Phage arabinosyl-hydroxy-cytosine DNA modifications result in distinct evasion and sensitivity responses to phage defense systems

Mahler, Marina; Cui, Liang; Smith, Leah M.; Wandera, Katharina G.; Dietrich, Oliver; Mayo-Muñoz, David; Balamkundu, Seetharamsing; Jackson, Simon A.; Brouns, Stan J.J.; More Authors

DOI
10.1016/j.chom.2025.06.005

Publication date
2025

Document Version
Final published version

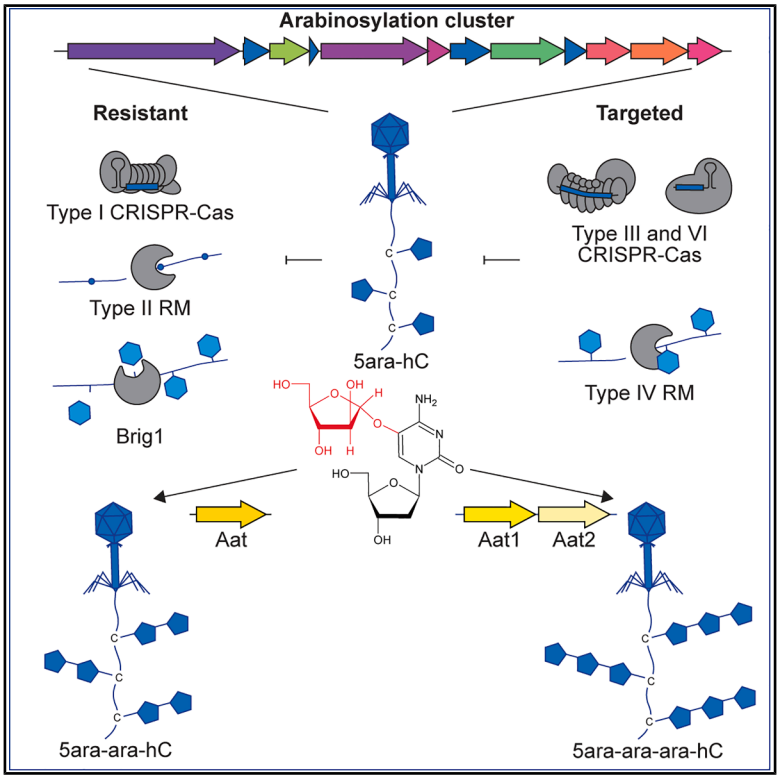
Published in
Cell Host and Microbe

Article

Cell Host & Microbe

Phage arabinosyl-hydroxy-cytosine DNA modifications result in distinct evasion and sensitivity responses to phage defense systems

Graphical abstract



Authors

Marina Mahler, Liang Cui, Leah M. Smith, ..., Peter C. Dedon, Stan J.J. Brouns, Peter C. Fineran

Correspondence

liangcui@smart.mit.edu (L.C.), peter.fineran@otago.ac.nz (P.C.F.)

In brief

Mahler et al. describe a phage DNA modification that adds arabinose to cytosines via a hydroxy linkage. This 5ara-hC modification can be further modified to double and triple arabinosylation by phage-encoded arabinosyl-5ara-hC transferases. Arabinosylated phages are protected from DNA-targeting CRISPR-Cas and RM, but they remain sensitive to RNA-targeting CRISPR-Cas and glycosylation-dependent RM.

Highlights

- Phages substitute cytosines with an unpredicted 5-arabinose-hydroxy-C in their DNA
- 5ara-hC DNA results in evasion or sensitivity to different phage defense systems
- Phage-encoded arabinosyl-5ara-hC transferases add additional arabinose to 5ara-hC
- Single-, double, or triple ara-hC was found on genomes of diverse *Straboviridae* phages

Mahler et al., 2025, Cell Host & Microbe *33*, 1173–1190
July 9, 2025
https://doi.org/10.1016/j.chom.2025.06.005

CellPress

Cell Host & Microbe

CellPress

Article

Phage arabinosyl-hydroxy-cytosine DNA modifications result in distinct evasion and sensitivity responses to phage defense systems

Marina Mahler,[1,2,3,4,13] Liang Cui,[5,*] Leah M. Smith,[1,3,4,12] Katharina G. Wandera,[1,12] Oliver Dietrich,[1,12] David Mayo-Muñoz,[1,3,4,6,12,14] Seetharamsing Balamkundu,[7,12] Megan En Lee,[7] Hong Ye,[7] Chuan-Fa Liu,[7] Junzhou Wu,[5] Juby Mathew,[8,15] Jeremy Dubrulle,[1] Lucia M. Malone,[1,16] Simon A. Jackson,[1,3,4,6,17] Antony J. Fairbanks,[8,9] Peter C. Dedon,[5,10] Stan J.J. Brouns,[2,11] and Peter C. Fineran[1,3,4,6,18,*]

[1]Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand
[2]Department of Bionanoscience, Delft University of Technology, Delft, the Netherlands
[3]Maurice Wilkins Centre for Molecular Biodiscovery, University of Otago, Dunedin, New Zealand
[4]Genetics Otago, University of Otago, Dunedin, New Zealand
[5]Singapore-MIT Alliance for Research and Technology, Antimicrobial Resistance Interdisciplinary Research Group, Campus for Research Excellence and Technological Enterprise, Singapore
[6]Bioprotection Aotearoa, University of Otago, Dunedin, New Zealand
[7]School of Biological Sciences, Nanyang Technological University, Singapore
[8]Biomolecular Interaction Centre, University of Canterbury, Christchurch, New Zealand
[9]School of Physical and Chemical Sciences, University of Canterbury, Christchurch, New Zealand
[10]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
[11]Kavli Institute of Nanoscience, Delft, the Netherlands
[12]These authors contributed equally
[13]Present address: Biozentrum, University of Basel, Basel, Switzerland
[14]Present address: Department of Plant and Environmental Science, University of Copenhagen, Frederiksberg, Denmark
[15]Present address: Syft technologies, Christchurch, New Zealand
[16]Present address: Brightlands Masstricht Health Campus BV, Maastricht, the Netherlands
[17]Present address: School of Pharmacy and Biomedical Sciences, University of Waikato, Hamilton, New Zealand
[18]Lead contact
*Correspondence: liangcui@smart.mit.edu (L.C.), peter.fineran@otago.ac.nz (P.C.F.)
https://doi.org/10.1016/j.chom.2025.06.005

SUMMARY

Bacteria encode diverse anti-phage systems, such as CRISPR-Cas and restriction modification (RM), which limit infection by targeting phage DNA. We identified a DNA modification in phages, i.e., 5-arabinosyl-hydroxy-cytosine (5ara-hC), which adds arabinose to cytosines via a hydroxy linkage and protects phage from DNA targeting. The hydroxy linkage was common among arabinoslyated phages, with some arabinosylated phages encoding arabinose-5ara-hC transferases (Aat) that add a second or third arabinose to DNA. DNA arabinosylation enables evasion from DNA-targeting type I CRISPR-Cas and type II RM systems. However, arabinosylated phages remain sensitive to RNA-targeting CRISPR-Cas (type III and VI) and promiscuous type IV restriction endonucleases. 5ara-hC enables evasion of glycosylase defenses that target phages with glucosylated hydroxymethyl cytosines, and 5ara-ara-hC protects against some defenses capable of targeting 5ara-hC-modified phages. Collectively, this work identifies DNA modifications that enable phages to evade multiple defenses yet remain vulnerable to some systems that target RNA or modified nucleobases.

INTRODUCTION

Bacteria are under constant phage pressure[1,2] and protect themselves using diverse defenses, including restriction-modification (RM) and CRISPR-Cas systems.[3,4] CRISPR-Cas immunity relies on viral DNA memories from past infections, termed spacers, which are stored in the bacterial CRISPR array.[5,6] Immunity occurs upon transcription and processing of the CRISPR array into CRISPR RNAs (crRNAs), around which Cas proteins assemble to form an interference complex that recognizes nucleic acids complementary to the crRNA (termed protospacers).[7,8] Depending on the CRISPR-Cas type, target recognition results in protection through phage genome cleavage or activation of ancillary proteins that induce cell death or dormancy and result in abortive infection.[9]

Phages employ various mechanisms to evade bacterial defenses.[10–12] Some phages encode anti-CRISPR (Acr) proteins that inhibit CRISPR-Cas by multiple mechanisms, such as

blocking interference,[13–15] mimicking DNA targets[16–18] or degrading signaling molecules.[19] In addition, some jumbo phages form a nucleus-like structure during infection, which protects their genome from DNA-targeting defenses.[20–22] Furthermore, masking DNA with modifications can block defense systems that recognize and cleave DNA.[3,10,23,24]

Tailed double-stranded DNA (dsDNA) phages of the *Caudoviricetes* class harbor the largest diversity of naturally occurring, chemically modified deoxynucleotides of all forms of life.[24,25] Phage DNA modifications are usually synthesized by phage enzymes that modify the nucleotides pre- or post-DNA replication.[26–28] Phages often modify nucleotides by adding a nucleophilic molecule, such as a hydroxymethyl group, to the nucleobase, which serves as the reactive site for further DNA hyper-modifications. Some phages completely substitute one of the four nucleobases with a non-canonical version,[29,30] while others modify a fraction of the nucleotides[31,32] or modify specific DNA motifs.[33,34] Phage DNA modifications protect against cleavage by diverse RM systems,[23,24,35] which consist of a methyltransferase that methylates specific motifs on the DNA and a restriction endonuclease that cleaves the DNA at unmethylated sites.[36] However, bacteria can overcome this counter-defense mechanism by encoding type IV RM systems that recognize and cleave DNA in a modification-dependent manner.[37–40]

The best-characterized DNA modification is the complete replacement of cytosines with 5-α- or β-glucosyl-hydroxymethylcytosine (5ghmC) on the *Escherichia coli* (*E. coli*) phage T4 genome by a combination of pre- and post-replication modification steps.[29,41–43] The 5ghmC modification protects T4 from restriction endonucleases; provides various degrees of resistance against DNA-targeting type I, II, and V CRISPR-Cas systems;[44–47] and provides diverse innate nuclease-based defense systems.[48] In contrast, T4 infection was inhibited by the RNA-targeting type VI CRISPR-Cas system,[49] and the modified phage DNA can be cleaved by ghmC-dependent type IV restriction endonucleases such as GmrSD.[50] Except for the work on the 5ghmC modification of T4 in the *E. coli* model system and its implications for defense and counter-defense, the consequences of other phage DNA modifications regarding bacterial defense evasion are currently unknown.

Here, we identified 5-arabinosyl-hydroxy-cytosine (5ara-hC) modifications in phages LC53 and 92A1, which infect *Serratia* sp. ATCC 39006 (hereafter *Serratia*)[51] and *Serratia* strain 95, respectively. We identified additional phages with double (*E. coli* phages RB69, Bas46, and Bas47) and triple (*Acinetobacter baumannii* phage Maestro) arabinosylation of hydroxy-cytosines due to the function of phage-encoded arabinose-5ara-hC transferases (Aat). These modifications provided protection against DNA-targeting defenses (type I CRISPR-Cas and type II RM) but were vulnerable to type IV RM and RNA-targeting CRISPR-Cas (type III and VI), and they were identified in numerous phages in genomic and metagenomic databases. Hydroxy-cytosine arabinose modifications enabled evasion of DNA glycosylases that target the 5ghmC in other phages (e.g., T4). Finally, we demonstrated that two, but not one, arabinose units protected phages from defense provided by a large clinical plasmid. In summary, we have identified a series of DNA modifications that assist phages in overcoming defense systems.

RESULTS

Phage LC53 DNA is arabinosylated and insensitive to type I CRISPR-Cas and restriction enzymes

The type I-E, I-F, and III-A CRISPR-Cas systems are present in *Serratia*,[52] the host of the *Winklervirus* LC53.[51] To search for mechanisms used by phages to evade CRISPR-Cas, we generated plasmids with mini-arrays containing type I-E, I-F, and III-A spacers that target phage LC53 (Figure 1A). Spacer expression with the native DNA-targeting type I-E and I-F *cas* operons in *Serratia* provided no protection against phage LC53 (Figure 1B). By contrast, a spacer for the native RNA-targeting type III-A system reduced phage plaquing (Figure 1B). All spacers were confirmed functional, as they provided interference in conjugation efficiency assays against plasmids bearing targets of these spacers (Figures 1C and 1D). Therefore, phage LC53 possesses a mechanism to evade the DNA-targeting type I CRISPR-Cas systems, while remaining sensitive to RNA-targeting type III CRISPR-Cas.

Phage LC53 belongs to the *Tevenvirinae* subfamily within the *Straboviridae* family, is related to *E. coli* phage T4, and encodes all core T4-like genes.[51] Since T4 modifies its DNA, we examined LC53 for homologs involved in DNA modification (Figure 1E; Table S1). T4 substitutes the cytosine pool with hydroxymethylated cytosines (5hmC) and incorporates 5hmC into the newly synthesized phage DNA during replication[41,42,53] (Figure 1F). T4 α- and β-glucosyl transferases subsequently add glucose post-replication to 5hmC, generating 5ghmC[29,43] (Figure 1F). DNA modification genes are in a specific genomic region between two T4 core genes (DNA polymerase and a recombination protein). Even though this region is not conserved between T4 and LC53 (Figure 1E), LC53 encodes homologs of the dCTPase and the dCMP hydroxymethylase (HMase) (25% amino acid [aa] identity), which are responsible for the pre-replicative substitution of the dCTP nucleotides for 5hmdCMP (Figures 1F and S1A; Table S1). The 5hmdCMP molecules are phosphorylated by phage- and host-encoded kinases and incorporated into the DNA by the phage DNA polymerase during replication.[24] LC53 (and its *Serratia* host) encodes homologs for all the proteins involved in the modification of dC to 5hmdC, except homologs of the T4 glucosyl transferases, which add glucose to 5hmdC (Figure 1F; Table S1).

To investigate whether LC53 has modified DNA and is protected from RM systems, we tested the susceptibility of phage DNA to restriction enzymes with different modification requirements. LC53 DNA was protected from cleavage by MspI, a type II enzyme that cuts DNA containing C, mC, and hmC at the second C of CCGG (Figure 1G). By contrast, LC53 DNA was degraded by the type IV restriction enzyme AbaSI, which is reported to be ghmC-dependent[54,55] (Figure 1H). The restriction-enzyme response profile of LC53 was the same as that of glucosylated T4 DNA and differed from *Serratia* phages PCH45 and JS26, which contain non-modified cytosines (Figures 1G, 1H, and S1B). Therefore, LC53 DNA likely contains 5hmdC, which is glucosylated either by enzymes distinct from the T4 glucosyl transferases (Figure 1E; Table S1) or glycosylated with an alternative sugar. The latter possibility would indicate that AbaSI has a wider sugar-modified DNA cleavage profile than previously reported.[54,55]

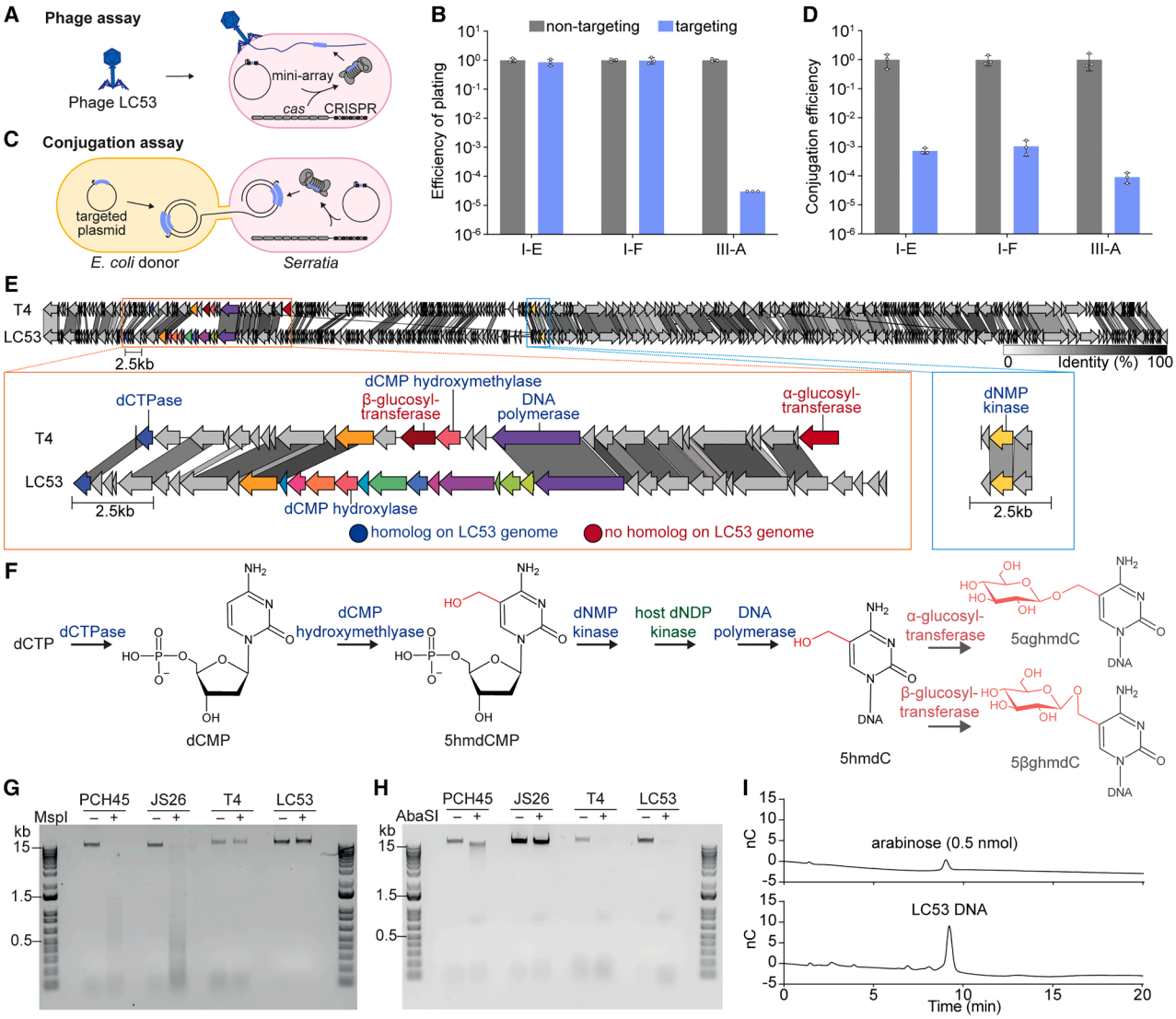


Figure 1. Phage LC53 evades DNA-targeting CRISPR-Cas systems through arabinosylated DNA

(A) Schematic of CRISPR-Cas phage interference assay (EOP). A similar protospacer region in gp37 was targeted by the native type I-E (sp1), I-F (sp1), and III-A (sp1) CRISPR-Cas systems with a spacer from a mini-array plasmid.

(B) EOP of LC53 on a targeting strain (blue bars) compared with a non-targeting control (gray bars). When no single plaques were visible (e.g., type III-A targeting), EOP was calculated as one plaque in the first dilution that did not show any clearing of the bacterial lawn.

(C) Schematic of conjugation efficiency assay of a plasmid with a 500-bp phage sequence containing the protospacer targeted by the native CRISPR-Cas systems with a spacer from a mini-array plasmid.

(D) Conjugation efficiency calculated as transconjugants to recipients and normalized to the non-targeting control (gray bars). In (B) and (D), data are the mean of biological triplicates ± standard deviation (SD), and individual data points are shown.

(E) Whole-genome comparison of T4 and LC53 with loci encoding T4 DNA modification genes highlighted. Proteins with amino acid identities ≥30% are linked in a white to black gradient corresponding to increased homology. Modification genes labeled in blue font encode homologs in LC53, while no homologs could be found for the genes labeled in red font.

(F) DNA modification pathway of T4 to substitute dC with 5ghmC. Enzymes in blue have homologs in LC53, whereas those in red have no identified homologs in LC53. Corresponding LC53 homologs and identities are summarized in Table S1.

(G) Restriction digest of *Serratia* phages PCH45, JS26, LC53, and *Escherichia* phage T4 DNA, with MspI that cleaves dC-, mdC-, and hmdC-containing DNA.

(H) Restriction digest of the same phage DNA with the ghmC-dependent type IV restriction enzyme AbaSI.

(I) HPAE-PAD traces of arabinose standard (top) and hydrolyzed sugar from LC53 DNA (bottom). nC, nanocoulombs.

