow, H. Comparative genomics reveals close genetic relationships between phages from dairy bacteria and pathogenic Streptococci: evolutionary implications for prophage–host interactions. *Virology* **288**, 325–341 (2001).
35. Meeske, A. J. et al. A phage-encoded anti-CRISPR enables complete evasion of type VI-A CRISPR–Cas immunity. *Science* **369**, 54–59 (2020).
36. Bondy-Denomy, J. et al. A unified resource for tracking Anti-CRISPR names. *CRISPR J.* **1**, 304–305 (2018).
37. Perez-Casal, J., Caparon, M. G. & Scott, J. R. Mry, a trans-acting positive regulator of the M protein gene of *Streptococcus pyogenes* with similarity to the receptor proteins of two-component regulatory systems. *J. Bacteriol.* **173**, 2617–2624 (1991).
38. Nussenzweig, P. M., McGinn, J. & Marraffini, L. A. Cas9 cleavage of viral genomes primes the acquisition of new immunological memories. *Cell Host Microbe* **26**, 515–526.e6 (2019).
39. Feiner, R. et al. A new perspective on lysogeny: prophages as active regulatory switches of bacteria. *Nat. Rev. Microbiol.* **13**, 641–650 (2015).
40. Shapiro, J. A. Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. *Proc. Natl Acad. Sci. USA* **76**, 1933–1937 (1979).
41. Taylor, A. L. Bacteriophage-induced mutation in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **50**, 1043–1051 (1963).
42. Davidson, A. R. et al. Anti-CRISPRs: protein inhibitors of CRISPR–Cas systems. *Annu. Rev. Biochem.* **89**, 309–332 (2020).
43. Banks, D. J., Beres, S. B. & Musser, J. M. The fundamental contribution of phages to GAS evolution, genome diversification and strain emergence. *Trends Microbiol.* **10**, 515–521 (2002).
44. McShan, W. M., McCullor, K. A. & Nguyen, S. V. The bacteriophages of *Streptococcus pyogenes*. *Microbiol Spectr.* **7**, 7.3.8 (2019).
45. Osuna, B. A. et al. Critical anti-CRISPR locus repression by a bi-functional Cas9 inhibitor. *Cell Host Microbe* **28**, 23–30.e5 (2020).
46. Marino, N. D. et al. Discovery of widespread type I and type V CRISPR-Cas inhibitors. *Science* **362**, 240–242 (2018).
47. Yin, Y., Yang, B. & Entwistle, S. Bioinformatics identification of anti-CRISPR loci by using homology, guilt-by-association, and CRISPR self-targeting spacer approaches. *mSystems* **4**, e00455-19 (2019).
48. Osuna, B. A. et al. *Listeria* phages induce Cas9 degradation to protect lysogenic genomes. *Cell Host Microbe* **28**, 31–40.e9 (2020).
49. Varble, A., Meaden, S., Barrangou, R., Westra, E. R. & Marraffini, L. A. Recombination between phages and CRISPR-cas loci facilitates horizontal gene transfer in staphylococci. *Nat. Microbiol.* **4**, 956–963 (2019).
50. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
51. Garneau, J. R., Depardieu, F., Fortier, L. C., Bikard, D. & Monot, M. PhageTerm: a tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. *Sci. Rep.* **7**, 8292 (2017).
52. Nelson, D., Schuch, R., Chahales, P., Zhu, S. & Fischetti, V. A. PlyC: a multimeric bacteriophage lysin. *Proc. Natl Acad. Sci. USA* **103**, 10765–10770 (2006).
53. Goldberg, G. W., Jiang, W., Bikard, D. & Marraffini, L. A. Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature* **514**, 633–637 (2014).
54. Kreiswirth, B. N. et al. The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage. *Nature* **305**, 709–712 (1983).
55. Pall, G. S. & Hamilton, A. J. Improved northern blot method for enhanced detection of small RNA. *Nat. Protoc.* **3**, 1077–1084 (2008).
56. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
57. Couvin, D. et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* **46**, W246–W251 (2018).
58. Arndt, D. et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44**, W16–W21 (2016).
59. Zimmermann, L. et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237–2243 (2018).

Acknowledgements

We thank A. J. Meeske for bioinformatic assistance with the RNA-seq analysis. Support for this work was provided by the National Institutes of Health Director's Pioneer Award 1DP1GM128184-01 and the Burroughs Wellcome Fund PATH Award to L.A.M. L.A.M. is an Investigator of the Howard Hughes Medical Institute. A.V. was supported by the Arnold and Mabel Beckman Postdoctoral Fellowship. J.T.R. was supported by the Boehringer Ingelheim Fonds PhD fellowship. Funding for V.A.F. and E.C. was from the Rockefeller University Laboratory Support. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

A.V. and L.A.M. conceived the study. E.C., C.W.E. and V.A.F. aided with experimental design. A.V. performed all the experiments with the help of J.F. and A.K. E.C. and P.M. assisted with bioinformatic analysis. E.C. and C.W.E. provided assistance, strains and reagents for *S. pyogenes* experiments. J.T.R. provided reagents. A.V. and L.A.M. wrote the manuscript with the help of the other authors.

