Title:

Chromosomal curing drives prophage evolution and bacterial epidemiology

Author list:

Min Jung Kwun[1], Alexandru V. Ion[1], Katinka J. Apagyi[1] & Nicholas J. Croucher[1,*]

Affiliations:

[1] MRC Centre for Global Infectious Disease Analysis, Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, London, W12 0BZ, UK

* Correspondence to n.croucher@imperial.ac.uk

Abstract

Many integrative mobile genetic elements (MGEs) are sensitive to RecA-DNA filaments, suggesting their activation may enable them to evade deletion through homologous recombination (HR). This was studied using pneumococcal prophage. Most pneumococci contained prophage that inhibited transformation-mediated HR through chromosomal modification. These included a commonly-targeted *attB* that reduced transformation by modifying the small cytoplasmic RNA. The prophage were almost universally regulated by C1-type or ImmAR-type systems. Representatives with either type were efficiently deleted by HR. Mutations that impaired a C1-regulated prophage increased this deletion rate, reflecting this element being activated by RecA-DNA filaments imported during transformation. ImmAR-regulated prophage instead responded to transient stimuli by excising as deletion-resistant pseudolysogens, with varying RecA sensitivity, gradually committing to lysing cells following longer-duration signals. These processes were modulated by interactions with the DNA-binding protein DprA. Therefore bacterial diversification though HR is impeded by MGEs that selfishly prevent themselves from being eliminated from the chromosome.

Introduction

The population of viruses that infect bacteria encompasses both bacteriophage, the most abundant biological entities worldwide [1,2], and prophage, dormant sequences found in most bacterial genomes [3]. Both population genomic and metagenomic data has revealed these viruses to be highly diverse [4]. Despite this variation, the induction of prophage to generate phage by the genotoxic streptomycete-produced antibiotic mitomycin C (MMC) is common across the global microbiome, including samples from animal and human microbiotas [5,6], soil [7,8], freshwater [9], seawater [10], and subseafloor sediments [11]. Correspondingly, MMC induction of prophage and other types of mobile genetic element (MGE) [12,13] are observed in most bacterial phyla [3,14] and archaea [15,16]. Hence the capacity to respond to MMC is disseminated across a genetically diverse set of MGEs found within full taxonomic and ecological range of prokaryotic hosts.

This broad distribution is the result of convergent evolution of multiple regulatory mechanisms [13,15–18] that detect the RecA-coated single-stranded DNA (ssDNA) lesions generated after cells attempt to replicate DNA crosslinked by MMC [19]. In many bacteria, these RecA-ssDNA structures catalyse the autodigestion of a LexA transcriptional repressor, activating the DNA repair genes of the SOS response [20]. These structures also activate the LexA-like C1 repressor protein of *Escherichia coli* phage λ, the archetypal MMC-inducible MGE [21]. *E. coli* prophage 186 is instead activated by LexA cleavage derepressing expression of Tum, which inhibits a repressor [22]. Similarly, the CTX□ prophage of *Vibrio cholerae* and GIL01 prophage of *Bacillus thuringiensis* are directly repressed by LexA binding the prophage in conjunction with a phage protein [17,23]. By contrast, prophage SSV1 of the

archaeon *Saccharolobus shibatae* is proposed to be repressed by RadA, the archaeal orthologue of RecA, binding the prophage until it is removed through a higher-affinity interaction with ssDNA [15]. The sensitivity of other MGEs to MMC depends on regulatory machinery that structurally resembles the IrrE/PprI-DdrO regulators [24] that control DNA repair genes in the *Deinococcaceae* and *Thermaceae* families [25,26]. These use a bipartite system that features a ImmR transcriptional repressor [12] that can be degraded or displaced by an MGE-encoded ImmA protease [13,24].

Yet prophage remain MMC-sensitive even in the bacterium *Streptococcus pneumoniae* [27] (the pneumococcus), which lacks LexA [20], a canonical SOS response [28], or any well-characterised IrrE/PprI-DdrO regulators [26]. Instead, the equivalent pneumococcal response to MMC is to induce competence for natural transformation [29], enabling the bacteria to import exogenous DNA for integration into the chromosome through homologous recombination (HR) [30]. Although pneumococci frequently contain prophage [31,32] and other integrative MGEs [31,33–35], both β-lactam resistance [36] and vaccine evasion [37] occur through transformation. The requisite competence machinery includes a ComY (or ComG) pilus, able to contact exogenous DNA [38]; a ComE pore complex, which enables the import of single-stranded DNA (ssDNA) fragments, and ssDNA-binding proteins, such as DprA [39] and SsbB [40], which protect and store intracellular ssDNA. Their expression is coordinated by the quorum sensing pheromone competence stimulating peptide [41] (CSP), which induces competence in a fraction of cells, as the overall population engages in “bet hedging” [42–44]. Furthermore, pneumococcal competence is only transient, as after ~20 minutes is it shut down by DprA [45].

Despite this extensive regulation, the primary evolutionary advantage of transformation remains controversial [30,46].

One hypothesis proposes the main function of transformation is “chromosomal curing”, the elimination of deleterious MGEs from the genome through HR [47]. This is supported by the high efficiency with which non-essential sequences can be removed from the pneumococcal chromosome by transformation, which is not matched by an equivalent ability to insert equivalently-sized loci [48]. Hence transformation is asymmetric, implying it has the potential to cure cells of MGEs integrated into their chromosome, with little risk of inserting an additional element. Hence there is a theoretical basis for the intragenomic conflict between prophage and the competence machinery [47].

Yet integrative MGEs can only be deleted by HR when they are inserted into the chromosome at their *attB* (bacterial attachment) site, flanked by regions with similarity to the imported DNA. However, following activation, most prophage excise from the genome as extrachromosomal circular forms [17,49] that cannot be removed by HR. In the CI-regulated phage λ, this excision is traditionally regarded as the first stage of a rapid, irreversible transition from a lysogenic to lytic life cycle [21]. In other phage, the phage may be detected in a non-replicating circularised form, termed a pseudolysogen [49,50]. The low copy number renders such a form unstable [51–55], meaning pseudolysogeny is typically regarded as a stalled phase associated with starvation conditions [2,49,50]. Nevertheless, it has the potential to affect the efficiency of deletion by HR.

In this work, we test whether transformation is an efficient mechanism for deleting prophage despite their sensitivity to the ssDNA the competence machinery imports. This identifies the key interfaces in the evolutionary arms race between prophage and their hosts, and how this affects overall pneumococcal epidemiology.

Results

Most pneumococcal genomes contain prophage inhibiting transformation

Analysis of 20,047 published Global Pneumococcal Sequencing (GPS) project genomes [56] (Text S1, Fig. S1) estimated 14,113 were infected by 22,682 prophage-like sequences (Fig. 1A, S2), with the number per cell being Poisson distributed (Fig. 1B). These were distributed across 16 *attB* sites (Fig. 1A, S3-S5; Table S2), although the majority of insertions (19,255; 84.9%) occurred at the four most common *attB* sites [57,58]. The most frequently-affected *attB* site, $attB_{OXC}$, result in a modification of *ccnC* [59] that enhances the activity [60] of the competence-suppressing non-coding RNA it encodes [61]. Similarly, the fourth most commonly-targeted site disrupts the *comYC* competence pilus gene, necessary for transformation [62,63]. Among the rarer *attB* sites identified in this analysis, two were highly likely to inhibit transformation. One inserted between the promoter and translation start site of *yjbK*, a positive regulator of competence [43]. The other, previously identified in *Streptococcus mitis*, disrupts the *ssbB* gene [64], required for efficient recombination when high concentrations of donor DNA are available [40].

A phylogeny of the integrases targeting these *attB* sites identified seven clades of closely-related sequences (Fig. S6). This tree was consistent with each competence-

inhibiting *attB* site being targeted by an independently-evolved integrase, which has subsequently diversified to target rarer *attB* sites not predicted to limit DNA uptake. The exceptions were $attB_{MM1}$ [58], which duplicates the stop codon of *whiA* without affecting the encoded protein's sequence, and $attB_{scr}$, within the 3' half of the *scr* gene, encoding the small cytoplasmic RNA (scRNA). Genomic and transcript analyses of prophage integrations at $attB_{scr}$ found these insertions generated a modified scr_{attL} allele (Fig. 1C, Fig. S7). To test whether this affected transformation, *scr* was disrupted using the selectable and counter-selectable Janus cassette in *S. pneumoniae* R6. This caused a significant reduction in transformation efficiency, associated with a substantial growth defect (Fig. 1D, E). Restoring the cellular sequence (scr_{attB}) reversed both changes. However, replacing the cassette with scr_{attL} only repaired the growth defect, while the transformation efficiency remained ~25-fold lower than that of R6 (Fig. 1D, E). Therefore prophage insertions at $attB_{scr}$ inhibit transformation, without affecting host growth rate, by modifying the scRNA.

The $attB_{scr}$ site contained an insertion in 23.7% of the analysed pneumococcal genomes. It is ubiquitous across some common pneumococcal genotypes, including PMEN1 [33,62], due to the conservation of a prophage remnant at $attB_{scr}$ [63], which can be described as the Competence Inhibiting Prophage Remnant (CIPhR). Yet PMEN1 has diversified extensively through transformation [62]. PCR amplification of DNA from a PMEN1 isolate demonstrated that the CIPhR was able to excise and reintegrate into the bacterial chromosome (Fig. S7). Hence the native *scr* allele is restored when CIPhR is extrachromosomal, whereas the integrated form of the element will protect itself from being deleted through HR [60]. Therefore, including both CIPhR and full-length prophage, the majority of analysed pneumococcal

genomes (11,346 isolates; 56.6%; Fig. S2) contained MGE insertions at *attB* sites that inhibited transformation.

Pneumococcal prophage are primarily regulated by two types of system

Re-analysis of a previously-defined set of 538 prophage-derived sequences [65] found full-length elements contained one of two mutually-exclusive regulatory systems: a C1-like protein, or an ImmAR-like system (Text S1, Fig. S8). A search of the GPS isolates found 10,069 (50.2%) contained an ImmAR-type regulator, and 8,612 (43.0%) contained a C1-type regulator. The distribution of these systems across *attB* sites in monolysogenic isolates (Text S1, Fig. 1A) was heterogeneous (χ^2 test, $p < 10^{-16}$). The ImmAR-type systems dominated at the three most commonly-infected sites (att_{OXC}, att_{MM1} and $attB_{scr}$), whereas C1-type systems dominated at the sites that disrupted protein-coding genes that were part of the competence machinery (att_{comYC} and att_{ssbB}). This suggested that the ImmAR-type and C1-type regulators differed in their interaction with cellular transformation.

Consequently, three model systems were developed in which the prophage were active and the cells were transformable (Table S1). *S. pneumoniae* RMV8 [66] hosted a prophage encoding a C1-type regulator inserted at $attB_{\mathrm{OXC}}$, a chromosomal alteration that does not affect exogenously-induced competence [60] (Fig. 2A). This phage's lysogeny module also included a gene annotated as *thaL*, for "transmembrane HIRAN-domain protein associated with lysogeny" (Fig. 2B). *S. pneumoniae* 08B02743 [47] hosted a prophage encoding an ImmAR-type regulator inserted at att_{MM1}, the lysogeny module of which included a gene annotated as *szaL*, for "surface-exposed zinc metalloprotein associated with lysogeny". Finally, *S.*

pneumoniae RMV4 [66] contained a prophage encoding an ImmAR-type regulator inserted at att_{MM1}, the lysogeny module of which that included a gene annotated as *spaL*, for “surface-exposed zinc protein associated with lysogeny”. The 08B02743 and RMV4 prophage were highly similar outside of their regulatory loci, where RMV8 only shared sequence similarity in the structural and lytic regions of the prophage.

To improve the reproducibility of these systems [43,59,60], the phase-variable restriction-modification systems [66,67] were deleted or fixed in a stable arrangement (Text S2, Fig. S9-10). The prophage imposed a fitness cost on their host cells, as mutants in which the viral sequence was replaced with a Janus cassette grew to a higher cell density in stationary phase than the corresponding lysogens (Fig. 2C). Targeted disruption of the prophage integrase (*int*), replication (*rep*) and lytic (*lyt*) genes with a chloramphenicol acetyltransferase (*cat*) resistance gene resulted in growth phenotypes intermediate between the genotypes with intact and removed prophage, consistent with these mutations impairing phage replication. All three prophage were strongly activated by MMC (Fig. 2D, S11). Hence these three systems represent naturally transformable cells containing MMC-inducible viruses encoding the two most common regulatory strategies employed by pneumococcal prophages.

Pneumococcal prophage are efficiently deleted by transformation

To test whether transformation is an efficient mechanism for curing prophage from bacterial genomes, a base substitution causing rifampicin resistance was introduced into each mutant in which the prophage had been replaced with a Janus cassette.

These double mutants provided donor genomic DNA that could replace the prophage, resulting in aminoglycoside resistance, or transfer an unlinked single nucleotide polymorphism, resulting in rifampicin resistance. The ratio of transformation with these two resistance phenotypes therefore enabled quantification of the efficiency with which prophage were deleted, relative to the transfer of a single base substitution [48]. This ratio was measured two hours after the addition of the DNA, as this allowed for the resistance phenotypes to be expressed, but minimised the time for replication and selection to distort the ratio away from that reflecting the relative rates of mutation import (Fig. S12). For all three prophage systems, the median ratio of deletion to base substitution acquisition was above one (Fig. 2E). This contrasted with the rate at which transformation was able to insert complete prophage sequences, which was estimated as >10,000-fold lower than the rate at which base substitutions were imported (Fig. S13). Therefore, transformation is an efficient mechanism for removing, but not importing, prophage.

