7-Deazaguanine modifications protect phage DNA from host restriction systems

Geoffrey Hutinet 1*, Witold Kot 2, Liang Cui 3, Roman Hillebrand 4,12, Seetharamsingh Balamkundu 3, Shanmugavel Gnanakalai 3, Ramesh Neelakandan 3, Alexander B. Carstens 2, Chuan Fa Lui 5, Denise Tremblay 6,7, Deborah Jacobs-Sera 8, Mandana Sassanfar 9, Yan-Jiun Lee 10, Peter Weigele 10, Sylvain Moineau 6,7, Graham F. Hatfull 8, Peter C. Dedon 3,4, Lars H. Hansen 2 & Valérie de Crécy-Lagard 1,11*

Genome modifications are central components of the continuous arms race between viruses and their hosts. The archaeosine base (G+), which was thought to be found only in archael tRNAs, was recently detected in genomic DNA of Enterobacteria phage 9g and was proposed to protect phage DNA from a wide variety of restriction enzymes. In this study, we identify three additional 2′-deoxy-7-deazaguanine modifications, which are all intermediates of the same pathway, in viruses: 2′-deoxy-7-amido-7-deazaguanine (dADG), 2′-deoxy-7-cyano-7-deazaguanine (dPreQ0) and 2′-deoxy-7-aminomethyl-7-deazaguanine (dPreQ1). We identify 180 phages or archaeal viruses that encode at least one of the enzymes of this pathway with an overrepresentation (60%) of viruses potentially infecting pathogenic microbial hosts. Genetic studies with the Escherichia phage CAjan show that DpdA is essential to insert the 7-deazaguanine base in phage genomic DNA and that 2′-deoxy-7-deazaguanine modifications protect phage DNA from host restriction enzymes.
In the continuous battle between bacteria and phages, both entities are constantly evolving defenses and counterattack mechanisms. To escape these defenses, phages have developed multiple strategies, and one of the most widespread strategies is to modify their DNA. For example, the genomic DNA of Escherichia coli phage T4 contains the nucleobase glucosyl-hydroxymethylcytosine, which inhibits the restriction-modification (RM) and clustered regularly interspersed short palindromic repeat (CRISPR)–CRISPR-associated (Cas) systems.

The increased availability of complete phage genome sequences has led to recent discoveries of novel complex DNA modifications, such as 2′-deoxy-5-hydroxymethyluracil derivatives in Enterobacteria phage M6, Salmonella phage Vil, and Delfia phage phi W-14, and 2′-deoxyarchaeosine (dG+) in Enterobacteria phage 9g.

Two 7-deazaguanine modifications, 2′-deoxy-7-amido-7-deazaguanosine (dADG) and the 2′-deoxyribonucleoside analog of archaeosine, which were previously thought to be present only in tRNA as queuosine (Q) in bacteria and archaeosine (G) in archaea, were recently discovered in bacteria and phage DNA, respectively, by combining in silico data mining and experimental validation. As shown in Fig. 1, 7-cyano-7-deazaguanine (preQ0) is synthesized from GTP by four enzymes (FolE, QueD, QueE, and QueC) and is the key intermediate in both the Q and G+ pathways. tRNA-guanine-transglycosylases (TGT in bacteria, arcTGT in archaea) are the signature enzymes in the Q and G+ tRNA modification pathways, as they exchange the targeted guanines with 7-deazaguanine precursors. In archaea, preQ0 is directly incorporated into tRNA by arcTGT before being further modified by different types of amidotransferases (ArcS, Gat-QueC, or QueF-L). In bacteria, preQ0 is reduced to 7-amino-2′-deoxy-7-deazaguanine (preQ1) by QueF18 before TGT incorporates it in tRNA, where it is further modified to Q in two steps (Fig. 1).

The presence of homologs of Q synthesis genes has long been predicted to G in tRNAs at position 34 (blue dashed square). In bacteria, preQ0 is transferred to position 15 of tRNA before being modified to G+ (red dashed square). The bases found in phage DNA in this study are in red. Molecule abbreviations: guanosine triphosphate (GTP), dihydronopterin triphosphate (H3NTP), 6-carboxy-5,6,7,8-tetrahydropterin (CPH4), 7-carboxy-7-deazaguanine (CDG), 7-amido-7-deazaguanine (ADG), 7-cyano-7-deazaguanine (preQ0), 7-aminomethyl-7-deazaguanine (preQ1), queuosine (Q), and archaeosine (G+).

**Fig. 1** Queuosine and archeosine synthesis pathways. preQ0 is synthesized from GTP in both bacteria and archaea through FolE, QueD, QueE, and QueC, as shown. In most bacteria, four more enzymatic steps lead to the insertion of Q in tRNAs at position 34 (blue dashed square). In archaea, preQ0 is transferred to position 15 of tRNA before being modified to G+ (red dashed square). The bases found in phage DNA in this study are in red. Molecule abbreviations: guanosine triphosphate (GTP), dihydronopterin triphosphate (H3NTP), 6-carboxy-5,6,7,8-tetrahydropterin (CPH4), 7-carboxy-7-deazaguanine (CDG), 7-amido-7-deazaguanine (ADG), 7-cyano-7-deazaguanine (preQ0), 7-aminomethyl-7-deazaguanine (preQ1), queuosine (Q), and archaeosine (G+).
dG\(^+\), in *E. coli* DNA. Because the presence of dG\(^+\) confers resistance to EcoRI digestion\(^{29}\), we used restriction profiles as a first indication for the presence of modifications in plasmid DNA. The two phage genes were both cloned into pBAD24 and pBAD33. EcoRI cuts pBAD24 once and pBAD33 twice, as shown in the digestion profiles of plasmids extracted from *E. coli* cotransformed with the two empty plasmids (Fig. 2b, c, lane 1). Because the *gat-queC* and *dpdA* genes of phage 9g lack EcoRI sites, the restriction profiles of plasmids extracted from *E. coli* derivatives cotransformed with an empty plasmid and a plasmid containing one of the two genes are shifted by the insert sizes (Fig. 2c, lanes 2, 3, 5, and 6). An additional band corresponding to the uncut plasmid was observed only for plasmids extracted from strains expressing both *gat-queC* and *dpdA* genes (Fig. 2c, lanes 4 and 7, and Fig. 2b, white arrows). As a supplemental control, we digested the same combination of plasmids with Psii (TTA\(^-\)) and EcoRI (Supplementary Fig. 1). The single digestion by Psii linearized all these plasmids, and the plasmids encoding both *dpdA* and *gat-queC* of phage 9g were again partially resistant to EcoRI digestion (red arrows in Supplementary Fig. 1).