Competing interests

L.A.M. and A.V. are co-inventors on a patent application filed by The Rockefeller University relating to work in this study. L.A.M. is a founder and advisor of Intellia Therapeutics, Eligo Biosciences and CRISPR Biotechnologies. The remaining authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41564-021-00996-8.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-021-00996-8.

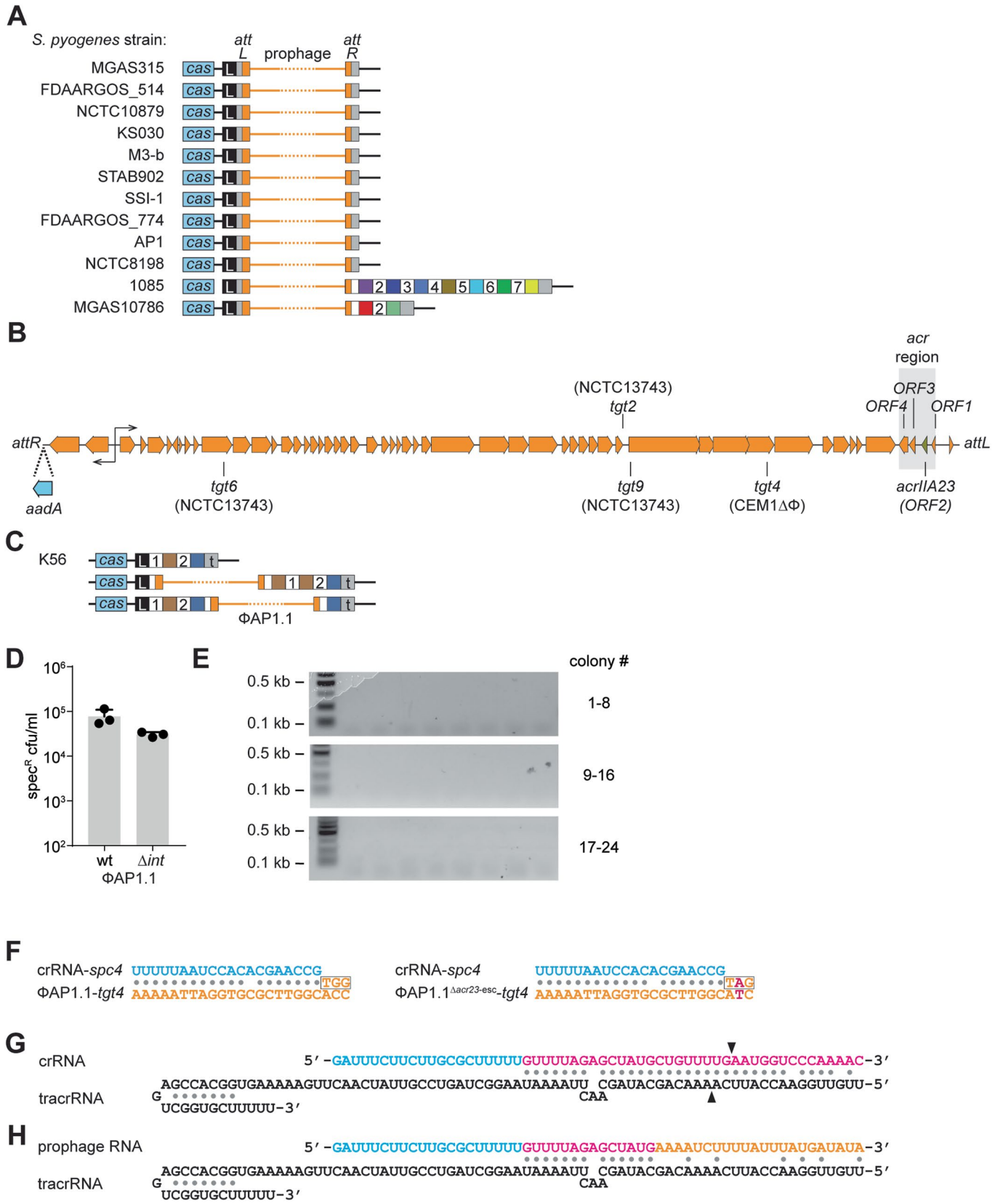
Correspondence and requests for materials should be addressed to Andrew Varble or Luciano A. Marraffini.