These same deletion ratios were also calculated for the impaired prophage mutants, to test whether the intact phage were sometimes escaping deletion (Fig. 2E). In the □RMV8 system, these mutations significantly increased the efficiency of prophage deletion ~3-10 fold. This was primarily driven by a higher rate of prophage deletion, rather than a change in rifampicin resistance acquisition (Fig. S14). The high deletion rate of the □RMV8 *int*::*cat* mutant, which should remain embedded within its *attB* site, demonstrated efficient integration of the aminoglycoside resistance marker was not dependent on the excision of the prophage from the host chromosome. To confirm this higher rate was associated with the *int* mutation, the deletion rate was replicated in an *int*::Janus mutant, in which □RMV8 deletion was assayed by

measuring replacement by an *ermB* macrolide resistance marker. The deletion rate then fell after the counter-selectable Janus cassette was replaced through restoration of the functional *int* gene (Fig. 2E). Hence the rate of deletion of the C1-regulated □RMV8 prophage depended on the ability of the prophage to activate.

By contrast, removal of the *int* gene decreased the efficiency with which the □08B02743 prophage was deleted. Other mutations impairing □08B02743 or □RMV4 did not change the efficiency with which transformation removed these elements. To test whether this was attributable to the ImmAR-type regulator, the *szaL* gene of □08B02743 was replaced with a *cat* gene, enabling the transfer of the regulatory locus of this prophage into the □RMV8 background to generate □RMV8$_{immARcat}$ (Fig. 2B). This hybrid prophage was inducible by MMC (Fig. 2D, S10), suggesting it was functional. This transfer did not substantially affect the rate at which the intact prophage was deleted, which remained similar to that of both the parental phage (Fig. 2E). In contrast to □RMV8, disruption of the *int* gene did not significantly elevate the deletion rate of □RMV8$_{immARcat}$. This difference did not seem to be attributable to changes outside the prophage (Fig. S15), implicating the ImmAR-type regulators as the cause. Yet, the ImmA protease itself did not appear to have a direct role in suppressing the deletion of ImmAR-regulated prophage, as an □08B02743 *immA*::*cat* mutant was just as efficiently removed from the genome as the intact prophage (Fig. 2E). Therefore the rate of deletion of the ImmAR-regulated prophage appeared to be independent of their ability to activate.

One hypothesis explaining these differences between the regulatory systems is that there is a difference in their spatial sensitivity to the HR events enabled by

transformation. As RecA-ssDNA nucleoprotein filaments are expected to interact with the C1-type protein, itself often tightly associated with the phage regulatory locus [21], it seemed likely that activation might be maximal when DNA specifically targeted the prophage element, as proposed for *Bacillus subtilis* 105 [68]. This would contrast with the bipartite ImmAR-type regulators, in which ImmA could encounter stimuli throughout the cell, before interacting with prophage-bound ImmR. Hence impairment of the prophage encoding ImmAR-type regulators would reduce prophage-driven lysis of cells acquiring the Janus cassette or rifampicin resistance mutation equally. By contrast, impairment of RMV8 would only reduce the frequency of Janus cassette acquisition, as the prophage would not respond to HR importing rifampicin resistance at a distant locus. This was tested by transforming cells with PCR amplicons targeted at different parts of the prophage. Although PCR amplicons targeting the regulatory region did appear to activate RMV8 more effectively than those annealing to other parts of the prophage, the same pattern was observed in RMV8$_{immARcat}$ (Fig. S16). Hence the apparent spatial sensitivity of prophage activation was likely an artefact of local variation in recombination rate, as observed in other assays [48]. Therefore the prophage deletion ratios were not explained by regulators' differential spatial sensitivity to HR events.

C1-regulated pneumococcal prophage activate in response to CSP

An alternative hypothesis was that C1-type regulators responded more sensitivity, or rapidly, to competence induction than ImmAR-type regulators. This could explain the similarity in the deletion rates across all intact prophage as C1-regulated prophage being activated equally in cells deleting the prophage or acquiring rifampicin resistance, whereas ImmAR-regulated prophage did not activate in either. This

would account for the differential deletion rates of mutated C1-regulated prophage as being a consequence of the elevated proportion of them being too slow to lyse cells before being deleted by transformation (Fig. 2C), if they were targeted by imported DNA. Nevertheless, they could still lyse host cells that became competent, and acquired rifampicin resistance, but failed to eliminate the prophage. Hence the deletion rate would rise because more cells survived when transformation removed the prophage, but the same fraction of bacteria that acquired rifampicin resistance would be lost. To test this hypothesis, the activation of the four prophage systems, including □RMV8$_{immARcat}$, was studied through assaying gene expression after exposure to MMC or competence induction by CSP for two hours (Fig. 3A), corresponding to the period allowed for transformation-mediated deletion.

While MMC did not affect *recA* expression, transcription of this gene was induced by CSP in all systems except □08B02743. This likely reflects variation in the stochastic bet hedging process; while MMC will induce ssDNA lesions uniformly across the population, CSP will only induce competence in a variably-sized fraction of the population [42–44]. Nevertheless, □RMV8 responded more strongly to CSP than MMC, consistent with the C1-regulated prophage being most sensitive to competence induction. First, the *int* gene was elevated 60 min post-CSP, followed by ~10-100-fold upregulation of the replication and lysis genes by 120 min post-CSP. These kinetics are consistent with C1-regulated prophage evading deletion through activation. This contrasted with □RMV8$_{immARcat}$, in which there was a smaller increase in the expression of regulatory and replication genes 120 min post-CSP, but no rise in lytic gene expression relative to untreated cells. Yet both □RMV8 and □RMV8$_{immARcat}$ responded near-identically to MMC, with 2-10-fold rises in the

expression of all prophage genes by 120 min post-MMC. Hence only the C1-regulated □RMV8 was induced by CSP to express the genes necessary to lyse host cells, corresponding with the more efficient deletion of this prophage when these loci were disrupted by mutation (Fig. 2E).

Consistent with the hypothesised lower sensitivity of the ImmAR-regulated prophage to competence induction, □08B02743 showed an even smaller transcriptional response to CSP and MMC than □RMV8$_{immARcat}$. As the only difference between their regulatory systems with the absence of *szaL* in the hybrid prophage, this gene was disrupted. However, this mutation neither affected the induction of □08B02743 by MMC, nor failure of the host cell to upregulate *recA* post-CSP (Fig. S17). The □RMV4 system also did not exhibit a substantial response to MMC or CSP, with activation instead rising over time as population density increased. This suggests the phage may be sensitive to a quorum sensing signal, which are commonly employed by pneumococci [69], or nutrient availability. Either mechanism would also explain why □RMV4 mutations affected stationary phase cell density, rather than exponential growth rate (Fig. 2C).

Three sets of qPCR assays were used to understand how these gene expression differences related to prophage excision and activation (Fig. 3B). The *attB*-to-*attL* ratio quantified the prevalence of bacteria in which the prophage had excised, relative to those in which a prophage remained integrated. The extent of excised prophage replication was measured as the concentration of *attP* relative to chromosomal DNA, estimated using *rpoA* copy number. The same ratio was also measured in the culture supernatant, to quantify the release of virions relative to cell

lysis, as it is not possible to use plaques to enumerate phage that infect encapsulated pneumococci [70]. Consistent with the gene expression data, □RMV8 exhibited the strongest response to CSP (Fig. 3C). This prophage excised from the chromosome in the first 60 min after the addition of CSP, before replicating, and lysing host cells by 120 min. MMC triggered a similar, albeit weaker, response. Hence competence is a natural stimulus for the activation of C1-regulated pneumococcal prophage.

The ImmAR-regulated prophage exhibited a different response to both stimuli. MMC induced all three to excise from the chromosome after 60 min, with the concentration of circular □08B02743 rising ~100-fold, despite none of them showing a strong transcriptional response to this stimulus at this timepoint. However, the absence of a substantial rise in the concentration of extracellular phage DNA implied there was limited cell lysis and virion release by 120 min. This would be consistent with increasing proportions of the ImmAR-regulated prophage populations becoming pseudolysogens in response to MMC. The response of both □RMV4 and □RMV8$_{immARcat}$ to CSP was similar to their behaviour after exposure to MMC. Therefore the unresponsiveness of □08B02743 to CSP likely reflected the limited upregulation of competence genes in this host (Fig. 3A). The divergence of the □RMV8$_{immARcat}$ response from that of □RMV8 supported the attribution of this behavioural change to the ImmAR-type system itself. Therefore ImmAR-regulated prophage excised in response to MMC or CSP, but frequently stalled in a pseudolysogenic state.

ImmAR-regulated prophage exhibit differential responses to RecA

The stronger response of □RMV8 to CSP, which increases *recA* expression, than MMC, which generates ssDNA lesions without stimulating *recA* transcription (Fig 3A), is likely attributable to its dependence on RecA. The similar responses of the ImmAR-type regulators to both stimuli suggested they were less dependent on RecA for activation. To test this difference, the induction of lysogens by MMC was assayed in *recA*$^-$ cells (Fig. 2D, S11). The C1-regulated □RMV8 was dependent on RecA to efficiently lyse cells in response to MMC. However, the ImmAR-regulated prophage all lysed a high proportion of host cells in the absence of RecA. Although MMC appeared to have a diminished effect on the activation of □08B02743, this was mainly a consequence of the *recA*$^-$ mutant exhibiting a substantial growth defect in the absence of MMC (Fig. S18). This suggested ImmAR-type regulators could function independently of, or potentially be repressed by, RecA.

To test for evidence of RecA-mediated repression, the *attB*-to-*attL* ratio was measured in wild type cells and *recA*::Janus mutants of all four prophage systems, as excision from the genome was the most reproducible signal of □08B02743 activity (Fig. 4A). Excision of the C1-regulated □RMV8 required RecA and the prophage integrase. In contrast, both □08B02743 and □RMV8$_{immARcat}$ increased their excision from the genome when *recA* was disrupted, consistent with RecA repressing these prophage. Yet excision of □RMV4 was dependent on RecA. This variation between the ImmAR-regulated prophage could not be attributed to SzaL or SpaL. Instead, excision of □08B02743 was abrogated when *immA* was disrupted. Additionally, transfer of the □08B02743 lysogeny regulation locus into □RMV4 generated □RMV4$_{immARcat}$, the excision of which was also increased in a *recA*$^-$ host.

Therefore the activation of excision by ImmAR-type regulators varied in its sensitivity to RecA.

This suggested ImmAR-regulated prophage integrated multiple signals when activating different behaviours. This is supported by excision not being a prerequisite for efficient lytic gene expression, as all the *int*::*cat* mutants responded to MMC at least as strongly as the intact prophage (Fig. 2D). Quantitative PCR assays demonstrated this was not associated with *in situ* replication (Fig. S19). Significantly, the phage genes for lysis, rather than for lysogeny, were orientated with the coding bias of the pneumococcal genome at all 16 *attB* sites, facilitating their expression even while the prophage were still chromosomally integrated (Fig. S3-5). The separate regulation of excision and lysis added further complexity to the role of RecA in MMC induction of prophage, with □RMV8 being RecA-dependent, □08B02743 activating through a RecA-independent mechanism, while RecA was necessary only for the excision of □RMV4 from the chromosome.

RecA can repress or activate prophage expression

As the regulatory role of RecA is difficult to compare between 08B02743 and other isolates, owing to the differences in the expression of *recA* (Fig. 3A), the protein's effects on □RMV8 and □RMV8$_{immARcat}$ were tested. Quantifying gene expression in *recA*::Janus mutants (Fig. S20), and standardising these to pre-exposure expression levels in *recA*$^{+}$ cells (Fig. 4B), demonstrated that □RMV8 genes were repressed by 10-100-fold in *recA*$^{-}$ cells, whereas □RMV8$_{immARcat}$ genes were ~5-fold upregulated. Both CSP and MMC failed to activate □RMV8 *recA*::Janus, whereas □RMV8$_{immARcat}$ *recA*::Janus still responded to MMC, suggesting activation was driven by ssDNA

independently of RecA. Hence RecA's repressive effect could be an indirect consequence of the protein competing for ssDNA-based structures that activated the prophage. Therefore to test whether the effect of RecA changed with its concentration, the *recA* genes of □RMV8 and □RMV8$_{immARcat}$ were engineered to be under control of a *lacI* repressor sequence that was inducible by isopropyl β-D-1-thiogalactoside (IPTG; Fig. S21). While IPTG caused a small decrease in cell growth in unsupplemented media for both prophage systems, it enhanced the effects of MMC on □RMV8, while limiting the induction of □RMV8$_{immARcat}$ (Fig. 4C). This effect of IPTG was dose-dependent and not observed when *recA* was expressed from its native promoter (Fig. S22). Hence replacing a C1-type regulator with an ImmAR-type regulator reversed the sensitivity of a prophage to RecA concentration.

The synergistic activation of □RMV8 by RecA overexpression and MMC were consistent with RecA-ssDNA nucleoprotein filaments being the trigger for prophage activation. This suggested the activation of C1-type regulator by CSP may by driven by the import of ssDNA through the ComE competence pore. Therefore □RMV8 gene expression was quantified in *comEA*::Janus mutants, which lack a functional pore (Fig. 4D). Despite *recA* still being upregulated after CSP addition, the □RMV8 genes failed to respond to the induction of competence in the *comEA*::Janus mutant. Assaying the excision and replication of the prophage confirmed induction by CSP depended on the import of DNA, whereas the response to MMC-induced ssDNA lesions did not (Fig. 4E, S23).

ImmAR-regulated prophage adopt divergent HR-sensitive activation strategies

The RecA-independent activation of 08B02743's ImmAR-type regulators by MMC (Fig. 4B) suggested these proteins may be activated by ssDNA alone, as for the ImmA orthologue PprI [71], or by the association of a different protein with ssDNA. Hence the effects of other HR-associated DNA-binding proteins to activate this prophage were assayed (Fig. S24). The most substantial change in phenotype was an increased sensitivity to both MMC and CSP in a *dprA*::Janus mutant (Fig. 5A). These responses were reduced in a prophage-cured mutant (Fig. S24), indicating they were not simply a growth defect resulting from misregulation of competence [45]. This response may result from the competence-induced DprA protein itself acting as a repressor, or activation of the prophage by the prolonged competence period in the absence of DprA-mediated shutdown (Fig. 5B). An *immA*$^-$ *dprA*$^-$ double mutant was less responsive to MMC, but retained hypersensitivity to CSP, indicating this mechanism was independent of ImmAR (Fig. 5A). Yet disruption of *dprA* did not increase induction of RMV8 or RMV4 (Fig. S24), indicating CSP-associated lysis was associated with the 08B02743-specific SzaL. Correspondingly, a *szaL*$^-$ *dprA*$^-$ double mutant was less susceptible to induction by CSP (Fig. 5A). Furthermore, a *szaL*$^-$ single mutant replicated the MMC hypersensitivity of the *dprA*$^-$ mutant (Fig. 5A), implying 08B02743 encoded a secondary regulation mechanism that regulates cell lysis in response to competence induction.