To find glycosylation genes, we searched the LC53 DNA modification locus and identified homologs in other *Straboviridae* phages (e.g., *E. coli* RB69 and *Serratia* 92A1). *E. coli* RB69 and *Serratia* 92A1 phages were reported to modify 5hmdC to 5-arabinosyl-hydroxymethyl-2′-deoxycytidine (5ara-hmdC),[56,57] leading us to hypothesize that LC53 contained this DNA modification. To test this, we compared the retention time of hydrolyzed phage DNA with different hexose and pentose standards by

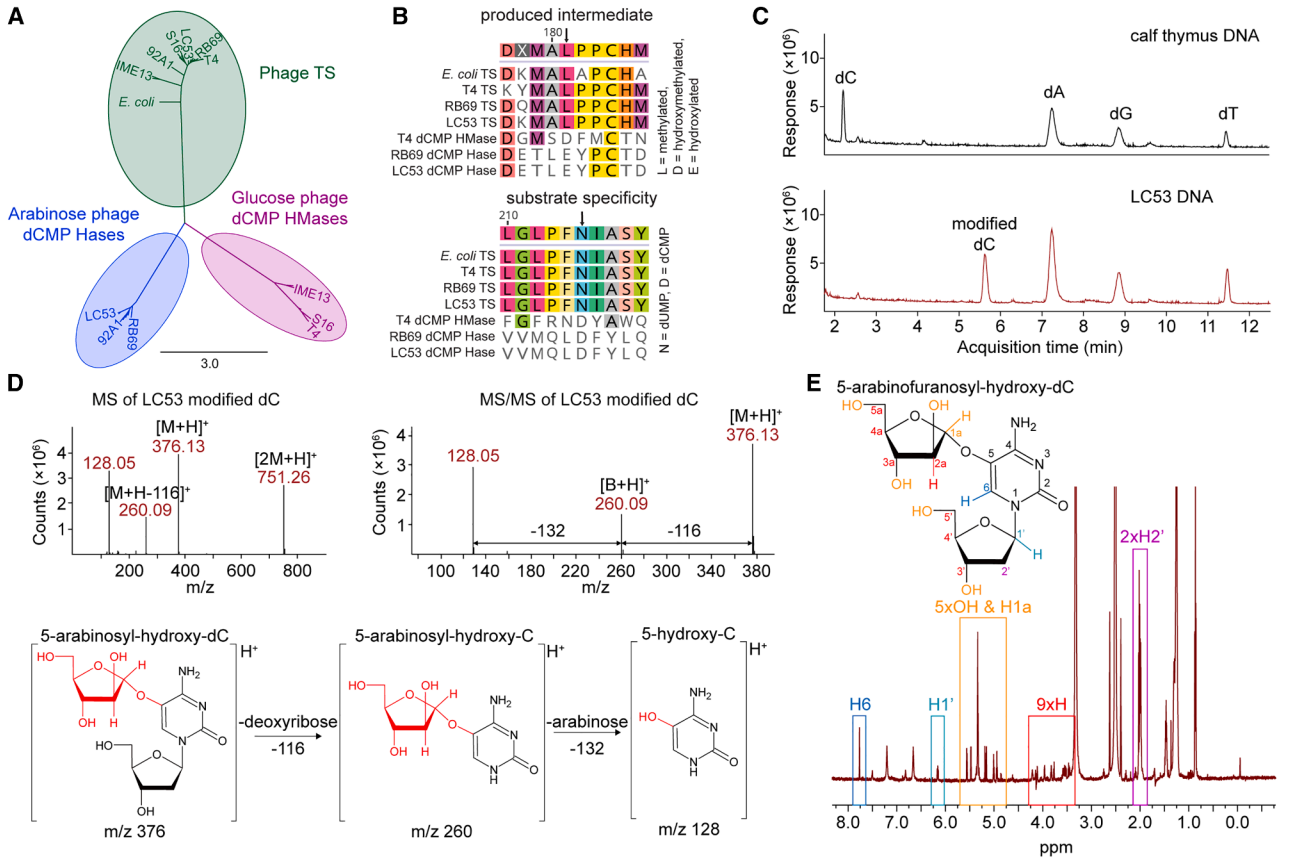


Figure 2. Cytosines on LC53 DNA are substituted with 5-arabinosyl-hydroxy-cytosines

(A) Phylogeny of phage-encoded TS, dCMP HMase, and dCMP Hases of arabinosylated phages. Representatives are labeled (accession numbers in Table S2). The tree was rooted with the *E. coli* ThyA TS.

(B) Amino acid alignment of *E. coli* TS and phage-encoded TS family enzymes around the residues that form a gate to the active site and determine the substrate (dUMP or dCMP) and the intermediate produced (methylated or hydroxymethylated).

(C) LC-MS/MS of LC53 nucleosides in comparison to unmodified nucleosides from calf thymus DNA.

(D) MS and MS/MS spectra of modified dC from LC53 DNA with proposed chemical structures corresponding to the obtained *m/z* ratios.

(E) Structure of 5-arabinofuranosyl-hydroxy-dC and corresponding NMR spectrum. Protons and corresponding peaks are color-coded. The anomeric proton of arabinose (H1a) was observed at 5.14 ppm.

high-performance anion-exchange chromatography with pulsed amperometric detection (HPAE-PAD). The hydrolyzed LC53 DNA contained a sugar with a similar retention time to arabinose (Figure 1I), and co-injection of hydrolyzed phage DNA with arabinose resulted in a single major peak at the same retention time (Figure S1C). As a control, we confirmed the presence of glucose on T4 DNA (Figure S1D). In summary, LC53 DNA contains an arabinose modification that differs from T4 5ghmC and is insensitive to DNA-targeting type I CRISPR-Cas systems and type II restriction enzymes. By contrast, LC53 is vulnerable to type IV restriction and an RNA-targeting type III CRISPR-Cas system.

LC53 DNA has an arabinose-hydroxy-cytosine modification

Next, we examined how LC53 DNA is modified with arabinose. We initially predicted that LC53 would have 5ara-hmdC modifications, similar to 5-glucosyl-hmdC on T4. In T4, dCMP HMases convert dCMP to 5hmdCMP (Figure 1F). These dCMP HMases belong to the large thymidylate synthases (TSs)/dCMP HMase superfamily, which are key enzymes of pre-replication pyrimidine modifications.[24] To examine protein similarity, we aligned predicted dCMP HMases from LC53, T4 and related phages, and as a predicted outgroup their TS, which these phages use to catalyze the methylation of dUMP to dTMP.[24] As expected, the phage-encoded TSs all clustered with the *E. coli* TS (Figure 2A). By contrast, phages encoding genes for arabinosylation or glucosylation of DNA have dCMP HMases that cluster in two distinct groups for arabinosylated and glucosylated phages, respectively (Figures 2A and S1A). We show below that the proteins originally assigned as dCMP HMases in LC53 and other arabinosylated phages are instead dCMP hydroxylases (dCMP Hases), and this nomenclature will be used hereafter. We compared the LC53 dCMP Hase sequence with that of the *E. coli* TS and other phage enzymes (Figure 2A) and observed that both groups of enyzmes (HMases/Hases) are predicted to modify the same substrate (dCMP) (Figure 2B).

To evaluate these predictions and to directly detect any DNA modifications and their frequency, we performed liquid

chromatography-tandem mass spectrometry (LC-MS/MS) of the 2′-deoxyribonucleosides obtained from nuclease- and phosphatase-digested LC53 DNA. Contrary to expectations, the cytosines were completely substituted by a modification that did not correspond to 5ara-hmdC (Figures 2C and 2D). The modification mass in the positive ion mode was 376 Da, which corresponded with 5-arabinosyl-hydroxy-2′-deoxycytidine (5ara-hdC, *m/z* [mass to charge] of 376) and is 14 Da less than 5ara-hmdC (*m/z* 390). High-resolution MS of the protonated $[M+H]^+$ ion (*m/z* 376.1362) of the modification matched well with the theoretical mass of the $[M+H]^+$ ion of 5ara-hdC (*m/z* 376.1356, mass error = 1.59 ppm). Analysis of the MS/MS fragmentation pattern supported the identification of 5ara-hdC as the modification (Figure 2D). The collision-induced dissociation (CID) MS/MS spectrum of the modification showed two major $[M+H]^+$ fragment ions at *m/z* 260 and *m/z* 128, corresponding to the ions with the loss of 2′-deoxyribose (*m/z* 260) and the loss of both 2′-deoxyribose and arabinose (*m/z* 128), respectively. As a control, we analyzed T4 DNA and confirmed complete cytosine substitution with 5α- and 5β-ghmC (Figures 2B and S2A).

Next, 5ara-hdC was confirmed and the stoichiometry of arabinose determined by purifying the modified nucleosides by HPLC and performing nuclear magnetic resonance (NMR) (Figures 2E and S2C–S2G). In the ^{1}H NMR spectrum, the cytosine H6 was detected at 7.56 ppm as a singlet peak, and the anomeric proton of deoxyribose and arabinose was observed at 6.15 (H1′) and 5.14 (H1a) ppm, respectively, through the analysis of deuterium exchange ^{1}H NMR, COSY, and NOESY ^{1}H NMR. The arabinose is either a 1,2-*trans* alpha-D or a 1,2-*trans* beta-L arabinofuranose since H1a was observed as a doublet with a small coupling constant (J = 0.6 Hz).[58] Therefore, proteins similar to T4 dCMP HMases in LC53 and other arabinosylated phages are dCMP Hases (Figure 2A). In summary, LC53 DNA is modified with 5ara-hC, a previously unreported modification.

Phage DNA arabinosylation impedes DNA-targeting type I CRISPR-Cas systems

To further investigate the protection provided by 5ara-hC on the LC53 genome against DNA-targeting CRISPR-Cas systems, we engineered multiple spacers for the two native type I CRISPR-Cas systems of *Serratia* to target different phage genes. The spacers were individually expressed from a plasmid CRISPR mini-array, while the native chromosomal *cas* genes encoded the interference complexes. In agreement with our initial results with a single spacer (Figure 1), LC53 completely evaded type I-F CRISPR-Cas targeting for all nine spacers tested (Figure S3A). LC53 also evaded type I-E targeting, with six spacers fully evaded and two partially evaded; however, one spacer provided protection (Figure S3B). The position and number of modified cytosines may influence type I-E evasion, but no clear position bias was apparent (Figure S3C). We hypothesize that 5ara-hC provides total evasion from type I-F targeting due to modification of the double-stranded GG/CC protospacer adjacent motif (PAM), which must be recognized by the type I-F interference complex for targeting.[59] Indeed, LC53 fully evaded type I-F CRISPR-Cas, even if the only modified cytosines were in the PAM, with others in non-base-paired positions or on the non-targeted strand (Figure S3C). In summary, the 5ara-hC modification of LC53 facilitates evasion from the DNA-targeting type I CRISPR-Cas systems in *Serratia*.

Arabinosylated phages remain vulnerable to RNA-targeting CRISPR-Cas

The 5ara-hC-based evasion of DNA-targeting defense is consistent with the vulnerability of LC53 to type III-A RNA-targeting CRISPR-Cas (Figure 1). Since our initial result was obtained using a single spacer, we further investigated RNA targeting by designing three type III spacers targeting putative early (gp46, DNA polymerase), middle (gp53, endonuclease), and late (gp124, endolysin) expressed LC53 genes (Figure S3D). As predicted, DNA arabinosylation did not protect LC53 from type III targeting, as all spacers provided full protection in liquid culture and reduced or completely prevented the phage from plaquing on solid media (Figures 3A, 3B, and S3E). We hypothesized that RNA-targeting type VI (Cas13a) CRISPR-Cas systems would also inhibit infection by arabinosylated phages. Therefore, we expressed the *Leptotrichia buccalis* type VI-A system from a plasmid with spacers targeting LC53 regions similar to those targeted by the type III system (Figure S3D). Notably, type VI CRISPR-Cas provided full or partial protection against LC53 infection depending on the spacer (Figures 3A, 3B, and S3E), demonstrating that both class 1 and class 2 RNA-targeting CRISPR-Cas systems are effective against the DNA-modified LC53 phage.

The type III-A complex binds RNA complementary to the crRNA,[60] which activates the HD (a single-stranded DNA [ssDNA] nickase)[61,62] and palm (cyclic-oligo-adenylate [cOA] synthase)[63,64] domains of Cas10 (Figure 3C). The accessory dsDNase NucC recognizes cyclic tri-adenylate (cA_3) and, in response, degrades phage and host DNA.[65] Immunity is deactivated when the Cas7 cleaves the target RNA, freeing the type III complex, which switches off Cas10.[66] To determine how type III-A CRISPR-Cas provided protection against a phage with a hypermodified genome, we tested defense against LC53 in different type III mutants. Mutation of the RNA cleavage activity of Cas7 subunits or the HD domain in Cas10 had no effect on defense. However, mutation of the Cas10 palm domain or deletion of *nucC* abrogated phage immunity (Figures 3C and 3D). Therefore, type III immunity against LC53 requires target RNA recognition, cA_3 synthesis, and subsequent activation of NucC that results in dsDNA degradation.

The requirement of NucC for defense led us to test whether NucC cleaves the 5ara-hC-modified DNA. While NucC with added cA_3 fully degraded *Serratia* genomic DNA and the DNA of unmodified phages (JS26 and PCH45), LC53 and T4 DNA were still largely intact (Figure S3F). This demonstrates that the 5ara-hC or 5ghmC modification inhibits NucC DNA binding and/or cleavage but that there may be some phage DNA genomic regions accessible to NucC due to low cytosine abundance. However, as chromosomal DNA is still sensitive to degradation by NucC (Figure S3F), this is likely the major contributor to phage resistance. This is similar to how the type III-A system provides defense against a nucleus-forming jumbo phage, despite the inaccessibility of phage DNA within the phage nucleus.[22,65]

Many type III systems have an accessory non-specific RNase (e.g., Csm6) rather than a DNase,[67] so we swapped NucC for Csm6. Csm6 provided LC53 resistance, highlighting that non-specific collateral RNA cleavage is sufficient to provide

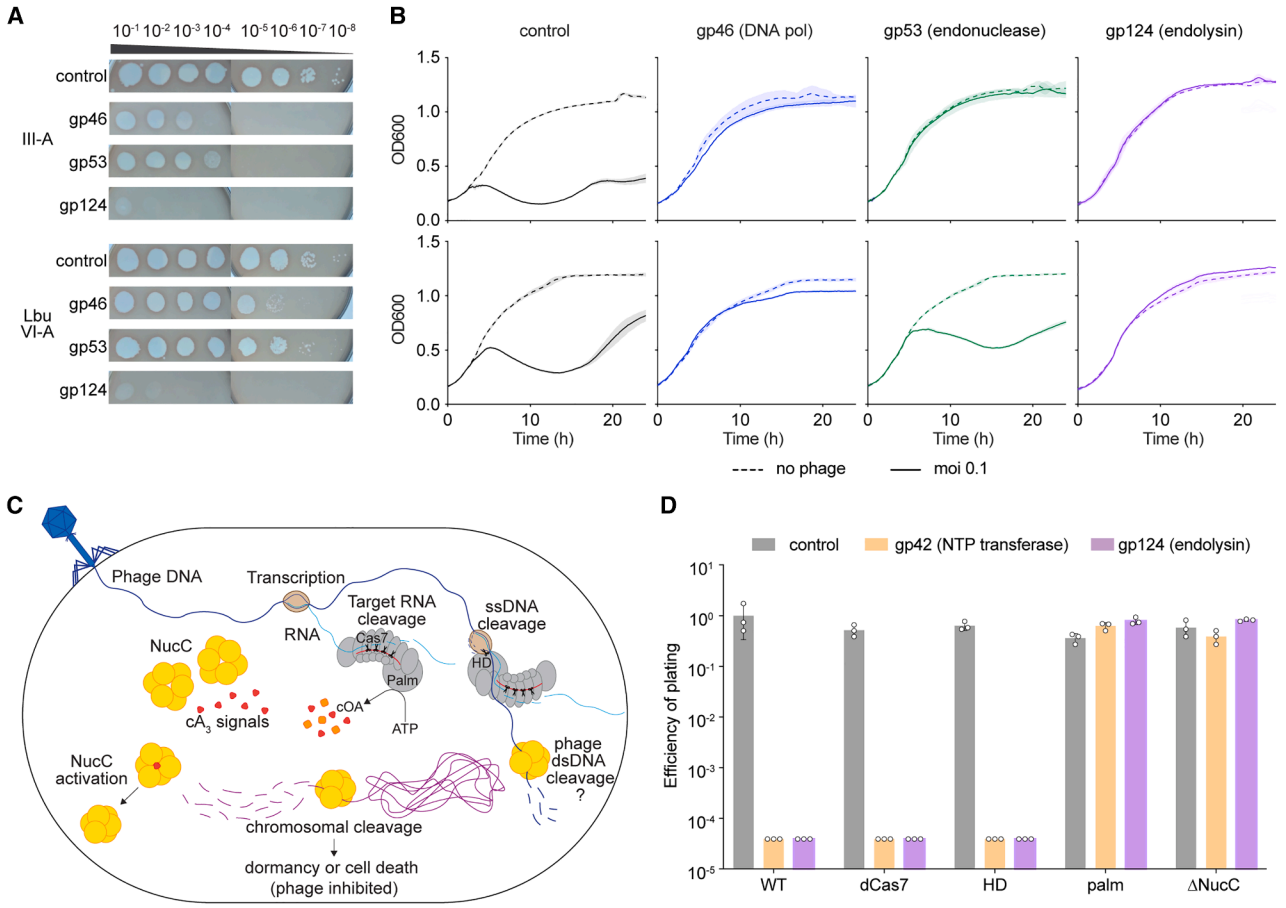


Figure 3. RNA-targeting type III-A or VI-A CRISPR-Cas systems provide immunity against LC53 with arabinosylated DNA

(A and B) (A) Phage infection on plates and (B) liquid culture (moi = 0.1) by LC53 on *Serratia* strains, each expressing a spacer targeting a phage transcript (DNA polymerase, endonuclease, or endolysin) from a plasmid-borne mini-array for the corresponding CRISPR-Cas system (type III-A or a heterologous LbuCas13a). In (A), representative spot serial phage dilutions are shown.

(C) Schematic of *Serratia* type III-A CRISPR-Cas mechanism.

(D) EOP of LC53 on different catalytical type III mutants with a spacer targeting the NTP transferase or the endolysin. Bars are the mean of biological triplicates ± standard deviation, and individual data points are shown.

resistance against modified phages (Figure S3G). This non-specific RNase, Csm6, can cause a dormancy-type response[68] similar to type VI (i.e., Cas13) CRISPR-Cas systems, which also trigger collateral RNase activity.[69] Overall, LC53 remains vulnerable to RNA-targeting type III and VI CRISPR-Cas due to their collateral DNA or RNA cleavage, which likely causes cell dormancy or death, limiting phage propagation.

Arabinosylated phages add one or two arabinose molecules to 5-hydroxy-cytosine

We were interested in whether 5ara-hC occurs in other arabinosylated phages. Only two other phages with arabinosylated DNA are reported. For *E. coli* phage RB69, arabinose was detected but the base modified and the chemical linkage were unknown.[56] For *Serratia* phage 92A1, the absence of unmodified dC nucleotides was demonstrated, and genes for arabinosylation of 5ara-hmC were predicted.[57] Therefore, to determine whether arabinosylated phages have 5ara-hC rather than 5ara-hmC, we identified potentially arabinosylated phages by examining their DNA modification locus (Figure 4A). As described earlier, the dCMP Hases of known (RB69 and 92A1) and predicted arabinosylated phages are highly conserved and distinct from dCMP HMases in glucosylated phages (Figures 2A, 2B, and S1A). In agreement, most genes in the modification loci were conserved between the predicted arabinosylated phages (Figure 4A), suggesting that they have 5ara-hC-modified DNA. To directly test for 5ara-hC, we used LC-MS/MS to examine *Serratia* phage 92A1 DNA (*Straboviridae* distantly related to LC53 [68.6% nt id]), revealing the same 5ara-hdC as LC53 (Figure S4A). Next, we explored if 5ara-hC is specific to *Serratia* phages by examining *E. coli* phages Bas46, Bas47,[70] and RB69 (highly similar to one another and 68.8% nt id to LC53), which belong to the *Mosigvirus* genus within the *Tevenvirinae* subfamily of *Straboviridae* (Table S3). LC-MS/MS indicated these phages have the same hydroxy linkage, as a peak corresponding to 5ara-hdC was detected (Figures 4B, S4B, and S4C). This demonstrates that arabinosylated phages use hydroxylated rather than hydroxymethylated cytosines.