Peer review information *Nature Microbiology* thanks Joseph Bondy-Denomy and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

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

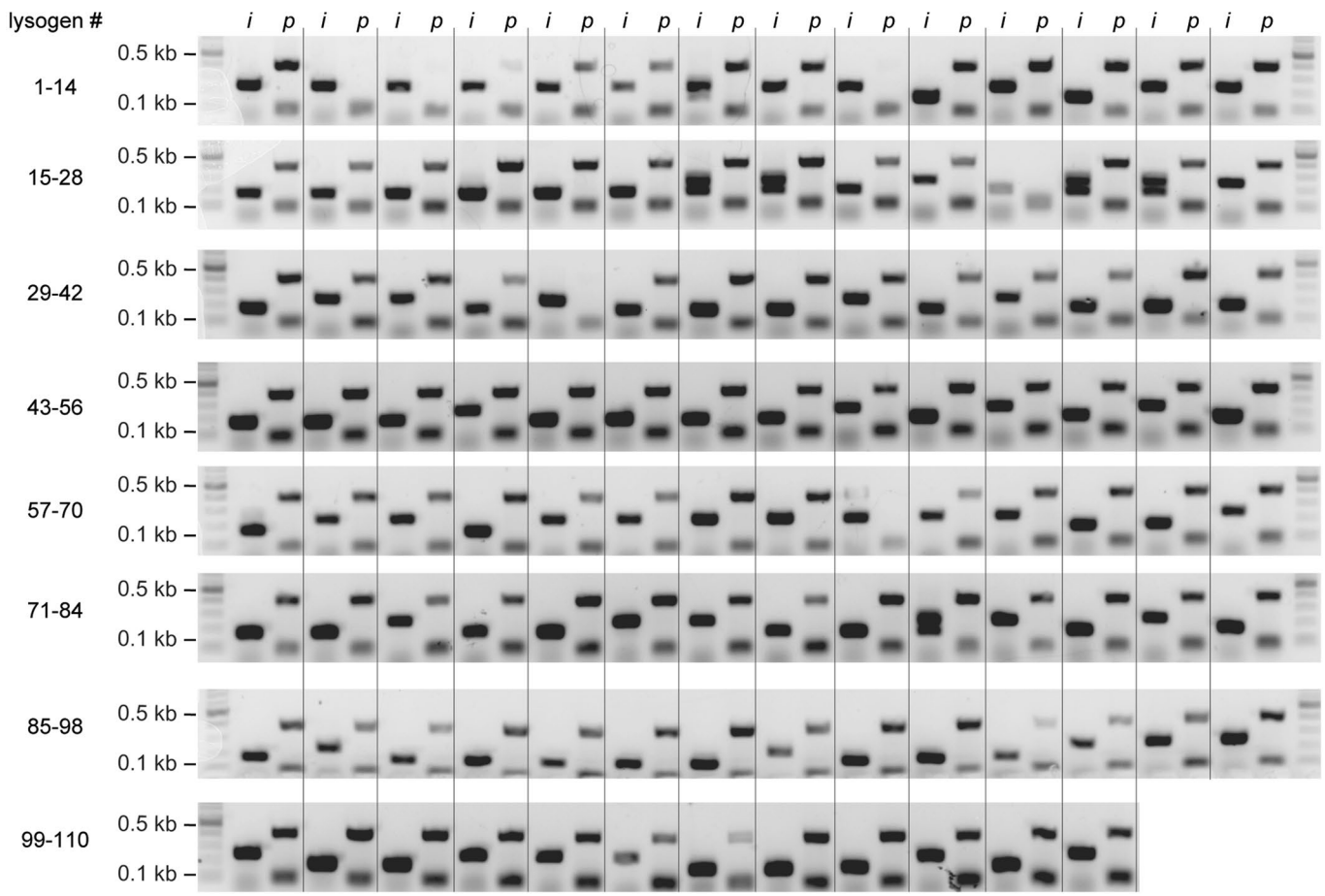
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2021