Consistent with the effects of CSP, SzaL did not have a substantial effect on phage gene expression (Fig. S17). Therefore its effect on the topological state of 08B02743 was tested after MMC and CSP exposure in *recA*$^-$ and *dprA*$^-$ cells (Fig. 5C). Neither DprA nor SzaL had a substantial effect on 08B02743 excision or replication, while both processes were detectably higher in the presence of MMC

and the absence of RecA. However, the *dprA*⁻ and *szaL*⁻ mutations were associated with a decreased ratio of prophage to chromosomal DNA in the culture supernatant. The double mutant exhibited the same phenotype as the single mutants, again consistent with the effects of DprA being mediated through SzaL. Given the faster MMC-induced cell density decline in the *szaL*⁻ and *dprA*⁻ mutants (Fig. 5A), this suggested SzaL moderated cell lysis in the presence of DprA, regulating the transition from intracellular replication to virion release.

The equivalent approach was used to test whether DprA and SpaL formed an analogous signalling pathway in □RMV4 (Fig. 5D). The *spaL*⁻ mutation increased the activation of the prophage by CSP. However, rather than the transient induction observed in □08B02743, this activation persisted as cell density rose. Correspondingly, although *spaL* disruption had little effect on prophage gene expression (Fig. S17), loss of this gene eliminated the apparent cell density-dependence of □RMV4 excision, replication and lysis (Fig. 5E). While the *spaL*⁻ mutation had a substantial effect on □RMV4 activation in a *recA*⁻ background, there was little evidence of an interaction with DprA, the loss of which abrogated cell lysis by □RMV4 (Fig. 5D). Correspondingly, the *dprA*⁻ mutant still underwent population density-dependent excision and replication, but the apparent lack of lysis meant viral DNA was difficult to detect in the supernatant (Fig. 5E). Therefore these similar ImmAR-regulated prophage employed divergent activation strategies that integrated information from co-regulators and the HR protein DprA, in addition to RecA, to alter the rates of pseudolysogeny and virion release.

ImmAR-type regulators require longer signal duration to cause cell lysis

The combination of signals required to release virions of the ImmAR-regulated prophage suggested they would be slower to commit to cell lysis than the C1-regulated □RMV8. Nevertheless, both types of prophage lysed cells with similar kinetics when continuously exposed to MMC (Fig. 2D). To test whether the time taken to decide to lyse the host cell differed between the four prophage systems, they were each exposed to MMC pulses of 20 min (equivalent to period for which the competence machinery is induced), 60 min (the timepoint at which prophage were observed to have excised from the chromosome), and 120 min (the timepoint at which cells were lysed by □RMV8). The washed cells were then grown in unsupplemented media (Fig. 6A, Fig. S25).

These experiments demonstrated that □RMV8 committed to lysing host cells between 20 and 60 min, with the additional 60 min MMC exposure causing little additional activation. However, the three ImmAR-regulated prophage increasingly committed to host cell lysis throughout the 120 min. This effect was observed at two different MMC concentrations, and was not an artefact of □RMV8 being more sensitive to MMC in continuous culture (Fig. S26). Therefore the C1-type regulators and ImmAR-type regulators prophage are associated with two different strategies for controlling the activation of phage (Fig. 6B).

Discussion

This study demonstrates that pneumococcal transformation is a highly efficient mechanism for the deletion of prophage (Fig. 2E), resulting in an intragenomic conflict in which integrative MGEs can benefit from reducing their rate of removal by the cellular HR machinery. The three characterised natural prophage each exhibited

distinct strategies for evading deletion. The C1-regulated □RMV8 activated its lytic cycle in response to a short pulse of MMC (Fig. 6A), or the transient induction of competence by CSP (Fig. 3). As only a small proportion of pneumococci both become competent after CSP administration [42,43] and import DNA, the detected magnitude of phage induction suggests a very high sensitivity to RecA-ssDNA filaments. Phage were excised from the genome after 60 min, and had activated transcription, replicated and lysed their host cells by 120 min. Such lysogens that activated in competent cells were described as “naïve” in *Bacillus subtilis* (Fig. 6B), as their induction was characterised as a maladaptive response to the competence-associated triggering of the SOS response [72]. However, the high sensitivity of the C1-type regulator to RecA-ssDNA filaments can instead be understood as an adaptation that enables prophage to escape HR-mediated elimination, which targets vertical inheritance, by rapidly transitioning to a lifecycle that enables horizontal transmission [73]. The speed of this response appears crucial for the prophage to evade deletion by HR, suggesting a finely-balanced race between MGEs and the cellular recombination machinery. Yet the RecA-ssDNA filaments appear late in the transformation process [30]. Hence it is likely the C1-type repressors recognise these structures because they are highly-conserved as the point of convergence [74] between the otherwise variable competence machinery, found across many bacteria [30] and archaea [75], as well as other HR mechanisms [76,77] that could drive prophage deletion.

By contrast, the ImmAR-regulated prophage exhibited a “cautious” strategy (Fig. 6B). After 60 min, both □RMV4 and □RMV8$_{immARcat}$ excised from the genome in response to CSP or MMC (Fig. 3C). The magnitudes of these responses were similar at a

population level, which suggests frequent, rapid excision of prophage in the subset of cells that induce competence. Although □08B02743 only responded to MMC, this is likely an artefact of the atypically low competence induction in this system (Fig. 5B), because □RMV8$_{immARcat}$ responded more strongly despite sharing the same ImmAR-type regulators (Fig. 3C). However, none of the ImmAR-regulated prophage activated transcription of viral genes by 120 min (Fig. 3A). This is consistent with the longer MMC pulse required to induce these viruses (Fig. 6A), suggesting the excised pseudolysogens may reintegrate once the activation stimulus is removed. Temporarily becoming a pseudolysogen, to evade a time-limited risk of removal by HR, before reinserting to the chromosome to ensure stable inheritance after cell division would likely to be an optimal means of maximising vertical inheritance in a transiently competent host (Fig. 6B). Such behaviour is difficult to rationalise in the absence of the risk of deletion by HR.

Therefore both the naïve and cautious strategies likely evolved under selection to evade chromosomal curing, explaining the similar efficiency of deletion of all the wild type prophage (Fig. 2E). However, the co-existence of the two strategies across pneumococci suggests they are advantageous in different situations. The most likely distinction is that cautious prophage may survive without being forced into lysing their host in monolysogens (~40% of pneumococci), whereas they risk their host cell being lysed before they have generated virions in polylysogens (~30% of pneumococci) that also contain a naïve prophage (Fig. 1B). Furthermore, within the cautious phage, there were differences in the effect of both RecA on excision (Fig. 4A) and DprA on lysis (Fig. 5A,D), consistent with differing sensitivity to ssDNA-based structures. Such unpredictability in phage responses may be key to

preventing cells from suppressing the signals needed for individual prophage to activate, as might have otherwise been possible by 08B02743 limiting *recA* upregulation during competence (Fig. 5B). Therefore the multiple pressures shaping phage evolution can explain the co-existence of these distinct strategies between competing prophage (Fig. 6B).

The detailed characterisation of representative prophage in this analysis precludes the sample size necessary to determine whether the naïve and cautious strategies always correlate with the C1-type and ImmAR-type regulators, respectively. Nevertheless, a transition from a naïve to cautious strategy occurred when □RMV8$_{immARcat}$ was generated from □RMV8 (Fig. 2E, 3, 4B). This is consistent with the change depending on a single locus. These differences are also explicable through the biochemistry of the regulators: while *E. coli* phage λ C1 is considered to undergo an irreversible switch [21], the ImmA-type protease activator of at least one streptococcal phage is less stable than its paired ImmR repressor [78], suggesting that after an inducing stimulus is removed, an increased ratio of ImmR to ImmA may limit the induction of the phage. Furthermore, the phage regulatory proteins' contrasting response to inducing signals can also explain their differential distribution between *attB* sites (Fig. 1A). C1-type regulators dominate at *attB* sites where prophage inhibit transformation through causing a loss-of-function disruption to a protein-coding gene [62,64] (e.g. *attB*$_{comYC}$, *attB*$_{ssbB}$), whereas ImmAR-type regulators dominate where competence is inhibited through gain-of-function mutations in RNA genes [60] (e.g. *attB*$_{scr}$, *attB*$_{OXC}$). This is consistent with ImmAR-type regulators being adapted to reintegration, as translation of a restored protein-coding gene would likely enable competence at the time the prophage reinserted,

whereas the gain-of-function suppressors remain active even when they are only part of a mixed RNA population [60].

This proposed co-evolution between regulator and *attB* site is supported by the CIPhR element, a prophage remnant encoding an ImmAR-type regulator that inhibits transformation when reversibly integrated into the scRNA gene. This inhibition of competence through an MGE-modified regulatory RNA is likely to be common, as it has already been observed in both pneumococci [60] and Legionella species [79]. Unlike active pneumococcal prophage, CIPhR is stable in multiple pneumococcal strains [31], which are nevertheless able to detectably diversify through transformation, in contrast with those in which prophage disrupt *comYC* [37] or transposons disrupt *comEC* [34]. Hence rather than abrogating competence completely, its selfish inhibition of HR when integrated makes it an active phage regulatory switch [80], similar to the reversibly-integrating *comK*-interrupting element in *Listeria monocytogenes* [81].

This likely represents a specific example of how the complex regulation of pneumococcal competence [43] has emerged through the co-evolutionary arms race with phage. The highly-conserved DprA protein [30,82] appears to be a key interface in this interaction, consistent with related proteins having anti-phage functions [83]. By coating ssDNA imported by the competence machinery [82], DprA delays the formation of RecA-ssDNA filaments that activate C1-regulated prophages, while also shortening the duration of competence [45], to limit the induction of slowly-activating ImmAR-regulated prophages. In 08B02743, DprA also appears to moderate the bet hedging that determines the heterogeneous distribution of competence throughout

pneumococcal populations, such that its loss enabled a sufficiently widespread competence induction that a substantial proportion of the population underwent prophage-associated lysis (Fig. 5A). Therefore it is likely that the bet hedging induction of competence in only a subset of pneumococcal populations [42–44], and the rapid shutdown of the transient state [45], reflects the trade-offs between the benefits of chromosomal curing of MGEs by HR, and potential for activating prophage.

Hence prophage limit pneumococcal transformation through their targeting of *attB* sites, their elevated rates of activation in competent cells, and by selecting for genotypes that hedge their bets against the risks of inducing competence. This has substantial implications for pneumococcal epidemiology, as transformation is required for the acquisition of resistance [34,36] and evasion of vaccine-induced immunity [34]. Such impacts of the co-evolutionary arms race between phage and RecA-mediated HR will be ubiquitous across prokaryotes, given the broad distribution of MMC-sensitive prophage across microbiomes [5–11], and the plethora of cellular DNA exchange mechanisms that depend on HR. Its impact may be even more extreme in other species, as the competence machinery is broadly conserved [30], yet often difficult to induce *in vitro*, suggesting a strong pressure to suppress its activation. Hence as both MMC-responsive prophage [15,84–86] and competence, let alone cellular HR [30,75–77], are more taxonomically broadly distributed than the SOS response [20], it is likely that deletion of integrated genomic parasites was the original selection pressure for the evolution of RecA- and ssDNA-sensitive regulators such as C1 and ImmAR. Hence this ancient global conflict between MGEs and their

hosts has determined, and will continue to shape, the evolution and epidemiology of many bacterial pathogens.

Methods

Bacterial isolates and sequence data

The isolates from which all mutants were derived are described in Table S1, associated with the accession codes for relevant sequence data. All mutant genotypes used in this work are listed in Table S3. These are based on the previously-characterised *S. pneumoniae* RMV8 [60,66]; *S. pneumoniae* 08B02743 [47]; *S. pneumoniae* RMV4 [66]; *S. pneumoniae* 11930 [62], or *S. pneumoniae* R6 [87]. Derivatives of each were generated that lacked the phase-variable *ivr* and *tvr* loci [66,67,88], encoding restriction-modification systems, using the primers detailed in Table S4.

Bioinformatic and statistical analysis of prophage sequences

The bioinformatic analysis of prophage sequences is detailed in Text S1. In summary, prophage were identified through BLASTN [89] comparison of 20,047 Global Pneumococcal Sequencing (GPS) project isolates [56] with a manually-curated database of 63 known prophage sequences. This enabled the identification of 16 *attB* sites (Table S2). A further BLASTN search identified instances were these *attB* sites were disrupted.

An independent analysis of prophage sequences identified in a smaller dataset [31] enabled the identification of C1-type and ImmA-type regulatory proteins. These were used to generate hidden Markov models using MAFFT [90] and HMMer3 [91], which were then used to classify proteins encoded by the GPS isolates. These analyses

were highly consistent when combined to generate the overall analysis of prophage distribution.

The scRNA structures were predicted using MXfold2 [92] and Rfam [93], and plotted using RRNA [94]. Other plots and statistical analyses were undertaken in R [95] using the tidyverse [96], cowplot [97], ggpubr [98], ggtree [99], genoPlotR [100], circlize [101] and fitdistrplus [102] libraries.

Culturing and growth assays

S. pneumoniae were grown at 35°C in 5% CO_2. Culturing on solid media used Todd-Hewitt broth supplemented with 0.5% yeast extract (Millipore) and 1.5% agar (Sigma-Aldrich). Culturing in liquid media used a 2:3 ratio mixture of Todd-Hewitt broth (Sigma-Aldrich) with 0.5% yeast extract (Sigma-Aldrich) and Brain-Heart Infusion media (Sigma-Aldrich), as described previously [60]. For growth assays, $2x10^4$ cells from titrated frozen stocks stored in 10% glycerol (Sigma-Aldrich) were used to inoculate 200 µl liquid media in 96-well microtiter plates at 35°C in 5% CO_2. The optical density at the wavelength of 600 nm (OD_{600}) was measured at 30 min intervals using a FLUOstar Omega microplate reader (BMG LABTECH). At least three independent biological replicates of each genotype were grown in separate wells for each analysis.