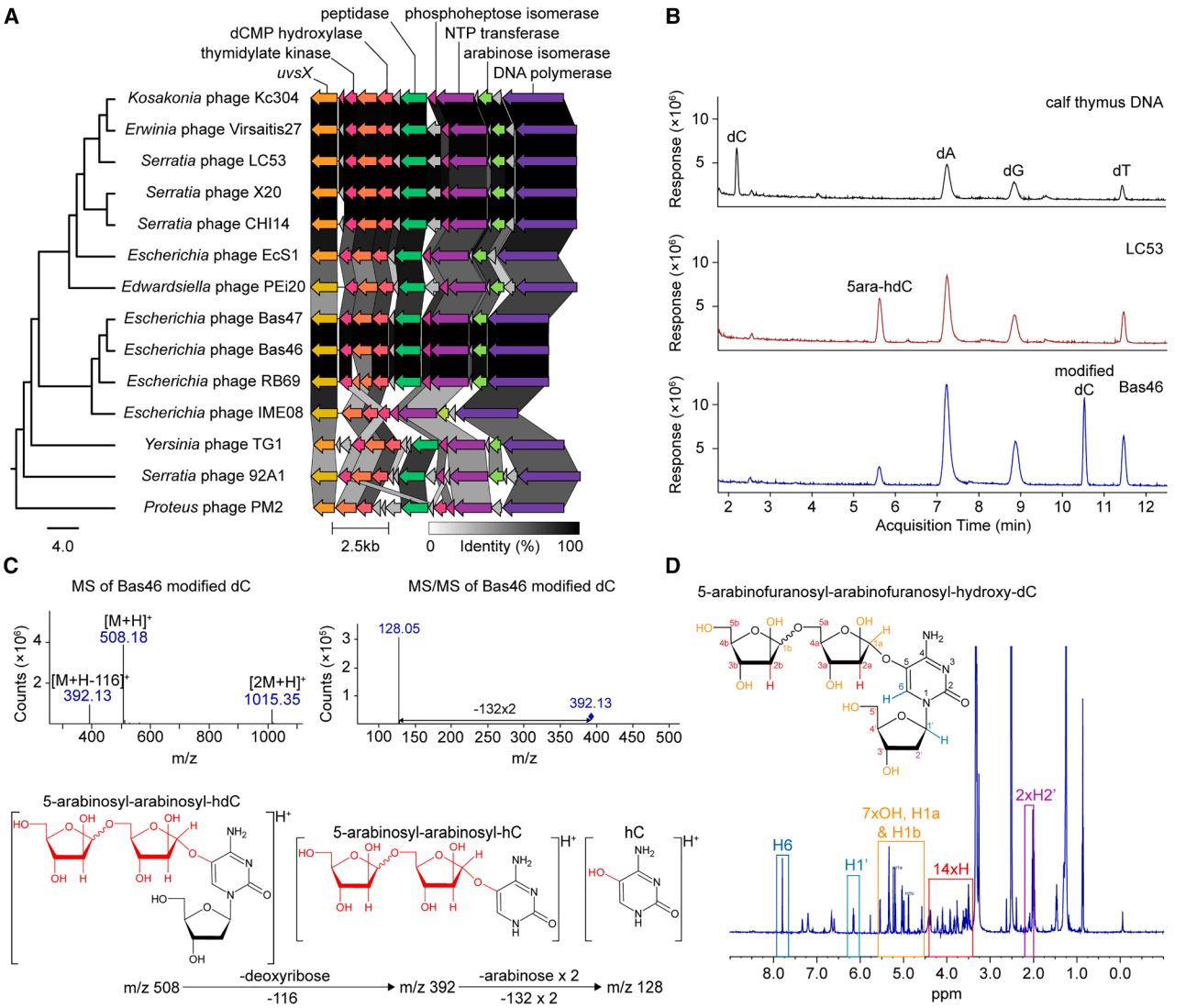


Figure 4. Some arabinosylated phages add a second arabinose yielding 5ara-ara-hC

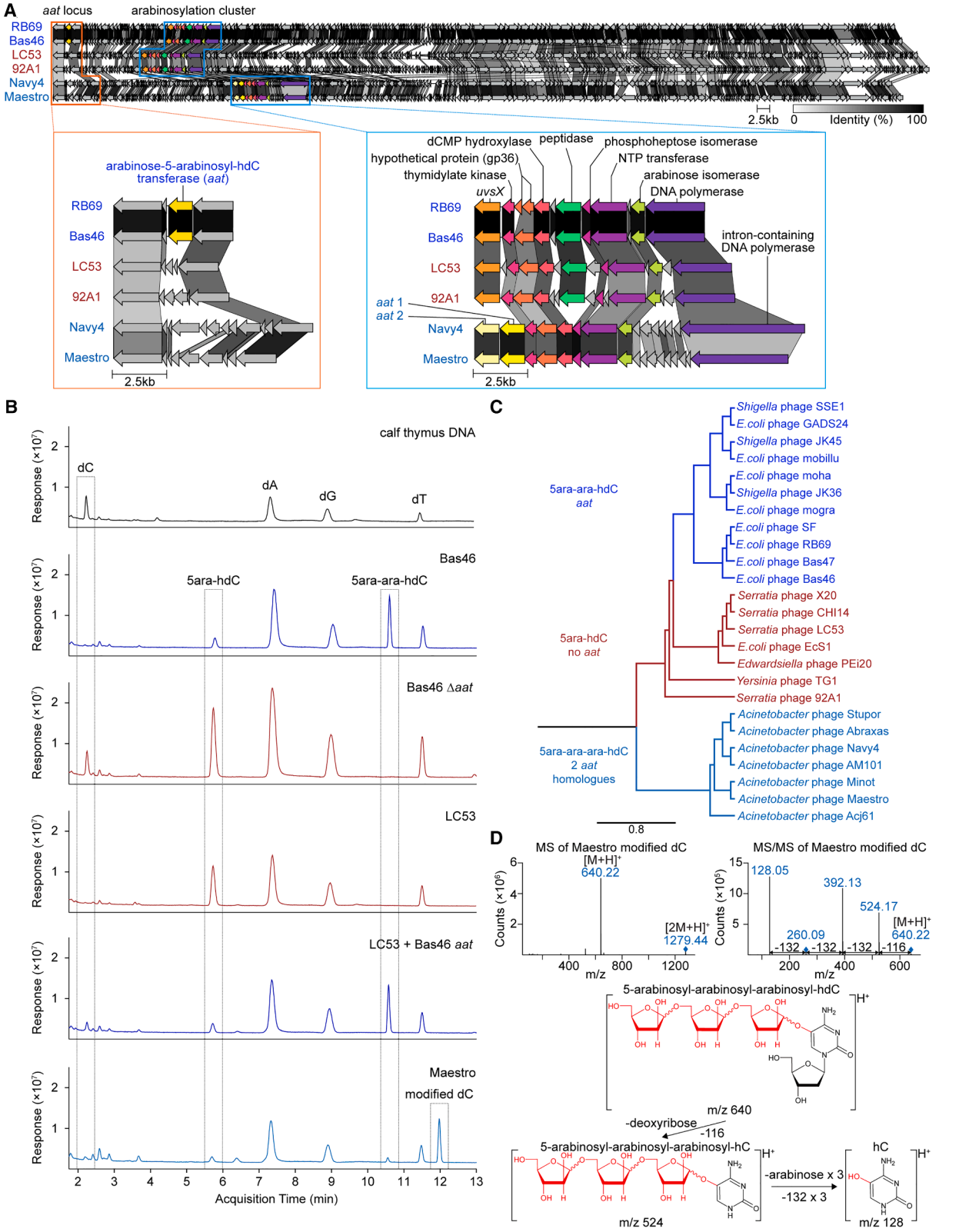
(A) Phylogenetic tree of genomes of arabinosylated phages. The arabinosylation cluster of these phages is shown, with protein conservation represented with a white-black gradient. Details of phages are provided in Table S3.

(B) LC-MS/MS trace of *E. coli* phage Bas46 compared with 5ara-hC-modified LC53 DNA and unmodified calf thymus DNA.

(C) MS and MS/MS spectra of modified dC nucleosides from Bas46 DNA with chemical structures corresponding to the *m/z* ratios from MS and MS/MS.

(D) Structure of 5-arabinofuranosyl-arabinofuranosyl-hydroxy-dC and corresponding NMR spectrum. Protons and corresponding peaks are color-coded. The anomeric protons of arabinose one (H1a) and arabinose two (H1b) were observed at 5.24 and 4.88 ppm, respectively. The linking position between the second and the first arabinose is likely a 1-5 linkage.

Surprisingly, the LC-MS/MS of these *E. coli* phages differed from that of LC53 as the 5ara-hdC peak was at low abundance, and an additional, more abundant peak was detected (Figures 4B, S4B, and S4C). The modification mass in the positive ion mode was 508 Da, which corresponds to the mass of an additional arabinose (132 Da), compared with 5ara-hdC (*m/z* 376) (Figure 4C). The high-resolution MS of the protonated $[M+H]^+$ ion of the modification (*m/z* 508.1781) matches the theoretical mass of the $[M+H]^+$ ion of 5ara-ara-hdC (*m/z* 508.1779, mass error = 0.39 ppm). The CID MS/MS spectrum showed $[M+H]^+$ fragment ions corresponding to the loss of 2′-deoxyribose (*m/z* 392) and a further two arabinoses (*m/z* 128). The modified nucleosides were purified by HPLC and analyzed by NMR (Figures 4D and S4D–S4H). In ^{1}H NMR spectrum, the cytosine H6 was detected at 7.80 ppm as a singlet peak, and the anomeric protons of deoxyribose and arabinoses were observed at 6.15 ppm (H1′), 5.24 ppm (H1a), and 4.88 ppm (H1b), respectively, through the analysis of deuterium exchange ^{1}H NMR, COSY, and NOESY ^{1}H NMR. Similar to the arabinose in 5ara-hdC, H1a of 5ara-ara-hdC has a very small coupling constant (broad singlet), indicating the first arabinose is either a 1,2-*trans* alpha-D or a 1,2-*trans* beta-L arabinofuranose. By contrast, H1b was a doublet with a coupling constant of 4.6 Hz, indicating the second arabinose has a different stereochemistry, possibly



(legend on next page)

either a 1,2-*cis* beta-D or a 1,2-*cis* alpha-L arabinofuranose.[71,72] In summary, arabinosylated phages add one or two arabinose molecules to cytosines on the phage DNA via a hydroxy linkage.

Arabinose-5ara-hC transferases add a second or third arabinose to arabinosylated DNA

To identify the protein that adds the second arabinose to generate 5ara-ara-hC, we compared the genomes of arabinosylated phages (Figure 5A). One gene was present in Bas46, RB69, and other *Mosigviruses* but absent in singly arabinosylated phages belonging to other genera within the *Straboviridae* (e.g., LC53 and 92A1) (Figure 5A). This protein was previously predicted to be a 5hmdC-arabinosyl transferase,[56] but we show below that it transfers the second arabinose unit onto 5ara-hC and therefore named it arabinose-5ara-hC transferase (Aat). To determine if Aat synthesizes 5ara-ara-hC, we exploited a recombination and type VI (Cas13a) RNA-targeting CRISPR-Cas counter-selection strategy to delete *aat* (GenBank: QXV76947.1) from Bas46 and generated phage Bas46 Δ*aat* (Figure S5A). The ability of Cas13a to target the 5ara-ara-hC-modified *E. coli* phages (Figure S5B) further supports that the arabinose modifications (single or double) do not protect from RNA-targeting defenses. Since Bas46 Δ*aat* DNA contained only a single arabinose (5ara-hdC) when analyzed by LC-MS/MS (Figure 5B), we concluded that Aat incorporates a second arabinose unit onto 5ara-hC. Furthermore, by propagating LC53 on *Serratia* containing a plasmid that overexpressed *aat* from Bas46 or Bas47, the resulting phage DNA was modified predominantly with 5ara-ara-hdC instead of 5ara-hdC (Figures 5B and S5C). This demonstrates modularity within this modification pathway and the ease of changing the modification status of the phage DNA via replication on an engineered host strain. The phylogeny of arabinosylated phages revealed that 5ara-hC-modified phages lacking *aat* cluster together, whereas the 5ara-ara-hC-modified phages of the *Mosigvirus* genus form an independent clade (Figure 5C). It is therefore likely that all arabinosylated phages use the same enzymes to substitute cytosines with 5ara-hC during DNA replication and that the *Mosigviruses* have evolved a further 5ara-ara-hC modification due to the Aat enzyme.

Interestingly, we identified that some *Acinetobacter* phages of two different genera within the *Twarogvirinae* subfamily (*Straboviridae*) encode two Aats. The two Aats (Aat1 and Aat2) share approximately 50% aa identity with each other, and each Aat group is phylogenetically distinct from the *Mosigvirus* Aat (i.e., three different clades—Aat, Aat1, and Aat2; Figure S5D; Tables S3 and S4). Acinetobacter phage Maestro DNA contained modified cytosines distinct from LC53 and Bas46 phages (Figure 5B), which were demonstrated by MS to contain three pentose units attached to hC (Figure 5D). The mass of the modification in the positive ion mode was 640 Da, which is 132 Da more than the mass of 5ara-ara-hdC (*m/z* 508). The conservation of phage genes involved in generating the single and double arabinosylation of DNA with the *Twarogvirinae* phages suggests that the triple pentose-hydroxy-cytosine modification of Maestro contains three arabinoses (Figure 5A). Therefore, we denote this modification 5ara-ara-ara-hdC. The high-resolution MS of the protonated $[M+H]^+$ ion (*m/z* 640.2205) matches well with the theoretical mass of the $[M+H]^+$ ion of 5ara-ara-ara-hdC (*m/z* 640.2201, mass error = 0.62 ppm). The CID MS/MS spectrum showed $[M+H]^+$ fragment ions at *m/z* 524, *m/z* 392, *m/z* 260, and *m/z* 128, corresponding to ions with the loss of 2′-deoxyribose (*m/z* 524), the loss of 2′-deoxyribose and one pentose (*m/z* 392), the loss of 2′-deoxyribose and two pentoses (*m/z* 260), and the loss of 2′-deoxyribose and three pentoses (*m/z* 128), respectively (Figure 5D).

One of the *Twarogvirinae* Aat homologs likely adds an arabinose to 5ara-hC (generating 5ara-ara-hC), while the second Aat adds the third arabinose to make 5ara-ara-ara-hC. The separate phylogenetic clustering of the two Aats suggests they have different substrate specificities. We aimed to investigate the role of the two *Twarogvirinae* Aats by complementing LC53 or Bas46 Δ*aat*, but expressing the Aats individually or together was toxic for *Serratia* and *E. coli* and/or interfered with phage replication. In summary, arabinosylated phages have arabinose linked to cytosines via a hydroxy group to generate 5ara-hC, and when present, Aat enzyme(s) further modify 5ara-hC to 5ara-ara-hC and 5ara-ara-ara-hC.

Proposed phage DNA cytosine arabinosylation pathway(s)

We predict 5ara-hC to be essential for LC53 replication since trials to delete the dCTPase, dCMP Hase, NTP transferase, or the glycosyltransferase (gp36) by homologous recombination and type III or VI CRISPR-Cas counter-selection strategies all resulted in viral extinction. In addition, investigating gene function by in *trans* modification of a phage (e.g., T4 C mutant) was hindered either by being unable to obtain correct clones, by toxicity during expression, or by interference with phage replication. Examining the different gene functions and their role in the 5ara-hC formation would therefore require *in vitro* approaches.

Although the chemical pathway for 5ara-hC synthesis and the enzymes responsible could not be directly tested *in vivo*, we propose a model based on the conservation of genes in the arabinosylation cluster with T4 and our results (Figure 6; Table S5). Bacterial DNA degradation, following phage infection, and dCTPase activity result in the accumulation of dCMP that can be hydroxylated, likely via the dCMP Hase and the U32 family

Figure 5. Some phages encode Aat that add a second or third arabinose to the 5ara-hC-modified DNA

(A) Genome comparison of 5-ara-ara-hC- (RB69 and Bas46), 5-ara-hC- (LC53 and 92A1), and 5ara-ara-ara-hC-modified (Navy4 and Maestro) phages (top) with the modification loci highlighted (below). Hypothetical protein (gp36) is split into two genes in RB69 with both labeled. Protein conservation is represented by a black-white gradient. Details of genes encoding Aats are provided in Table S4.

(B) LC-MS/MS of Bas46, Bas46 Δ*aat* mutant, LC53, LC53 complemented with Bas46 *aat*, and Maestro compared with unmodified calf thymus DNA.

(C) Whole-genome phylogeny of representative arabinosylated phages indicating the presence (blue) or absence (red) of *aat* gene(s). Details of phages shown are provided in Table S3.

(D) MS and MS/MS spectra of the modified dC from Maestro DNA with chemical structures corresponding to the obtained *m/z* ratios. The linking position between the third and the second arabinose is likely a 1–5 linkage but needs confirmation.

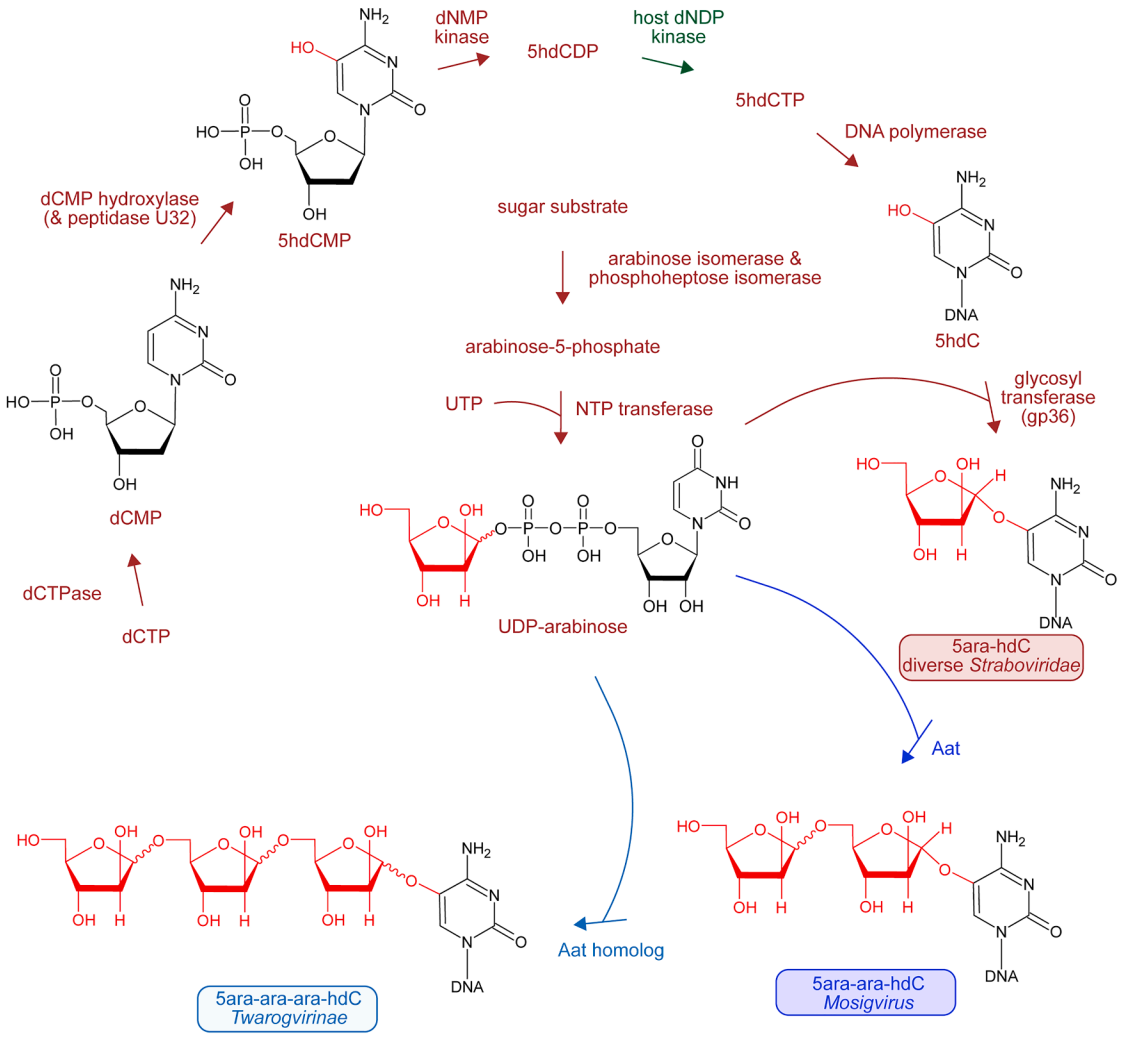


Figure 6. Proposed arabinosylation pathway

Pathway with phage- and host-encoded enzymes to generate the 5ara-hC modification, which occurs in diverse members of the *Straboviridae* family, the 5ara-ara-hC modification of *Mosigvirus* genus, and the 5ara-ara-hC modification of *Twarogvirinae* subfamily. Accession numbers of representative enzymes for each modification are listed in Table S5.

peptidase.[73,74] The related *E. coli* RlhA U32 peptidase hydroxylates cytosines in 23S rRNA,[74] and the *E. coli* TrhP U32 peptidase contributes to 5-hydroxyuridine formation during tRNA modification.[73] Therefore, the dCMP Hase and U32 peptidase are likely involved in dCMP hydroxylation. Interestingly, these peptidases were not detected in *Acinetobacter Twarogvirinae* phages, so their role awaits testing. The 5hdCMP is subsequently phosphorylated by phage and host-encoded kinases to generate 5hdCTP, which is incorporated in DNA by the phage DNA polymerase.