Analysis of dG\(^+\), dADG, dPreQ\(_0\), and dPreQ\(_1\) profiles by liquid chromatography-coupled triple quadrupole mass spectrometry (LC-MS/MS, quantification results in Table 1, mean ± standard deviation based on two or three replicates) revealed that plasmid DNA extracted from strains expressing only *dpdA* contained dPreQ\(_0\) with 790 ± 8 modifications per 10\(^6\) nucleotides; 0.316 ± 0.0028% of the Gs. No modifications were detected in strains expressing only *gat-queC* (Table 1). Taken together, these results showed that dG\(^+\) but not preQ\(_0\) confers resistance to EcoRI and that the phage 9g pathway that inserts dG\(^+\) in its viral DNA can be transferred to modify *E. coli* genomic DNA.
Interestingly, whereas we had failed to complement the Q⁻ phenotype of the *E. coli ΔqueC* strain when expressing the *gat-queC* gene of phage 9g, the EcoRI resistance phenotype caused by 7-deazaguanine insertion in strains expressing both *dpdA* and *gat-queC* of phage 9g was still observed in a Δ*queC* background (Fig. 2c, lanes 8 and 9) but not in a Δ*queD* background (Fig. 2c, lanes 10 and 11). Furthermore, only dG⁺ modification was observed in the DNA of the Δ*queC* strains by LC-MS/MS (Table 1), with similar amounts as in the wild type (WT; 13,750 modifications per 10⁶ nucleotides, 5.5% of the Gs, and 23,000 ± 17,000 modifications per 10⁶ nucleotides, 9.2 ± 7% of the Gs). This suggests that the Gat-QueC protein can produce preQ₀ but that it is channeled to the putative DNA-modifying enzyme DpdA and not to the tRNA-modifying pathway enzyme QueF.

Finally, we tested whether the *E. coli TGT* was required for DpdA activity in *E. coli*, as the active forms of TGT enzymes are known to be dimers. This did not seem to be the case, as the restriction phenotype was still observed in the Δtg background (Fig. 2c, lanes 12 and 13).

### A wide variety of phages encode dG⁺ synthesis proteins

We identified another subfamily of DpdA, renamed DpdA2, encoded by the *Vibrio* phage nt-1 by investigating genes flanking the preQ₀ biosynthesis gene cluster. Indeed, DpdA2 (YP_008125322) of phage nt-1 is not detected when using *Enterobacteria* phage 9g DpdA as a query in PSI-BLAST. This DpdA2 family does not possess the conserved histidine found at position 196. However, some similarities with members of the TGT enzymes family were detected by HHpred, with a confidence score of 100%.

An in silico search for phages that could harbor 7-deazaguanine derivatives in their genomic DNA revealed a total of 182 viruses deposited in GenBank that were found to encode a DpdA/DpdA2 homolog and/or at least a G⁻ amidotransferase. Among these viruses, 16 were archaeal viruses and 3 were *Bacteriophage* phages, while 16 were archaeal viruses and 3 were eukaryotic viruses. The eukaryotic viruses only encode FoaE, which is most likely linked to the folate pathway. Analyses of the presence/absence patterns of the predicted Q/G⁺ biosynthesis genes led to a classification of these viruses into various groups and, in some cases, predicted the nature of the 7-deazaguanine base modification. It is important to note that the DpdA objective is to encode homologs to one of the three amidotransfereases involved in the last steps of G⁺ synthesis: ArcS₁⁵, QueF-L₁⁶ (or QueF), or a glutamine amidotransferase (Gat) domain fused to the canonical QueC. These phages likely modify their DNA with dG⁺, as does phage 9g. It should be noted that the discrimination between the QueF-L homologs, predicted to produce the G⁺ base from preQ₀, and QueF homologs, predicted to produce preQ₁ from preQ₀, is difficult to establish based only on sequence similarity. Therefore, the phages encoding these proteins might harbor dG⁺ or preQ₁ (or both). Of note, this viral group includes a *Pseudomonas aeruginosa* phage that was isolated; the genome of this phage was sequenced in this study, and the phage was named *Pseudomonas* phage Quinobequin P09 (description in Supplementary Information).

The second group includes 40 phages and is represented by *E. coli* phage CAjan (NC_028776) and *Mycobacterium* phage Rosebush (AY129334) in Fig. 3. These phages encode a homolog of one of the two types of DpdA and of the preQ₀ synthesis enzymes (FoaE, QueD, QueE, and QueC), but they are missing an amidotransferase. As such, we predicted that these phages modify their DNA with preQ₀ or ADG, similar to the bacteria that contain the *dpd* cluster. *Mycobacterium* phage Bigger (KU728633), which is only missing a gene coding for QueC, was added to this group even if it could be modified by the 8Q₀ and ADG, similar to the bacteria that contain the *dpd* cluster. *Mycobacterium* phage Bigger (KU728633), which is only missing a gene coding for QueC, was added to this group even if it could be modified by the QueC substrate (7-carboxy-7-deazaguanine, see Fig. 1). The uncultured phage clone 7AX₂ (MF417872) was also added to this group because it lacks *queC*, although this may be due to the incomplete genome sequence of this phage. In addition, we cannot exclude that this phage encodes an amidotransferase.

The third group is currently the largest, as it contains 76 phages, including *Salmonella* phage 7-11 (NC_015938) and *Mycobacterium* phage Orion (DQ398046), as shown in Fig. 3. These phages encode DpdA but no G⁻ or preQ₀ biosynthesis protein homologs. At this stage, their genome modification status, if any, is difficult to predict. Phages in this group could rely on preQ₀ synthesized by the host or on the uptake of exogenous 7-deazapurine precursors. Some phages do encode homologs of YhhQ, the preQ₀ transporter, but there is no correlation with any specific group of phages. The large size of this group compared to the others might be caused by the relatively large number of Mycobacteriophages in the Virus database due to the massive phage isolation and sequencing effort of PhagesDB and the SEA-PHAGES project.

The last group is composed of 48 phages encoding proteins of the preQ₀/G⁺ pathway but not DpdA. These phages could boost the production of the Q precursor to increase the level of Q in the host tRNA and increase translation efficiency. However, it is possible that 7-deazaguanines are inserted in their DNA in a DpdA-independent pathway, as there is a recent report that the genomes of *Campylobacter* phages of this group are heavily modified by dADG. Similarly, the Halovirus HVT-1

### Table 1 DNA modifications identified by mass spectrometry in the plasmids shown Fig. 2b.

| Lane in Fig. 2b | Background | 9g gene in pBAD24 | 9g gene in pBAD33 | dADG per 10⁶ nt | dPreQ₀ per 10⁶ nt | dPreQ₁ per 10⁶ nt | dCDG per 10⁶ nt | dG⁻ per 10⁶ nt |
|-----------------|------------|------------------|------------------|----------------|----------------|----------------|----------------|----------------|
| 1               | MG1655     | None             | None             | <6             | <6             | <6             | <6             | <6             |
| 2               | MG1655     | dpdA             | None             | <6             | 790 ± 8        | <6             | <6             | <6             |
| 3               | MG1655     | dpdA             | dpdA             | <6             | <6             | <6             | <6             | 45,000 ± 25,000 |
| 4               | MG1655     | gat-queC         | gat-queC         | 77 ± 7⁺        | <6             | <6             | <6             | <6             |
| 5               | MG1655     | gat-queC         | None             | <6             | <6             | <6             | <6             | 6.5 ± 0.5      |
| 6               | MG1655     | dtubes           | dtpa             | 84 ± 26        | <6             | <6             | 22,750 ± 17,250 |
| 7               | MG1655     | gat-queC         | gat-queC         | <6             | <6**           | <6**           | <6**           | 13,750**       |
| 8               | MG1655     | ΔqueC            | gat-queC         | <6             | <6             | <6             | 45,000 ± 25,000 |
| 9               | MG1655     | ΔqueC            | gat-queC         | <6             | <6             | <6             | 23,000 ± 17,000 |

All values represent the mean ± standard deviation for three replicate analyses, except asterisk (*), mean ± standard deviation for five replicate analyses, and double asterisks (**), single analysis.
(NC_020158), presented in Fig. 3, may have found another way to insert the modifications and should harbor either dPreQ1 or dG+, as it encodes the QueF, or QueF-like, protein.