For measuring prophage induction by different stimuli, pneumococcal cells were first grown in 10 ml of mixed media at 35°C in 5% CO_2 overnight. A 750 µl sample was used to inoculate a fresh 10 ml liquid culture, which was maintained under the same conditions until it reached an OD_{600} of 0.2, measured using an Ultrospec 10

photospectrometer (Amersham Biosciences). Two hundred microlitre samples of this early exponential phase culture were then transferred into individual wells of 96-well microtitre plates. Individual wells were then supplemented with stimuli with the potential to induce prophage: 0.75 µl 0.1 mg ml^{-1} CSP1, or 0.5 mg ml^{-1} CSP2 (AnaSpec; Table S1); 1 µl 500 mM calcium chloride (Sigma-Aldrich); 2 µl 0.5 µg ml^{-1} mitomycin C (Cayman Chemical Company); 2µl 9% aqueous hydrogen peroxide solution (LCM Group); 2 µl 0.02 M $ZnCl_2$ (Sigma-Aldrich); 5 µl 0.02 M sodium citrate tribasic dihydrate (Sigma-Aldrich), or 5 µl 0.2 M manganese (II) chloride tetrahydrate (Sigma-Aldrich). The optical density of each well was then measured as described above.

Generation and sequencing of DNA constructs

To extract intracellular DNA, cells were pelleted by centrifugation for 10 min at 4000 *g*, followed by resuspension in 480 µL lysis buffer (Promega) and 120 µL 30 mg mL^{-1} lysozyme (Promega), and 30 min incubation at 35 °C. Genomic DNA was then extracted from cells using the Wizard Genomic DNA purification kit (Promega). Whole genome sequences were generated using a PromethION device (Oxford Nanopore Technologies; Table S1).

To quantify the concentration of phage DNA released into the medium from lysed cells, a 2 ml culture sample was spun down for 10 °C for 10 min at 13,000 *g*. The supernatants were then filtered through 0.22 µm cellulose filters (Sartorius) and stored at -80°C. The frozen supernatants were thawed and diluted in a 1:1 ratio with water, then incubated at 95 °C for 10 min. This freezing, thawing and heating cycle was repeated once more, to ensure release of phage DNA from virions. This

processed supernatant was then used directly as the template for subsequent analyses.

PCR amplification used either Dreamtaq (Sigma-Aldrich), for establishing the size of amplicons, or Platinum PCR SuperMix High Fidelity (Invitrogen), for generating constructs for mutagenesis or sequencing. The oligonucleotides (Invitrogen) used to prime PCR amplifications are listed in Table S4. For the construction of mutants, the upstream (or left) flanking region was amplified with oligonucleotides that added an ApaI restriction site at the 5' end of the right primer. The downstream (or right) flanking region was amplified with oligonucleotides that added a BamHI restriction site at the 5' end of the left primer. The appropriate resistance marker (a Janus cassette, or an *aph3'*, *cat* or *ermB* gene) was amplified using primers that carried ApaI and BamHI sites at the 5' ends of the left and right primers, respectively.

Individual DNA constructs were extracted from agarose gels after electrophoresis using the QIAquick PCR & Gel Cleanup kit (Qiagen). DNA concentrations were quantified using dsDNA BR Assay Kit for the Qubit (Qiagen). PCR amplicons were digested using the appropriate restriction enzymes (Promega), according to the manufacturer's instructions. The digested DNA molecules were ligated using T4 DNA ligase (Sigma-Aldrich) at room temperature overnight, and the full construct amplified by PCR. This DNA amplicon was checked through agarose gel electrophoresis or dideoxy terminator sequencing (Genewiz), depending on its final use.

Transformation and mutagenesis

Competence was induced by addition of the appropriate CSP pherotype (Table S1) at early exponential phase, defined as an OD_{600} of 0.2-0.25. A 1 mL sample of the culture was mixed with 5 µl 500 mM calcium chloride and the transforming DNA (1 µg for genomic DNA, or 100 ng for a PCR construct). After 2 h incubation, cells were plated on the appropriate selective media. To select transformed cells, solid media were supplemented with the appropriate antibiotic. For the Janus cassette or *aph3*' marker, kanamycin (Sigma-Aldrich) at 400µg ml^{-1} was used. For the *ermB* marker, erythromycin (Sigma-Aldrich) at 1µg ml^{-1} was used. For the *cat* marker, chloramphenicol (Sigma-Aldrich) at 4µg ml^{-1} was used. For the alteration of *rpoB*, rifampicin (Fisher Scientific) at 4µg ml^{-1} was used.

For the assays of prophage deletion, donor genotypes were constructed by selecting for the replacement of a prophage by an appropriate resistance marker, then adding a rifampicin resistance base substitution. The recipient cells were transformed with 1 µg of the genomic DNA. The successful transformants were selected on agar plates supplemented with either rifampicin or kanamycin.

Extraction of RNA for gene expression quantification

Cells were grown in 10 ml of media overnight and used to inoculate six fresh 10 ml cultures. At an OD_{600} of 0.2, a total of 4 ml was collected across all cultures to represent the pre-treatment timepoint. Two cultures were maintained as untreated controls; two were supplemented with 25 µl CSP (1 mg ml^{-1} CSP1, or 0.5 mg ml^{-1} CSP2; Table S1), and two were supplemented with 100 µl mitomycin C. At 20, 60 and 120 min post-treatment, 4 ml of cell culture was collected to represent each condition. All samples were immediately treated with RNA protect at a 1:1 ratio. RNA

was then extracted using the SV Total RNA Isolation System (Promega), according to the manufacturer's recommendations.

Nucleic acid analysis with quantitative real-time PCR

The concentration and quality of all extracted RNA samples was assessed using a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific). A 0.2 µg sample of RNA was then treated with amplification-grade DNase I (Invitrogen) and used to template a reverse transcription reaction using the First-Strand III cDNA synthesis kit (Invitrogen). Each reaction used 100 units of the SuperScript III reverse transcriptase, 1 µL of 100 µM random hexamer primers (Thermo Fisher Scientific) and 1 µL of 10 mM dNTP mix (Bioline). This mixture was incubated at 50 °C for 30 min, then 55°C for a further 30 min, prior to reverse transcriptase inactivation through heating at 70 °C for 15 min.

Quantitative real-time PCR amplifications were run using MicroAmp™ Optical 96-well reaction plates (Thermo Fisher Scientific) and the QuantStudio™ 7 Flex Real-Time PCR System (Applied Biosystems). Samples were denatured at 50 °C for 2 min, prior to 40 amplification cycles of denaturation at 95 °C for 15 s, followed by annealing and elongation at 60 °C for 1 min. Validation of the results used melt curve analysis of each amplicon, and standard curves that were generated for each DNA or RNA amplicon of interest on each plate. For each analysis, three technical replicate measures were made of three or more biological replicates.

Data Availability

The raw data are available from a Github repository (https://github.com/nickjcroucher/prophage_transformation_analysis). The hidden Markov models used for identifying C1-type and ImmAR-type regulator proteins are available from FigShare (https://doi.org/10.6084/m9.figshare.26425075). The sequence data used in this study are available from the European Nucleotide Archive with the accession codes listed in Table S1, with additional information available from FigShare (https://doi.org/10.6084/m9.figshare.26425042 and https://doi.org/10.6084/m9.figshare.26425066).

Code Availability

The code used in this analysis is available from https://github.com/nickjcroucher/prophage_transformation_analysis.

Acknowledgements

This work was supported by a Sir Henry Dale fellowship jointly funded by Wellcome and the Royal Society (grant 104169/Z/14/A), and the UK Medical Research Council and Department for International Development (grant MR/R015600/1).

Competing interests

NJC has consulted for Antigen Discovery Inc., Merck and Pfizer, and has received an investigator-initiated award from GlaxoSmithKline.

References

1. Chevallereau A, Pons BJ, van Houte S, Westra ER. Interactions between bacterial and phage communities in natural environments. Nat Rev Microbiol. 2022;20: 49–62. doi:10.1038/s41579-021-00602-y
2. Wommack KE, Colwell RR. Virioplankton: viruses in aquatic ecosystems. Microbiol Mol Biol Rev. 2000;64: 69–114. doi:10.1128/MMBR.64.1.69-114.2000
3. Silveira CB, Luque A, Rohwer F. The landscape of lysogeny across microbial community density, diversity and energetics. Environ Microbiol. 2021;23: 4098–4111. doi:https://doi.org/10.1111/1462-2920.15640
4. Moura de Sousa JA, Pfeifer E, Touchon M, Rocha EPC. Causes and Consequences of Bacteriophage Diversification via Genetic Exchanges across Lifestyles and Bacterial Taxa. Mol Biol Evol. 2021;38: 2497–2512. doi:10.1093/molbev/msab044
5. Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, et al. The human gut virome: Inter-individual variation and dynamic response to diet. Genome Res . 2011;21: 1616–1625. Available: http://genome.cshlp.org/content/21/10/1616.abstract
6. Kim M-S, Bae J-W. Lysogeny is prevalent and widely distributed in the murine gut microbiota. ISME J. 2018;12: 1127–1141. doi:10.1038/s41396-018-0061-9
7. Roy K, Ghosh D, DeBruyn JM, Dasgupta T, Wommack KE, Liang X, et al. Temporal Dynamics of Soil Virus and Bacterial Populations in Agricultural and Early Plant Successional Soils . Frontiers in Microbiology . 2020. Available: https://www.frontiersin.org/articles/10.3389/fmicb.2020.01494
8. Williamson KE, Radosevich M, Smith DW, Wommack KE. Incidence of

lysogeny within temperate and extreme soil environments. Environ Microbiol. 2007;9: 2563–2574. doi:https://doi.org/10.1111/j.1462-2920.2007.01374.x

9. Maurice CF, Mouillot D, Bettarel Y, De Wit R, Sarmento H, Bouvier T. Disentangling the relative influence of bacterioplankton phylogeny and metabolism on lysogeny in reservoirs and lagoons. ISME J. 2011;5: 831–842. doi:10.1038/ismej.2010.181

10. Paul JH. Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? ISME J. 2008;2: 579–589. doi:10.1038/ismej.2008.35

11. Engelhardt T, Sahlberg M, Cypionka H, Engelen B. Induction of prophages from deep-subseafloor bacteria. Environ Microbiol Rep. 2011;3: 459–465. doi:https://doi.org/10.1111/j.1758-2229.2010.00232.x

12. Auchtung JM, Lee CA, Garrison KL, Grossman AD. Identification and characterization of the immunity repressor (ImmR) that controls the mobile genetic element ICE*Bs*1 of *Bacillus subtilis*. Mol Microbiol. 2007;64: 1515–1528. doi:https://doi.org/10.1111/j.1365-2958.2007.05748.x

13. Bose B, Auchtung JM, Lee CA, Grossman AD. A conserved anti-repressor controls horizontal gene transfer by proteolysis. Mol Microbiol. 2008;70: 570–582. doi:https://doi.org/10.1111/j.1365-2958.2008.06414.x

14. Knowles B, Bailey B, Boling L, Breitbart M, Cobián-Güemes A, del Campo J, et al. Variability and host density independence in inductions-based estimates of environmental lysogeny. Nat Microbiol. 2017;2: 17064. doi:10.1038/nmicrobiol.2017.64

15. Fusco S, Aulitto M, Iacobucci I, Crocamo G, Pucci P, Bartolucci S, et al. The interaction between the F55 virus-encoded transcription regulator and the

RadA host recombinase reveals a common strategy in Archaea and Bacteria to sense the UV-induced damage to the host DNA. Biochim Biophys Acta - Gene Regul Mech. 2020;1863: 194493. doi:https://doi.org/10.1016/j.bbagrm.2020.194493

16. Chen B, Chen Z, Wang Y, Gong H, Sima L, Wang J, et al. ORF4 of the Temperate Archaeal Virus SNJ1 Governs the Lysis-Lysogeny Switch and Superinfection Immunity. J Virol. 2020;94: 10.1128/jvi.00841-20. doi:10.1128/jvi.00841-20

17. McLeod SM, Kimsey HH, Davis BM, Waldor MK. CTXφ and *Vibrio cholerae*: exploring a newly recognized type of phage–host cell relationship. Mol Microbiol. 2005;57: 347–356. doi:https://doi.org/10.1111/j.1365-2958.2005.04676.x

18. Sauer RT, Ross MJ, Ptashne M. Cleavage of the lambda and P22 repressors by recA protein. J Biol Chem. 1982;257: 4458–4462. doi:10.1016/S0021-9258(18)34744-6

19. Payne-Dwyer AL, Syeda AH, Shepherd JW, Frame L, Leake MC. RecA and RecB: probing complexes of DNA repair proteins with mitomycin C in live *Escherichia coli* with single-molecule sensitivity. J R Soc Interface. 2022;19: 20220437. doi:10.1098/rsif.2022.0437

20. Erill I, Campoy S, Barbé J. Aeons of distress: an evolutionary perspective on the bacterial SOS response. FEMS Microbiol Rev. 2007;31: 637–656. doi:10.1111/j.1574-6976.2007.00082.x

21. Ptashne M, Johnson AD, Pabo CO. A genetic switch in a bacterial virus. Sci Am. 1982;247: 128-130,132,134-140. doi:10.1038/scientificamerican1182-128

22. Shearwin KE, Brumby AM, Egan JB. The Tum Protein of Coliphage 186 Is an

Antirepressor. J Biol Chem. 1998;273: 5708–5715. doi:https://doi.org/10.1074/jbc.273.10.5708

23. Fornelos N, Butala M, Hodnik V, Anderluh G, Bamford JK, Salas M. Bacteriophage GIL01 gp7 interacts with host LexA repressor to enhance DNA binding and inhibit RecA-mediated auto-cleavage. Nucleic Acids Res. 2015;43: 7315–7329. doi:10.1093/nar/gkv634