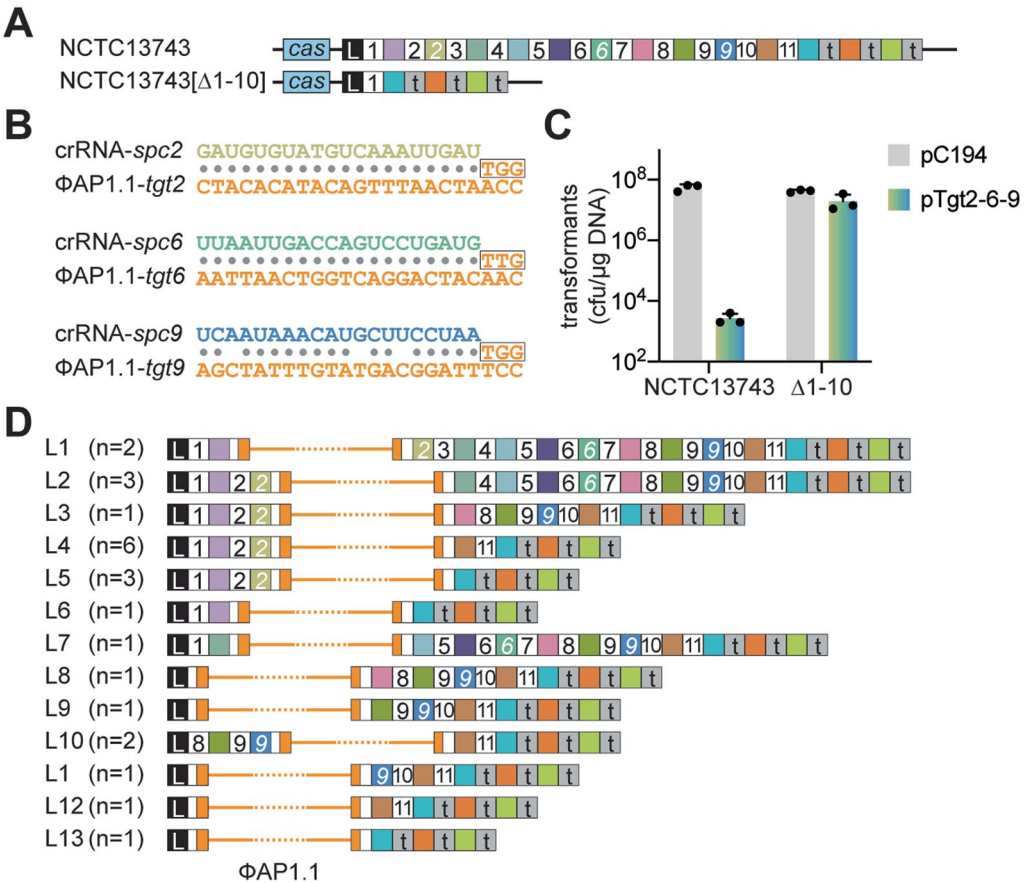


Extended Data Fig. 1 | See next page for caption.

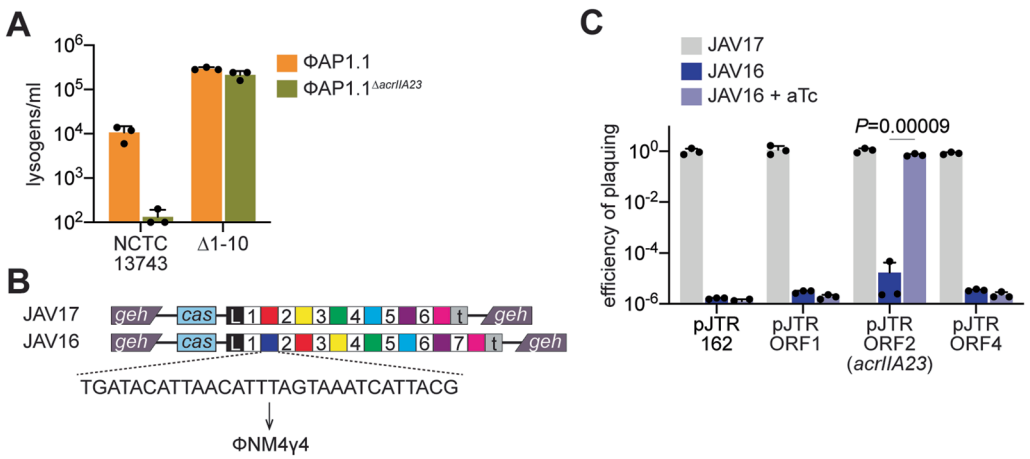
Extended Data Fig. 1 | ΦAP1.1 integration into *S. pyogenes* type II-A CRISPR loci. (**a**) ΦAP1.1-like phages integrated into the type II-A CRISPR locus of *S. pyogenes* strains. (**b**) Schematic of the ΦAP1.1 genome. Targets for the spacers found in different strains are shown on top or bottom depending on the phage DNA strand that is targeted. The *acr* region and its four ORFs are highlighted in grey. The *aadA* gene inserted to confer spectinomycin resistance to ΦAP1.1 lysogens is shown in light blue. (**c**) Spacer duplication events observed in *S. pyogenes* K56(ΦAP1.1) lysogens. (**d**) To determine if the tyrosine site-specific recombinase present in the ΦAP1.1 genome is required for its insertion into the CRISPR repeats, we deleted the integrase gene, generating ΦAP1.1$^{\Delta int}$. We infected streptococci with both wild-type and int phages and enumerated spectinomycin-resistant colonies. Interestingly, the number of colony-forming units (cfu) obtained was similar for both infections. However, in the absence of the integrase, the colonies lost antibiotic resistance over time, and we were unable to detect PCR products for the *attL* site or the phage itself, indicating that infection resulted in the transient transfer of antibiotic resistance but failed to produce lysogenic integration into the CRISPR array. Enumeration of spectinomycin-resistant cfu per ml of *S. pyogenes* CEMΔΦ[Δ2-5] cultures infected with either wild-type ΦAP1.1 or ΦAP1.1$^{\Delta int}$. Mean + STD of 3 biological replicates are reported. Interestingly, the number of colony-forming units (cfu) obtained was similar for both infections. However, in the absence of the integrase, the colonies lost antibiotic resistance over time (not shown). (**e**) Agarose gel electrophoresis of the *attL* PCR product obtained after the amplification shown in Fig. 1B, using DNA extracted from 24 spectinomycin-resistant colonies that resulted from infection of *S. pyogenes* CEMΔΦ[Δ2-5] cultures with ΦAP1.1$^{\Delta int}$. Collected from three biological replicates. (**f**) Sequences of the *S. pyogenes* CEM1ΔΦ *spc4* crRNA annealed to its targets in ΦAP1.1 and ΦAP1.1$^{\Delta acrIIA23\text{-esc}}$. The PAM nucleotides are boxed. The escape mutation of ΦAP1.1$^{\Delta acrIIA23\text{-esc}}$ is in red. (**g**) Base pairs formed between the crRNA and tracrRNA of *S. pyogenes* (grey dots). Blue sequence, spacer; pink sequence, repeat; black sequence, tracrRNA. Black arrows indicate RNase III cleavage sites required for crRNA processing. (**h**) Same as (**G**) but showing the possible base pairs formed by the prophage transcript and the tracrRNA. Lack of annealing with the prophage-derived sequences (orange) prevents RNase III cleavage and thus processing of the upstream crRNA.