Phages containing FolE and QueC singletons were discarded from further analysis because FolE is shared between folate and preQ0 synthesis13, while QueC is also part of a superfamily of ATPases37, making their precise role difficult to identify.

All the phages identified above are members of the Caudovirales order and are distributed into various families: Siphoviridae (95), Myoviridae (23), Ackermannviridae (20), and Podoviridae (3). For the Archaeal viruses, we identified 12 members of the Ligamenvirales order and 2 of the Bicaudaviridae family (Supplementary Data 2).

**Detailed analysis of phage 7-deazaguanine synthesis proteins.** To evaluate the isofunctionality of the studied protein families, sequence similarity networks (SSNs) were generated. Proteins in the same cluster should share the same function38. Several of the 7-deazaguanine biosynthesis proteins are part of protein families that are known to harbor subgroups with different functions that could impede functional annotations using only PSI-BLAST.
Fig. 4 Protein similarity networks. Source data are provided as a Source Data file. a DpdA/Tgt protein network, each node is a group of proteins identical at 90%, and each edge presents an alignment score ≥15. The TGTs of archaea, bacteria, and eukaryotes are shown in dark green, green, and light green, respectively. Bacterial DpdA are shown in light red. The phage DpdA are separated depending on the gene content of phages: in red, DpdA in genomes encoding the G\(^+\) pathway; in orange, the preQ0 pathway; in yellow, the genomes with only dpdA; in dark blue, DpdA2 with G\(^+\) pathway; and in light blue, DpdA2 with a preQ0 pathway. The arrow shows clusters of nodes specific to a clade of a bacterial host (1 is Mycobacterium and 2 is \(\gamma\)-Proteobacteria). b QueC protein network, with a threshold alignment score of 44. In light red, the QueC from bacteria; in dark red, the QueC from phages that encode a DpdA; in orange, the QueC from phages that are not encoding a DpdA; and in yellow, the QueC from phages that are not encoding a DpdA. c QueF protein network with an alignment score threshold of 10. In light red, the bacterial QueF; and in dark blue, the phage protein identified as QueF. In light blue, the archaeal QueF-L; and in dark blue, the phage protein identified as QueF-L.

Most phage QueC proteins do not cluster with bacterial QueC proteins when the BLAST threshold score is sufficient to separate QueC from the Gat-QueC groups (Fig. 4b). However, when a lower threshold score is used, the QueC and Gat-QueC proteins can be connected (Supplementary Fig. 2A). This is not the case for the QueC proteins encoded as singletons in phages, such as Bacillus phage SP-15 and Salmonella phage SEPI (Supplementary Data 1), suggesting that even though the proteins were identified as QueC by HHpred, they may be part of a functionally unrelated subgroup of the N-type ATP pyrophosphatases superfamily. Finally, phage and archaeal Gat-QueC proteins form a single cluster, strengthening their functional association. HHpred predicted that the QueF family proteins encoded by phages are, for most of them, closer to the archaeal QueF-L proteins than to the bacterial QueF proteins (see Supplementary Data 1). However, they clustered with bacterial QueF proteins in the SSNs (Fig. 4c). Further experimental studies are required to determine whether the phage QueF proteins are nitrile reductases or amidotransferases (Fig. 1).

SNNs for the FolE, QueD, QueE, and ArcS families are shown in Supplementary Fig. 2B–E. The phage proteins cluster nicely with their bacterial and archaeal homologs, reinforcing the initial functional annotations.

The host may participate in phage DNA modification. To study the interaction between phages containing 7-deazaguanine-related genes and their bacterial hosts, we gathered metadata on the hosts and their habitat using RefSeq\(^{39}\) and the Globi database\(^{40}\) and analyzed the distribution of Q, G\(^+\), and dADG synthesis genes in these organisms (see Supplementary Data 2 and 3). Interestingly, 106 of the collected phages (~60%) infect a host strain that is the model for a known bacterial pathogen (Supplementary Data 2), where only ~9% of all the double-stranded DNA (dsDNA) viruses from the Virus-Host database\(^{41}\) infect a strain related to pathogens (data not shown), making our sample six to seven times more enriched compared to a random sampling. No clear environment was found for the archaeal hosts. All phage hosts predicted to modify their DNA with G\(^+\) possess the pathway to produce Q in tRNA. Curiously, the hosts of phages coding for a QueF-L and a 9g DpdA homolog do not encode the preQ0 biosynthetic pathway (QueDEC, see Fig. 1) but encode the specific preQ0 transporter YhhQ\(^{33}\) and the rest of the Q pathway (QueFAG and TGT, Fig. 1). Conversely, all the hosts of the DpdA2-encoding phages encode the full Q pathway.
There is no clear pattern for the bacterial hosts of phages encoding both DpdA and the whole preQ0 pathway. Most of them encode the full Q pathway enzymes except for Streptococcus pneumoniae, which lacks the preQ0 pathway genes; Rhodococcus erythropolis, which encodes only TGT; and Mycobacteria, which possess none of these genes.

The hosts of the phages encoding only DpdA also encode the full set of Q synthesis enzymes except the Clostridium species, which lack the preQ0 pathway genes, and the Mycobacterium genus, which possesses none of these genes. Sulfolobus were not referenced in PubSEED, but by performing a BLASTp search with default parameters and the genes listed in Supplementary Table 1 as queries, we identified all G+ pathway genes (Supplementary Table 2). Hence, the 7-deazaguanine intermediates produced by these hosts, Clostridium and Mycobacterium excluded, might be used by phages that lack the biosynthesis proteins to produce a 7-deazaguanine precursor.

Finally, the hosts of the phages that do not encode a DpdA homolog but encode the preQ0 pathway proteins all encode the full Q synthesis pathway.

A few bacterial hosts, such as 46 different strains of E. coli, Haloarcula vallismortis, and Vibrio harveyi 1DA3, also harbor homologs of the bacterial DpdA, which are known to modify bacterial DNA by either dPreQ0 or dADG.

Different 7-deazaguanine modifications in distinct phages. To test our predictions on the nature of phage DNA modifications, a set of phages from each group were selected (Fig. 3), and their genomic DNAs were extracted for mass spectrometric analysis (Table 2, mean ± standard deviation based on two replicates). No 2'-deoxyqueuosine (dQ) was found in any of the tested samples, but dPreQ0 and dG+ were identified in all the tested samples.