24. de Groot A, Siponen MI, Magerand R, Eugénie N, Martin-Arevalillo R, Doloy J, et al. Crystal structure of the transcriptional repressor DdrO: insight into the metalloprotease/repressor-controlled radiation response in *Deinococcus*. Nucleic Acids Res. 2019;47: 11403–11417. doi:10.1093/nar/gkz883

25. Ludanyi M, Blanchard L, Dulermo R, Brandelet G, Bellanger L, Pignol D, et al. Radiation response in *Deinococcus deserti*: IrrE is a metalloprotease that cleaves repressor protein DdrO. Mol Microbiol. 2014;94: 434–449. doi:https://doi.org/10.1111/mmi.12774

26. Sadowska-Bartosz I, Bartosz G. Antioxidant defense of *Deinococcus radiodurans*: how does it contribute to extreme radiation resistance? Int J Radiat Biol. 2023;99: 1803–1829. doi:10.1080/09553002.2023.2241895

27. Martin B, García P, Castanié MP, Claverys JP. The *recA* gene of *Streptococcus pneumoniae* is part of a competence-induced operon and controls lysogenic induction. Mol Microbiol. 1995;15: 367–379. doi:10.1111/j.1365-2958.1995.tb02250.x

28. Claverys J-P, Prudhomme M, Mortier-Barrière I, Martin B. Adaptation to the environment: *Streptococcus pneumoniae*, a paradigm for recombination-mediated genetic plasticity? Mol Microbiol. 2000;35: 251–259. doi:https://doi.org/10.1046/j.1365-2958.2000.01718.x

29. Prudhomme M, Attaiech L, Sanchez G, Martin B, Claverys J-P. Antibiotic Stress Induces Genetic Transformability in the Human Pathogen *Streptococcus pneumoniae*. Science (80-). 2006;313: 89–92. doi:10.1126/science.1127912

30. Johnston C, Martin B, Fichant G, Polard P, Claverys JP. Bacterial transformation: Distribution, shared mechanisms and divergent control. Nat Rev Microbiol. 2014;12: 181–196. doi:10.1038/nrmicro3199

31. Croucher NJ, Coupland PG, Stevenson AE, Callendrello A, Bentley SD, Hanage WP. Diversification of bacterial genome content through distinct mechanisms over different timescales. Nat Commun. 2014;5: 5471.

32. Brueggemann AB, Harrold CL, Rezaei Javan R, Van Tonder AJ, McDonnell AJ, Edwards BA. Pneumococcal prophages are diverse, but not without structure or history. Sci Rep. 2017;7: 42976. doi:10.1038/srep42976

33. Croucher NJ, Walker D, Romero P, Lennard N, Paterson GK, Bason NC, et al. Role of conjugative elements in the evolution of the multidrug-resistant pandemic clone *Streptococcus pneumoniae*^{Spain23F} ST81. J Bacteriol. 2008/12/31. 2009;191: 1480–1489. doi:10.1128/jb.01343-08

34. D'Aeth JC, van der Linden MPG, McGee L, de Lencastre H, Turner P, Song J-H, et al. The role of interspecies recombination in the evolution of antibiotic-resistant pneumococci. Elife. 2021;10: e67113. doi:10.7554/eLife.67113

35. Martínez-Rubio R, Quiles-Puchalt N, Martí M, Humphrey S, Ram G, Smyth D, et al. Phage-inducible islands in the Gram-positive cocci. ISME J. 2017;11: 1029–1042. doi:10.1038/ismej.2016.163

36. Dowson CG, Coffey TJ, Kell C, Whiley RA. Evolution of penicillin resistance in *Streptococcus pneumoniae*; the role of *Streptococcus mitis* in the formation of

a low affinity PBP2B in *S. pneumoniae*. Mol Microbiol. 1993/08/01. 1993;9: 635–643. doi:10.1111/j.1365-2958.1993.tb01723.x

37. Croucher NJ, Chewapreecha C, Hanage WP, Harris SR, McGee L, van der Linden M, et al. Evidence for soft selective sweeps in the evolution of pneumococcal multidrug resistance and vaccine escape. Genome Biol Evol. 2014;6: 1589–1602. doi:10.1093/gbe/evu120

38. Sandra M, Marie-Stephanie A, Priyanka N, Birgitta H-N. Gram-Positive Type IV Pili and Competence. Microbiol Spectr. 2019;7: 10.1128/microbiolspec.psib-0011–2018. doi:10.1128/microbiolspec.psib-0011-2018

39. Johnston CHG, Hope R, Soulet A-L, Dewailly M, De Lemos D, Polard P. The RecA-directed recombination pathway of natural transformation initiates at chromosomal replication forks in the pneumococcus. Proc Natl Acad Sci. 2023;120: e2213867120. doi:10.1073/pnas.2213867120

40. Attaiech L, Olivier A, Mortier-Barrière I, Soulet A-L, Granadel C, Martin B, et al. Role of the Single-Stranded DNA–Binding Protein SsbB in Pneumococcal Transformation: Maintenance of a Reservoir for Genetic Plasticity. PLOS Genet. 2011;7: e1002156. Available: https://doi.org/10.1371/journal.pgen.1002156

41. Håvarstein LS, Coomaraswamy G, Morrison DA. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*. Proc Natl Acad Sci. 1995;92: 11140 LP – 11144. doi:10.1073/pnas.92.24.11140

42. Veening J-W, Smits WK, Kuipers OP. Bistability, epigenetics, and bet-hedging in bacteria. Annu Rev Microbiol. 2008;62: 193–210. doi:10.1146/annurev.micro.62.081307.163002

43. Kwun MJ, Ion A V, Oggioni MR, Bentley SD, Croucher NJ. Diverse regulatory pathways modulate bet hedging of competence induction in epigenetically-differentiated phase variants of *Streptococcus pneumoniae*. Nucleic Acids Res. 2023;51: 10375–10394. doi:10.1093/nar/gkad760

44. Prudhomme M, Johnston CHG, Soulet A-L, Boyeldieu A, De Lemos D, Campo N, et al. Pneumococcal competence is a populational health sensor driving multilevel heterogeneity in response to antibiotics. Nat Commun. 2024;15: 5625. doi:10.1038/s41467-024-49853-2

45. Mirouze N, Bergé MA, Soulet A-L, Mortier-Barrière I, Quentin Y, Fichant G, et al. Direct involvement of DprA, the transformation-dedicated RecA loader, in the shut-off of pneumococcal competence. Proc Natl Acad Sci U S A. 2013;110: E1035-44. doi:10.1073/pnas.1219868110

46. Redfield RJ. Genes for breakfast: The have-your-cake and-eat-it-too of bacterial transformation. J Hered. 1993;84: 400–404. doi:10.1093/oxfordjournals.jhered.a111361

47. Croucher NJ, Mostowy R, Wymant C, Turner P, Bentley SD, Fraser C. Horizontal DNA Transfer Mechanisms of Bacteria as Weapons of Intragenomic Conflict. Barton NH, editor. PLOS Biol. 2016;14: e1002394. doi:10.1371/journal.pbio.1002394

48. Apagyi KJ, Fraser C, Croucher NJ. Transformation asymmetry and the evolution of the bacterial accessory genome. Mol Biol Evol. 2018;35: 575–581. doi:10.1093/gbe/evs009

49. Mäntynen S, Laanto E, Oksanen HM, Poranen MM, Díaz-Muñoz SL. Black box of phage–bacterium interactions: exploring alternative phage infection strategies. Open Biol. 2021;11: 210188. doi:10.1098/rsob.210188

50. Łoś M, Węgrzyn G. Pseudolysogeny. Adv Virus Res. 2012;82: 339–349. doi:10.1016/B978-0-12-394621-8.00019-4

51. Ogunseitan OA, Sayler GS, Miller R V. Dynamic interactions of *Pseudomonas aeruginosa* and bacteriophages in lake water. Microb Ecol. 1990;19: 171–185. doi:10.1007/BF02012098

52. Ripp S, Miller R V. The role of pseudolysogeny in bacteriophage-host interactions in a natural freshwater environment. Microbiology. 1997;143: 2065–2070. doi:https://doi.org/10.1099/00221287-143-6-2065

53. Cenens W, Mebrhatu MT, Makumi A, Ceyssens P-J, Lavigne R, Van Houdt R, et al. Expression of a Novel P22 ORFan Gene Reveals the Phage Carrier State in *Salmonella Typhimurium*. PLOS Genet. 2013;9: e1003269. Available: https://doi.org/10.1371/journal.pgen.1003269

54. Sakaguchi Y, Hayashi T, Kurokawa K, Nakayama K, Oshima K, Fujinaga Y, et al. The genome sequence of *Clostridium botulinum* type C neurotoxin-converting phage and the molecular mechanisms of unstable lysogeny. Proc Natl Acad Sci U S A. 2005;102: 17472–17477. doi:10.1073/pnas.0505503102

55. Lood R, Collin M. Characterization and genome sequencing of two *Propionibacterium acnes* phages displaying pseudolysogeny. BMC Genomics. 2011;12: 198. doi:10.1186/1471-2164-12-198

56. Gladstone RA, Lo SW, Lees JA, Croucher NJ, van Tonder AJ, Corander J, et al. International genomic definition of pneumococcal lineages, to contextualise disease, antibiotic resistance and vaccine impact. EBioMedicine. 2019;43: 338–346. doi:10.1016/j.ebiom.2019.04.021

57. Garriss G, Henriques-Normark B. Lysogeny in *Streptococcus pneumoniae*. Microorganisms. 2020. doi:10.3390/microorganisms8101546

58. Romero P, Croucher NJ, Hiller NL, Hu FZ, Ehrlich GD, Bentley SD, et al. Comparative genomic analysis of ten *Streptococcus pneumoniae* temperate bacteriophages. J Bacteriol. 2009;191: 4854–4862. doi:10.1128/JB.01272-08

59. Furi L, Crawford LA, Rangel-Pineros G, Manso AS, De Ste Croix M, Haigh RD, et al. Methylation warfare: interaction of pneumococcal bacteriophages with their host. J Bacteriol. 2019; JB.00370-19. doi:10.1128/JB.00370-19

60. Kwun MJ, Ion A V, Cheng H-C, D'Aeth JC, Dougan S, Oggioni MR, et al. Post-vaccine epidemiology of serotype 3 pneumococci identifies transformation inhibition through prophage-driven alteration of a non-coding RNA. Genome Med. 2022;14: 144. doi:10.1101/2022.09.21.508813

61. Halfmann A, Kovács M, Hakenbeck R, Brückner R. Identification of the genes directly controlled by the response regulator CiaR in *Streptococcus pneumoniae*: Five out of 15 promoters drive expression of small non-coding RNAs. Mol Microbiol. 2007;66: 110–126. doi:10.1111/j.1365-2958.2007.05900.x

62. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, et al. Rapid pneumococcal evolution in response to clinical interventions. Science (80-). 2011/01/29. 2011;331: 430–434. doi:10.1126/science.1198545

63. Croucher NJ, Hanage WP, Harris SR, McGee L, van der Linden M, de Lencastre H, et al. Variable recombination dynamics during the emergence, transmission and "disarming" of a multidrug-resistant pneumococcal clone. BMC Biol. 2014;12: 49. doi:10.1186/1741-7007-12-49

64. Denapaite D, Brückner R, Nuhn M, Reichmann P, Henrich B, Maurer P, et al. The Genome of *Streptococcus mitis* B6 - What Is a Commensal? PLoS One. 2010;5: e9426. Available: https://doi.org/10.1371/journal.pone.0009426

65. Croucher NJ, Finkelstein JA, Pelton SI, Mitchell PK, Lee GM, Parkhill J, et al. Population genomics of post-vaccine changes in pneumococcal epidemiology. Nat Genet. 2013;45: 656–663. doi:10.1038/ng.2625

66. Kwun MJ, Oggioni MR, De Ste Croix M, Bentley SD, Croucher NJ. Excision-reintegration at a pneumococcal phase-variable restriction-modification locus drives within- and between-strain epigenetic differentiation and inhibits gene acquisition. Nucleic Acids Res. 2018;46: 11438–11453. doi:10.1093/nar/gky906

67. Manso AS, Chai MH, Atack JM, Furi L, De Ste Croix M, Haigh R, et al. A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. Nat Commun. 2014;5: 5055. doi:10.1038/ncomms6055

68. Garro AJ. DNA-Mediated Prophage Induction in Bacillus subtilis Lysogenic for φ105c4. J Virol. 1973;12: 18–24. doi:10.1128/jvi.12.1.18-24.1973

69. Aggarwal SD, Yesilkaya H, Dawid S, Hiller NL. The pneumococcal social network. PLoS Pathog. 2020;16: e1008931. doi:10.1371/journal.ppat.1008931

70. López R, García E. Recent trends on the molecular biology of pneumococcal capsules, lytic enzymes, and bacteriophage. FEMS Microbiol Rev. 2004;28: 553–580. doi:10.1016/j.femsre.2004.05.002

71. Lu H, Chen Z, Xie T, Zhong S, Suo S, Song S, et al. The *Deinococcus* protease PprI senses DNA damage by directly interacting with single-stranded DNA. Nat Commun. 2024;15: 1892. doi:10.1038/s41467-024-46208-9

72. McVeigh RR, Yasbin RE. Phenotypic differentiation of “smart” versus “naive” bacteriophages of *Bacillus subtilis*. J Bacteriol. 1996;178: 3399–3401. doi:10.1128/jb.178.11.3399-3401.1996

73. Kwun MJ, Oggioni MR, Bentley SD, Fraser C, Croucher NJ. Synergistic

Activity of Mobile Genetic Element Defences in *Streptococcus pneumoniae*. Genes (Basel). 2019;10: 707.