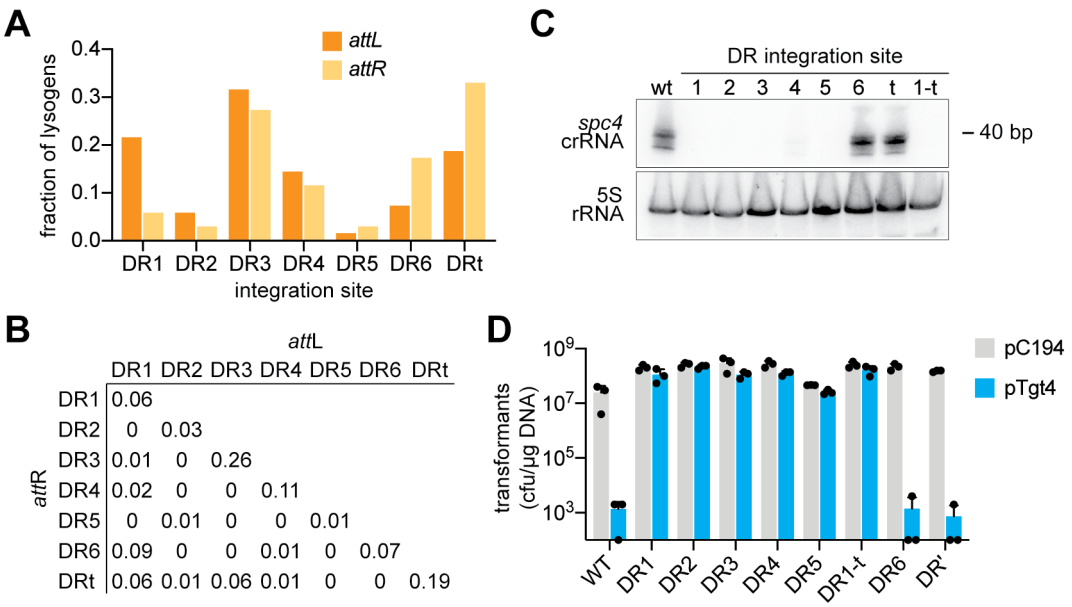
Extended Data Fig. 2 | ΦAP1.1 lysogenization location. We decided to look for integration of ΦAP1.1 in other genomic sites with the 8-bp sequence of the *attB* present on the CRISPR repeat. Because strain K56 is not sequenced and therefore we could not identify such sequences, we decided to perform experiments with strain SF370, which harbors an additional 27 sites with the *attB* sequence. To do this, we infected strain CEM1ΔΦ[Δ2-5], which lacks the prophages as well as most spacers of the CRISPR array (in particular *spc4*, which targets ΦAP1.1) present in SF370. We isolated and analyzed 110 spectinomycin-resistant colonies and looked for the presence of the phage (*p*) and its integration *attL* site (*i*) via PCR. We found that all prophages were inserted into the CRISPR array (in different repeats, generating different sizes for the *attL* PCR products). This result indicates a strong preference for integration into the repeat sequences and also suggests that regions outside of the 8-bp *attB* site are important for integration. Agarose gel electrophoresis of the *attL* (i) or phage (p) PCR products obtained after selection of individual spectinomycin-resistant ΦAP1.1 lysogens. PCRs for individual lysogens are divided by black lines.