Table 2 DNA modifications identified by mass spectrometry in the different phages.

| Phage/virus Accession # | Phage/virus name | Phage/virus GC content | Prediction based on gene content | dPreQ0 per 106 nt | dADG per 106 nt | dG+ per 106 nt | dPreQ1 per 106 nt | dQ per 106 nt |
|------------------------|------------------|------------------------|--------------------------------|------------------|---------------|---------------|------------------|-------------|
| NC_028776              | Escherichia phage CAjan | 44.70%                 | dPreQ0                          | 70,628 ± 2445     | <6            | <6            | <6               | <6          |
| NC_020158              | Halovirus HVT-1   | 58.30%                 | None/dG+                        | <6               | 152 ± 3        | 22 ± 1         | 88,607 ± 3014    | <6          |
| NC_008197              | Mycobacterium phage Orion | 66.50%                 | None                            | <6               | <6            | <6            | <6               | <6          |
| NC_004684              | Mycobacterium phage Rosebush | 69.00%                 | dPreQ0                          | 96,530 ± 2529    | 9 ± 1         | <6            | <6               | <6          |
| NC_015938              | Salmonella phage 7-11 | 64.10%                 | None/PreQ0                      | <6               | 50 ± 2         | <6            | <6               | <6          |
| NC_015274              | Streptococcus phage Dp-1 | 40.30%                 | dPreQ0/dG+                      | <6               | <6            | <6            | 3389 ± 184       | <6          |
| NC_021529              | Vibrio phage nt-1  | 41.30%                 | dG+                             | 232 ± 4          | 72 ± 2         | 44 ± 1         | <6               | <6          |

All values represent the mean ± deviation of the mean for two analyses.

Then we investigated phages of the second group that encode both a DpdA and the four proteins of the preQ0 biosynthesis pathway but no amidotransferase homolog. Mycobacterium phage Rosebush was found to harbor dPreQ0 in its DNA (96,530 ± 2529 modifications per 10^6 nucleotides, ~28 ± 1% of the Gs), as does Escherichia phage CAjan (70,628 ± 2445 modifications per 10^6 nucleotides, ~32 ± 1% of the Gs). However, Mycobacterium phage Rosebush was also found to harbor a negligible amount of dADG (9 ± 1 modifications per 10^6 nucleotides, ~0.003 ± 0.0003% of the Gs).

The genomic DNA of Salmonella phage 7–11 and Mycobacterium phage Orion from the third group of phages, which only encode a DpdA, were also analyzed by LC-MS/MS. Mycobacterium phage Orion lacked any 7-deazaguanine modifications in its DNA. This result was expected, as none of the phage nor the host encode for the preQ0 biosynthesis pathway (Mycobacterium smegmatis, see Supplementary Data 3). However, Salmonella phage 7–11 was unexpectedly modified by dADG (50 ± 2 modifications per 10^6 nucleotides, ~0.02 ± 0.0009% of the Gs), suggesting that the phage encoded a protein responsible for the oxidation of preQ0.

Finally, Halovirus HVT-1, which encodes the four proteins of the preQ0 biosynthesis pathway and a Que-F-L homolog but no DpdA, contained mainly dPreQ1 (88,607 ± 3014 modifications per 10^6 nucleotides, ~30 ± 1% of the Gs) but also relatively small amounts of dADG and dG+ (152 ± 3 modifications per 10^6 nucleotides, ~0.05 ± 0.001% of the Gs, and 22 ± 1 modifications per 10^6 nucleotides, ~0.008 ± 0.0003% of the Gs, respectively). As its host, H. vallismortis harbors a DpdA homolog, and it is possible that the host DpdA inserts preQ0 in Halovirus HVT-1 DNA before it is further modified to dPreQ1 or dG+ by the viral Que-F-L or to dADG by another unidentified protein.

dpdA is essential for DNA modification. To evaluate the role of the 7-deazaguanine modifications in phages, we used the Escherichia phage CAjan as a genetic model. CAjan is a virulent phage belonging to the Seoravirus genus of the Siphoviridae family with many similarities with Enterobacteria phage 9g, particularly within the 7-deazaguanine modification pathway. Using the CRISPR-Cas9 genome editing technology, we generated a CAjan derivative with an inactive allele of the dpdA gene (Supplementary Fig. 3A). The presence of this allele was confirmed by PCR and sequencing (Supplementary Fig. 3B). The LC-MS/MS analysis of the DNA of the mutated phage showed a complete lack of 7-deazaguanine modifications (Table 2).
The DNA modifications protect DNA from restriction enzymes. The different modifications present in the phages analyzed above may lead to distinct resistance patterns to host defense mechanisms, such as RM systems. To test this hypothesis, phage DNA preparations were digested with a set of restriction enzymes that had been shown to be totally or partially inactivated in the presence of the dG\(^+\) modification\(^{29}\). As a control, we reproduced the results published with *Enterobacteria* phage 9g DNA (Fig. 5a); no digestion was observed with BamHI, EcoRI, EcoRV, and SwaI, while it was partially restricted with BstXI, HaeIII, MluI, NdeI, and PciI.

*Mycobacterium* phage Rosebush DNA that carries preQ\(_0\) showed a slightly different pattern of resistance. The restriction profiles for BamHI, BstXI, and EcoRV were identical to those of *Enterobacteria* phage 9g. However, Rosebush DNA was fully sensitive to HaeIII, MluI, and PciI and resisted NdeI degradation (Fig. 5b). EcoRI and SwaI could not be tested because the corresponding sites are absent in the *Mycobacterium* phage Rosebush genome.

Though *Escherichia* phage CAjan DNA carries the same modification as *Mycobacterium* phage Rosebush DNA, differences in the restriction patterns were observed (Fig. 5c). Indeed, while EcoRI and SwaI fully digested this DNA preparation, BamHI digested it only partially, and HaeIII did not cut at all. These differences could be explained by the additional small amount of dADG present in *Mycobacterium* phage Rosebush DNA, by the differences in modification density potentially affecting accessibility to the restriction sites, or by the presence of another undetected modification. In comparison, the ∆dpdA mutant of CAjan, lacking any modifications, was fully digested by all the tested restriction enzymes (Fig. 5d), formally linking the presence of the dpdA gene and the dG\(^+\) modification to the restriction resistance phenotype.

Last but not least, Halovirus HVTV-1 DNA that carries mainly dPreQ\(_1\) was found to resist restriction by all enzymes tested, even those that lack guanine in the recognition site (Fig. 5e and Supplementary Fig. 4). It is possible that this virus has other modifications that help resist restriction and, if not dPreQ\(_1\), is the...
best modification for protection from restriction enzymes identified in this study.

Discussion
In a previous study, we identified two 7-deazaguanine modifications in DNA: dADG in bacteria and dG+ in phages. Here we added two modifications, dPreQ_0 and dPreQ_0, both found in phages. Similar to the result of Szymanski’s group on Campylobacter phages, we also detected dADG in phage genomes. We identified the genes involved in the synthesis of these different modifications. FolE, QueD, and QueE from Enterobacteria phage 9 g were shown to functionally replace their E. coli orthologs (Fig. 2a), and their clustering in SSNs (Supplementary Fig. 2) leaves no doubt on the isofunctionality of these families. No individual phage QueC was tested, but the strong clustering of bacterial, archaeal, and phage QueC proteins in SSNs also point to identical functions. One exception may be the singleton encoded QueC-like protein, found in Escherichia phage ECML-4 (YP_009101458 in NC_025446) or Mycobacterium phage Muddy (YP_008408902 in NC_0202054), which is likely a member of another subfamily of the N-type ATP pyrophosphatases superfamily.

Most 7-deazaguanine-containing phage genomes also harbor a gene coding for a DpdA homolog. As with its bacterial homolog, the phage DpdA introduces PreQ_0 in DNA (Fig. 2c, Table 1), most likely through a base exchange mechanism similar to its TGT reaction, such as the bacterial DpdC. Further biochemical characterization will be required to explore these hypotheses. One cannot exclude the possibility that the small amount of dADG detected in Vibrio phage nt-1, Halovirus HVTV-1, Mycobacterium phage Rosebush, and Escherichia phage CAjan could be the result of the natural oxidation of dPreQ_0.