74. Rocha EPC, Cornet E, Michel B. Comparative and Evolutionary Analysis of the Bacterial Homologous Recombination Systems. PLOS Genet. 2005;1: e15. Available: https://doi.org/10.1371/journal.pgen.0010015

75. Fonseca DR, Halim MFA, Holten MP, Costa KC. Type IV-Like Pili Facilitate Transformation in Naturally Competent Archaea. J Bacteriol. 2020;202: 10.1128/jb.00355-20. doi:10.1128/jb.00355-20

76. Lang AS, Zhaxybayeva O, Beatty JT. Gene transfer agents: phage-like elements of genetic exchange. Nat Rev Microbiol. 2012;10: 472–482. doi:10.1038/nrmicro2802

77. Chen J, Quiles-Puchalt N, Chiang YN, Bacigalupe R, Fillol-Salom A, Chee MSJ, et al. Genome hypermobility by lateral transduction. Science (80-). 2018;362: 207–212. doi:10.1126/science.aat5867

78. Koberg S, Mohamed MDA, Faulhaber K, Neve H, Heller KJ. Identification and characterization of cis- and trans-acting elements involved in prophage induction in *Streptococcus thermophilus* J34. Mol Microbiol. 2015;98: 535–552. doi:10.1111/mmi.13140

79. Durieux I, Ginevra C, Attaiech L, Picq K, Juan P-A, Jarraud S, et al. Diverse conjugative elements silence natural transformation in *Legionella* species. Proc Natl Acad Sci. 2019;116: 18613 LP – 18618. doi:10.1073/pnas.1909374116

80. Feiner R, Argov T, Rabinovich L, Sigal N, Borovok I, Herskovits AA. A new perspective on lysogeny: prophages as active regulatory switches of bacteria. Nat Rev Microbiol. 2015;13: 641–650. doi:10.1038/nrmicro3527

81. Loessner MJ, Inman RB, Lauer P, Calendar R. Complete nucleotide sequence, molecular analysis and genome structure of bacteriophage A118 of *Listeria monocytogenes*: implications for phage evolution. Mol Microbiol. 2000;35: 324–340. doi:https://doi.org/10.1046/j.1365-2958.2000.01720.x

82. Claverys J-P, Martin B, Polard P. The genetic transformation machinery: composition, localization, and mechanism. FEMS Microbiol Rev. 2009;33: 643–656. doi:10.1111/j.1574-6976.2009.00164.x

83. Burroughs AM, Zhang D, Schäffer DE, Iyer LM, Aravind L. Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. Nucleic Acids Res. 2015;43: 10633–10654. doi:10.1093/nar/gkv1267

84. Schleper C, Kubo K, Zillig W. The particle SSV1 from the extremely thermophilic archaeon *Sulfolobus* is a virus: demonstration of infectivity and of transfection with viral DNA. Proc Natl Acad Sci U S A. 1992;89: 7645–7649. doi:10.1073/pnas.89.16.7645

85. Prangishvili D, Vestergaard G, Häring M, Aramayo R, Basta T, Rachel R, et al. Structural and Genomic Properties of the Hyperthermophilic Archaeal Virus ATV with an Extracellular Stage of the Reproductive Cycle. J Mol Biol. 2006;359: 1203–1216. doi:https://doi.org/10.1016/j.jmb.2006.04.027

86. Götz D, Paytubi S, Munro S, Lundgren M, Bernander R, White MF. Responses of hyperthermophilic crenarchaea to UV irradiation. Genome Biol. 2007;8: R220. doi:10.1186/gb-2007-8-10-r220

87. Lanie JA, Ng WL, Kazmierczak KM, Andrzejewski TM, Davidsen TM, Wayne KJ, et al. Genome sequence of Avery's virulent serotype 2 strain D39 of *Streptococcus pneumoniae* and comparison with that of unencapsulated

laboratory strain R6. J Bacteriol. 2006/10/17. 2007;189: 38–51. doi:JB.01148-06 [pii] 10.1128/JB.01148-06 [doi]

88. Croucher NJ, Campo JJ, Le TQ, Pablo J V, Hung C, Teng AA, et al. Genomic and panproteomic analysis of the development of infant immune responses to antigenically-diverse pneumococci. Nat Commun. 2024;15: 355. doi:10.1038/s41467-023-44584-2

89. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10: 421. doi:10.1186/1471-2105-10-421

90. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013;30: 772–780. doi:10.1093/molbev/mst010

91. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol. 2011;7: e1002195. doi:10.1371/journal.pcbi.1002195

92. Sato K, Akiyama M, Sakakibara Y. RNA secondary structure prediction using deep learning with thermodynamic integration. Nat Commun. 2021;12: 941. doi:10.1038/s41467-021-21194-4

93. Kalvari I, Nawrocki EP, Ontiveros-Palacios N, Argasinska J, Lamkiewicz K, Marz M, et al. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. Nucleic Acids Res. 2021;49: D192–D200. doi:10.1093/nar/gkaa1047

94. Bida JP, Maher LJ 3rd. Improved prediction of RNA tertiary structure with insights into native state dynamics. RNA. 2012;18: 385–393. doi:10.1261/rna.027201.111

95. R Core Team. R: A Language and Environment for Statistical Computing. R

Found Stat Comput. 2019. doi:10.1007/978-3-540-74686-7

96. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. J Open Source Softw. 2019;4: 1686. doi:10.21105/joss.01686

97. Wilke C. cowplot: Streamlined Plot Theme and Plot Annotations for “ggplot2.” 2020.

98. Kassambara A. ggpubr: “ggplot2” Based Publication Ready Plots. 2020. Available: https://cran.r-project.org/package=ggpubr

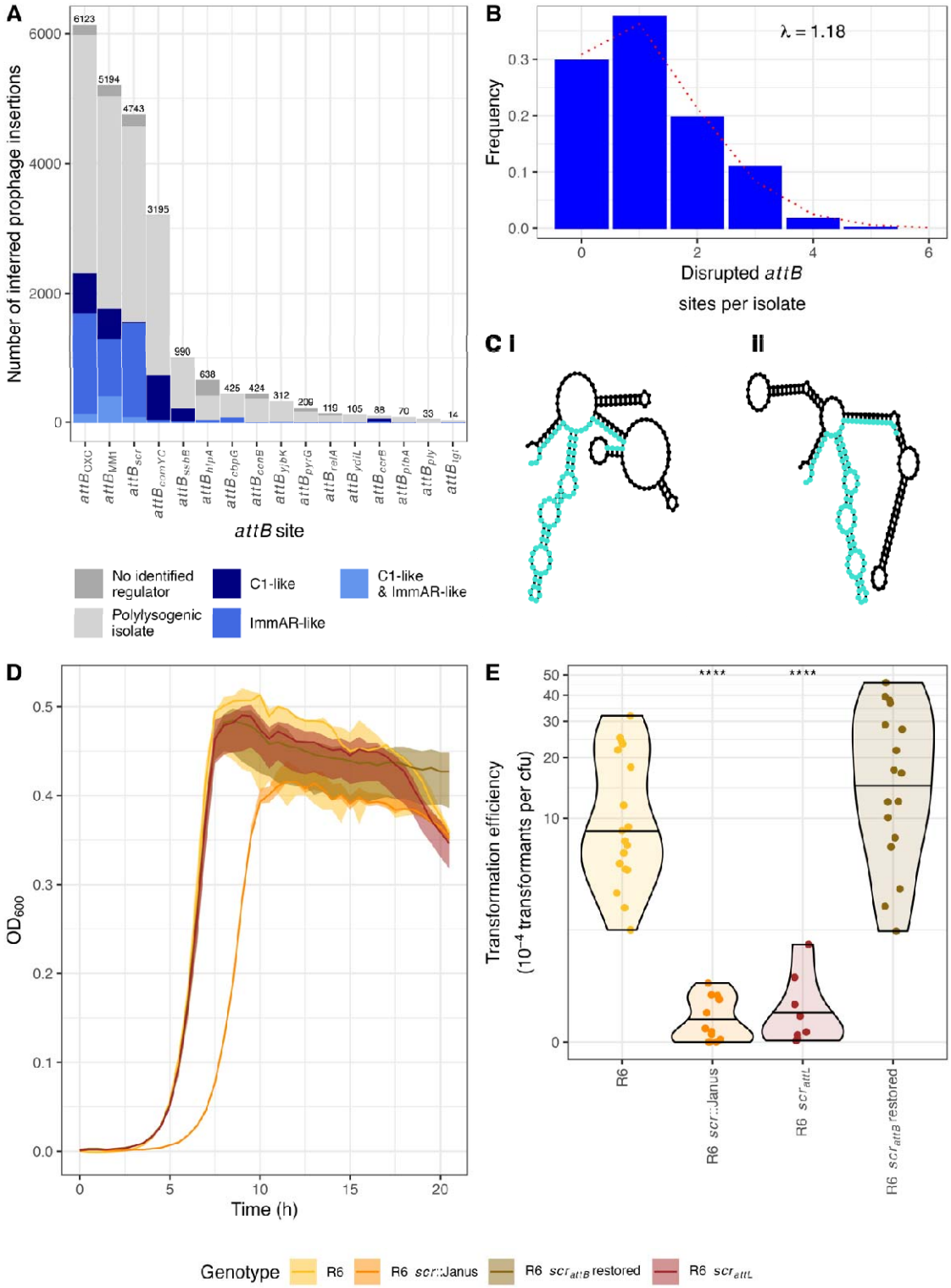
99. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017;8: 28–36. doi:10.1111/2041-210X.12628

100. Guy L, Roat Kultima J, Andersson SGE. genoPlotR: comparative gene and genome visualization in R. Bioinformatics. 2010;26: 2334–2335. doi:10.1093/bioinformatics/btq413

101. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize implements and enhances circular visualization in R . Bioinformatics. 2014;30: 2811–2812. doi:10.1093/bioinformatics/btu393

102. Delignette-Muller ML, Dutang C. fitdistrplus: An R Package for Fitting Distributions. J Stat Softw. 2015;64: 1–34. doi:10.18637/jss.v064.i04

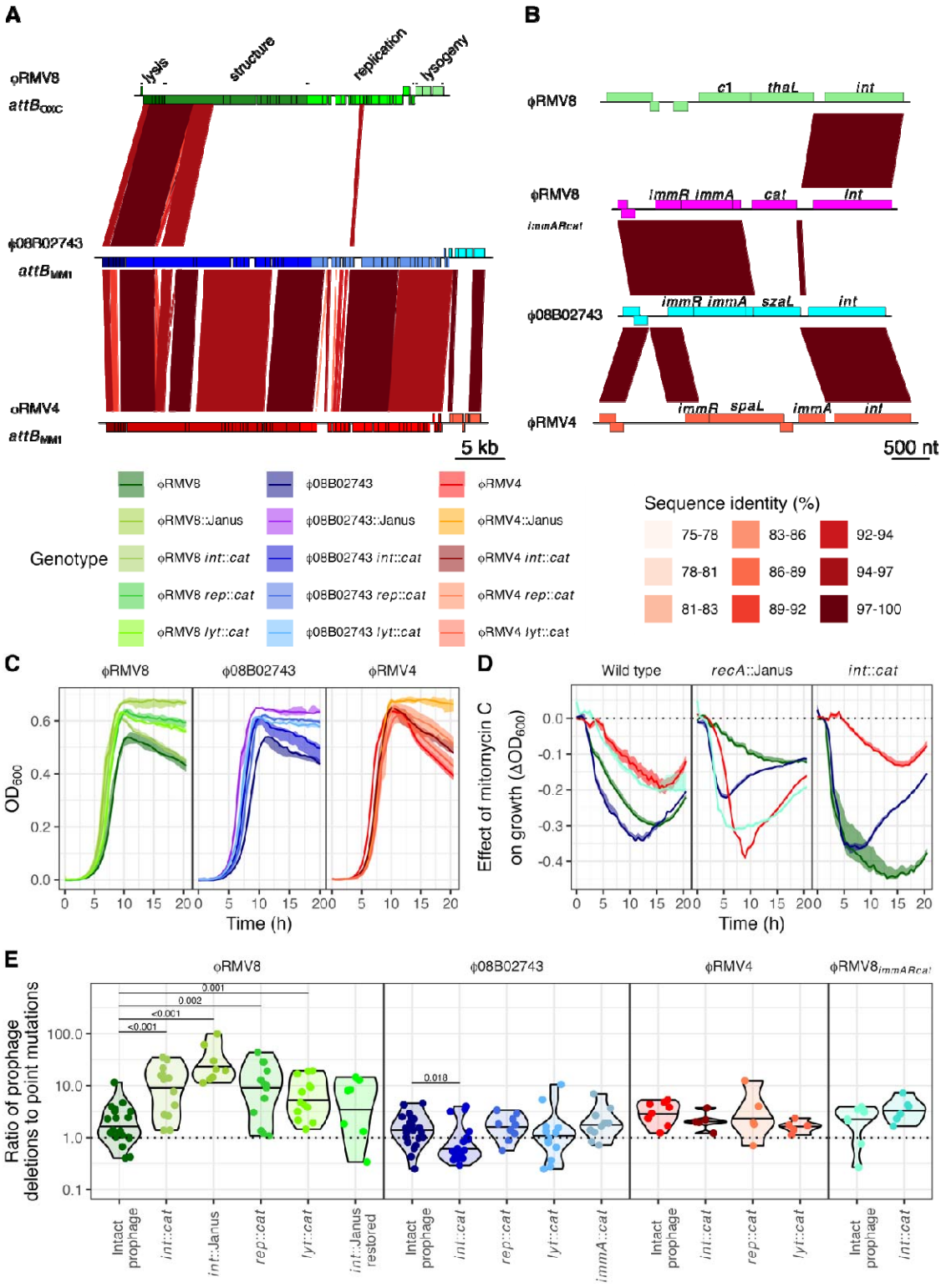
1 **Figures**


A
Number of inferred prophage insertions
attB site
No identified regulator
Polylysogenic isolate
C1-like
ImmAR-like
C1-like & ImmAR-like
B
λ = 1.18
Frequency
Disrupted attB sites per isolate
C I
II
D
OD600
Time (h)
E
Transformation efficiency (10⁻⁴ transformants per cfu)
R6
R6 scr::Janus
R6 scrattL
R6 scrattB restored
Genotype
R6
R6 scr::Janus
R6 scrattB restored
R6 scrattL

2

3

Figure 1 Transformation inhibition by prophage integration. (A) Frequency of prophage insertion at the 16 identified pneumococcal *attB* sites across the 20,047 isolates of the Global Pneumococcal Sequence project collection. Bars are coloured by the inferred prophage regulatory system detected, possible in monolysogenic isolates; otherwise, they are coloured grey, as it was not feasible to reliably associate regulators and *attB* sites in draft assemblies of polylysogenic isolates. **(B)** Number of disrupted *attB* sites in a set of isolates corrected for the effects of population structure by randomly selecting one genome from each Global Pneumococcal Sequence Cluster. **(C)** Predicted secondary structure of the scRNA in (i) *S. pneumoniae* R6, which expresses scr_{attB}; (iii) *S. pneumoniae* ATCC 700669, which expresses scr_{attL}. In each panel, the turquoise sequence highlights the region matching the signal recognition particle RNA Rfam domain. **(D)** Effect of *scr* modification on pneumococcal growth. *S. pneumoniae* R6, with its phase-variable *ivr* restriction-modification locus removed, was compared to a mutant in which the *scr* gene was replaced by a Janus cassette, and two additional genotypes in which this cassette was replaced by the prophage-modified scr_{attL} sequence, or the original *scr* (scr_{attB}) gene was restored. The growth curve displays the optical density at 600 nm (OD_{600}) across three biological replicates for each genotype. The solid line shows the median value, and the shaded area represents the full range of measurements, at each time point. **(E)** Effect of *scr* modification on the transformation efficiency of the R6 genotypes profiled in (D).