We predict that arabinose is added after 5hdC incorporation into the phage DNA, based on 5ghmC generation in T4; however, we cannot rule out that arabinose is added pre-replication. UDP-arabinose synthesis and transfer to 5hdC likely involve multiple phage proteins. Arabinose-5-phosphate (ara-5-P) is likely generated from a sugar substrate via the ara-5-P isomerase (arabinose isomerase) and/or the phosphoheptose isomerase, both belonging to the sugar isomerase (SIS) domain protein family.[75,76] Interestingly, the arabinose isomerase is absent in some phages containing the arabinosylation cluster (e.g., PM2; Figure 5A) and was not essential since a LC53 mutant with a premature stop codon in this gene still contained 5ara-hdC when assessed by LC-MS/MS (Figure S6A). Bacterial arabinose isomerases involved in lipopolysaccharide synthesis (e.g., KdsD in *Enterobacteriaceae* hosts like *Serratia*)[77] may provide redundancy for the phage arabinose isomerase during 5ara-hC DNA modification. In the next step, an NTP transferase domain-containing protein (NTP transferase) is predicted to combine UTP with ara-5-P to generate UDP-arabinose. Indeed, structural prediction of the NTP transferase indicates a function as a dual sugar-1-phosphate nucleotidyl transferase (Figure S6B).

Incorporation of arabinose into the 5hdC-modified DNA is predicted to involve an uncharacterized protein encoded downstream of the dCMP Hase (gp36 for LC53). Structural prediction

of this protein suggested its role as a glycosyltransferase (Figure S6B). Therefore, UDP-arabinose likely serves as the substrate for the hypothetical protein (gp36) to be transferred to 5hdC to generate 5ara-hdC-modified phage DNA (Figure 6). UDP-arabinose is also the predicted substrate for the Aat enzymes to add the additional arabinose to generate the 5ara-ara-hC and 5ara-ara-ara-hC modification. In summary, we propose a pathway for the generation of different arabinose modifications in phage DNA.

Distribution of arabinosylated phages in genomic databases

To investigate how widespread these phage genome modifications are, we searched phage and metagenomic databases for proteins involved in DNA arabinosylation and glucosylation (Figures 7A and S7A). Within phages (INPHARED),[78] we predicted glucosylation in 464 phages and arabinosylation in 215 phages with various combinations of *aat* genes (Figure 7A). Within metagenomes (IMG/VR),[79] we predicted an additional 1,758 glucosylated and 30 arabinosylated phages (Figure S7B). We assessed which taxonomic phage groups contain glycosylated genomes (Figure 7B). Whereas phages that lacked possible glycosylation genes are taxonomically diverse, glucosylated phages belonged to *Tequatroviruses* (i.e., T4 phages) and *Jiaodaviruses* (T-even phages that typically infect *Klebsiella*). Phages that contain arabinosylation genes, but no Aat enzymes, belong to diverse genera that infect members of the *Enterobacteriaceae* (Figures 7B and S7C). When searching for phages with possible double arabinosylation (i.e., an Aat similar to Bas46/RB69), we identified similar *Mosigviruses* predicted to infect *Escherichia* and *Shigella* (Figures 7B and S7C). Phages containing Aat1 and Aat2, and therefore likely triple arabinosylation, belong to a few different genera with *Acinetobacter* hosts (Figures 7B and S7C). Phages with one or more Aats, but which could not be specifically assigned as Aat, Aat1, or Aat2 members, belong to five major genera and infect mostly *Escherichia*, *Acinetobacter*, and other members of the *Enterobacteriaceae*. In summary, many phage genera are predicted to possess single or multiple arabinosylation modifications, and these phages infect *Enterobacteriaceae*, many of which are important pathogens.

DNA arabinosylation and number of arabinoses affect defense responses

The existence of 5ara-hC-, 5ara-ara-hC-, and 5ara-ara-ara-hC-modified DNA and their different chemistry to 5ghmC led us to hypothesize that they evolved to evade different phage defense systems. We exploited *E. coli* phage Bas46 and the Δ*aat* mutant to directly compare defense system responses to isogenic phages with one or two ara groups on their DNA. First, we tested two different type IV (GmrSD) RM systems in *E. coli*,[81] which had strong defense against both Bas46 and Bas46 Δ*aat* (Figure 7C) and T4.[81] The vulnerability of 5ara-ara-hC DNA to type IV restriction was confirmed *in vitro* by Bas46 DNA cleavage with AbaSI, while the DNA was protected from the type II restriction enzyme MspI (Figures S7D and S7E). Therefore, both single or double arabinosylated DNA protects from type II RM while remaining vulnerable to type IV restriction. AbaSI is derived from *A. baumannii*[55] and was also active against DNA from *A. baumannii* phage Maestro containing 5ara-ara-ara-hC-modified DNA (Figure S7F). Our results demonstrate that AbaSI is active on a much wider range of glycosylated (glucosylated and arabinosylated) DNA substrates than previously known.

We predicted that defenses with differential responses based on arabinosylation or glucosylation would exist. A DNA glycosylase enzyme (Brig1) that recognizes and excises α-5ghmC bases from T4 DNA to provide resistance was recently discovered.[82] We tested Brig1 and three other diverse homologs for their ability to provide phage resistance against arabinosylated phages Bas46 (5ara-ara-hC), Bas46 Δ*aat* (5ara-hC), and T4 (5ghmC) (Figures 7D and S7G). As expected, Brig1 provided resistance against T4, and another homolog also provided T4 resistance. No Brig1 homologs provided defense against Bas46 or Bas46 Δ*aat*, providing direct evidence for defense evasion via different phage DNA modifications, with either 5ara-ara-hC or 5ara-hC sufficient to evade Brig1 activity against phages with 5ghmC DNA. We hypothesize that proteins that specifically recognize arabinosylated DNA await discovery.

We further hypothesize that 5ara-ara-hC evolved in phages via *aat* acquisition to overcome some defenses effective against 5ara-hC-modified phages. Since no defenses with this specificity are known, we screened natural clinical plasmids from different *Enterobacteriaceae* in *E. coli* for defense against Bas46 and Bas46 Δ*aat*. A plasmid from *Klebsiella pneumoniae* (*K. pneumoniae*) provided minimal defense in *E. coli* against the 5ara-ara-hC Bas46 phage, with only a small plaque size reduction. In comparison, this plasmid provided stronger defense against the 5ara-hC Bas46 Δ*aat* phage, with tiny plaques and a reduction in efficiency of plating (EOP) on plates (Figure 7C), which was more striking (approximately 10^2 reduction) when the top agar concentration was increased (Figure S7H). T4 (5ghmC-modified) was the most sensitive phage to the defense on this plasmid (Figure 7C). Therefore, this natural plasmid contains a defense most effective against 5ghmC-modified T4 than 5ara-hC-modified phages, whereas the presence of *aat* and the resulting 5ara-ara-hC modification improved phage infection. The plasmid is large (>200 kb) with multiple predicted defenses and many genes of unknown function. Thus, the system responsible remains to be characterized in a future study. In summary, distinct modifications have evolved to allow phages to overcome defenses with different specificities for the modification (e.g., 5ghmC vs. 5ara-hC) or length (5ara-hC vs. 5ara-ara-hC).

DISCUSSION

Many bacterial defense systems rely on the recognition of foreign phage DNA to elicit immunity.[4,83] Therefore, modification of nucleobases enables phages to evade diverse DNA-sensing and -targeting defenses. We have discovered phage families with either single, double, or triple arabinosylated cytosines. These DNA modifications involve arabinosylation of 5-hydroxy cytosine (5ara-hC, 5ara-ara-hC, or 5ara-ara-ara-hC) rather than hydroxymethylated cytosines reported for *E. coli* phage T4. This hydroxy linkage between the pyrimidine and sugar was confirmed by NMR. *Serratia* phage LC53, which harbors the 5ara-hC modification, has complete or partial evasion of the native DNA-targeting type I-F and I-E CRISPR-Cas systems. Furthermore, 5ara-hC, 5ara-ara-hC, and 5ara-ara-ara-hC DNA were insensitive to cleavage by type II restriction endonucleases

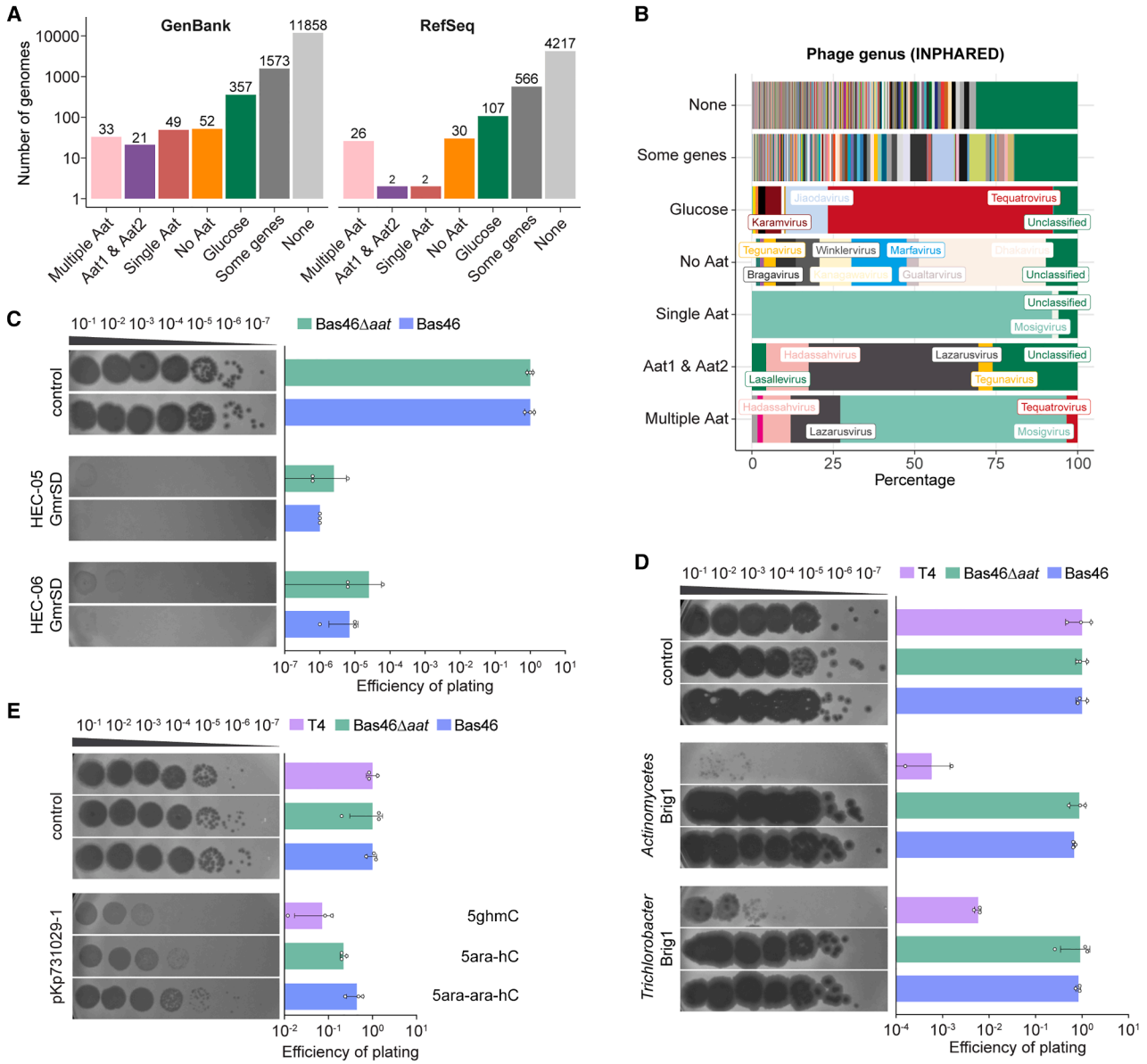


Figure 7. The distribution and effects on phage defense of DNA arabinosylation

(A) Number of predicted genomes (INHPARED database) for phages with different modification components present in the genome (color, none; no similar genes detected, some genes; an incomplete set of modification genes identified, glucose; glucosylation predicted, no Aat; single arabinosylation predicted, single Aat; double arabinosylation predicted, Aat1 and Aat2; triple arabinosylation predicted based, multiple Aat; either more than two Aats or two Aats from the same group (Aat1 or Aat2).

(B) Percentage of each phage genus (color) for each modification group predicted in the INPHARED database. Genera representing over 3% of a modification group (not "none") are indicated by name. Predictions of modifications in metagenomes (IMG/VR) and the bacteria hosts of phages are provided in Figure S7.

(C) EOP of phage infection by Bas46 and the Bas46 Δ*aat* mutant on *E. coli* expressing two different type IV (GmrSD) RM systems (HEC-05 [similar to BrxU][80] and HEC-06) (0.35% w/v agar).

(D) EOP of phage infection by T4, Bas46, and the Bas46 Δ*aat* mutant on *E. coli* expressing two different Brig1 systems (0.35% w/v agar).

(E) EOP of phage infection by T4, Bas46, and the Bas46 Δ*aat* mutant on *E. coli* carrying a plasmid from *K. pneumoniae* (MSRN731029). In (C)–(E), representative spot serial phage dilutions are shown, and the data presented are the means ± SD with individual biological replicates shown.

but were cut by the type IV restriction endonuclease AbaSI. Additionally, 5ara-hC and 5ara-ara-hC phages were vulnerable *in vivo* to two different GmrSD systems. Therefore, the spectrum of substrates accessible to AbaSI (and GmrSD systems) is broader than previously known, suggesting that these enzymes evolved to suppress invaders with a range of modifications. Notably, 5ara-hC and 5ara-ara-hC (and presumably 5ara-ara-ara-hC) phages were vulnerable to RNA-targeting CRISPR-Cas defenses (type III and VI). Importantly, we demonstrated that 5ara-hC and 5ara-ara-hC modifications enable these phages

to evade DNA glycosylases that were able to target the 5ghmC modification of T4. Finally, we identified a large defense plasmid that provides the strongest defense against 5ghmC-modified T4, less defense against phages with 5ara-hC-modified DNA, and minimal defense against 5ara-ara-hC-modified phages. Overall, we have discovered a family of previously unknown DNA modifications and demonstrated their importance in evasion of multiple bacterial defenses, while remaining vulnerable to others.

Almost all reported DNA modifications at the 5-position of cytosine, such as 5-methylcytosine (5mC) and 5hmC, are formed through a C–C bond. 5-hydroxycytosine (5hC), in which the hydroxy group is linked directly to the pyrimidine ring through a C–O bond, has been viewed as an oxidative damage product of cytosine. However, one study has indicated that 5hC could potentially be formed enzymatically because all phage N-17 cytosines were replaced with 5hC.[84] In our study, we detected the hydroxy linkage as a common feature of phages with arabinosylated DNA, which enabled us to update the modification of RB69 and 92A1 to 5ara-ara-hC and 5ara-hC, respectively.[56,57] We also demonstrated that Bas46 and Bas47 phages have DNA with 5ara-ara-hC modifications. Indeed, the dCMP Hases of arabinosylated phages cluster separately from and share low similarity with dCMP HMases of glucosylated phages (e.g., T4). Therefore, enzymes of the TS family not only methylate or hydroxymethylate pyrimidines[85,86] but can also hydroxylate cytosines (i.e., the arabinosylated phage dCMP Hases). Interestingly, a pentose-hdC linkage was recently described for a *Shewanella* phage.[87] We demonstrated that the Aat enzyme is responsible for the addition of the second arabinose to generate 5ara-ara-hC in *Mosigviruses*. This enzyme was previously incorrectly described as the arabinosyl-hmC transferase.[56] Our experimentally verified role for Aat agrees with the absence of *aat* from the genomes of single-arabinosylated phages. Our prediction of *Twarogvirinae* phages with two Aat enzymes and the demonstration that they add two additional sugars further support our assignment of the role of Aat.

We demonstrated that 5ara-hC-modified DNA fully or partially protected phage LC53 from DNA-targeting CRISPR-Cas systems. These findings expand on the earlier results that 5ghmC modification enables phage T4 to completely or partially evade DNA-targeting CRISPR-Cas systems.[44–46,88] For T4, the more open structure of Cas12 (type V) was less affected by modifications than type I-E or Cas9 (type II).[45,46] T4 was not completely resistant to type I and II,[44,45,88] which is similar to our finding that the 5ara-hC modification of LC53 did not provide complete resistance to targeting by all type I-E spacers. This observed variability for T4 and LC53 likely depends on the position, strand, and number of modified cytosines in the target DNA. Importantly, we show that the 5ara-hC-modified LC53 phage completely evades the DNA-targeting type I-F system, which we ascribe to the strict requirement for a GG/CC double-stranded PAM for recognition and DNA cleavage.[59] Based on these findings, we predict that the 5ghmC modification of phage T4 will also result in complete evasion of type I-F systems. Since the RNA transcripts of arabinosylated phages remain accessible, 5ara-hC and 5ara-ara-hC phages were vulnerable to RNA-targeting CRISPR-Cas systems (type III and VI). Thus, we exploited type VI targeting in counter-selection strategies for phage mutant generation, highlighting the utility of RNA-targeting systems for engineering phages containing modified DNA.

Overall, we found an unpredicted hydroxy linkage-based DNA modification in arabinosylated phages that can be found on cytosines with one, two, or three arabinose units. These modifications likely evolved to aid in the evasion of CRISPR-Cas, RM, Brig1, and other DNA-targeting defenses. The larger 5ara-ara-hC and 5ara-ara-ara-hC modifications might have arisen to provide greater steric hindrance of DNA-targeting bacterial defenses, such as that provided by the *K. pneumoniae* clinical plasmid. Despite diverse modifications in phages, some type IV RM[89] and END[90] nucleases are highly flexible toward chemically distinct DNA modifications. Our findings highlight the importance of chemical confirmation of bioinformatically predicted phage DNA modifications and suggest that the diversity of natural non-canonical nucleobases on phage genomes is higher than previously predicted.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Peter C. Fineran (peter.fineran@otago.ac.nz).

Materials availability

Materials will be made available by the authors upon reasonable request and, if necessary, due to commercial interest under a material transfer agreement.

Data and code availability

- The NMR data for 5-arabinofuranosyl-hydroxy-dC and 5-arabinofuranosyl-arabinofuranosyl-hydroxy-dC have been deposited in the Natural Products Magnetic Resonance Database (NP-MRD; www.np-mrd.org) and can be found at NP0351009 (5-arabinofuranosyl-hydroxy-dC) and NP0351010 (5-arabinofuranosyl-arabinofuranosyl-hydroxy-dC).
- Code used for the analysis of DNA modification genes in phages is provided in the GitHub repository https://github.com/fineranlab/phage-genome-arabinosylation. The repository contains a conda YAML file specifying the programs to install for the analysis. The program versions used in the analysis are provided at the end of each Jupyter notebook. Accession numbers for phage genomes and proteins are available as PDF in the supplemental information and in excel format in the GitHub repository. The accession numbers of phages from INPHARED with the predicted DNA modification and gene annotations are available under github.com/fineranlab/phage-genome-arabinosylation/docs/INPHARED_predicted-genomes.xlsx.
- Any additional information required to reanalyze the data reported in this work is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was funded by Te Pūtea Rangahau a Marsden, Marsden Fund, Royal Society of New Zealand (RSNZ), Te Apārangi, Bioprotection Aotearoa (Tertiary Education Commission, NZ), and P.C.F. was supported by a James Cook Research Fellowship (RSNZ, Te Apārangi). M.M., D.M.-M., J.D., and L.M.M. were supported by University of Otago Doctoral scholarships. M.M. received Maurice Wilkins Centre Category 4 (travel) Funding. K.G.W. was supported by a Feodor Lynen Research Fellowship from the Alexander von Humboldt Stiftung and S.A.J. by a Sir Charles Hercus Health Research Fellowship from the Health Research Council. L.C. and P.C.D. were supported by the National Research Foundation of Singapore under the Singapore-MIT Alliance for Research and Technology Antimicrobial Resistance IRG and by the Agilent ACT-UR program (4658 and 4773). This work was also supported by the European Research Council (ERC) CoG under the European Union's Horizon 2020 research and innovation program (grant no. 101003229 to S.J.J.B.). We thank Benjamin Adler (Doudna laboratory, University of California Berkeley) for pBA559 and pBA560, Adair Borges (Arcadia Science) for *Serratia* strain 92 and phage 92A1,

Alexander Harms (ETH Zürich) for the BASEL phages, Jason Gill (Texas A&M) for *A. baumannii* TP1 and phage Maestro, and the Félix d'Hérelle Reference Center (Université Laval) for phage RB69. We thank Witold Kot (University of Copenhagen) for guidance with nanopore sequencing trials of sugar-modified DNA, Robert Fagerlund for help with NucC purification, Leighton Payne for GmrSD plasmids, Suzanne Warring for helpful discussions, and staff of the Genome Analysis Service (GAS), Otago for Sanger sequencing.