The discrepancy observed between the SSNs and HHpred predictions for the QueF/QueF-L homologs was resolved by analyzing Streptococcus phage Dp-1 and Halovirus HVTV-1 DNA. HHpred analysis predicted that a homolog of the archaeal QueF-L, which synthesizes G+-tRNA from the preQ_0-tRNA, was encoded by these phages, whereas the SSN analysis predicted that this same protein was part of a group of bacterial QueF proteins (Fig. 4) that synthesize preQ_0 from the free preQ_0 base. We found that Streptococcus phage Dp-1 and Halovirus HVTV-1 were modified by dPreQ_0, confirming the SSN prediction. However, it is unclear whether the reduction occurs on free preQ_0, similar to the bacterial QueF proteins, and then the free base preQ_0 is inserted by DpdA or if the phage QueF is able to modify the DNA-bound dPreQ_0, as does the archaeal QueF-L with tRNA. However, Halovirus HVTV-1 contains mainly dPreQ_0, but also a small amount of dADG and dG+. It is possible that the QueF-L transitions between its function as an amidohydrolase to an amidotransferase, but one cannot rule out that the host ArcS could catalyze the reaction, although the PUA domain specific for tRNA binding makes it highly unlikely.

The combination of comparative genomic analyses and experimental validations has allowed pathways for the insertion of dPreQ_0, dPreQ_0, and dG+ in phage genomes to be predicted (Fig. 6). The presence of the minimal set of FolE, QueD, QueE, QueC, and DpdA proteins leads to the insertion of dPreQ_0, as observed in Mycobacterium phage Rosebush and Escherichia phage CAjan genomes (Table 2 and Fig. 3). The replacement of QueC by Gat-QueC leads to the introduction of dG+ (Fig. 2c, Table 1 and previous study). However, it is not known whether Gat-QueC converts preQ_0 into G+ before or after it is inserted into DNA. The function of ArcS homologs in phages/viruses is less clear. Indeed, Vibrio phage nt-1 encodes an ArcS homolog, and its DNA contains mainly dPreQ_0 but also dG+ and dADG (Table 2 and Fig. 3). ArcS was the first G+ synthase identified in archaea. Based on the phage and archaeal ArcS cluster in the SSNs (Supplementary Fig. 2), it is possible that some phage ArcS protein evolved to perform not only an amidotransferase reaction, such as the archaeal ArcS, but also an amidohydrolase reaction, such as the bacterial DpdC.

The discrepancy observed between the SSNs and HHpred predictions for the QueF/QueF-L homologs was resolved by analyzing Streptococcus phage Dp-1 and Halovirus HVTV-1 DNA. HHpred analysis predicted that a homolog of the archaeal QueF-L, which synthesizes G+-tRNA from the preQ_0-tRNA, was encoded by these phages, whereas the SSN analysis predicted that this same protein was part of a group of bacterial QueF proteins (Fig. 4) that synthesize preQ_0 from the free preQ_0 base. We found that Streptococcus phage Dp-1 and Halovirus HVTV-1 were modified by dPreQ_0, confirming the SSN prediction. However, it is unclear whether the reduction occurs on free preQ_0, similar to the bacterial QueF proteins, and then the free base preQ_0 is inserted by DpdA or if the phage QueF is able to modify the DNA-bound dPreQ_0, as does the archaeal QueF-L with tRNA. However, Halovirus HVTV-1 contains mainly dPreQ_0, but also a small amount of dADG and dG+. It is possible that the QueF-L transitions between its function as an amidohydrolase to an amidotransferase, but one cannot rule out that the host ArcS could catalyze the reaction, although the PUA domain specific for tRNA binding makes it highly unlikely.

The combination of comparative genomic analyses and experimental validations has allowed pathways for the insertion of dPreQ_0, dPreQ_0, and dG+ in phage genomes to be predicted (Fig. 6). The presence of the minimal set of FolE, QueD, QueE, QueC, and DpdA proteins leads to the insertion of dPreQ_0, as observed in Mycobacterium phage Rosebush and Escherichia phage CAjan genomes (Table 2 and Fig. 3). The replacement of QueC by Gat-QueC leads to the introduction of dG+ (Fig. 2c, Table 1 and previous study). However, it is not known whether Gat-QueC converts preQ_0 into G+ before or after it is inserted into DNA. The function of ArcS homologs in phages/viruses is less clear. Indeed, Vibrio phage nt-1 encodes an ArcS homolog, and its DNA contains mainly dPreQ_0 but also dG+ and dADG (Table 2 and Fig. 3). ArcS was the first G+ synthase identified in archaea. Based on the phage and archaeal ArcS cluster in the SSNs (Supplementary Fig. 2), it is possible that some phage ArcS protein evolved to perform not only an amidotransferase reaction, such as the archaeal ArcS, but also an amidohydrolase reaction, such as the bacterial DpdC. Further biochemical characterization will be required to explore these hypotheses. One cannot exclude the possibility that the small amount of dADG detected in Vibrio phage nt-1, Halovirus HVTV-1, Mycobacterium phage Rosebush, and Escherichia phage CAjan could be the result of the natural oxidation of dPreQ_0.

The discrepancy observed between the SSNs and HHpred predictions for the QueF/QueF-L homologs was resolved by analyzing Streptococcus phage Dp-1 and Halovirus HVTV-1 DNA. HHpred analysis predicted that a homolog of the archaeal QueF-L, which synthesizes G+-tRNA from the preQ_0-tRNA, was encoded by these phages, whereas the SSN analysis predicted that this same protein was part of a group of bacterial QueF proteins (Fig. 4) that synthesize preQ_0 from the free preQ_0 base. We found that Streptococcus phage Dp-1 and Halovirus HVTV-1 were modified by dPreQ_0, confirming the SSN prediction. However, it is unclear whether the reduction occurs on free preQ_0, similar to the bacterial QueF proteins, and then the free base preQ_0 is inserted by DpdA or if the phage QueF is able to modify the DNA-bound dPreQ_0, as does the archaeal QueF-L with tRNA. However, Halovirus HVTV-1 contains mainly dPreQ_0, but also a small amount of dADG and dG+. It is possible that the QueF-L transitions between its function as an amidohydrolase to an amidotransferase, but one cannot rule out that the host ArcS could catalyze the reaction, although the PUA domain specific for tRNA binding makes it highly unlikely.

The combination of comparative genomic analyses and experimental validations has allowed pathways for the insertion of dPreQ_0, dPreQ_0, and dG+ in phage genomes to be predicted (Fig. 6). The presence of the minimal set of FolE, QueD, QueE, QueC, and DpdA proteins leads to the insertion of dPreQ_0, as observed in Mycobacterium phage Rosebush and Escherichia phage CAjan genomes (Table 2 and Fig. 3). The replacement of QueC by Gat-QueC leads to the introduction of dG+ (Fig. 2c, Table 1 and previous study). However, it is not known whether Gat-QueC converts preQ_0 into G+ before or after it is inserted into DNA. The function of ArcS homologs in phages/viruses is less clear. Indeed, Vibrio phage nt-1 encodes an ArcS homolog, and its DNA contains mainly dPreQ_0 but also dG+ and dADG (Table 2 and Fig. 3). ArcS was the first G+ synthase identified in archaea. Based on the phage and archaeal ArcS cluster in the SSNs (Supplementary Fig. 2), it is possible that some phage ArcS protein evolved to perform not only an amidotransferase reaction, such as the archaeal ArcS, but also an amidohydrolase reaction, such as the bacterial DpdC. Further biochemical characterization will be required to explore these hypotheses. One cannot exclude the possibility that the small amount of dADG detected in Vibrio phage nt-1, Halovirus HVTV-1, Mycobacterium phage Rosebush, and Escherichia phage CAjan could be the result of the natural oxidation of dPreQ_0.