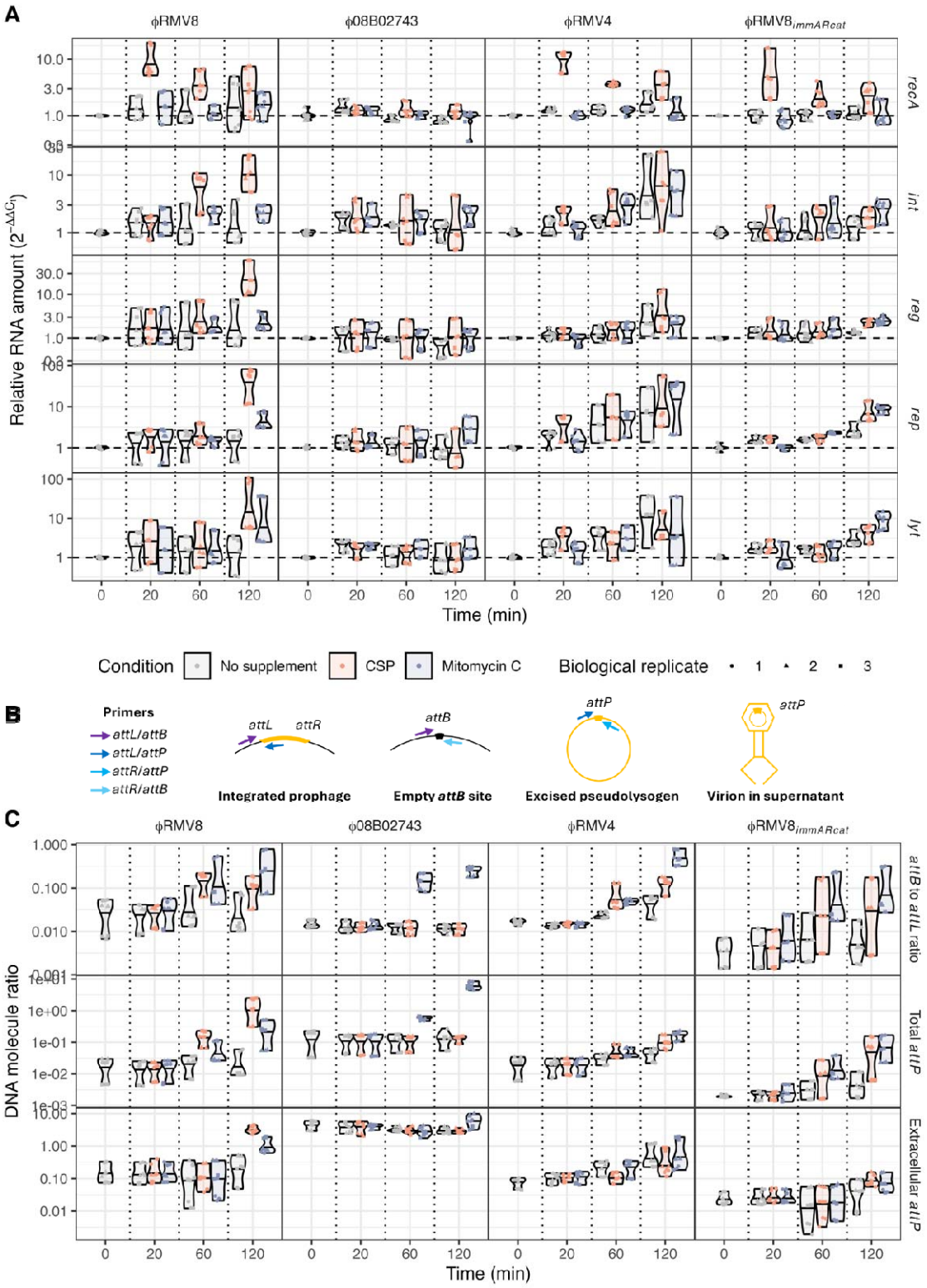

A
lysis
structure
replication
lysogeny
ϕRMV8
attBcoxC
ϕ08B02743
attBMM1
ϕRMV4
attBMM1
5 kb
B
ϕRMV8
c1
thaL
int
ϕRMV8
immR immA
cat
int
immARcat
ϕ08B02743
immR immA
szaL
int
ϕRMV4
immR
spaL
immA
int
500 nt
Genotype
ϕRMV8
ϕRMV8::Janus
ϕRMV8 int::cat
ϕRMV8 rep::cat
ϕRMV8 lyt::cat
ϕ08B02743
ϕ08B02743::Janus
ϕ08B02743 int::cat
ϕ08B02743 rep::cat
ϕ08B02743 lyt::cat
ϕRMV4
ϕRMV4::Janus
ϕRMV4 int::cat
ϕRMV4 rep::cat
ϕRMV4 lyt::cat
Sequence identity (%)
75-78
78-81
81-83
83-86
86-89
89-92
92-94
94-97
97-100
C
ϕRMV8
ϕ08B02743
ϕRMV4
OD600
Time (h)
D
Wild type
recA::Janus
int::cat
Effect of mitomycin C on growth (ΔOD600)
Time (h)
E
ϕRMV8
ϕ08B02743
ϕRMV4
ϕRMV8immARcat
Ratio of prophage deletions to point mutations
0.001
0.002
<0.001
<0.001
0.018
Intact prophage
int::cat
int::Janus
rep::cat
lyt::cat
int::Janus restored
immA::cat

1

2

Figure 2 Efficient deletion of divergent prophage by transformation. (A) Alignment of the prophage, each labelled with its name and *attB* site. Protein coding sequences (CDS) are shown as boxes above the line, if encoded on the top strand, or above the line, if encoded on the bottom strand. Their colour indicates whether they were associated with lysogeny regulation, phage replication, virion structure or host cell lysis. The red bands linking the sequences mark regions of sequence similarity identified by BLASTN. **(B)** Alignment of lysogeny regulation loci. CDSs are annotated, and similarity between sequences indicated by bands, as in panel (A). **(C)** Growth curves comparing lysogens with cells cured of the prophage, and mutants with disruptions of the integrase (*int*), replication (*rep*) or lytic (*lyt*) loci, as described in Fig. 1D. **(D)** The effect of mitomycin C on the growth of prophage systems. The solid lines show the difference between the median OD_{600} of cells grown in mitomycin C and the median OD_{600} of the same genotype grown in parallel in unsupplemented media. The shaded area shows the range of OD_{600} values observed in cultures containing mitomycin C, similarly adjusted by subtracting the median OD_{600} of untreated cultures. **(E)** Deletion of prophage by transformation. The horizontal axis indicates the genotype of the host cells that were transformed with DNA from a donor genotype of the corresponding prophage system, in which the prophage was replaced by a Janus cassette, and the *rpoB* gene contained a base substitution conferring rifampicin resistance. The exception was the □RMV8 *int*::Janus mutant, which was instead removed by donor DNA in which the prophage was replaced by an *ermB* macrolide resistance marker. Each point represents an independent biological replicate in which the ratio of prophage replacement to base substitution acquisition was quantified through selection on appropriate media. The violin plots summarise the distribution and indicate the median value with a

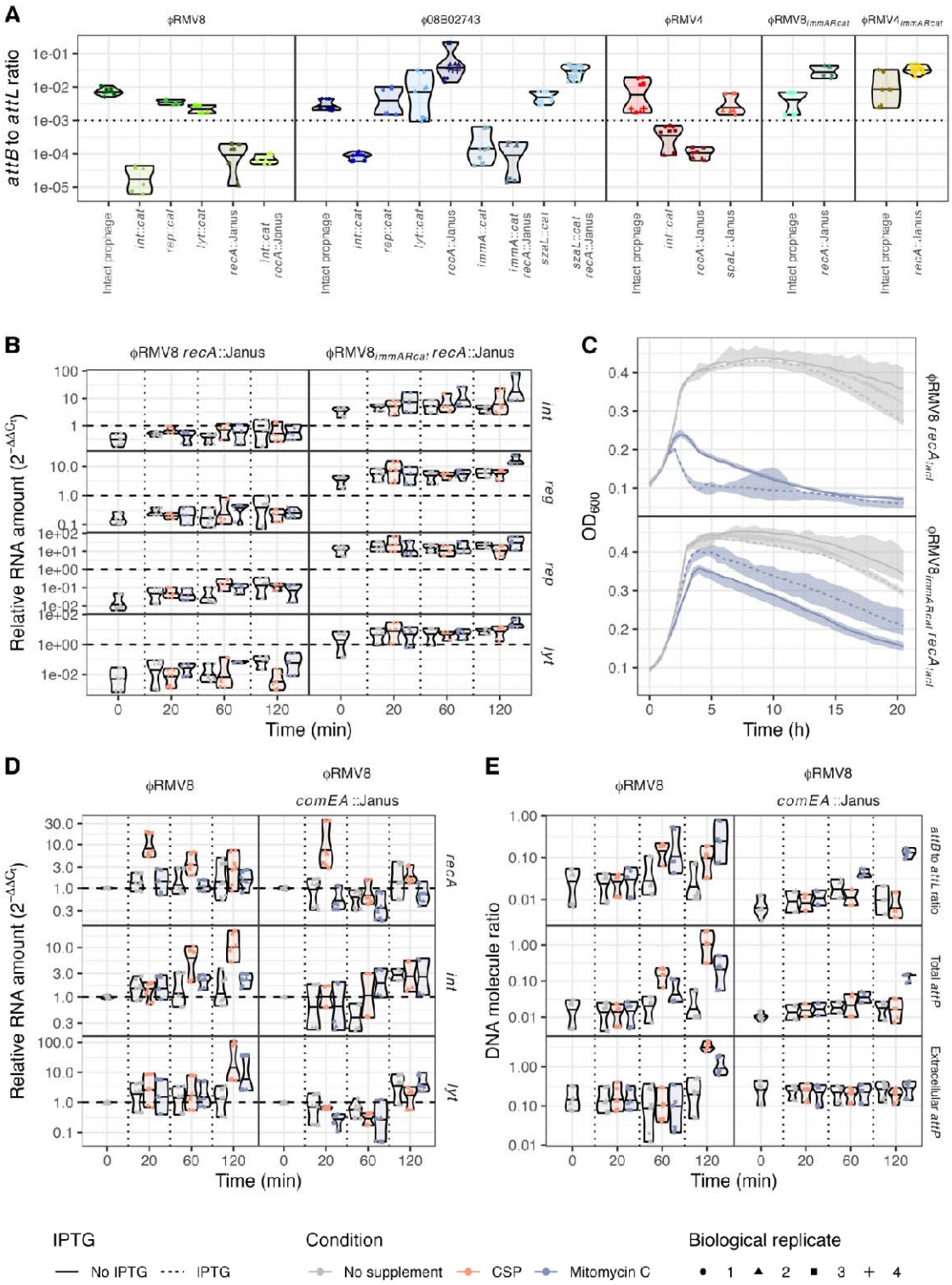
horizontal line. The vertical axis has a logarithmic scale, and the dashed line shows the value at which prophage are deleted as efficiently as base substitutions are exchanged. The efficiency with which each mutant prophage was deleted was compared to the equivalent values for the intact prophage of the same system using a two-tailed Wilcoxon rank sum test, with a Holm-Bonferroni correction for multiple testing within each panel. Significant differences are annotated by horizontal bars, associated with the calculated *p* value.