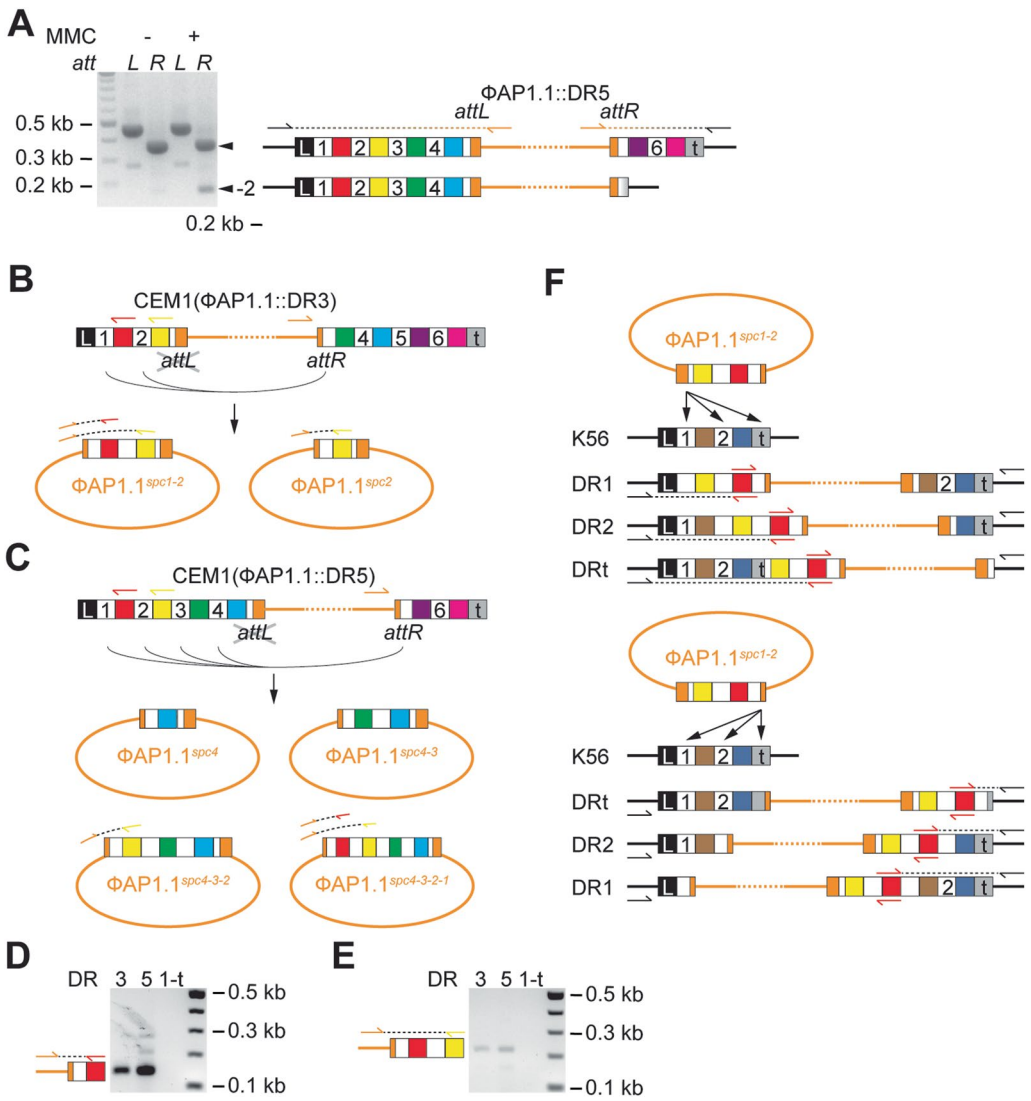
Extended Data Fig. 3 | ΦAP1.1 lysogenization into the *S. pyogenes* NCTC13743 type II-A CRISPR locus. (**a**) *S. pyogenes* NCTC13743 type II-A CRISPR-*cas* locus, wild-type and the mutant lacking spacer-repeat units 1 through 10, [Δ1-10]. The 't' indicates three degenerate repeats that contain mutations from the repeat consensus sequence. (**b**) Sequences of the *S. pyogenes* NCTC13743 *spc2*, *spc6* and *spc9* crRNAs annealed to their targets in ΦAP1.1. The PAM nucleotides are boxed. (**c**) We verified that the type II-A CRISPR system present in NCTC13743 is capable of restricting a plasmid containing the ΦAP1.1 target sties. Transformation efficiency of the pC194 plasmid or a modified version harboring the three targets for *spc2*, *spc6* and *spc9*, pTgt2-6-9 from ΦAP1.1, after electroporation of wild-type or [Δ1-10] *S. pyogenes* NCTC13743 competent cells. Mean + STD of 3 biological replicates are reported. These results demonstrate that the type II-A CRISPR system present in this strain is capable of restricting a plasmid containing the three target sequences present in ΦAP1.1 (**d**) Diagrams of the type II-A locus of 25 ΦAP1.1 lysogens in *S. pyogenes* NCTC13743, reconstructed after sequencing of their *attL* and *attR* sites.