AUTHOR CONTRIBUTIONS

Conceptualization, M.M. and P.C.F.; formal analysis, M.M., L.C., L.M.S., K.G.W., O.D., D.M.-M., S.B., J.M., L.M.M., S.A.J., A.J.F., P.C.D., S.J.J.B., and P.C.F.; funding acquisition, P.C.F., S.J.J.B., P.C.D., and L.C.; investigation, M.M., L.C., L.M.S., K.G.W., O.D., D.M.-M., S.B., M.E.L., H.Y., C.-F.L., J.M., and L.M.M.; project administration, P.C.F.; resources, J.D.; software, O.D.; supervision, L.M.S., S.A.J., S.J.J.B., and P.C.F.; validation and visualization, M.M., L.C., O.D., S.B., and P.C.F.; writing – original draft, M.M. and P.C.F.; and writing – review and editing, all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Bacteria, phages, primers, plamids and growth conditions
  - Preparation of phage stocks and titration
- METHOD DETAILS
  - Bioinformatic analysis of modification genes
  - Phage genomic DNA extraction
  - Restriction digests of phage DNA
  - HPAE-PAD analysis of phage DNA
  - LC-MS analysis of purified phage DNA
  - NMR analysis of 5-arabinofuranosyl-hydroxy-deoxycytidine (5ara-hdC)
  - NMR analysis of 5-arabinofuranosyl-arabinofuranosyl-hydroxy-deoxycytidine (5ara-ara-hdC)
  - Arabinosylation of 5ara-hC phage DNA *in trans*
  - Plasmid expression of anti-phage spacers
  - Phage resistance efficiency of plating assay
  - Phage resistance infection time course
  - Conjugation efficiency assay
  - Expression of Cas13 and anti-phage spacers
  - Investigation of type III-A CRISPR-Cas protection mechanism
  - *In vitro* NucC cleavage assay
  - Exchange of type III-A ancillary nuclease
  - Phage engineering using type VI CRISPR-Cas counter-selection
  - Isolation of arabinose isomerase mutant phage
  - Identification of phages with similar DNA modifications in genomic databases
  - Brig1 cloning and the *Klebsiella pneumoniae* clinical plasmid
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.chom.2025.06.005.

Received: October 26, 2024
Revised: May 11, 2025
Accepted: June 5, 2025
Published: June 26, 2025

REFERENCES

1. d'Hérelle, F. (1917). Sur un microbe invisible antagoniste des bacilles dysentérique. Acad. Sci. Paris *165*, 373–375.
2. Suttle, C.A. (2007). Marine viruses–major players in the global ecosystem. Nat. Rev. Microbiol. *5*, 801–812. https://doi.org/10.1038/nrmicro1750.
3. Georjon, H., and Bernheim, A. (2023). The highly diverse antiphage defence systems of bacteria. Nat. Rev. Microbiol. *21*, 686–700. https://doi.org/10.1038/s41579-023-00934-x.
4. Mayo-Muñoz, D., Pinilla-Redondo, R., Birkholz, N., and Fineran, P.C. (2023). A host of armor: Prokaryotic immune strategies against mobile genetic elements. Cell Rep. *42*, 112672. https://doi.org/10.1016/j.celrep.2023.112672.
5. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. Science *315*, 1709–1712. https://doi.org/10.1126/science.1138140.
6. Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology (Reading) *151*, 2551–2561. https://doi.org/10.1099/mic.0.28048-0.
7. Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R., et al. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat. Struct. Mol. Biol. *18*, 529–536. https://doi.org/10.1038/nsmb.2019.
8. Brouns, S.J.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J.H., Snijders, A.P.L., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960–964. https://doi.org/10.1126/science.1159689.
9. Watson, B.N.J., Steens, J.A., Staals, R.H.J., Westra, E.R., and van Houte, S. (2021). Coevolution between bacterial CRISPR-Cas systems and their bacteriophages. Cell Host Microbe *29*, 715–725. https://doi.org/10.1016/j.chom.2021.03.018.
10. Mayo-Muñoz, D., Pinilla-Redondo, R., Camara-Wilpert, S., Birkholz, N., and Fineran, P.C. (2024). Inhibitors of bacterial immune systems: discovery, mechanisms and applications. Nat. Rev. Genet. *25*, 237–254. https://doi.org/10.1038/s41576-023-00676-9.
11. Davidson, A.R., Lu, W.T., Stanley, S.Y., Wang, J., Mejdani, M., Trost, C.N., Hicks, B.T., Lee, J., and Sontheimer, E.J. (2020). Anti-CRISPRs: Protein Inhibitors of CRISPR-Cas Systems. Annu. Rev. Biochem. *89*, 309–332. https://doi.org/10.1146/annurev-biochem-011420-111224.
12. Malone, L.M., Birkholz, N., and Fineran, P.C. (2021). Conquering CRISPR: how phages overcome bacterial adaptive immunity. Curr. Opin. Biotechnol. *68*, 30–36. https://doi.org/10.1016/j.copbio.2020.09.008.
13. Meeske, A.J., Jia, N., Cassel, A.K., Kozlova, A., Liao, J., Wiedmann, M., Patel, D.J., and Marraffini, L.A. (2020). A phage-encoded anti-CRISPR enables complete evasion of type VI-A CRISPR-Cas immunity. Science *369*, 54–59. https://doi.org/10.1126/science.abb6151.
14. Bondy-Denomy, J., Pawluk, A., Maxwell, K.L., and Davidson, A.R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature *493*, 429–432. https://doi.org/10.1038/nature11723.
15. Pawluk, A., Bondy-Denomy, J., Cheung, V.H.W., Maxwell, K.L., and Davidson, A.R. (2014). A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of *Pseudomonas aeruginosa*. mBio *5*, e00896. https://doi.org/10.1128/mBio.00896-14.
16. Liu, L., Yin, M., Wang, M., and Wang, Y. (2019). Phage AcrIIA2 DNA Mimicry: Structural Basis of the CRISPR and Anti-CRISPR Arms Race. Mol. Cell *73*, 611–620.e3. https://doi.org/10.1016/j.molcel.2018.11.011.
17. Shin, J., Jiang, F., Liu, J.J., Bray, N.L., Rauch, B.J., Baik, S.H., Nogales, E., Bondy-Denomy, J., Corn, J.E., and Doudna, J.A. (2017). Disabling

Cas9 by an anti-CRISPR DNA mimic. Sci. Adv. *3*, e1701620. https://doi.org/10.1126/sciadv.1701620.

18. Guo, T.W., Bartesaghi, A., Yang, H., Falconieri, V., Rao, P., Merk, A., Eng, E.T., Raczkowski, A.M., Fox, T., Earl, L.A., et al. (2017). Cryo-EM Structures Reveal Mechanism and Inhibition of DNA Targeting by a CRISPR-Cas Surveillance Complex. Cell *171*, 414–426.e12. https://doi.org/10.1016/j.cell.2017.09.006.
19. Athukoralage, J.S., McMahon, S.A., Zhang, C., Grüschow, S., Graham, S., Krupovic, M., Whitaker, R.J., Gloster, T.M., and White, M.F. (2020). An anti-CRISPR viral ring nuclease subverts type III CRISPR immunity. Nature *577*, 572–575. https://doi.org/10.1038/s41586-019-1909-5.
20. Chaikeeratisak, V., Nguyen, K., Khanna, K., Brilot, A.F., Erb, M.L., Coker, J.K.C., Vavilina, A., Newton, G.L., Buschauer, R., Pogliano, K., et al. (2017). Assembly of a nucleus-like structure during viral replication in bacteria. Science *355*, 194–197. https://doi.org/10.1126/science.aal2130.
21. Mendoza, S.D., Nieweglowska, E.S., Govindarajan, S., Leon, L.M., Berry, J.D., Tiwari, A., Chaikeeratisak, V., Pogliano, J., Agard, D.A., and Bondy-Denomy, J. (2020). A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases. Nature *577*, 244–248. https://doi.org/10.1038/s41586-019-1786-y.
22. Malone, L.M., Warring, S.L., Jackson, S.A., Warnecke, C., Gardner, P.P., Gumy, L.F., and Fineran, P.C. (2020). A jumbo phage that forms a nucleus-like structure evades CRISPR-Cas DNA targeting but is vulnerable to type III RNA-based immunity. Nat. Microbiol. *5*, 48–55. https://doi.org/10.1038/s41564-019-0612-5.
23. Samson, J.E., Magadán, A.H., Sabri, M., and Moineau, S. (2013). Revenge of the phages: defeating bacterial defences. Nat. Rev. Microbiol. *11*, 675–687. https://doi.org/10.1038/nrmicro3096.
24. Weigele, P., and Raleigh, E.A. (2016). Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses. Chem. Rev. *116*, 12655–12687. https://doi.org/10.1021/acs.chemrev.6b00114.
25. Sood, A.J., Viner, C., and Hoffman, M.M. (2019). DNAmod: the DNA modification database. J. Cheminform. *11*, 30. https://doi.org/10.1186/s13321-019-0349-4.
26. Lee, Y.J., Dai, N., Müller, S.I., Guan, C., Parker, M.J., Fraser, M.E., Walsh, S.E., Sridar, J., Mulholland, A., Nayak, K., et al. (2022). Pathways of thymidine hypermodification. Nucleic Acids Res. *50*, 3001–3017. https://doi.org/10.1093/nar/gkab781.
27. Burke, E.J., Rodda, S.S., Lund, S.R., Sun, Z., Zeroka, M.R., O'Toole, K.H., Parker, M.J., Doshi, D.S., Guan, C., Lee, Y.J., et al. (2021). Phage-encoded ten-eleven translocation dioxygenase (TET) is active in C5-cytosine hypermodification in DNA. Proc. Natl. Acad. Sci. USA *118*, e2026742118. https://doi.org/10.1073/pnas.2026742118.
28. Pyle, J.D., Lund, S.R., O'Toole, K.H., and Saleh, L. (2024). Virus-encoded glycosyltransferases hypermodify DNA with diverse glycans. Cell Rep. *43*, 114631. https://doi.org/10.1016/j.celrep.2024.114631.
29. Lehman, I.R., and Pratt, E.A. (1960). On the structure of the glucosylated hydroxymethylcytosine nucleotides of coliphages T2, T4, and T6. J. Biol. Chem. *235*, 3254–3259. https://doi.org/10.1016/S0021-9258(20)81347-7.
30. Zhou, Y., Xu, X., Wei, Y., Cheng, Y., Guo, Y., Khudyakov, I., Liu, F., He, P., Song, Z., Li, Z., et al. (2021). A widespread pathway for substitution of adenine by diaminopurine in phage genomes. Science *372*, 512–516. https://doi.org/10.1126/science.abe4882.
31. Lee, Y.J., Dai, N., Walsh, S.E., Müller, S., Fraser, M.E., Kauffman, K.M., Guan, C., Corrêa, I.R., Jr., and Weigele, P.R. (2018). Identification and biosynthesis of thymidine hypermodifications in the genomic DNA of widespread bacterial viruses. Proc. Natl. Acad. Sci. USA *115*, E3116–E3125. https://doi.org/10.1073/pnas.1714812115.
32. Cui, L., Balamkundu, S., Liu, C.F., Ye, H., Hourihan, J., Rausch, A., Hauß, C., Nilsson, E., Hoetzinger, M., Holmfeldt, K., et al. (2023). Four additional natural 7-deazaguanine derivatives in phages and how to make them. Nucleic Acids Res. *51*, 9214–9226. https://doi.org/10.1093/nar/gkad657.
33. Olsen, N.S., Nielsen, T.K., Cui, L., Dedon, P., Neve, H., Hansen, L.H., and Kot, W. (2023). A novel Queuovirinae lineage of *Pseudomonas aeruginosa* phages encode dPreQ0 DNA modifications with a single GA motif that provide restriction and CRISPR Cas9 protection in vitro. Nucleic Acids Res. *51*, 8663–8676. https://doi.org/10.1093/nar/gkad622.
34. Kot, W., Olsen, N.S., Nielsen, T.K., Hutinet, G., de Crécy-Lagard, V., Cui, L., Dedon, P.C., Carstens, A.B., Moineau, S., Swairjo, M.A., et al. (2020). Detection of preQ0 deazaguanine modifications in bacteriophage CAjan DNA using Nanopore sequencing reveals same hypermodification at two distinct DNA motifs. Nucleic Acids Res. *48*, 10383–10396. https://doi.org/10.1093/nar/gkaa735.
35. Hutinet, G., Kot, W., Cui, L., Hillebrand, R., Balamkundu, S., Gnanakalai, S., Neelakandan, R., Carstens, A.B., Fa Lui, C., Tremblay, D., et al. (2019). 7-Deazaguanine modifications protect phage DNA from host restriction systems. Nat. Commun. *10*, 5442. https://doi.org/10.1038/s41467-019-13384-y.
36. Tock, M.R., and Dryden, D.T.F. (2005). The biology of restriction and anti-restriction. Curr. Opin. Microbiol. *8*, 466–472. https://doi.org/10.1016/j.mib.2005.06.003.
37. Kelleher, J.E., and Raleigh, E.A. (1991). A novel activity in *Escherichia coli* K-12 that directs restriction of DNA modified at CG dinucleotides. J. Bacteriol. *173*, 5220–5223. https://doi.org/10.1128/jb.173.16.5220-5223.1991.
38. Waite-Rees, P.A., Keating, C.J., Moran, L.S., Slatko, B.E., Hornstra, L.J., and Benner, J.S. (1991). Characterization and expression of the *Escherichia coli* Mrr restriction system. J. Bacteriol. *173*, 5207–5219. https://doi.org/10.1128/jb.173.16.5207-5219.1991.
39. Mulligan, E.A., and Dunn, J.J. (2008). Cloning, purification and initial characterization of *E. coli* McrA, a putative 5-methylcytosine-specific nuclease. Protein Expr. Purif. *62*, 98–103. https://doi.org/10.1016/j.pep.2008.06.016.
40. Mulligan, E.A., Hatchwell, E., McCorkle, S.R., and Dunn, J.J. (2010). Differential binding of *Escherichia coli* McrA protein to DNA sequences that contain the dinucleotide m5CpG. Nucleic Acids Res. *38*, 1997–2005. https://doi.org/10.1093/nar/gkp1120.
41. Munro, J.L., and Wieberg, J.S. (1968). Evidence that gene 56 of bacteriophage T4 is a structural gene for deoxycytidine triphosphatase. Virology *36*, 442–446. https://doi.org/10.1016/0042-6822(68)90169-4.
42. Lamm, N., Wang, Y., Mathews, C.K., and Rüger, W. (1988). Deoxycytidylate hydroxymethylase gene of bacteriophage T4. Nucleotide sequence determination and over-expression of the gene. Eur. J. Biochem. *172*, 553–563. https://doi.org/10.1111/j.1432-1033.1988.tb13925.x.
43. Kornberg, S.R., Zimmerman, S.B., and Kornberg, A. (1961). Glucosylation of deoxyribonucleic acid by enzymes from bacteriophage-infected *Escherichia coli*. J. Biol. Chem. *236*, 1487–1493. https://doi.org/10.1016/S0021-9258(18)64202-4.
44. Bryson, A.L., Hwang, Y., Sherrill-Mix, S., Wu, G.D., Lewis, J.D., Black, L., Clark, T.A., and Bushman, F.D. (2015). Covalent Modification of Bacteriophage T4 DNA Inhibits CRISPR-Cas9. mBio *6*, e00648. https://doi.org/10.1128/mBio.00648-15.
45. Vlot, M., Houkes, J., Lochs, S.J.A., Swarts, D.C., Zheng, P., Kunne, T., Mohanraju, P., Anders, C., Jinek, M., van der Oost, J., et al. (2018). Bacteriophage DNA glucosylation impairs target DNA binding by type I and II but not by type V CRISPR-Cas effector complexes. Nucleic Acids Res. *46*, 873–885. https://doi.org/10.1093/nar/gkx1264.
46. Liu, Y., Dai, L., Dong, J., Chen, C., Zhu, J., Rao, V.B., and Tao, P. (2020). Covalent Modifications of the Bacteriophage Genome Confer a Degree of Resistance to Bacterial CRISPR Systems. J. Virol. *94*, e01630–e01620. https://doi.org/10.1128/JVI.01630-20.
47. Dong, J., Chen, C., Liu, Y., Zhu, J., Li, M., Rao, V.B., and Tao, P. (2021). Engineering T4 Bacteriophage for in Vivo Display by Type V CRISPR-Cas Genome Editing. ACS Synth. Biol. *10*, 2639–2648. https://doi.org/10.1021/acssynbio.1c00251.

48. Wang, S., Sun, E., Liu, Y., Yin, B., Zhang, X., Li, M., Huang, Q., Tan, C., Qian, P., Rao, V.B., et al. (2023). Landscape of New Nuclease-Containing Antiphage Systems in *Escherichia coli* and the Counterdefense Roles of Bacteriophage T4 Genome Modifications. J. Virol. *97*, e0059923. https://doi.org/10.1128/jvi.00599-23.

49. Adler, B.A., Hessler, T., Cress, B.F., Lahiri, A., Mutalik, V.K., Barrangou, R., Banfield, J., and Doudna, J.A. (2022). Broad-spectrum CRISPR-Cas13a enables efficient phage genome editing. Nat. Microbiol. *7*, 1967–1979. https://doi.org/10.1038/s41564-022-01258-x.

50. Bair, C.L., and Black, L.W. (2007). A type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs. J. Mol. Biol. *366*, 768–778. https://doi.org/10.1016/j.jmb.2006.11.051.

51. Mahler, M., Malone, L.M., van den Berg, D.F., Smith, L.M., Brouns, S.J. J., and Fineran, P.C. (2023). An OmpW-dependent T4-like phage infects *Serratia* sp. ATCC 39006. Microb. Genom. *9*, mgen000968. https://doi.org/10.1099/mgen.0.000968.

52. Patterson, A.G., Jackson, S.A., Taylor, C., Evans, G.B., Salmond, G.P.C., Przybilski, R., Staals, R.H.J., and Fineran, P.C. (2016). Quorum Sensing Controls Adaptive Immunity through the Regulation of Multiple CRISPR-Cas Systems. Mol. Cell *64*, 1102–1108. https://doi.org/10.1016/j.molcel.2016.11.012.

53. Flaks, J.G., and Cohen, S.S. (1959). Virus-induced acquisition of metabolic function. I. Enzymatic formation of 5-hydroxymethyldeoxycytidylate. J. Biol. Chem. *234*, 1501–1506. https://doi.org/10.1016/S0021-9258(18)70038-0.

54. Wang, H., Guan, S., Quimby, A., Cohen-Karni, D., Pradhan, S., Wilson, G., Roberts, R.J., Zhu, Z., and Zheng, Y. (2011). Comparative characterization of the PvuRts1I family of restriction enzymes and their application in mapping genomic 5-hydroxymethylcytosine. Nucleic Acids Res. *39*, 9294–9305. https://doi.org/10.1093/nar/gkr607.

55. Borgaro, J.G., and Zhu, Z. (2013). Characterization of the 5-hydroxymethylcytosine-specific DNA restriction endonucleases. Nucleic Acids Res. *41*, 4198–4206. https://doi.org/10.1093/nar/gkt102.

56. Thomas, J.A., Orwenyo, J., Wang, L.X., and Black, L.W. (2018). The Odd "RB" Phage-Identification of Arabinosylation as a New Epigenetic Modification of DNA in T4-Like Phage RB69. Viruses *10*, 313. https://doi.org/10.3390/v10060313.

57. Borges, A.L., Dutton, R.J., McDaniel, E.A., Radkov, A., Reiter, T., and Weiss, E.C.P. (2023). Isolation of a Phage with an Arabinosylated Genome from a Cheese Microbial Community (Arcadia Book Company Science). https://doi.org/10.57844/arcadia-743p-ty94.

58. Pathak, A.K., Pathak, V., Suling, W.J., Riordan, J.R., Gurcha, S.S., Besra, G.S., and Reynolds, R.C. (2009). Synthesis of deoxygenated alpha(1–>5)-linked arabinofuranose disaccharides as substrates and inhibitors of arabinosyltransferases of *Mycobacterium tuberculosis*. Bioorg. Med. Chem. *17*, 872–881. https://doi.org/10.1016/j.bmc.2008.11.027.

59. Rollins, M.F., Schuman, J.T., Paulus, K., Bukhari, H.S.T., and Wiedenheft, B. (2015). Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from *Pseudomonas aeruginosa*. Nucleic Acids Res. *43*, 2216–2222. https://doi.org/10.1093/nar/gkv094.

60. Staals, R.H.J., Zhu, Y., Taylor, D.W., Kornfeld, J.E., Sharma, K., Barendregt, A., Koehorst, J.J., Vlot, M., Neupane, N., Varossieau, K., et al. (2014). RNA targeting by the type III-A CRISPR-Cas Csm complex of *Thermus thermophilus*. Mol. Cell *56*, 518–530. https://doi.org/10.1016/j.molcel.2014.10.005.

61. Samai, P., Pyenson, N., Jiang, W., Goldberg, G.W., Hatoum-Aslan, A., and Marraffini, L.A. (2015). Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. Cell *161*, 1164–1174. https://doi.org/10.1016/j.cell.2015.04.027.

62. Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, Č., and Siksnys, V. (2016). Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. Mol. Cell *62*, 295–306. https://doi.org/10.1016/j.molcel.2016.03.024.

63. Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G., and Siksnys, V. (2017). A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. Science *357*, 605–609. https://doi.org/10.1126/science.aao0100.

64. Niewoehner, O., Garcia-Doval, C., Rostøl, J.T., Berk, C., Schwede, F., Bigler, L., Hall, J., Marraffini, L.A., and Jinek, M. (2017). Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. Nature *548*, 543–548. https://doi.org/10.1038/nature23467.

65. Mayo-Muñoz, D., Smith, L.M., Garcia-Doval, C., Malone, L.M., Harding, K.R., Jackson, S.A., Hampton, H.G., Fagerlund, R.D., Gumy, L.F., and Fineran, P.C. (2022). Type III CRISPR-Cas provides resistance against nucleus-forming jumbo phages via abortive infection. Mol. Cell *82*, 4471–4486.e9. https://doi.org/10.1016/j.molcel.2022.10.028.

66. Rouillon, C., Athukoralage, J.S., Graham, S., Grüschow, S., and White, M.F. (2018). Control of cyclic oligoadenylate synthesis in a type III CRISPR system. eLife *7*, e36734. https://doi.org/10.7554/eLife.36734.

67. Hoikkala, V., Graham, S., and White, M.F. (2024). Bioinformatic analysis of type III CRISPR systems reveals key properties and new effector families. Nucleic Acids Res. *52*, 7129–7141. https://doi.org/10.1093/nar/gkae462.

68. Rostøl, J.T., and Marraffini, L.A. (2019). Non-specific degradation of transcripts promotes plasmid clearance during type III-A CRISPR-Cas immunity. Nat. Microbiol. *4*, 656–662. https://doi.org/10.1038/s41564-018-0353-x.

69. Meeske, A.J., Nakandakari-Higa, S., and Marraffini, L.A. (2019). Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. Nature *570*, 241–245. https://doi.org/10.1038/s41586-019-1257-5.

70. Maffei, E., Shaidullina, A., Burkolter, M., Heyer, Y., Estermann, F., Druelle, V., Sauer, P., Willi, L., Michaelis, S., Hilbi, H., et al. (2021). Systematic exploration of *Escherichia coli* phage-host interactions with the BASEL phage collection. PLoS Biol. *19*, e3001424. https://doi.org/10.1371/journal.pbio.3001424.

71. Ayers, B., Long, H., Sim, E., Smellie, I.A., Wilkinson, B.L., and Fairbanks, A.J. (2009). Stereoselective synthesis of beta-arabino glycosyl sulfones as potential inhibitors of mycobacterial cell wall biosynthesis. Carbohydr. Res. *344*, 739–746. https://doi.org/10.1016/j.carres.2009.02.006.

72. Suthagar, K., and Fairbanks, A.J. (2016). Synthesis and anti-mycobacterial activity of glycosyl sulfamides of arabinofuranose. Org. Biomol. Chem. *14*, 1748–1754. https://doi.org/10.1039/c5ob02317c.

73. Sakai, Y., Kimura, S., and Suzuki, T. (2019). Dual pathways of tRNA hydroxylation ensure efficient translation by expanding decoding capability. Nat. Commun. *10*, 2858. https://doi.org/10.1038/s41467-019-10750-8.

74. Kimura, S., Sakai, Y., Ishiguro, K., and Suzuki, T. (2017). Biogenesis and iron-dependency of ribosomal RNA hydroxylation. Nucleic Acids Res. *45*, 12974–12986. https://doi.org/10.1093/nar/gkx969.

75. Eidels, L., and Osborn, M.J. (1974). Phosphoheptose isomerase, first enzyme in the biosynthesis of aldoheptose in *Salmonella typhimurium*. J. Biol. Chem. *249*, 5642–5648. https://doi.org/10.1016/S0021-9258(20)79775-9.

76. Bateman, A. (1999). The SIS domain: a phosphosugar-binding domain. Trends Biochem. Sci. *24*, 94–95. https://doi.org/10.1016/s0968-0004(99)01357-2.

77. Sommaruga, S., Gioia, L.D., Tortora, P., and Polissi, A. (2009). Structure prediction and functional analysis of KdsD, an enzyme involved in lipopolysaccharide biosynthesis. Biochem. Biophys. Res. Commun. *388*, 222–227. https://doi.org/10.1016/j.bbrc.2009.07.154.

78. Cook, R., Brown, N., Redgwell, T., Rihtman, B., Barnes, M., Clokie, M., Stekel, D.J., Hobman, J., Jones, M.A., and Millard, A. (2021). INfrastructure for a PHAge REference Database: Identification of Large-Scale Biases in the Current Collection of Cultured Phage Genomes. Phage (New Rochelle) *2*, 214–223. https://doi.org/10.1089/phage.2021.0007.

79. Camargo, A.P., Nayfach, S., Chen, I.A., Palaniappan, K., Ratner, A., Chu, K., Ritter, S.J., Reddy, T.B.K., Mukherjee, S., Schulz, F., et al. (2023). IMG/VR v4: an expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata. Nucleic Acids Res. *51*, D733–D743. https://doi.org/10.1093/nar/gkac1037.

80. Picton, D.M., Luyten, Y.A., Morgan, R.D., Nelson, A., Smith, D.L., Dryden, D.T.F., Hinton, J.C.D., and Blower, T.R. (2021). The phage defence island of a multidrug resistant plasmid uses both BREX and type IV restriction for complementary protection from viruses. Nucleic Acids Res. *49*, 11257–11273. https://doi.org/10.1093/nar/gkab906.

81. Payne, L.J., Hughes, T.C.D., Fineran, P.C., and Jackson, S.A. (2024). New antiviral defences are genetically embedded within prokaryotic immune systems. Preprint at bioRxiv. https://doi.org/10.1101/2024.01.29.577857.

82. Hossain, A.A., Pigli, Y.Z., Baca, C.F., Heissel, S., Thomas, A., Libis, V.K., Burian, J., Chappie, J.S., Brady, S.F., Rice, P.A., et al. (2024). DNA glycosylases provide antiviral defence in prokaryotes. Nature *629*, 410–416. https://doi.org/10.1038/s41586-024-07329-9.

83. Aframian, N., and Eldar, A. (2023). Abortive infection antiphage defense systems: separating mechanism and phenotype. Trends Microbiol. *31*, 1003–1012. https://doi.org/10.1016/j.tim.2023.05.002.

84. Kchromov, I.S., Sorotchkina, V.V., Nigmatullin, T.G., and Tikchonenko, T. I. (1980). A new nitrogen base 5-hydroxycytosine in phage N-17 DNA. FEBS Lett. *118*, 51–54. https://doi.org/10.1016/0014-5793(80)81216-6.

85. Carreras, C.W., and Santi, D.V. (1995). The catalytic mechanism and structure of thymidylate synthase. Annu. Rev. Biochem. *64*, 721–762. https://doi.org/10.1146/annurev.bi.64.070195.003445.

86. Song, H.K., Sohn, S.H., and Suh, S.W. (1999). Crystal structure of deoxycytidylate hydroxymethylase from bacteriophage T4, a component of the deoxyribonucleoside triphosphate-synthesizing complex. EMBO J. *18*, 1104–1113. https://doi.org/10.1093/emboj/18.5.1104.

87. Brandt, D., Dörrich, A.K., Persicke, M., Leonhard, T., Haak, M., Nölting, S., Ruwe, M., Schmid, N., Thormann, K.M., and Kalinowski, J. (2024). Discovery of a pentose as a cytosine nucleobase modification in *Shewanella* phage Thanatos-1 genomic DNA mediating enhanced resistance towards host restriction systems. Preprint at bioRxiv. https://doi.org/10.1101/2024.02.27.582347.

88. Yaung, S.J., Esvelt, K.M., and Church, G.M. (2014). CRISPR/Cas9-mediated phage resistance is not impeded by the DNA modifications of phage T4. PLoS One *9*, e98811. https://doi.org/10.1371/journal.pone.0098811.

89. Rodriguez-Rodriguez, L., Pfister, J., Schuck, L., Martin, A.E., Mercado-Santiago, L.M., Tagliabracci, V.S., and Forsberg, K.J. (2025). Metagenomic selections reveal diverse antiphage defenses in human and environmental microbiomes. Preprint at bioRxiv. https://doi.org/10.1101/2025.02.28.640651.

90. Yee, W.-X., Lee, Y.-J., Klein, T.A., Wirganowicz, A., Gabagat, A.E., Csörgő, B., Makarova, K.S., Koonin, E.V., Weigele, P.R., and Bondy-Denomy, J. (2025). END nucleases: Antiphage defense systems targeting multiple hypermodified phage genomes. Preprint at bioRxiv. https://doi.org/10.1101/2025.03.31.646159.

91. Liu, M., Hernandez-Morales, A., Clark, J., Le, T., Biswas, B., Bishop-Lilly, K.A., Henry, M., Quinones, J., Voegtly, L.J., Cer, R.Z., et al. (2022). Comparative genomics of Acinetobacter baumannii and therapeutic bacteriophages from a patient undergoing phage therapy. Nat. Commun. *13*, 3776. https://doi.org/10.1038/s41467-022-31455-5.

92. Durfee, T., Nelson, R., Baldwin, S., Plunkett, G., 3rd, Burland, V., Mau, B., Petrosino, J.F., Qin, X., Muzny, D.M., Ayele, M., et al. (2008). The complete genome sequence of Escherichia coli DH10B: insights into the biology of a laboratory workhorse. J. Bacteriol. *190*, 2597–2606. https://doi.org/10.1128/JB.01695-07.

93. Thoma, S., and Schobert, M. (2009). An improved Escherichia coli donor strain for diparental mating. FEMS Microbiol. Lett. *294*, 127–132. https://doi.org/10.1111/j.1574-6968.2009.01556.x.

94. Jackson, S.A., Fellows, B.J., and Fineran, P.C. (2020). Complete Genome Sequences of the *Escherichia coli* Donor Strains ST18 and MFDpir. Microbiol. Resour. Announc. *9*, e01014–e01020. https://doi.org/10.1128/MRA.01014-20.

95. Thomson, N.R., Crow, M.A., McGowan, S.J., Cox, A., and Salmond, G.P. (2000). Biosynthesis of carbapenem antibiotic and prodigiosin pigment in Serratia is under quorum sensing control. Mol. Microbiol. *36*, 539–556. https://doi.org/10.1046/j.1365-2958.2000.01872.x.

96. Malone, L.M., Hampton, H.G., Morgan, X.C., and Fineran, P.C. (2022). Type I CRISPR-Cas provides robust immunity but incomplete attenuation of phage-induced cellular stress. Nucleic Acids Res. *50*, 160–174. https://doi.org/10.1093/nar/gkab1210.

97. Yeh, L.S., Hsu, T., and Karam, J.D. (1998). Divergence of a DNA replication gene cluster in the T4-related bacteriophage RB69. J. Bacteriol. *180*, 2005–2013. https://doi.org/10.1128/JB.180.8.2005-2013.1998.

98. Yoon, S.H., Ha, S.M., Lim, J., Kwon, S., and Chun, J. (2017). A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Leeuwenhoek *110*, 1281–1286. https://doi.org/10.1007/s10482-017-0844-4.

99. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

100. Gilchrist, C.L.M., and Chooi, Y.H. (2021). Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. Bioinformatics *37*, 2473–2475. https://doi.org/10.1093/bioinformatics/btab007.

101. Holm, L., Laiho, A., Törönen, P., and Salgado, M. (2023). DALI shines a light on remote homologs: One hundred discoveries. Protein Sci. *32*, e4519. https://doi.org/10.1002/pro.4519.

102. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780. https://doi.org/10.1093/molbev/mst010.

103. Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321. https://doi.org/10.1093/sysbio/syq010.

104. Meier-Kolthoff, J.P., and Göker, M. (2017). VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics *33*, 3396–3404. https://doi.org/10.1093/bioinformatics/btx440.

105. Eddy, S.R. (2011). Accelerated Profile HMM Searches. PLoS Comput. Biol. *7*, e1002195. https://doi.org/10.1371/journal.pcbi.1002195.

106. Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113. https://doi.org/10.1186/1471-2105-5-113.

107. Team, R.C. (2021). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

108. Wickham, H., and Bryan, J. (2025). readxl: Read Excel Files. R package version 1.4.5. https://github.com/tidyverse/readxl.

109. Wickham, H. (2023). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.5.1. https://github.com/tidyverse/stringr.

110. Wickham, H., François, R., Henry, L., Müller, K., and Vaughan, D. (2023). dplyr: A Grammar of Data Manipulation. R package version 1.1.4. https://dplyr.tidyverse.org.

111. Wickham, H., Vaughan, D., and Girlich, M. (2024). tidyr: Tidy Messy Data. R package version 1.3.1. https://CRAN.R-project.org/package=tidyr.

112. Winter, D., Chamberlain, S., and Guangchun, H. (2020). Rentrez: 'Entrez' in R. https://cran.r-project.org/web/packages/rentrez/.

113. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag New York).

114. Wilkins, D. (2020). gggenes: Draw Gene Arrow Maps in 'ggplot2'. R package version 0.4.1. https://wilkox.org/gggenes/.

115. Slowikowski, K. (2024). ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. https://github.com/slowkow/ggrepel.

116. Wilke, C. (2024). cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version 1.1.3. https://wilkelab.org/cowplot/.

117. Kolde, R. (2018). pheatmap: Pretty heatmaps. R package version 1.0.12.

118. Pagès, H., Aboyoun, P., Gentleman, R., and DebRoy, S. (2025). Biostrings: efficient manipulation of biological strings. R package version 2.76.0.

119. Paradis, E., and Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics *35*, 526–528. https://doi.org/10.1093/bioinformatics/bty633.

120. Wang, L.G., Lam, T.T.Y., Xu, S., Dai, Z., Zhou, L., Feng, T., Guo, P., Dunn, C.W., Jones, B.R., Bradley, T., et al. (2020). Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. Mol. Biol. Evol. *37*, 599–603. https://doi.org/10.1093/molbev/msz240.

121. Xu, S., Li, L., Luo, X., Chen, M., Tang, W., Zhan, L., Dai, Z., Lam, T.T., Guan, Y., and Yu, G. (2022). Ggtree: A serialized data object for visualization of a phylogenetic tree and annotation data. Imeta *1*, e56. https://doi.org/10.1002/imt2.56.

122. Kluyver, T., Ragan-Kelley, B., Perez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., et al. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. In Positioning and Power in Academic Publishing: Players, Agents and Agendas, F. Loizides and B. Schmidt, eds., pp. 87–90.

123. Kluyver, T., Angerer, P., Schulz, J., and Ram, K. (2025). IRkernel: Native R Kernel for the 'jupyter notebook'. R package version 1.3.2.9000. https://github.com/irkernel/irkernel.

124. Payne, L.J., Todeschini, T.C., Wu, Y., Perry, B.J., Ronson, C.W., Fineran, P.C., Nobrega, F.L., and Jackson, S.A. (2021). Identification and classification of antiviral defence systems in bacteria and archaea with PADLOC reveals new system types. Nucleic Acids Res. *49*, 10868–10878. https://doi.org/10.1093/nar/gkab883.

125. Bodenhofer, U., Bonatesta, E., Horejš-Kainrath, C., and Hochreiter, S. (2015). msa: an R package for multiple sequence alignment. Bioinformatics *31*, 3997–3999. https://doi.org/10.1093/bioinformatics/btv494.

126. Duprey, A., Taib, N., Leonard, S., Garin, T., Flandrois, J.P., Nasser, W., Brochier-Armanet, C., and Reverchon, S. (2019). The phytopathogenic nature of *Dickeya aquatica* 174/2 and the dynamic early evolution of *Dickeya* pathogenicity. Environ. Microbiol. *21*, 2809–2835. https://doi.org/10.1111/1462-2920.14627.

127. Borges, A.L. (2024). Propagating *Serratia* phage 92A1 and host. protocols.io. https://doi.org/10.17504/protocols.io.81wgbxb3qlpk/v1.

128. Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059–3066. https://doi.org/10.1093/nar/gkf436.

129. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649. https://doi.org/10.1093/bioinformatics/bts199.

130. Lefort, V., Longueville, J.E., and Gascuel, O. (2017). SMS: Smart Model Selection in PhyML. Mol. Biol. Evol. *34*, 2422–2424. https://doi.org/10.1093/molbev/msx149.

131. Schrödinger, L. (2015). The PyMOL Molecular Graphics System, [Version 1.8].

132. Del Sal, G., Manfioletti, G., and Schneider, C. (1989). The CTAB-DNA precipitation method: a common mini-scale preparation of template DNA from phagemids, phages or plasmids suitable for sequencing. BioTechniques *7*, 514–520.

133. Sambrook, J., and Russell, D.W. (2001). Molecular Cloning: a Laboratory Manual, Third Edition (Cold Spring Harbor Laboratory Press).

134. Thiaville, J.J., Kellner, S.M., Yuan, Y., Hutinet, G., Thiaville, P.C., Jumpathong, W., Mohapatra, S., Brochier-Armanet, C., Letarov, A.V., Hillebrand, R., et al. (2016). Novel genomic island modifies DNA with 7-deazaguanine derivatives. Proc. Natl. Acad. Sci. USA *113*, E1452–E1459. https://doi.org/10.1073/pnas.1518570113.

135. Hampton, H.G., McNeil, M.B., Paterson, T.J., Ney, B., Williamson, N.R., Easingwood, R.A., Bostina, M., Salmond, G.P.C., and Fineran, P.C. (2016). CRISPR-Cas gene-editing reveals RsmA and RsmC act through FlhDC to repress the SdhE flavinylation factor and control motility and prodigiosin production in *Serratia*. Microbiology (Reading) *162*, 1047–1058. https://doi.org/10.1099/mic.0.000283.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
AbaSI	NEB	Cat # R0665
Acetonitrile	Fisher Scientific	Cat #A9561
5-Aminolevulinic acid (ALA)	ACROS ORGANICS	Cat # 103920050
AMPure XP Reagent	Beckman Coulter	Cat # A63880
c-triAMP (cA_3)	Biolog Life Science Institute	Cat # C 362-005
D-Arabinose	Sigma	Cat # A3131-100G
Benzonase Nuclease	Sigma	Cat #E8263
Hexadecyltrimethylammonium bromide (CTAB)	Sigma	Cat # 52365-50G
DNase I	Roche Diagnostics	Cat #10104159001
Ethylenediaminetetra-acetic acid (EDTA)	Ajax Finechem	Cat # AJA180-500G
Formic acid	Fisher Scientific	Cat #A117
L-Fucose	Biosynth	Cat # MF06710
D-Galactose	Biosynth	Cat # MG05201
D-Glucose	Sigma	Cat # G5767-5KG
D-Mannose	Biosynth	Cat # MM06704
MspI	NEB	Cat # R0106L
N-Acetyl-D-galactosamine	Biosynth	Cat # MA04390
N-Acetyl-D-glucosamine	Biosynth	Cat # MA00834
NucC	Mayo-Muñoz et al.[65]	N/A
Phenol/Chloroform/Isoamyl alcohol (25:24:1)	Thermo Scientific	Cat # 327111000
Phosphatase	Sigma	Cat #P5521
Phosphodiesterase	Sigma	Cat #P3243
Proteinase K	Thermo Scientific	Cat # EO0492
D-Ribose	Sigma Aldrich	Cat # R1757-100G
RNase A	Roche Diagnostics	Cat #10109142001
Sodium Dodecyl Sulfate (SDS)	Invitrogen™	Cat #15525-017
Sodium Hydroxide	Supelco	Cat #1064821000
Trifluoroacetic acid (TFA)	Merck	Cat # MER302031
D-Xylose	Sigma Aldrich	Cat # X1500-500G
Critical commercial assays		
GFX™ PCR DNA and Gel Band Purification Kit	Cytiva	Cat # GE28-9034-71
Qubit™ dsDNA BR Assay Kit	Invitrogen™	Cat # Q32850
Qubit™ dsDNA HS Assay Kit	Invitrogen™	Cat # Q32851
Zyppy™ Plasmid Miniprep Kit	Zymo Research	Cat # D4020
Experimental models: Organisms/strains		
Bacteria		
Acinetobacter baumannii propagation host for phage Maestro	Liu et al.[91]	TP1
E. coli test strain for Bas46 assays	Durfee et al.[92]	DH10β
E. coli propagation host for phage Bas46/47	Maffei et al.[70]	K-12 MG1655 ΔRM
E. coli Auxotrophic donor for biparental conjugation requires ALA	Thoma and Schobert[93] and Jackson et al.[94]	ST18

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Serratia sp. ATCC 39006 Lac- EMS mutant, denoted WT	Thomson et al.[95]	LacA
Serratia sp. ATCC 39006 Δ*nucC* mutant, LacA-derivative	Malone et al.[22]	PCF686
Serratia sp. ATCC 39006 Cas10 HD domain mutant, LacA-derivative	Malone et al.[22]	PCF690
Serratia sp. ATCC 39006 Cas10 Palm domain mutant, LacA-derivative	Malone et al.[22]	PCF691
Serratia sp. ATCC 39006 Type III-A dCas7 mutant, LacA-derivative	This study	PCF868
Serratia strain 92 Propagation host for *Serratia* phage 92A1	Borges[57]	Strain 92
Bacteriophages		
Serratia sp. ATCC 39006 phage LC53	Mahler et al.[51]	LC53
Serratia sp. ATCC 39006 phage LC53 *ara iso* mutant 8 (11 bp deletion (15 bp after start) in arabinose isomerase)	This work	N/A
Serratia sp. ATCC 39006 phage JS26	Malone et al.[96]	JS26
Serratia sp. ATCC 39006 phage PCH45	Malone et al.[22]	PCH45
Serratia strain 92 phage	Borges[57]	92A1
E. coli phage T4	Fagenbank	T4
E. coli phage RB69	Yey et al.[97]	RB69
E. coli phage Bas46	Maffei et al.[70]	Bas46
E. coli phage Bas46 Δ*aat*	This work	Bas46 Δ*aat*
E. coli phage Bas47	Maffei et al.[70]	Bas47
Acinetobacter phage Maestro	Jason Gill; unpub.	Maestro
Oligonucleotides		
See Table S6	IDT™	N/A
Recombinant DNA		
See Table S7	N/A	N/A
Software and algorithms		
Alphafold3	AlphaFold Server	https://alphafoldserver.com/
ANI calculator	Yoon et al.[98]	https://www.ezbiocloud.net/tools/ani
BLAST	Altschul et al.[99]	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Chromeleon™ 7 Chromatography Data System (CDS) Software	ThermoFisher Scientific	Version 7.2.7 October 2017; https://www.thermofisher.com/order/catalog/product/CHROMELEON7
Clinker	Gilchrist and Chooi[100]	https://cagecat.bioinformatics.nl/
Dali server	Holm et al.[101]	http://ekhidna2.biocenter.helsinki.fi/dali/
Geneious Prime	Dotmatics	https://www.geneious.com/
MAFFT	Katoh and Standley[102]	https://mafft.cbrc.jp/alignment/server/index.html
MassHunter Qualitative Analysis software	Agilent Technologies	https://www.agilent.com/en/product/software-informatics/mass-spectrometry-software/data-analysis/qualitative-analysis
PhyML	Guindon et al.[103]	http://www.atgc-montpellier.fr/phyml/
Prism 10.5.0	GraphPad	https://www.graphpad.com/
PyMOL	Schrödinger	https://www.pymol.org/
VICTOR	Meier-Kolthoff and Göker[104]	https://ggdc.dsmz.de/victor.php
HMMer 3.4	Eddy[105]	http://hmmer.org/
MUSCLE 5.3	Edgar[106]	https://drive5.com
R 4.4.2	R Core Team[107]	https://www.r-project.org/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Readxl	Wickham and Bryan[108]	https://readxl.tidyverse.org/
stringr	Wickham[109]	https://stringr.tidyverse.org/
dplyr	Wickham et al.[110]	https://dplyr.tidyverse.org/
tidyr	Wickham et al.[111]	https://tidyr.tidyverse.org/
rentrez	Winter et al.[112]	https://cran.r-project.org/web/packages/rentrez/index.html
ggplot2	Wickham[113]	https://ggplot2.tidyverse.org/
gggenes	Wilkins[114]	https://cran.r-project.org/web/packages/gggenes/index.html
ggrepel	Slowikowski[115]	https://ggrepel.slowkow.com/
cowplot	Wilke[116]	https://cran.r-project.org/web/packages/cowplot/index.html
pheatmap	Kolde[117]	https://cran.r-project.org/web/packages/pheatmap/index.html
BioStrings	Pagès et al.[118]	https://bioconductor.org/packages/release/bioc/html/Biostrings.html
ape	Paradis and Schliep[119]	https://cran.r-project.org/web/packages/ape/index.html
TreeIO	Wang et al.[120]	https://www.bioconductor.org/packages/release/bioc/html/treeio.html
ggTree	Xu et al.[121]	https://bioconductor.org/packages/release/bioc/html/ggtree.html
Jupyter	Kluyver et al.[122]	https://jupyter.org/
IRkernel	Kluyver et al.[123]	https://cran.r-project.org/web/packages/IRkernel/index.html
Padloc	Payne et al.[124]	https://padloc.otago.ac.nz/padloc/
msa	Bodenhofer et al.[125]	https://www.bioconductor.org/packages/release/bioc/html/msa.html

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Bacteria, phages, primers, plamids and growth conditions

Bacterial strains and phages used and generated in this study are summarised in the key resources table. *Serratia* sp. ATCC 39006 was recently classified as *Prodigiosinella confusarubida*,[126] but for consistency reasons with previous papers about the strain and the phages, we use the name *Serratia* in this work. Primers and plasmids used in this study are listed in Tables S6 and S7. Plasmids constructed were confirmed by Sanger sequencing, introduced into *E. coli* ST18 by heat shock transformation and subsequently conjugated into *Serratia* strains. Plasmids used for experiments in *E. coli* were introduced into the appropriate strain by heat shock transformation. *Serratia* strains were grown in lysogeny broth (LB) at 30°C, under shaking conditions (160 rpm) or on LB agar (LBA, 1.5% w/v) plates and incubated at 30°C. *E. coli* and *A. baumanii* were grown at 37°C but otherwise under the same conditions as *Serratia*. LBA overlays were prepared with 0.5% (w/v) agar unless otherwise stated. LB and LBA were supplemented with ampicillin (Ap, 100 μg/ml), chloramphenicol (Cm, 25 μg/ml), kanamycin (Km, 50 μg/ml), isopropyl-β-D-thiogalactopyranoside (IPTG, 0.1 mM), arabinose (Ara, 0.01-0.2% w/v) and 5-aminolevulinic acid (Ala, 50 μg/ml), when required.

Preparation of phage stocks and titration

Phage stocks were typically prepared as follows. A 5 ml culture of *Serratia* or *E. coli* was grown overnight, sub-cultured (100 μl) into 50 ml LB and then grown to an OD_{600} of ~0.2. The culture was infected with 10 μl of high titre (~5×10^{10} plaque forming units per ml (pfu/ml)) phage stock and grown again overnight. Phages were harvested by removing cell debris by centrifugation (3000 *g*, 20 min) and the phage-containing supernatant filtered (0.22 μm pore size). Phages were titrated by preparing a ten-fold serial dilution in LB or phage buffer (Tris Base (10 mM, pH 7.4), $MgSO_4$ (10 mM), gelatine (0.01% w/v)) and spotting 5 μl onto top agar overlays seeded with 100 μl *Serratia* overnight culture. Plates were incubated at 30°C overnight and plaques counted to calculate the phage titre as pfu/ml. If required, the phage stock was prepared by the double agar overlay method. Briefly, 4 ml of LB top agar containing 100 μl of bacterial overnight culture and 10 μl phage stock, at a concentration to produce semi-confluent lysis, were poured onto LBA plates and incubated overnight. Phages were harvested by pooling the overlays into a centrifuge tube, adding few drops of chloroform and cells lysed by vortexing. Cell debris was then removed by centrifugation (2000 *g*, 20 min) and the phages transferred into a sterile glass

universal. *Serratia* phage 92A1 was propagated by following the protocol provided by Arcadia Science.[127] *E. coli* phages (T4, RB69, Bas46 and Bas47) were replicated on DH10β, DH5α or K-12 ΔRM, *Serratia* phages LC53, PCH45 and JS26 were replicated on LacA, *Serratia* phage 92A1 was replicated on *Serratia* strain 92 and *A. baumanii* phage Maestro was replicated on TP1 unless otherwise stated. All phage stocks were stored at 4°C.

METHOD DETAILS

Bioinformatic analysis of modification genes

Whole phage genomes and specific loci were compared by visualising the similarity between different phages with the clinker tool for cluster comparison by setting the minimum alignment sequence identity to 0.3.[100] Identity of homologous DNA modification genes was investigated by protein BLAST (v2.12.0)[99] and MAFFT (v7) alignments (FFT_NS-i x1000 algorithm and JTT200 scoring matrix) of the appropriate amino acid sequences.[102,128] The phylogeny of diverse phage encoded thymidylate synthases and arabinosyl-ara-hC transferases was investigated by performing MAFFT (v7) (FFT_NS-i x1000 algorithm and JTT200 scoring matrix) alignments of the amino acid sequences in Geneious Prime (v2023.1.2)[102,128,129] and building a phylogenetic tree based on this alignment with PhyML (v3.0) using the Smart Model Selection.[103,130] Accession numbers of protein sequences used to build the phylogenetic trees of the thymidylate synthases and arabinosyl-ara-hC transferase are listed in Tables S2 and S4, respectively. Phylogenetic trees of phages based on whole genomes were built with VICTOR.[104] Accession numbers of phage genome sequences used to build the phylogenetic trees are listed in Table S3. Structures of selected phage proteins were predicted with AlphaFold3 (AlphaFold Server Beta) and the predicted structure was compared to existing protein structures with the DALI server.[101] Predicted protein structures were visualised in PyMOL Molecular Graphics System (v3.0, Schrödinger, LLC).[131] Genomic similarities between two phage genomes were calculated as the Orthologous Average Nucleotide Identity with USEARCH (OrthoANIu) with the ANI calculator.[98]

Phage genomic DNA extraction

Phage genomic DNA (gDNA) was either extracted by adapting the CTAB (cetyltrimethylammonium bromide) or the phenol-chloroform method.[132,133] Both methods required high-titre phage stocks with at least 10^{10} pfu/ml. For the CTAB method 2.5 ml of phage stock were mixed with RNase A (100 ng) and DNase I (100 U) and incubated at 37°C for 30 min. The nucleases were inactivated by the addition of 40 mM EDTA (pH 8.0) and the phage particles were degraded by incubation at 56°C for 20 min with 0.5 mg proteinase K. Next, 220 μl of 10% (w/v) CTAB in 4% (w/v) NaCl at 55°C were added. The DNA-CTAB complex was precipitated by cooling on ice for 15 min and pelleted by centrifugation (4000 *g* at 4°C for 5 min). The DNA was resuspended in 1.2 M NaCl and precipitated at -20°C overnight with 2 volumes of isopropanol. Precipitated DNA was concentrated by centrifugation (20,000 *g* at 4°C for 10 min) and washed twice with 500 μl 75% ethanol (centrifugation at 13,000 *g* at 4°C for 5 min). After the washing, the DNA was resuspended in 60 μl ultra-pure water.

For the phenol-chloroform method, 5 ml of phage stock were treated with DNase I and RNase A (1 μg/ml each) and incubated for 30 min at room temperature (RT). The reaction was stopped, and virions degraded by the addition of 0.5 M EDTA at pH 8.0 (20 mM), proteinase K (50 μg/ml) and SDS (0.5% w/v) and incubation at 56°C for at least one hour. Samples were cooled to RT before the DNA was extracted by adding an equal volume of phenol-chloroform-isoamyl-alcohol. The samples were centrifuged (3000 *g* for 10 min) and the clear supernatant was transferred to a new tube where an equal volume of chloroform was added. This step was repeated after centrifugation (3000 *g* for 10 min). The DNA obtained after two extractions with chloroform was precipitated by addition of 3 M sodium acetate at pH 5.0 (0.1 volume) and absolute ethanol (2.5 volumes). The precipitated DNA was stored at -20°C overnight and subsequently concentrated by centrifugation (14,000 *g* for 15 min). The pellet was washed in 70% ethanol and centrifuged (14,000 *g* for 5 min). The pellet was dried at RT for 5-10 min before the DNA was resuspended in 20-50 μl ultra-pure water.

DNA concentration and purity was assessed by nanodrop (NanoDrop One Microvolume UV-Vis Spectrophotometer) and by fluorimetry using the Qubit dsDNA BR or HS kit (Qubit 4 Fluorometer). One μl of phage DNA was run on a 1% agarose gel prepared in 1×TAE (Tris-acetate-EDTA) buffer and stained with ethidium bromide (EtBr) to check the quality of the DNA. If required for the subsequent analysis, the extracted phage DNA was further purified with AMPure XP beads (Beckman Coulter Life Sciences) following the manufacturer's instructions and upscaling the purification procedure depending on input and desired quantity of pure DNA. The purified DNA was eluted in ultra-pure water, and concentration and purity were again assessed by nanodrop and Qubit. The DNA was stored at 4°C for short term, at -20°C for long term or dried by speed vacuuming (Eppendorf Concentrator 5301 centrifugal vacuum concentrator) at 45°C if required.

Restriction digests of phage DNA

Phage DNA was quantified using a Qubit and approximately 100 ng used in restriction digests (25 μl) with MspI or AbaSI (both from NEB) according to the manufacturer's instructions. Water was added to the control samples instead of the enzyme. Digests were incubated and enzymes heat inactivated at the temperatures and for the duration suggested by the manufacturer. Lastly, 15 μl of the reaction were run at 100 V for 40 min on a 1% agarose gel prepared in 1×TAE buffer and stained with EtBr.

HPAE-PAD analysis of phage DNA

Previously extracted and purified phage DNA was hydrolysed and subjected to high-pH anion exchange chromatography with pulsed amperometric detection (HPAE-PAD) analysis as described previously.[56] Briefly, 15 μg of phage DNA were resuspended

in 300 μl sterile water and an equal volume of 4 M trifluoroacetic acid (TFA) was added. The mixture was heated to 100°C for 3 h to release the monosaccharides from the phage DNA, subsequently cooled to RT, lyophilised, and resuspended in 100 μl sterile water. Two sets of standards were prepared from commercially available monosaccharides (a mix of 6 hexoses – L-fucose, N-acetyl-D-galactosamine (GalNAc), N-acetyl-D-glucosamine (GlcNAc), D-galactose, D-glucose and D-mannose and a mix of 3 pentoses -D-arabinose, D-xylose and D-ribose). The stock solutions (100 nmol per monosaccharide in 10 ml sterile water) were treated with TFA (similar procedure as the phage DNA sample) to rule out any effect of the presence of residual acid in HPAE-PAD. The samples (phage DNA and standards) were subjected to HPAE-PAD analysis on a Dionex ICS-5000$^+$ DC Ion Chromatography system (software Chromeleon™ 7, version 7.2.7) with Gold, Carboquad (Carbohydrates, Quad potential) Waveform. The analysis was run as 12-100% C H20G2H5 (translates to isocratic 12 mM NaOH for 20 mins then a gradient of 12 to 100 mM NaOH for 2 min followed by hold at 100 mM for 5 mins) on a CarboPac™ PA 20 column and Dionex Amino Trap guard column at 30°C column and 15°C tray temperature and with a flow rate of 0.5 ml/min. Water was used as solvent A and 100 nM NaOH as solvent B. The reference electrode mode was set to AgCl and the used injection volume for the hydrolysed samples was 100 μl.

LC-MS analysis of purified phage DNA

Liquid chromatography-coupled mass spectrometry (LC-MS) analysis of phage DNA was performed following the previously published procedure with some adjustments.[134] Briefly, purified DNA (10 μg) was hydrolysed in 10 mM Tris-HCl (pH 7.9) with 1 mM $MgCl_2$ with Benzonase (20U), DNase I (4U), calf intestine phosphatase (17U) and phosphodiesterase (0.2U) for 16 h at ambient temperature. Following passage through a 10 kDa filter to remove proteins, the filtrate was analysed by LC-MS.

The LC-MS analysis was performed using Agilent 1290 ultrahigh pressure liquid chromatography system equipped with DAD and 6550 QTOF mass detector managed by a MassHunter workstation. The column used for the separation was a Waters ACQUITY HSS T3 column (2.1×150 mm, 1.8 μm). The oven temperature was set at 45°C. The gradient elution involved a mobile phase consisting of (A) 0.1% formic acid in water and (B) 0.1% formic acid in acetonitrile. The initial condition was set at 2% B. A 15 min linear gradient to 7% B was applied, followed by a 5 min gradient to 100% B which was held for 3 min, then returned to starting conditions over 0.5 min. Flow rate was set at 0.3 ml/min, and 2 μl of samples was injected. The electrospray ionisation mass spectra were acquired in positive ion mode. Mass data were collected between *m/z* 100 and 1000 at a rate of two scans per second. The electrospray ionization of the mass spectrometer was performed in positive ion mode with the following source parameters: drying gas temperature 250°C with a flow of 14 l/min, nebulizer gas pressure 40 psi, sheath gas temperature 350°C with a flow of 11 l/min, capillary voltage 3,500 V and nozzle voltage 500 V. Two reference masses were continuously infused to the system to allow constant mass correction during the run: *m/z* 121.0509 ($C_5H_4N_4$) and *m/z* 922.0098 ($C_{18}H_{18}O_6N_3P_3F_{24}$). Raw spectrometric data were analysed by MassHunter Qualitative Analysis software (Agilent Technologies, US) and the molecular features characterized by retention time, chromatographic peak intensity and accurate mass, were obtained by using the Molecular Feature Extractor algorithm.

NMR analysis of 5-arabinofuranosyl-hydroxy-deoxycytidine (5ara-hdC)

Phage LC53 DNA sample (2.5 mg) was enzymatically hydrolyzed and 5ara-hC was purified using HPLC. Purified 5ara-hdC was dissolved in DMSO-d_6 and recorded 1D and 2D ^{1}H NMR on a Bruker 600 MHz. Due to the low amount of compound, several impurities were seen along with the 5ara-hdC signals.

^{1}H NMR

H6 proton was seen as a singlet at 7.56 ppm and the anomeric proton of deoxyribose, H1′ was observed at 6.15 ppm. The five OH protons and anomeric proton of arabinose (H1a) was observed between 5.90-4.90 ppm. Protons of 2' CH_2 were observed between 1.90-2.10 ppm, remaining 9 protons were observed between 4.30-3.30 ppm. We found a triplet at 5.33 ppm with integration value of 1.53 protons, which could come from impurity.

Deuterium exchange ^{1}H NMR

We did a D_2O exchange and recorded ^{1}H NMR. Five protons between 5.90-4.90 ppm disappeared, confirming the presence of five OH protons. Triplet at 5.33 and a doublet at 5.14 ppm remained in the spectrum. To assign anomeric proton of arabinose (H1a) unambiguously, we recovered the sample from DMSO-D_6 and re-purified it using HPLC. When we recorded NMR for re-purified 5ara-hdC, the integral value of triplet at 5.33 ppm increased to 10 protons, which further confirms that it is from an impurity. Thus, the chemical shift of H1a proton should be at 5.14 ppm.

COSY ^{1}H NMR

The anomeric proton of deoxyribose (H1') has clear correlations to H2' protons (1.90-2.10 ppm). The H2' protons have a correlation to proton H3' (4.22 ppm) which has correlation to 3'-OH proton at 5.18 ppm. The H3' proton at 4.22 ppm has another correlation to H4' proton at 3.76 ppm, which in turn has correlation to the two H5' protons at 3.58-3.51 ppm. H5' protons have a correlation to 5'-OH proton at 5.00 ppm. We could not observe clear correlations between H1a and H2a protons and thus the remaining arabinose protons could not be assigned with confidence.

NOESY ^{1}H NMR

The aromatic proton H6 has NOE correlation to the anomeric protons H1' and H1a. The OH4a proton at 4.94 ppm has a strong correlation to OH3a proton 5.47 ppm, and the OH3a proton has a strong correlation to H3a proton at 4.1 ppm.