A
ϕRMV8
ϕ08B02743
ϕRMV4
ϕRMV8immARcat
recA
int
reg
rep
lyt
Relative RNA amount (2^-ΔΔCt)
Time (min)
Condition
No supplement
CSP
Mitomycin C
Biological replicate
1
2
3
B
Primers
attL/attB
attL/attP
attR/attP
attR/attB
attL
attR
attB
attP
Integrated prophage
Empty attB site
Excised pseudolysogen
Virion in supernatant
C
attB to attL ratio
Total attP
Extracellular attP
DNA molecule ratio
Time (min)

1

2

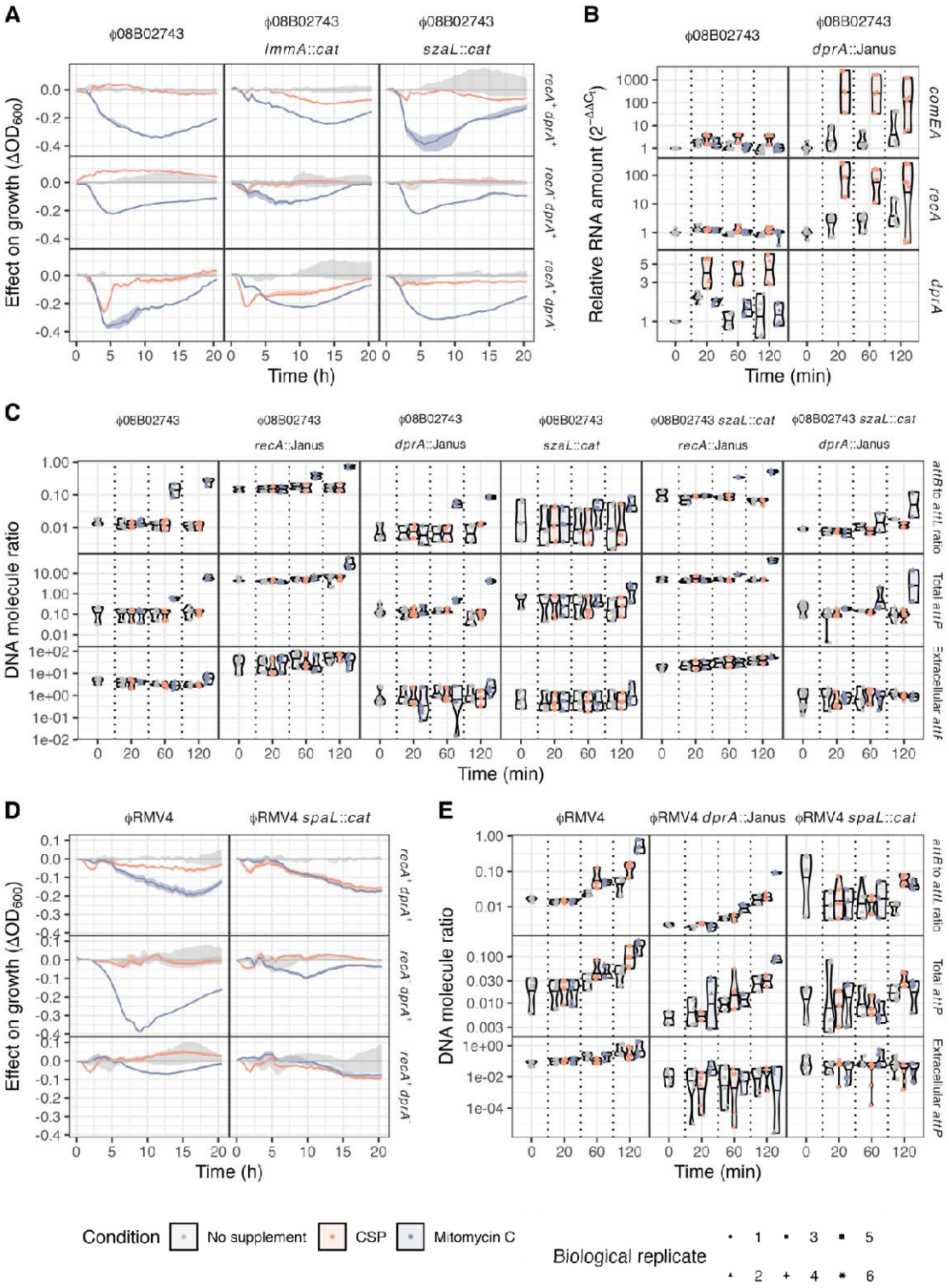
Figure 3 Activation of prophage in response to competence induction and exposure to mitomycin C. (A) Changes in RNA concentration measured by quantitative reverse transcriptase PCR. The expression of the *recA* gene and four prophage loci (*int*, *reg*, *rep* and *lyt*) were measured in each of the prophage systems during early exponential phase in rich media, and 20, 60 and 120 minutes after exposure to competence stimulating peptide (CSP), which induces competence, or mitomycin C (MMC). Expression was quantified as a change relative to the pre-exposure timepoint, standardised to the expression of *rpoA*. Each plot corresponds to a particular combination of a prophage system, defined by the column, and a locus, defined by the row. Different colours distinguish the treatment types, and vertical dashed lines separate the different timepoints. For each combination of locus, prophage system, timepoint and condition, three technical measurements were made for each of three biological replicates, as indicated by the shapes of the points. Their distribution is summarised by a violin plot, which indicates the median value with a horizontal line. **(B)** Diagram summarising the PCR amplicons used to analyse the topological changes associated with phage activation. **(C)** Changes in DNA topology measured by quantitative PCR. Data are shown as in panel (A), except that DNA concentrations were calculated as absolute ratios. The top row quantifies the excision of the prophage through comparing the ratio of the empty *attB* site, restored by prophage excision, to the *attL* site, formed by integrated prophage. The middle row shows the concentration of *attP* sites, formed by prophage excision and replication, from pelleted cells. The bottom row shows the concentration of *attP* sites in culture supernatant, released by host cell lysis.


A
ϕRMV8
ϕ08E02743
ϕRMV4
attB to attL ratio
Intact prophage
int::cat
rep::cat
lyt::cat
recA::Janus
int::cat recA::Janus
immA::cat
immA::cat recA::Janus
szaL::cat
szaL::cat recA::Janus
spaL::Janus
B
ϕRMV8 recA::Janus
Relative RNA amount (2^-ΔΔCt)
Time (min)
C
OD600
Time (h)
D
ϕRMV8
comEA::Janus
E
DNA molecule ratio
Total attP
Extracellular attP
IPTG
No IPTG
IPTG
Condition
No supplement
CSP
Mitomycin C
Biological replicate
1
2
3
4

1

2

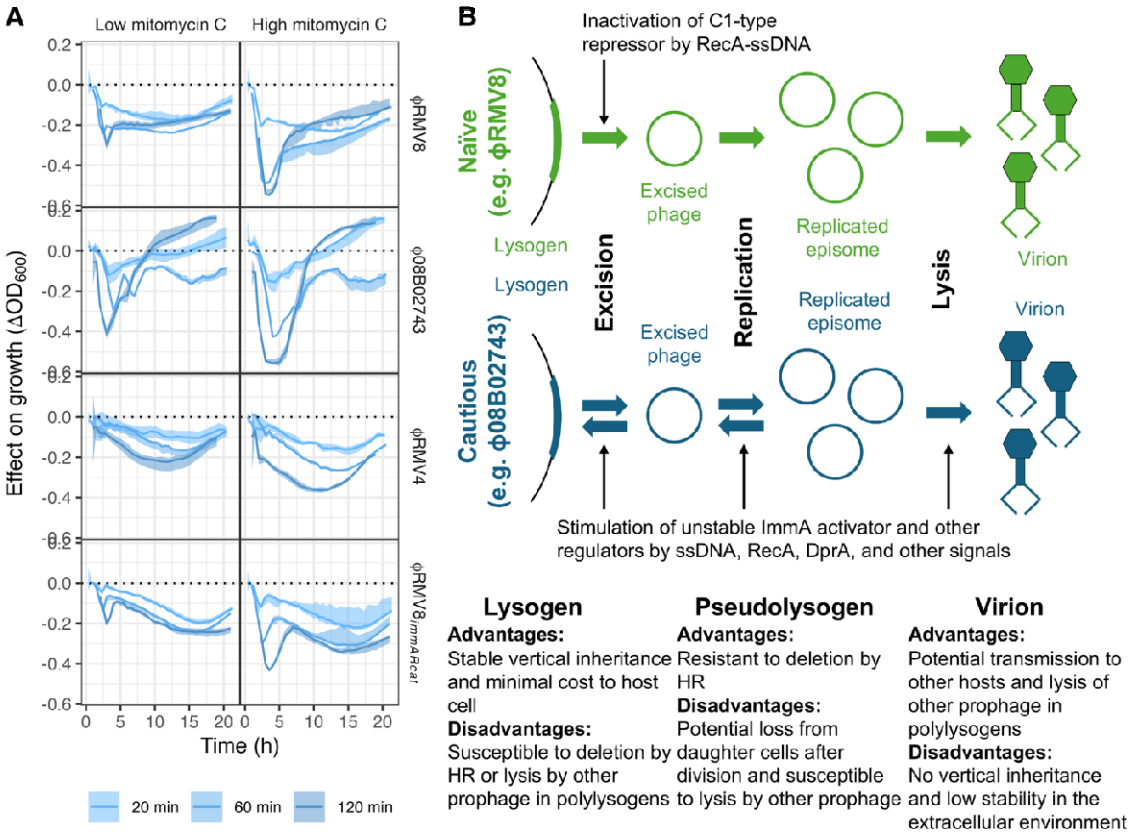
Figure 4 Effect of RecA on prophage activation. (A) The ratio of *attB* to *attL* sites, estimating the level of prophage excision, was measured by quantitative PCR. Data are shown as in Fig. 3C. The horizontal axis separates the tested genotypes, grouped according to the prophage system. The horizontal dashed line at 10^{-3} distinguishes the actively-excising intact prophage from those that appear to be stably integrated, such as the *int*::*cat* mutants. **(B)** Effect of RecA on prophage gene expression. The expression of the four prophage loci assayed in Fig. 3A were measured under equivalent conditions in *recA* ::Janus derivatives of the □RMV8 or □RMV8$_{immARcat}$ systems. The RNA levels were standardised relative to those in the equivalent prophage systems expressing *recA* under control of its native promoter. Therefore an expression level of one, indicated by the horizontal dashed line, indicates the pre-exposure level of expression in the wild type systems. **(C)** The effect of elevated RecA expression on prophage activation. The growth of bacteria with *recA* under the control of an IPTG-inducible *lacI* repressor binding site (*recA*$_{lacI}$), and carrying either □RMV8 or □RMV8$_{immARcat}$, is shown as in Fig. 1C. The colours distinguish cells grown in normal culture, and those exposed to mitomycin C. The line types distinguish bacteria grown in the presence of the *recA* inducer IPTG from those grown without this supplement. **(D)** The effect of single-stranded DNA import by the competence pore ComEA on the expression of □RMV8 genes. The data for *comEA*$^{+}$ wild type cells, shown in Fig. 3A, are compared to equivalent measurements in *comEA*::Janus mutants. **(E)** The effect of single-stranded DNA import by the competence pore ComEA on the excision of □RMV8. The data for *comEA*$^{+}$ wild type cells, shown in Fig. 3C, are compared to equivalent measurements in *comEA*::Janus mutants.


A
ϕ08B02743
ϕ08B02743 immA::cat
ϕ08B02743 szaL::cat
Effect on growth (ΔOD600)
recA+ dprA+
recA- dprA+
recA+ dprA-
Time (h)
B
ϕ08B02743
ϕ08B02743 dprA::Janus
Relative RNA amount (2-ΔΔCt)
comEA
recA
dprA
Time (min)
C
ϕ08B02743
ϕ08B02743 recA::Janus
ϕ08B02743 dprA::Janus
ϕ08B02743 szaL::cat
ϕ08B02743 szaL::cat recA::Janus
ϕ08B02743 szaL::cat dprA::Janus
DNA molecule ratio
attB to attL ratio
Total attP
Extracellular attP
Time (min)
D
ϕRMV4
ϕRMV4 spaL::cat
Effect on growth (ΔOD600)
recA+ dprA+
recA- dprA+
recA+ dprA-
Time (h)
E
ϕRMV4
ϕRMV4 dprA::Janus
ϕRMV4 spaL::cat
DNA molecule ratio
attB to attL ratio
Total attP
Extracellular attP
Time (min)
Condition
No supplement
CSP
Mitomycin C
Biological replicate
1
2
3
4
5
6

1

2

Figure 5 Effect of DprA and population density on prophage activation. (A) Effect of MMC and CSP on the growth of 08B02743, and derivatives in which the *immA* or *szaL* lysogeny module genes were disrupted. Each graph shows the deviation from the median growth of the genotype in unsupplemented media, as in Fig. 2D. The grey ribbon shows the variation in growth in unsupplemented media. The graphs are separated into columns, corresponding to the prophage genotype, and rows, according to whether the *recA* or *dprA* host cell ssDNA-binding proteins were intact. **(B)** The effect of MMC and CSP on the expression of the competence-associated genes in 08B02743, and a mutant in which *dprA* was disrupted. Data are shown as in Fig. 3A. **(C)** The effect of prophage regulators and host cell ssDNA-binding proteins on prophage excision and activation. The three rows quantify the excision, replication and release of the prophage, as in Fig. 3C. Each column corresponds to a different combination of mutations in the prophage-encoded regulators and host-encoded ssDNA-binding proteins. **(D)** Effect of MMC and CSP on the growth of RMV4, and a derivative in which the *spaL* lysogeny module gene was disrupted. Data are shown as in Fig. 2D. **(E)** The effect of prophage regulators and DprA on on intracellular prophage dynamics. The three rows quantify the excision, replication and release of the prophage, as in Fig. 3C.


A
Low mitomycin C
High mitomycin C
Effect on growth (ΔOD600)
Time (h)
φRMV8
φ08B02743
φRMV4
φRMV8immARcat
20 min
60 min
120 min
B
Inactivation of C1-type repressor by RecA-ssDNA
Naïve (e.g. φRMV8)
Lysogen
Excised phage
Replicated episome
Virion
Excision
Replication
Lysis
Lysogen
Excised phage
Replicated episome
Virion
Cautious (e.g. φ08B02743)
Stimulation of unstable ImmA activator and other regulators by ssDNA, RecA, DprA, and other signals
Lysogen
Advantages:
Stable vertical inheritance and minimal cost to host cell
Disadvantages:
Susceptible to deletion by HR or lysis by other prophage in polylysogens
Pseudolysogen
Advantages:
Resistant to deletion by HR
Disadvantages:
Potential loss from daughter cells after division and susceptible to lysis by other prophage
Virion
Advantages:
Potential transmission to other hosts and lysis of other prophage in polylysogens
Disadvantages:
No vertical inheritance and low stability in the extracellular environment

1

Figure 6 Contrasting prophage strategies in the response to inducing stimuli. **(A)** The effect of mitomycin C pulses of different durations on prophage activation. Each prophage system was exposed to a 20, 60 or 120 minute pulse of mitomycin C, either at a low (5 ng μl^{-1}) or high (10 ng μl^{-1}) concentration, prior to being transferred to fresh unsupplemented media. The line plots show the effect of the pulse on the subsequent growth of cells, with the colour corresponding to different pulse durations. The solid line shows the difference between the median post-pulse OD_{600} and the median OD_{600} of untreated cultures, grown in parallel. The shaded area shows the range of OD_{600} values observed in post-pulse cultures, similarly adjusted by subtracting the median OD_{600} of untreated cultures. (B) A summary of the differences between the "naïve" C1-regulated RMV8 prophage, and the "cautious" ImmAR-regulated 08B02743 and RMV4 prophage.