Extended Data Fig. 4 | AcrIIA23 inhibits the type II-A CRISPR-Cas response in *S. pyogenes* NCTC13743 and *S. aureus* RN4220. (**a**) Lysogenization rates of ΦAP1.1 or ΦAP1.1$^{\Delta acrIIA23}$ after infection of wild-type or [Δ1-10] *S. pyogenes* NCTC13743 cells. Mean + STD of 3 biological replicates are reported. (**b**) Integration of the *S. pyogenes* CEM1**ΔΦ** type II-A system into the lipase gene of *S. aureus* RNA4220 using the integrative vector pCL55, generating strain JAV17. A CRISPR locus with a spacer matching the *gp68* gene from the virulent staphylococcal phage ΦNM4γ4 was also integrated, generating JAV16. (**c**) JAV17 and JAV16 strains were transformed with each of the individual ΦAP1.1 ORFs under the control of an anhydrotetracycline (aTc)-inducible promoter, on the pJTR162 staphylococcal plasmid backbone, with the exception of *ORF3*, for which we did not obtain transformants. Strains harboring the pJTR-ORF1, pJTR-ORF2 and pJTR-ORF4 constructs were then infected with ΦNM4γ4 in the presence or absence of aTc, and pfus were counted. We found that none of the ORFs affected the immunity of the JAV16 strain in the absence of the inducer, and ΦNM4γ4 propagation was limited. In the presence of aTc, however, *ORF2* expression enabled high levels of viral replication, similar to those observed in the absence of type II-A CRISPR-Cas immunity after infection of the JAV17 strain. Efficiency of plaquing of phage ΦNM4γ4 on soft agar lawns of *S. aureus* JAV17 or JAV16, in the presence or absence of aTc, carrying the pJTR162 control plasmid or versions that express the different ORFs of the ΦAP1.1 anti-CRISPR locus. Mean + STD of 3 biological replicates are reported. Two-tailed unpaired t-test was used to calculate *P* value.



Extended Data Fig. 5 | ΦAP1.1$^{\Delta acrIIA23\text{-esc}}$ integration pattern. (**a**) Frequency of ΦAP1.1$^{\Delta acrIIA23\text{-esc}}$ lysogens ($n = 50$) carrying *attL* or *attR* sites in each of the DRs of the CEM1ΔΦ CRISPR locus. (**b**) Same as (**A**) but showing the combined distribution of *attL* or *attR* sites. (**c**) Northern blot analysis of *spc4* crRNAs, as well as 5 S rRNA, produced by wild-type *S. pyogenes* CEM1ΔΦ or its different ΦAP1.1 (DR1-5, 1-t) or ΦAP1.1$^{\Delta acrIIA23\text{-esc}}$ (DR6, DRt) lysogens. (**d**) Transformation efficiency of the pC194 plasmid or a modified version harboring a target for *spc4* from ΦAP1.1, pTgt4, after electroporation of the strains used in (**C**). Mean + STD of 3 biological replicates are reported.