The discrepancy observed between the SSNs and HHpred predictions for the QueF/QueF-L homologs was resolved by analyzing Streptococcus phage Dp-1 and Halovirus HVTV-1 DNA. HHpred analysis predicted that a homolog of the archaeal QueF-L, which synthesizes G+-tRNA from the preQ_0-tRNA, was encoded by these phages, whereas the SSN analysis predicted that this same protein was part of a group of bacterial QueF proteins (Fig. 4) that synthesize preQ_0 from the free preQ_0 base. We found that Streptococcus phage Dp-1 and Halovirus HVTV-1 were modified by dPreQ_0, confirming the SSN prediction. However, it is unclear whether the reduction occurs on free preQ_0, similar to the bacterial QueF proteins, and then the free base preQ_0 is inserted by DpdA or if the phage QueF is able to modify the DNA-bound dPreQ_0, as does the archaeal QueF-L with tRNA. However, Halovirus HVTV-1 contains mainly dPreQ_0, but also a small amount of dADG and dG+. It is possible that the QueF-L transitions between its function as an amidohydrolase to an amidotransferase, but one cannot rule out that the host ArcS could catalyze the reaction, although the PUA domain specific for tRNA binding makes it highly unlikely.
From a biological perspective, 7-deazaguanine modifications seem to dramatically decrease the susceptibility of phage genomes to host RM systems. RM systems are one of the major defense systems for bacteria to prevent invasion by foreign DNA. Phages evolved to escape these RM systems by different methods, including modification of their genomic DNA11,17-19. It was previously observed that the genome of Enterobacteria phage 9g contains dG+11 and is fully or partially resistant to a wide variety of restriction enzymes. We also observed that introducing dG+ modifications in the E. coli genome protected against cleavage by EcoRI (Fig. 2). These modifications might also block other DNA-binding proteins that require the nitrogen moiety at position 7 of the guanine to recognize their substrates, the most critical being sigma and transcription factors. However, phages only use the housekeeping sigma factor49, which has an AT-rich recognition sequence50, and encode their own transcription factors51.

Finally, the distribution of these modifications among phages seems to correlate with their host range, namely, bacterial pathogenic species. Interestingly, this was also observed in bacteria, where many pathogens harbor DADG modifications34. Although it is not clear how 7-deazaguanine modifications are spread through phage isolates, these modifications might give a selective advantage to pathogenic species. These 7-deazaguanine-modified phages are also most likely more adapted to propagate in hosts with modified DNA. We can only speculate on how bacteria evolve to counteract this specific antirestriction mechanism. As we were successful in deleting the dpdA gene from Escherichia phage CAjan using a CRISPR-Cas9 technique (see “Methods”), we know that these modifications do not provide resistance against the type II CRISPR-Cas system4. However, as the adaptive system of CRISPR-Cas recognizes the nitrogen in position 7 of the guanines in the PAM52, it is possible that these phages escape degradation by CRISPR-Cas by preventing the adaptation system from binding to its target DNA. One could also imagine that other means of defense, described in recent reviews3,4, provide an efficient protection mechanism against these phages or that some bacteria evolved means of defense yet to be discovered.

**Methods**

**Strains, phages, plasmids, and oligonucleotides.** The bacterial strains used in this study are listed in Supplementary Data 4. Phages are listed in Supplementary Data 5. Plasmids are listed in Supplementary Table 3, and plasmid constructions are described in Supplementary Information. Oligonucleotides are listed in Supplementary Data 6.

**Q detection in tRNA.** Overnight bacterial cultures were diluted 1/100-fold into 5 mL of LB supplemented with 0.4% arabinose and 100 µg/mL ampicillin and grown for 2 h at 37 °C. Cells were harvested by centrifugation at 16,000 × g for 1 min at 4 °C. Cell pellets were immediately resuspended in 1 mL of Trizol (Life Technologies, Carlsbad, CA). Small RNA samples were extracted using the PureLink™ total RNA Isolation Kit from Invitrogen (Carlsbad, CA) according to the manufacturer’s protocol. Purified RNAs were eluted in 50 µL of RNase-free water, and tRNA concentrations were measured with a NanoDrop® ND-1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA). Then 200 ng of RNA was mixed in a 10% acrylamide/bisacrylamide (29:1), Tris EDTA acetate (TAE) 1×, urea 8 M supplemented with 5 µg/mL 3-(acrylamido)-phenylboronic acid, as described in detail previously27. The migrated samples were transferred onto a Biodyne® B Nylon membrane (0.45 µm, Thermo Scientific, Rockford, IL). tRNA samples were detected using a 5′-biotin-CCCTCGGTTGACAGCCGAGG-3′ probe that anneals with tRNA_Cys (GUC) at a final concentration of 0.3 μM and the Chemiluminescent Nucleic Acid Detection Module Kit (Thermo Scientific, Rockford, IL), except that the first blocking buffer was changed to the DIG Easy Hyb buffer (Roche, Mannheim, Germany).

**Restriction assay for deazaquinine presence in plasmid DNA.** E. coli strains containing different variations of pBAD24 or pBAD33 (with or without dpdA or git-queC from Enterobacteria phage 9g, see Supplementary Information) were grown overnight in LB supplemented with ampicillin 100 µg/mL, chloramphenicol 20 µg/mL and 0.2% glucose at 37 °C. Each strain was diluted 100-fold in LB supplemented with ampicillin 100 µg/mL, chloramphenicol 20 µg/mL and 0.4% arabinose and grown for 6 h at 37 °C. Plasmids were extracted using the Qiagen QIAprep Spin Miniprep Kit, and 500 ng of plasmid was digested by EcoRI-HF (New England Biolabs, Ipswich MA) for 1 h at 37 °C in 20 µL of CutSmart buffer. The enzyme was inactivated by 20-min incubation at 80 °C. The samples were run on a 0.5% agarose gel and TAE 1×. The gel was then stained with 0.5 µg/mL ethidium bromide for 30 min, washed 3 times for 15 min in water, and visualized with the Azur Biosystem c200 Gel Doc system (Thermo Fisher Scientific, Waltham, MA, USA).

**Search for phage encoding Q and G+ biosynthesis proteins.** The Viruses nr database from NCBI was queried by three iterations of PSI-BLAST52, with the default set-up as previously suggested35, using the proteins referenced in Supplementary Table 1 known to be involved in Q or G+ biosynthesis, as well as DpdA from Enterobacteria phage 9g, predicted to be involved in the modification of phage DNA, and another DpdA2 from Vibrio phage nt-1, a representative set imported from the OMA database47. For the DpdA from bacteria, the protein sequences were imported from the genomes identified previously through PubSEED35. To generate the protein network, the sequences in fasta format were uploaded and analyzed online by the EFI-EST tool37. Each network was analyzed using the Cytoscape program37, and each family was clustered using the alignment score thresholds indicated in Fig. 3 and Supplementary Fig. 2.

**Identification of the host and their gene content.** The Virus-Host DB37 was used to obtain the host information for each phage identified in this study. For phages not referenced in this database, a manual investigation coupling RefSeq39 and the genome database from NCBI was queried by three iterations of PSI-BLAST53, with the preQ0-specific transporter YhhQ37 was also added. For each virus identified with at least one of these genes, a reverse analysis was performed to find the phage genome against the protein list) to ensure that no protein was missed during the first analysis. The annotations for each identified ortholog were verified by HHpred55.

**SSN generation.** For each protein family (FollC, QueD, QueE, QueC-Gat-QueC, QueF/QueFL, ArcC, and TGT), a representative set was imported from the OMA database56. For the DpdA from bacteria, the protein sequences were imported from the genomes identified previously through PubSEED35. To generate the protein network, the sequences in fasta format were uploaded and analyzed online by the EFI-EST tool37. Each network was analyzed using the Cytoscape program37, and each family was clustered using the alignment score thresholds indicated in Fig. 3 and Supplementary Fig. 2.

**Purification of phage and plasmid DNA.** The purification of each phage DNA in this study was performed specifically for each phage and is described in Supplementary Information.

**Mass spectrometric analysis.** DNA analysis was performed as previously described with several modifications39. Purified DNA (20 µg) was hydrolyzed in 10 mM Tris-HCl (pH 7.9) with 1 mM MgCl2 with benzonase (20 U), DNase I (4 U), calf intestine phosphatase (17 U), and phosphodiesterase (0.2 U) for 16 h at ambient temperature. Following passage through a 10-kDa filter to remove proteins, the filtrate was lyophilized and resuspended to a final concentration of 0.2 µg/µL (based on initial DNA quantity).

Quantification of the modified 2′-deoxynucleosides (dADG, dQ, dPreQ0, dPreQ1, and dG+′) and the four canonical 2′-deoxyribonucleosides (dA, dT, dG, and dC) was achieved by LC-MS/MS and an in-line diode array detector (LC- DAD), respectively. Aliquots of hydrolyzed DNA were injected into a Phenomenex Luna Omega Polar C18 column (2.1 x 100 mm, 1.6 µm particle size) equilibrated with 98% solvent A (0.1% v/v formic acid in water) and 2% solvent B (0.1% v/v formic acid in acetonitrile) at a flow rate of 0.25 mL/min and eluted with a 20 min gradient. The mobile phase started at 100% B for 10 min, 1 min ramp to 2% B for 10 min. The high-performance liquid chromatographic column was coupled to an Agilent 1290 Infinity DAD and an Agilent 6490 triple quadruple mass spectrometer (Agilent, Santa Clara, CA). The column was kept at 40 °C, and the autosampler was cooled at 4 °C. The ultraviolet wavelength of the DAD was set at 260 nm and the electrospray ionization of the mass spectrometer was performed in positive ion mode with the following source parameters: drying gas temperature, 200 °C with a flow of 14 L/min; nebulizer gas
pressure, 30 psi; sheath gas temperature, 400 °C with a flow of 11 L/min; capillary voltage, 3,000 V; and nozzle voltage, 800 V. Enzymes were inactivated by incubation at 80 °C for 20 min. The samples were run on a 0.7% agarose gel and TAE 1×. The gel was then stained for 30 min in 0.5 mM ethidium bromide, washed 3 times for 15 min in water, and visualized with the Azur Biosystem c200 Gel Doc system.

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

Data availability. Data supporting the findings of this work are available within the paper and its Supplementary files. A reporting summary for this article is available as a Supplementary Information file. The datasets generated and analyzed during the current study are available in Supplementary Information or from the corresponding author upon request. The source data are provided as a Source Data file.

Received: 1 May 2019; Accepted: 4 November 2019; Published online: 29 November 2019

References

1. Chapin, M. C., Chopin, A. & Bidnenko, E. Phage abortive infection in lactococci: variations on a theme. Curr. Opin. Microbiol. 8, 473–479 (2005).
2. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. Nat. Rev. Microbiol. 8, 317–327 (2010).
3. Golais, F., Hollý, J. & Vitkovská, J. Coevolution of bacteria and their viruses. Folia Microbiol. (Praha) 58, 177–186 (2013).
4. Makarova, K. S. et al. An updated evolutionary classification of CRISPR-Cas systems. Nat. Rev. Microbiol. 13, 722–736 (2015).
5. Ershova, A. S., Rusinov, I. S., Spirin, S. A., Karyagina, A. S. & Alexeevski, A. V. Role of restriction-modification systems in prokaryotic evolution and ecology. Biochemistry (Mosk.) 80, 1373–1386 (2015).
6. Samson, J. E., Magadan, A. H., Sabri, M. & Moineau, S. Revenge of the phages: defeating bacterial defences. Nat. Rev. Microbiol. 11, 675–687 (2013).
7. Borges, A. L., Davidson, A. R. & Bondy-Denomy, J. The discovery, mechanisms, and evolutionary impact of anti-CRISPRs. Annu. Rev. Virol. 4, 37–59 (2017).
8. Pawluš, A., Davidson, A. R. & Maxwell, K. L. Anti-CRISPR: discovery, mechanism and function. Nat. Rev. Microbiol. 16, 12–17 (2018).
9. Bryson, A. L. et al. Covalent modification of bacteriophage T4 DNA inhibits CRISPR-Cas9. MBio 6, e00648 (2015).
10. Flodman, K. et al. Type II restriction of bacteriophage DNA with ShmaU-derived base modifications. Front. Microbiol. 10, 1–13 (2019).
11. Thiaville, J. J. et al. Novel pyrimidine island modifies DNA with 7-deazaquinoline derivatives. Proc. Natl Acad. Sci. USA 113, E1452–E1459 (2016).
12. Reader, J. S., Metzgar, D., Schimmel, P. & De Crécy-Lagard, V. Identification of four genes necessary for biosynthesis of the modified nucleoside queuosine. J. Biol. Chem. 279, 6280–6285 (2004).
13. Phillips, G. et al. Biosynthesis of 7-deazaguanosine-modified tRNA nucleosides: a new role for GTP cyclohydrolase I. J. Bacteriol. 190, 7876–7884 (2008).
14. McCarty, R. M. & Bandaranid, V. Biosynthesis of pyrrolysoprimidines. Bioorg. Chem. 43, 15–25 (2012).
15. Phillips, G. et al. Discovery and characterization of an amidontransferase involved in the modification of archaeal tRNA. J. Biol. Chem. 285, 12706–12713 (2010).
16. Phillips, G. et al. Diversity of archaeosine synthesis in Crenarchaeota. ACS Chem. Biol. 7, 300–305 (2012).
17. Bon Ramos, A., Bao, L., Turner, B., de Crécy-Lagard, V. & Iwata-Reuyl, D. QueF-like, a non-homologous Archaeosine synthase from the Crenarchaeota. Biomolecules 7, 1–14 (2017).
18. Van Lanen, S. G. et al. From cyclohydrolase to oxidoreductase: discovery of nitrite reductase activity in a common fold. Proc. Natl Acad. Sci. USA 102, 4264–4269 (2005).
19. Steng, B., Reuter, K. & Kliege, G. Mechanism and substrate specificity of tRNA-guanine transglycosylases (TGTs): tRNA modifying enzymes from the three different kingdoms of life share a common catalytic mechanism. ChemBioChem 6, 1926–1939 (2005).
20. Van Lanen, S. G. & Iwata-Reuyl, D. Kinetic mechanism of the tRNA-modifying enzyme S-adenosylmethionine:tRNA ribosyltransferase-isomerase (QueA). Biochemistry 42, 5312–5320 (2003).
21. Miles, V. D., McCarty, R. M., Molnar, G. & Bandaranid, V. Discovery of epoxypseudouridine (oqu) reductase reveals parallels between halorespiration and tRNA modification. Proc. Natl Acad. Sci. USA 108, 7568–7572 (2011).
22. Zallot, R. et al. Identification of a novel epoxypseudouridine reductase family by comparative genomics. ACS Chem. Biol. 12, 844–851 (2017).
23. Carstens, A. B., Kot, W. & Hansen, L. H. Complete genome sequences of four novel Escherichia coli bacteriophages belonging to new phage groups. Genome Announc. 3, e00741–e00745 (2015).
24. Sabri, M. et al. Genome annotation and intraviral interactome for the streptococcus pneumoniae virulent phage Dp-1. J. Bacteriol. 193, 551–562 (2011).
25. Kot, W. et al. Complete genome sequence of Streptococcus pneumoniae virulent phage MS1. Genome Announc. 5, 9–10 (2017).
26. Pedulla, M. L. et al. Origins of highly mosaic mycobacteriophage genomes. Cell 113, 171–182 (2003).
27. Yuan, Y. et al. Identification of the minimal bacterial 2’d-deoxy-7-amido-7-deazaquinoline synthesis machinery. Mol. Microbiol. 110, 469–483 (2018).
28. Kulikov, E. et al. Genomic sequencing and biological characteristics of a novel Escherichia coli bacteriophage 9g, a putative representative of a new Siphoviridae genus. Viruses 6, 5077–5092 (2014).
29. Tsai, R., Corrêa, I. L. R., Xu, M. Y. & Xu, S. Y. Restriction and modification of deoxyarchaeosine (dG)+ containing phage 9 DNA. Sci. Rep. 7, 1–13 (2017).
30. Macáková, M., Boháčová, Š., Perliková, P., Poštová Slavětníková, L. & Hocek, M. Polymerase synthesis and restriction enzyme cleavage of DNA containing 7-substituted 7-deazaquinoline nucleobases. ChemBioChem 16, 2225–2236 (2015).
31. Hutinet, G., Swarlo, M. A. & De Crécy-Lagard, V. Deazaquinoline derivatives, examples of crosstalk between RNA and DNA modification pathways. RNA Biol. 14, 1175–1184 (2017).
32. Hanso, A. D. & Gregory, J. F. Synthesis and turnover of folates in plants. Curr. Opin. Plant Biol. 5, 244–249 (2002).
33. Zallot, R., Yuan, Y. & De Crécy-Lagard, V. The Escherichia coli COG1738 member YhhQ is involved in 7-cyano-deazaquinoline (preQ0) transport. Biomolecules 7, 1–13 (2017).
34. Russell, D. A. & Hatfull, G. F. PhagesDB: the actinobacteriophage database. Bioinformatics 33, 317–322 (2017).
35. Cimci, N. & Huang, R. H. Crystal structure of QueC from Bacillus subtilis: an enzyme involved in preQ0 biosynthesis. Proteins Struct. Funct. Genet. 72, 1084–1088 (2008).
36. Gerli, J. et al. Enzyme function in vitro: influence of sequence similarity tool (EII-EST): a web tool for generating protein sequence similarity networks. Biochem. Biophys. Acta 1854, 1019–1037 (2015).
39. O’Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).

40. Poelen, J. H., Simons, J. D. & Mungall, C. J. Global biotic interactions: an open infrastructure to share and analyze species-interaction datasets. *Ecol. Informatics* **24**, 148–159 (2014).

41. Mihrara, T. et al. Linking virus genomes with host taxonomy. *Viruses* **8**, 10–15 (2016).

42. Overbeek, R. et al. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* **33**, 5691–5702 (2005).

43. Carstens, A. B., Kot, W., Lametsch, R., Neve, H. & Hansen, L. H. Characterisation of a novel enterobacteria phage, CAjan, isolated from rat faeces. *Arch. Virol.* **161**, 2219–2226 (2016).

44. Lemay, M.-L., Renaud, A., Rousseau, G. & Moineau, S. Targeted genome editing of virulent phages using CRISPR-Cas9. *Bio-protocol* 7, 1–19 (2018).

45. Vourvahis, M. et al. Excretion and Metabolism of lersivirine (5-[(3,5-Diethyl-1-(2-hydroxyethyl)(3,5-14C2)-1H-pyrazol-4-yl)oxy]benzene-1,3-dicarbonitrile), a next-generation non-nucleoside reverse transcriptase inhibitor, after administration of [14C]lersivirine to healthy volunteers. *Drug Metab. Dispos.* **38**, 789–800 (2010).

46. Mei, X. et al. Crystal structure of the archaeosome synphase QacF-like–insights into amido transfer and tRNA recognition by the tunnel fold. *Proteins* **165**, 255–269 (2016).

47. Weigele, P. & Raleigh, E. A. Biosynthesis and function of modified bases in bacteria and their viruses. *Chem. Rev.* **116**, 12655–12687 (2016).

48. Lee, Y.-J. et al. Identification and biosynthesis of thymidine hypermodifications in the genomic DNA of widespread bacterial viruses. *Proc. Natl Acad. Sci. USA* **115**, E3116–E3125 (2018).

49. Nechaev, S. & Severinov, K. Bacteriophage-induced modifications of host RNA polymerase. *Annu. Rev. Microbiol.* **57**, 301–322 (2004).

50. Feklistov, A., Sharon, B. D., Darst, S. A. & Gross, C. A. Bacterial sigma factors: a historical, structural, and genomics perspective. *Annu. Rev. Microbiol.* **68**, 357–376 (2014).

51. Yang, H. et al. Transcription regulation mechanisms of bacteriophages. *Bioengineered* **5**, 300–304 (2014).

52. Gleditsch, D. et al. PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA Biol.* **16**, 504–517 (2019).

53. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

54. Lopes, A., Amarir-Boulham, J., Faure, G., Petit, M. A. & Gueruis, R. Detection of novel recombinases in bacteriophage genome unveils Rad52, Rad51 and Gp2.5 remote homologs. *Nucleic Acids Res.* **38**, 3952–3962 (2010).

55. Söding, J. Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**, 951–960 (2005).

56. Altenhoff, A. M. et al. The OMA orthology database in 2018: retrieving evolutionary relationships among all domains of life through richer web and programmatic interfaces. *Nucleic Acids Res.* **46**, D477–D485 (2018).

57. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2503 (2003).

58. Lemay, M. L., Tremblay, D. M. & Moineau, S. Genome engineering of virulent lactococcal phages using CRISPR-Cas9. *ACS Synth. Biol.* **6**, 1351–1358 (2017).

59. Kot, W., Vogensen, F. K., Sørensen, S. J. & Hansen, L. H. DPS—a rapid method for genome sequencing of DNA-containing bacteriophages directly from a single plaque. *J. Virol. Methods* **196**, 152–156 (2014).