HSQC NMR

We couldn't record ^{13}C NMR due to limited amount of sample, however we run HSQC spectrum. The carbon C5 at 129.3 ppm, the anomeric carbon of deoxyribose at 85.4 ppm, and the anomeric carbon of arabinose at 109.4 were observed. ^{1}H NMR (600 MHz,

DMSO-d_6) δ (ppm) 7.56 (s, 1H), 6.15 (q, J = 4.47 Hz, 1H), 5.55 (d, J = 4.8 Hz, 1H), 5.47 (d, J = 4.2 Hz, 1H), 5.18 (d, J = 4.2 Hz, 1H), 5.14 (d, J = 0.6 Hz, 1H), 5.00 (t, J = 4.8 Hz, 1H), 4.94 (t, J = 5.4 Hz, 1H), 4.22-4.18 (m, 1H), 4.11-4.09 (m, 1H), 3.97-3.94 (m, 1H), 3.82-3.80 (m, 1H), 3.76-3.75 (m, 1H), 3.58-3.51 (m, 3H), 3.46-3.43 (m, 1H), 2.10-2.07 (m, 1H), 2.02-1.96 (m, 1H).

NMR analysis of 5-arabinofuranosyl-arabinofuranosyl-hydroxy-deoxycytidine (5ara-ara-hdC)

Phage Bas46 DNA sample (2.5 mg) was enzymatically hydrolyzed and 5ara-ara-hdC was purified using HPLC. Purified 5ara-ara-hdC was dissolved in DMSO-d_6 and recorded 1D and 2D ^{1}H NMR on a Bruker 600 MHz. Due to low amount of the compound several impurities were seen along with the 5ara-ara-hdC signals.

^{1}H NMR

H6 proton was seen as a singlet peak at 7.80 ppm and the anomeric proton of deoxyribose H1′ was observed at 6.15 ppm. The seven OH protons and anomeric protons at both arabinoses (H1a and H1b) were observed between 4.55-5.54 ppm. Protons of 2' CH_2 were observed between 2.00-2.10 ppm, and the remaining 14 protons were observed between 3.46-4.40 ppm. We found a triplet at 5.33 ppm with integration value of 3.2 protons, which could come from an unknown impurity which was also found in the 5ara-hdC.

Deuterium exchange ^{1}H NMR

We did a D_2O exchange and recorded ^{1}H NMR. Seven protons between 4.55-5.54 ppm disappeared, confirming the presence of 7 OH protons. Peaks at 5.24 ppm (bs), 6.15 (dd, J = 6.6, 6.0 Hz) and 4.88 (d, J = 4.60 Hz) remained in the spectrum. Since the anomeric proton at arabinose for 5ara-hdC was observed at 5.14 ppm, the anomeric proton H1a for the first arabinose should be at 5.24 ppm, and the remaining 4.88 ppm should be the anomeric proton H1b for the second arabinose. These assignments were confirmed by NOESY ^{1}H NMR.

COSY ^{1}H NMR

The anomeric proton of ribose (H1') has clear correlations to H2' protons (2.00 and 2.10 ppm). The H2' protons have a correlation to proton H3' (4.20 ppm) which has correlation to 3'-OH proton at 5.18 ppm. The H3' proton at 4.20 ppm has another correlation to H4' proton at 3.77 ppm, which in turn has correlation to the two H5' protons at 3.55-3.57 ppm. H5' protons have a correlation to 5'-OH peak at 5.02 ppm. For arabinose, the anomeric proton of the second arabinose (H1b) (4.88 ppm) has a clear correlation to H2b proton (3.83 ppm), which in turn has a correlation to H2b-OH peak at 4.98 ppm. Due to strong background noise, we could not clearly determine the correlations of the other protons.

NOESY ^{1}H NMR

The H6 proton has a strong correlation with the anomeric proton H1a for the first arabinose at 5.24 ppm, and it also has correlations with H1′ and H2′ protons of the ribose. The anomeric proton of the second arabinose H1b (4.88 ppm) correlates with H2b (3.83 ppm).

HSQC NMR

We couldn't record ^{13}C NMR due to limited amount of sample, however we run HSQC spectrum. The anomeric carbon C1' for the ribose was at 85.6 ppm, the anomeric carbon C1a for the first arabinose was at 109.5 ppm which is generally expected for *O*-glycosylated sugars and the anomeric carbon C1b for the second arabinose was at 110.9 ppm. ^{1}H NMR (600 MHz, DMSO-d_6) δ (ppm) 7.80 (s, 1H), 6.15 (dd, J = 6.6, 6.0 Hz, 1H), 5.53 (d, J = 4.65 Hz, 1H), 5.24 (bs, 1H), 5.19 (s, 1H), 5.18 (d, J = 1.61, 1H), 5.02 (q, J = 5.41, 2H), 4.98 (d, J = 6.15, 1H), 4.88 (d, J = 4.6 Hz, 1H), 4.56 (t, J = 5.4 Hz, 1H), 4.37-4.36 (m, 1H), 4.21-4.19 (m, 1H), 4.10-4.07 (m, 1H), 3.92-3.91 (m, 1H), 3.83-3.80 (m, 1H), 3.77-3.72 (m, 2H), 3.63-3.50 (m, 4H), 3.50-3.44 (m, 3H), 2.10-2.07 (m, 1H), 2.01-1.96 (m, 1H).

Arabinosylation of 5ara-hC phage DNA *in trans*

To overexpress the Bas47 and 46 *aat* gene, the plasmids pPF3928 and pPF3931, respectively, were generated by adding the amplified gene (PF8103 and PF8104) to the PCR-amplified (PF6734 and PF6735) and DpnI-treated pPF781 backbone in a Gibson assembly reaction. The plasmids were conjugated into *Serratia* and cultures were grown over night with Ara 0.2% (w/v) to induce expression of Aat. The culture was diluted to OD_{600} 0.05 in 25 ml media supplemented with Ara 0.2% (w/v) and grown to OD_{600} 0.2. The culture was infected with LC53 at a multiplicity of infection (moi) of 0.2 and incubated over night to allow the phage to replicate. The phages were harvested by centrifugation and filtration and titrated on *Serratia* as described above. Phage DNA was extracted with the phenol-chloroform method, purified with AMPure XP beads and analysed by LC-MS as described above.

Plasmid expression of anti-phage spacers

Spacers to target phages were expressed from a mini-CRISPR array on plasmids. Briefly, pairs of reverse complement primers of spacer sequences each containing flanking BsaI restriction sites were annealed and cloned into the mini-CRISPR array entry vectors (pPF974, pPF975 and pPF976 for type I-E, I-F and type III-A, respectively) using BsaI and T4 DNA ligase (NEB). Spacers were expressed to crRNAs by the addition of IPTG and Km for plasmid maintenance.

Phage resistance efficiency of plating assay

Efficiency of plating (EOP) assays were performed to investigate the infectivity of the phages. LBA top agar (0.5% w/v, unless otherwise indicated) overlays were seeded with 100 μl bacterial overnight culture and poured on LBA plates. Ten-fold serial dilutions of high titre phage stock (between 10^{10} and 10^{11} pfu/ml) were spotted (5 μl) on the top agar overlays and plates were incubated overnight. The plaques were counted and the EOP calculated as the ratio of pfu/ml on the test strains to the pfu/ml on the control strain. The control strain was the empty vector control or a non-targeting spacer strain relevant to the test strains. For the GmrSD

experiments, the control was *E. coli* ΔRM without a plasmid. All EOP assays were plotted as mean ± SD with individual replicates shown.

Phage resistance infection time course

Bacterial overnight cultures were diluted to an OD_{600}=0.1 and distributed into wells of a 96-well microtiter plate. Phages were diluted to the appropriate concentration in LB and 5 μl were added to the wells to reach the desired moi. For the no phage control, 5 μl of LB were added instead of phages. The 96-well plates were incubated in a plate reader (VICTOR Nivo Multimode Microplate Reader or BioTek Epoch Microplate Spectrophotometer) for 24 h with continuous double orbital shaking (600 rpm) and OD_{600} measured every 10 min. All conditions were repeated in biological replicates and plotted as mean ± standard deviation.

Conjugation efficiency assay

Plasmids targeted by type I-E, I-F and III-A anti-phage spacers were generated by amplifying a 500 bp fragment of the phage DNA around the protospacer sequence. The targeted sequences were added to the PCR-amplified pPF781 (PF6734 and PF6735) backbone in a Gibson reaction. Target plasmids were transformed to *E. coli* ST18 and confirmed by sequencing. Expression of the targeted sequence was induced by the addition of Ara (0.02% w/v) and plasmids maintained with Cm. The interference with *Serratia* strains expressing type I-E, I-F and III-A spacers was investigated in a conjugation efficiency assay. Overnight cultures of the *E. coli* ST18 donor strains carrying the target plasmids or a non-targeted control (pPF781) and *Serratia* strains expressing the appropriate spacers were washed in LB twice and adjusted to OD_{600}=1. Donor and recipient strains were mixed in a 1:1 ratio, spotted on a filter (0.2 μm pore size) on LBA with 5-aminolevulinic acid (Ala) and incubated overnight at 30°C. Next, the filter paper was added to 1 ml phosphate-buffered saline (PBS) and the mating spot was resuspended. Ten-fold serial dilutions were performed in PBS and 10 μl spotted on LBA + Km for the total recipient count and on LBA + Km, IPTG, Cm and Ara (0.02% w/v) for the transconjugants count. Plates were incubated at 30°C for at least 40 h before colony forming units (cfu/ml) could be counted. Conjugation efficiency was expressed as transconjugants (cfu/ml) per total recipients (cfu/ml), measured in triplicates and plotted as mean ± standard deviation.

Expression of Cas13 and anti-phage spacers

For expression of the *Leptotrichia buccalis* (Lbu) type VI-A CRISPR-Cas system, the plasmid pBA559 (obtained from Benjamin Adler, UC Berkeley) encoding *Lbucas13a* under the control of the tetracycline promoter and a type VI mini-CRISPR array consisting of a repeat-spacer unit under the control of the constitutive promoter J23119 was used.[49] A mobilizable type VI spacer entry vector pPF3502 was constructed by adding the RP4 oriT to pBA559. The oriT was amplified from pPF976 (PF6879 and PF6880) and added to the PCR amplified (PF6877 and PF6878) and DpnI digested pBA559 backbone in a Gibson assembly reaction. Anti-phage type VI reverse complement primer pairs containing spacer sequences flanked with BsaI recognition sites were annealed, added to the pPF3502 entry vector in a restriction ligation reaction with BsaI and T4 DNA ligase. Type VI expression vectors were transformed into *E. coli* ST18, confirmed and conjugated into *Serratia*. Due to toxicity of Cas13 expression in *Serratia* even at very low inducer concentrations, all the phage resistance assays were performed without the addition of any inducer.

Investigation of type III-A CRISPR-Cas protection mechanism

Plasmids to express anti-phage LC53 spacers (pPF3348 and pPF3367) were conjugated into different *Serratia* type III-A mutant strains (PCF686 = ΔNucC mutant, PCF690 = HD mutant, PCF691 = Palm mutant that were generated previously[22] and PCF868 = dCas7 (deactivated Cas7) mutant. The dCas7 mutant (*cas7*^{D34A}) was constructed as follows. The up/downstream regions of *cas7* were cloned using primers pairs PF3750/PF3585 and PF4905/PF3751 respectively and *Serratia* WT colonies as DNA template; and primer pairs PF3589/PF3590 and gBlock PF3591 as DNA template. The inserts were cloned into pPF1117, previously digested with SalI and SphI, using Gibson assembly, resulting in plasmid pPF2398. The plasmid was conjugated from *E. coli* ST18 into *Serratia* to generate the site-directed mutant (PCF868) via allelic exchange mutagenesis and *sacB*/sucrose counterselection as described previously.[135] Involvement of these different type III proteins/domains in protection against LC53 was investigated in EOP assays.

In vitro NucC cleavage assay

Phage or host DNA (~100 ng) was incubated with 100 nM NucC (purified as described previously)[65] in 10 mM HEPES-NaOH pH 7.5, 100 mM KCl, 5% (v/v) glycerol, 1 mM DTT, 200 nM cA_3 (c-triAMP Biology Life Science Institute) and 10 mM $MgCl_2$ (total reaction volume of 8 μl). Samples were incubated at 30°C for 30 min and resolved on an 1.2% agarose gel prepared in 1×TAE buffer at 120 V for 40 min and stained with EtBr.

Exchange of type III-A ancillary nuclease

Previously generated plasmids for *in vivo* expression of NucC (pPF2503), Csm6 (pPF2505) and the empty vector control (pPF1618)[65] were conjugated in the *Serratia* ΔNucC mutant (PCF686) together with plasmids for type III-A crRNA expression to target the LC53 *NTP transferase* (pPF3367) or endolysin (pPF3348) genes, or the empty vector control (pPF976). Expression of the ancillary nucleases was induced by the addition of 0.1% (w/v) Ara to the growth media. Phage interference was investigated in EOP and growth curve assays.

Phage engineering using type VI CRISPR-Cas counter-selection

To delete the *arabinosyl-5ara-hC transferase* (*aat*) gene from the Bas46 genome a modified version of the previously described type VI based CRISPR-Cas counter-selection method was applied.[49] First, a recombination plasmid encoding a homologous repair template with the desired mutation and a selection plasmid expressing Cas13a and a type VI spacer that provides full protection against phage infection were generated. The recombination plasmid to delete Bas46 *aat* was generated by amplifying upstream (PF8105 and PF8106) and downstream (PF8107 and PF8108) fragments (~500 bp) for homologous recombination. These flanks were added to the PCR-amplified pPF976 (PF7510 and PF7511) backbone in a Gibson assembly reaction. *E. coli* DH10β carrying the *aat* recombination plasmid pPF3930 was grown in LB + Km. The selection plasmid pPF3924 was generated as described above by adding a type VI-A spacer (PF8114 and PF8115) targeting the mutated sequence of the *aat* gene on the phage genome to the type VI entry vector pPF3502. The targeting efficiency of the selection spacer was assessed in EOP and growth curve assays to ensure sufficient counter-selection of wild-type (WT) phages. For homologous recombination, phages were replicated on the *E. coli* DH10β carrying the recombination plasmid. An overnight culture of this recombination host was diluted to an OD_{600} of 0.05, 200 μl of the bacterial suspension were added to wells of a 96-well plate and infected with Bas46 at moi 0.01. The plate was incubated at 37°C for 5 h in a plate reader to allow phage replication. Phages were collected from wells displaying bacterial growth impairment/lysis indicating successful phage replication. Phages were harvested by centrifugation (2000 *g* for 20 min) and filtration (0.2 μm pore size) and the phage pool (containing a mixture of WT and recombinant phages) was titrated on *E. coli* DH10β. Next, counter-selection of the WT phages was performed by enrichment of the recombinant phages on *E. coli* DH10β carrying the CRISPR-Cas selection plasmid that targets the WT phages. An overnight culture of *E. coli* DH10β carrying the targeting plasmid pPF3924 was diluted to an OD_{600} of 0.05 and 200 μl were transferred to multiple wells of a 96-well plate and infected with the pool of recombinant and WT phages at moi 0.01. The 96-well plate was incubated in a plate reader at 37°C for 5 h and lysates of wells showing bacterial lysis were collected. Phages were again harvested by centrifugation and filtration before being titrated on the targeting (pPF3924) and non-targeting (pPF3502) *E. coli* DH10β strain. A decreased EOP on the targeting strain indicated uncomplete selection and presence of WT phages. Similar EOP results on targeting and non-targeting strains suggested the enrichment step had successfully selected for recombinant phages that are not targeted by type VI. Plaques were screened by PCR and sequencing to distinguish between type VI escape and recombinant phages. Recombinant phage stocks were generated by replicating phages isolated from positive plaques on the selection host (pPF3924).

Isolation of arabinose isomerase mutant phage

Phages with small deletions in the *arabinose isomerase* gene were obtained by liquid propagation of the phage on a host expressing a type III-A spacer targeting the start codon locus of the gene from a plasmid-borne mini-array (pPF3722). An overnight culture of this targeting host was diluted to an OD_{600} of 0.05 and 195 μl of the bacterial suspension were added to wells of a 96 well microtiter plate. The plate was incubated in a plate reader at 30°C with continuous double-orbital shaking (600 rpm) and OD_{600} measurements in ten minutes intervals. The cultures were grown for approximately 5 h until an OD_{600} of 0.3 was reached. Five μl of a phage dilution were added to the bacterial culture to reach moi 0.2. The plate was incubated in the plate reader overnight and OD_{600} measurements were continued. Impaired bacterial growth compared to the no phage control indicated the presence of type III-A escape phages. These phages were harvested and titrated on the targeting and non-targeting empty vector strain. Similar EOP on both strains gave further indication for the presence of escape phages and some plaques were picked from the targeting strain and dissolved in 50 μl phage buffer. Mutations were investigated by PCR amplification and sequencing of the *arabinose isomerase* gene from these plaques. A new stock of these escape mutants was obtained by using the isolated plaque for propagation of the phage on the targeting host (pPF3722).

Identification of phages with similar DNA modifications in genomic databases

Phage genome accession numbers were retrieved from the NCBI nucleotide database using the R package rentrez using the search terms '(phage[TITL] OR bacteriophage[TITL]) AND (genome[TITL] OR complete sequence[TITL])' and '"gbdiv_PHG"[prop]'. Accession numbers were matched to the INPHARED database (https://github.com/RyanCook94/inphared) entries.[78] Genome sequence files were retrieved from NCBI using the program *datasets* (version 16.37.0) and formatted using *dataformat*. The code for downloading genomes is available in the Jupyter notebook 'A_download-genomes.ipynb'. Viral genome fragments from metagenomes in IMG/VR v.4.1[79] (IMG_VR_2022-12-19_7) were downloaded as a file archive from https://genome.jgi.doe.gov/portal/IMG_VR/IMG_VR.home.html.

Phage genome metadata, annotations, and protein sequences were loaded into R. R packages readxl and Biostrings were used for data import. Data wrangling was performed using the tidyverse packages dplyr, tidyr. Visualizations were created using ggplot2 and gggenes. Gene annotations were harmonized using pattern matching by stringr. HMM profiles were built by hmmbuild from MUSCLE (version 5.3) alignments and used in hmmsearch against protein.faa files using HMMER (version 3.4) through the custom function *annotate_gene_by_hmm_profile*. For each protein involved in the proposed arabinosylation pathway we built an HMM profile from an alignment of representative protein sequences. Initially, we used sequences from seven phages (T4, 92A1, LC53, Bas46, RB69, Maestro, Navy4) to search for similar proteins in a training set of 35 phages (Table S3).

Subsequently, we built profiles from the training set to search INPHARED[78] and IMG/VR[79] and used gene patterns to predict genome modification. Genome glucosylation was predicted if at least two genes out of dCMP hydroxymethylase, alpha-glucosyl-transferase and beta-glucosyl-transferase were present in a genome. Genome arabinosylation was predicted if at least four genes out of dCMP hydroxylase, putative glycosyltransferase, arabinose isomerase, phosphoheptose isomerase and thymidylate kinase were

present in a genome. Distinct HMM profiles were built for Aat, Aat1, Aat2. 'No *aat*' was defined as arabinosylation without *aat*, *aat1* or *aat2* detected. 'Single *aat*' was defined as arabinosylation with one *aat*. 'Aat1 and Aat2' was defined as arabinosylation and one *aat1* and one *aat2*. Arabinosylation together with any other combination of *aat* genes was defined as 'multiple Aat'. The group 'some genes' was created to distinguish genomes for which any protein matched an HMM profile from genomes that had no similar proteins ('None'). The code for the detection of genes involved in arabinosyl-hydroxy-cytosine DNA modification and classification of genomes is available in the Jupyter notebook 'B_annotate_train_predict-INPHARED.ipynb'. The code for prediction of arabinosylation in the IMG/VR database is available in the Jupyter notebook 'C_predict-IMGVR.ipynb'.

Brig1 cloning and the *Klebsiella pneumoniae* clinical plasmid

Brig1 and three additional previously untested homologues were selected from a recent study.[82] The sequences were generated as gBlocks and cloned via Gibson assembly into the pQE-80LoriT backbone generated by PCR amplification with PF7427 and PF7428. An ampicillin-resistant clinical plasmid from *K. pneumoniae* MRSN 731029 was introduced into *E. coli* K-12 ΔRM. The pQE-80LoriT (Ap^R) plasmid was used as the control for all assays with Brig and the *Klebsiella* plasmid.

QUANTIFICATION AND STATISTICAL ANALYSIS

Efficiency of plating (EOP) and conjugation efficiency assays in Figures 1, 4, 7, S3, S5, and S7 were performed in at least biological triplicates. All data presented is the means ± standard deviation (SD) with individual biological replicates shown. Biological replicates are from cultures derived from individual bacterial colonies and treated as per the details described in the results and figure legends. Data were plotted in Prism and all details are provided in the legends.