Extended Data Fig. 6 | ΦAP1.1-mediated transduction of spacers from the CEM1ΔΦ to the K56 CRISPR locus. (**a**) Agarose gel electrophoresis of the *attL* or *attR* PCR products obtained after treatment of *S. pyogenes* CEM1ΔΦ(ΦAP1.1::DR5) lysogens with mitomycin C (MMC). Primers used and the expected products are shown in the diagrams to the right. Black arrowheads indicate PCR products that lack a number of the expected spacer-repeat units. Representative of three independent replicates. (**b**) Example of aberrant ΦAP1.1 excision in *S. pyogenes* CEM1ΔΦ(ΦAP1.1::DR3) lysogens, where DR1 or DR2 are used instead of *attL* for recombination with *attR*. ΦAP1.1-, *spc1*- and *spc2*-specific primers used to detect *attP* sites containing either of these spacers are shown as orange, red and yellow arrows, respectively. (**c**) Same as (**B**) but for a ΦAP1.1::DR5 lysogen. (**d**) Agarose gel electrophoresis of PCR products obtained after the amplification of the *attP* site of ΦAP1.1 phages induced with MMC from ΦAP1.1::DR3, ΦAP1.1::DR5 and ΦAP1.1::DR1-t lysogens, using *spc1*-specific primers. (**e**) Same as (**D**) but using *spc2*-specific primers. In both cases we found PCR products that indicated the presence of spacer-repeat units in the DNA of the viral particles, with the intensity of the amplicons indicating more *spc1*- than *spc2*-containing viral particles. These results suggest a possible preference for certain excision products. (**f**) Example of possible recombination of the 5′ end (top) or 3′ end (bottom) *attP* site of the ΦAP1.1$^{spc1\text{-}2}$ phage with the three *attB* sites present in the DR1, DR2 and DRt of *S. pyogenes* K56.

Organism	spacer/target alignement
Paenibacillus larvae subsp. larvae strain Eric_IV *Paenibacillus larvae* subsp. pulvifaciens strain CCM 38 *Paenibacillus larvae* subsp. larvae strain Eric_III *Paenibacillus larvae* subsp. pulvifaciens strain ATCC 13537	spacer ATTTTAATTTTCCGTTCCCCTCGTGTTGTTATCGT target TTTTTAATTTTCCGTTCCCCTCGTGTTGTTATCTA
Paenibacillus larvae subsp. pulvifaciens strain SAG 10367	spacer TATTGTTAAGTTGGCGAACCCACGGGATAACACCG target TATTGTTAAGTTGGCGAACCCACGGGATAACACCG
Pseudomonas aeruginosa strain H26023	spacer ACGACGCGACCAGAATCGCGATTTGCACCCAC target ACGACGCGACCAGAATCGCGATTTGCACCCAC
Dickeya aquatica strain 174/2	spacer GGGCGTCCGGGCCATCGTCGTGGTCAGCCATC target GGGCGTCCGGGCCATCGTCGTGGTCAGCCACC
Pseudomonas aeruginosa strain PA7790 *Pseudomonas aeruginosa* strain PA8281	spacer CCCGTCACGCTGTTGCCATCACGCAACTCGACGGT target CCCGTCACGCTGTTGCCATCACGCAAGTCGACCGT
Sulfurospirillum cavolei	spacer TGATAAGACACTCTACAACTCAGCGAGTGCGTCA target TGACAAAACACTCTACAACTCAGCCAGCGCGTCA

Extended Data Fig. 7 | Spacers targeting prophages integrated into other CRISPR-*cas* loci. Alignment of spacers described in Fig. 6 with their phage target sites; in the 5′ to 3′ orientation.

nature portfolio

Corresponding author(s): Luciano Marraffini

Last updated by author(s): Oct 8, 2021

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection
Data analysis	For analyzing lysogen, transformation, and plaquing experiments - used GraphPad Prism 7 For Phage sequencing: Used SPAdes 3.9 and PhageTerm 1.0.12. For RNA-seq analysis: Used bowtie2 2.4.4 and custom script in Python 3.9. For phage attP analysis: Used custom script in Python 3.9. For identification of phage in CRISPR-Cas loci: Used custom in script in Python 3.9, PHASTER, CRISPRCasFinder 1.1.2, PostgreSQL 9.5.24, and HHpred Custom scripts can be found at: https://github.com/Marraffini-Lab/Varble_etal_2021.git

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data from this study are available from the authors upon request. The raw data for the RNA-seq experiments in Figures 4A, 4B and Extended Data Figures 5D

and 5E can be found at the Sequence Read Archive (NIH), BioProject accession number PRJNA668016. Sequence of ΦAP1.1-spec is deposited under accession number MW168838. Numerical source data that underlie the graphs shown in Figs. 1, 2, 3 and 5, and Extended Data Figs. 1, 2, 3 and 4 are provided with the paper. Original gel or blot images in Figs. 1, 2 and 5 and Extended Data Figs. 1, 4, and 5 are provided with the paper. CRISPR-Cas databases were accessed from CRISPRCasFinder and the prophage database was accessed from PHASTER.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Calculations for determining sample size were not required for this study. Experiments were performed in at least three biological replicates of well-defined genetic variants. No data was excluded from this study.
Data exclusions	No data was excluded
Replication	All experimental findings were reliably replicated in at least three biological replicate
Randomization	Randomization was not relevant to this study. Experimental comparisons were made between well-defined genetic variants, selection biases should not apply to such experiments.
Blinding	Blinding was not relevant to this study. All data reported relied on quantitative outputs and therefore was not subject to bias on interpretation of results.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging