https://doi.org/10.1093/nar/gkad622

A novel *Queuovirinae* lineage of *Pseudomonas aeruginosa* phages encode dPreQ$_0$ DNA modifications with a single GA motif that provide restriction and CRISPR Cas9 protection *in vitro*

Nikoline S. Olsen [1,*], Tue K. Nielsen [1], Liang Cui [2], Peter Dedon [2,3], Horst Neve [4], Lars H. Hansen [1,*] and Witold Kot [1,*]

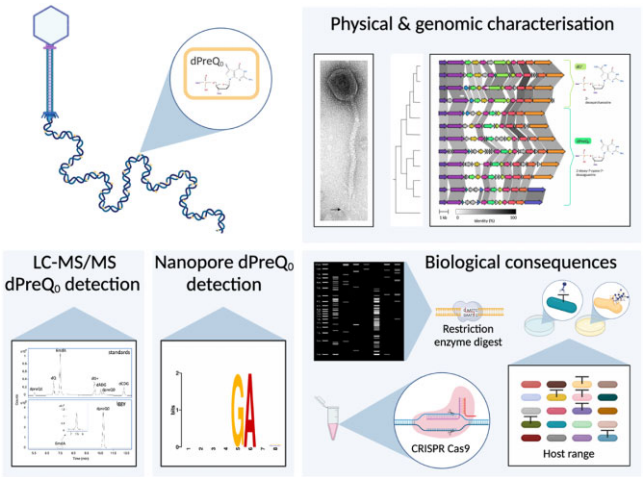
[1]Department of Plant and Environmental Sciences, University of Copenhagen, Frederiksberg C, Denmark, [2]Antimicrobial Resistance Interdisciplinary Research Group, Singapore-MIT Alliance for Research and Technology, Singapore, [3]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, US and [4]Department of Microbiology and Biotechnology, Max Rubner-Institut, Kiel, Germany

Received December 12, 2022; Revised June 02, 2023; Editorial Decision July 03, 2023; Accepted July 14, 2023

ABSTRACT

Deazaguanine DNA modifications are widespread in phages, particularly in those with pathogenic hosts. Pseudomonas phage iggy substitutes ∼16.5% of its genomic 2′-deoxyguanosine (G) with dPreQ$_0$, and the iggy deazaguanine transglycosylase (DpdA) is unique in having a strict GA target motif, not observed previously. The iggy PreQ$_0$ modification is shown to provide protection against both restriction endonucleases and Cas9 (when present in PAM), thus expanding our understanding of the deazaguanine modification system, its potential, and diversity. Phage iggy represents a new genus of Pseudomonas phages within the *Queuovirinae* subfamily; which have very little in common with other published phage genomes in terms of nucleotide similarity (<10%) and common proteins (<2%). Interestingly, shared similarity is concentrated in *dpdA* and preQ$_0$ biosynthesis genes. TEM imaging confirmed a siphovirus morphology with a prolate icosahedral head and a non-contractile flexible tail with one long central tail spike. The observed protective effect of the deazaguanine modification on the iggy DNA may contribute to its broad within-species host range. Phage iggy was isolated on *Pseudomonas aeruginosa* PAO1, but also infects PDO300, PAK, PA14, as well as 10 of 27 tested environmental isolates and 13 of 20 tested clinical isolates of *P. aeruginosa* from patients with cystic fibrosis.

GRAPHICAL ABSTRACT



INTRODUCTION

Bacteriophages (phages) have developed numerous strategies to evade bacterial defence mechanisms including the widest range of DNA modifications observed in any organism, with 21 distinct naturally occurring genomic 2′-deoxynucleotides identified so far (1). It has been verified that the deazaguanine DNA modification system provides phages with an advantage during infection as it protects against host restriction endonucleases (REs) *in vivo* (2). Many variations of the phage deazaguanine modification system have been observed, resulting in highly varying degrees of modification, from < 1% to complete

*To whom correspondence should be addressed. Tel: +45 26338069; Email: sno@plen.ku.dk
Correspondence may also be addressed to Lars H. Hansen. Tel: +45 28752053; Email: lhha@plen.ku.dk
Correspondence may also be addressed to Witold Kot. Email: wk@plen.ku.dk
Present address: Horst Neve, retired.

substitution of 2′-deoxyguanosine (dG) (2–5). It has so far been detected in tailed phages (*Caudoviricetes*) infecting a wide range of Gram-negative bacteria including *Proteobacteria, Firmicutes, Actinobacteria* and *Cyanobacteria*, as well as in viruses of *Archaea* (2–5). The tRNA-guanine transglycosylase paralog DpdA (DeoxyPurine in DNA) is responsible for the substitution of dG with 2-deoxy-7-cyano-7-deazaguanine ($dPreQ_0$) or its derivatives. While some phages encode only DpdA, such as Salmonella phage 7–11, which has ∼0.02% substitution of dG with 2′-deoxy-7-amido-7-deazaguanosine dADG (2), others, such as *Pseudomonas* phage iggy, encode both the GTP cyclohydrolase (FolE), the CPH_4 synthase (QueD), the CDG synthase (QueE), and the 7-cyano-7-deaza-guanine synthase (QueC). This gene set is responsible for the five-step $preQ_0$ biosynthetic pathway (GTP → H_2NTP → CPH4 → CDG → ADG → $preQ_0$) and is coupled with DpdA to introduce $preQ_0$ into DNA post-replication (4). The modification is expected to be sequence-specific, but a DpdA motif has so far only been determined for the *Escherichia coli* phage CAjan (GA and GGC) (6). Some phage genomes contain both 2′-deoxy-$preQ_0$ ($dPreQ_0$) and its derivatives, while others hold only $preQ_0$ derivatives (2). The exact pathways and reactions are not entirely defined, and it is unknown if $preQ_0$ is further modified prior to or following the substitution reaction. Both Mycobacterium phage Rosebush and Escherichia phage CAjan encode DpdA, FolE, and QueDEC and both have ∼25–28% of dG substituted with $dPreQ_0$, though Rosebush genomic DNA also contains 0.03% dADG (2). Escherichia phage 9g has a glutamine amidotransferase class II domain fused to QueC (Gat-QueC), which gives rise to ∼25% of dG being substituted with 2-deoxyarchaeosine (dG+) (7). Streptococcus phage Dp-1 and the *Archaea* virus Halovirus HVTV-1 (no verified DpdA) both encode the amidinotransferase QueF in addition to FolE and QueDEC and both have 2′-deoxy-7-aminomethyl-7-deazaguanine ($dPreQ_1$) substitutions (Dp-1: 4%, HVTV-1: 30%) (2). The DNA of Vibrio phage nt-1 encoding FolE, QueDEC and DpdA2 contains 0.1% $dPreQ_0$, 0.03% dADG and 0.02% dG^+ and the Campylobacter phages of the genera *Firehammervirus* encoding FolE, QueDEC, and no identified DpdA, completely substitute dG with dADG (2,3). However, the sequence distribution of the deazaguanine modifications in the genomes of these phages, and in the many phages encoding other variants of this gene cluster, is still unexplored. A study by Hutinet *et al.* (2019) detected homologues of deazaguanine DNA modification genes in 180 viruses, including 91 phages at the time classified as *Siphoviridae* (2) (this family is now renamed (8)). This number has since increased to at least 125 '*Siphoviridae'* phages of which 50 only encode DpdA. The majority of these belong to two distinct phylogenomic clusters, the *Mycobacterium* phages of the subfamily *Bclasvirinae* and the Gammaproteobacteria phages of the subfamily *Queuovirinae,* though the deazaguanine modification genes have also been detected in phages with *Myoviridae* morphology, such as those of the genera *Kleczkowskavirus* and *Firehammervirus* (3,9).

In this study on phage iggy infecting *Pseudomonas aeruginosa*, we use a synthetic whole genome amplification (WGA) made with canonical unmodified nucleotides, as a reference to detect deazaguanine modifications in the iggy genome by Nanopore sequencing. We also determine the exact modification by liquid chromatography-coupled triple quadrupole mass spectrometry (LC–MS/MS) and verify the protective nature of the modification by *in vitro* restriction enzyme (RE) and Cas9 ribonucleoprotein (RNP) complex analysis. The results broaden our understanding of the deazaguanine modification system and its diversity, not only regarding the (i) chemical composition of the base-substitutions, but also the (ii) assortment of dezaguanine modification motifs in phage genomic DNA. This expands the potential range of interactions between deazaguanine modifications and host defence systems, as demonstrated by the observed Cas9 RNP interaction. In addition, we propose a new genus within the *Queuovirinae* sub-family represented by iggy and the closely related but uncharacterised Pseudomonas phages PBPA162, Pa-X, Pa-V and Pa-oumiga (94.2–93.6%, BLASTn). These five phages share very limited nucleotide similarity (<10%, BLASTn) with all other published phage genomes except within *dpdA* and genes encoding $preQ_0$ biosynthesis. We determine the direct terminal repeats, epigenetics, phylogenetics, virion morphology, and primary receptors and assess the host range of phage iggy towards both reference, environmental, and clinical isolates of *P. aeruginosa*.

MATERIALS AND METHODS

Phage isolation and morphology

The isolation (HiTS method) of phage iggy from wastewater (Damhusåen, BIOFOS, Denmark, 2017) with *P. aeruginosa* PAO1 (NC_002516) as host was described previously (10). All assays were performed with LB-Miller with Ca^{2+} and Mg^{2+} (10 mM) and 4% agarose soft-agar overlays. Purification of phage lysate (0.5 L) and transmission electron microscopy (TEM) imaging were performed as described elsewhere (11).

DNA extraction

High titer lysate (∼10^{10} plaque-forming units (PFU) per ml) of iggy was purified by polyethylene glycol (PEG) precipitation (12), and the DNA was extracted by phenol-chloroform treatment (13). The purified DNA was used for Nanopore sequencing, WGA, LC–MS/MS analysis, RE digestions, and *in vitro* cleavage with Cas9 RNP complexes.

DNA sequencing, assembly and annotation

Phage iggy DNA, which was previously sequenced on the NextSeq platform, Mid Output v2 kit (300 cycles, Illumina), using the Nextera XT DNA kit (Illumina) as described elsewhere (10), was sequenced again on a Nanopore MinION R9.4.1 flowcell using the MinION Mk1B sequencing platform with MinKnow v3.5.5 and the Rapid Barcoding Sequencing kit (SQK-RBK004). Basecalling of raw data was performed with Guppy (v3.4.3). Basecalled reads were processed with the ont_fast5_api toolkit (v2.0.0). Mean read length was 2074 bp and coverage

295x. The Illumina reads were assembled *de novo* with CLC Genomics Workbench 22.0.2 (https://digitalinsights.qiagen.com/) and SPAdes 3.12.0. (14). Genome termini were determined by mapping of Nanopore reads to the whole genome sequence. Calling of ORFs, gene predictions and annotations were done with a customized RASTtk version 2.0 (15) workflow with GeneMark (16), followed by manual curation and verification using BLASTp (non-redundant protein sequences database) (17) and HHpred (PDB_mmCIF70_8_Apr, SCOPe70_2.07, and Pfam-A_v32.0 database) (18). The genome was screened for antibiotic resistance genes (ResFinder 3.1 (19)), restriction modification genes (Restriction-ModificationFinder-1.1 (REBASE) (20); 90% ID threshold, 60% minimum length), virulence factors, and temperate lifestyle genes (PhageLeads (21)). Lifestyle was further assessed by BACPHLIP (22). The iggy genome sequence was curated in Geneious Prime 2020.1.1 (http://www.geneious.com/) and is available at GenBank (MN029011.1) (23).

Phylogenomic and phylogenetic analysis

A total of 69 phage genomes with ≥1% query coverage to phage iggy were identified with BLASTn, and one was omitted from further analyses, as it appears to be a chimera of two phage genomes (Supplementary Table S1). Fifty-eight of these 68 phage genomes encode $preQ_0$ biosynthesis genes and 41 of these are members of the *Queuovirinae* subfamily. A phylogenetic tree (phylogeny.fr One Click mode; alignment by MUSCLE 3.8.31, refinement hereof by Gblocks 0.91b, Phylogeny by PhyML 3.1/3.0 aLRT (24)) was constructed with the terminase large subunit (*TerL*) using the 50 best amino acid matches (BLASTp, non-redundant protein sequences (nr), excluding Bacteria taxid:2). Three additional closely related *Pseudomonas* phages (Pa-X, Pa-V and Pa-oumiga; mash distances 0.024) were identified through the Inphared database (1Apr2023_genomes.fa.msh) (25) using Mash (26). These three phages are not annotated (registered as unverified organisms in the NCBI database), hence for these, *TerL* was manually identified in Geneious Prime 2020.1.1. Xanthomonas phage XAJ2 *TerL* (AMW36122.1) was included as outgroup. To investigate the phylogeny in a broader context, an amino acid-based whole-genome phylogeny with iggy and the 41 *Queuovirinae* against the 3433 phage genomes of the RefSeq database with (ViPTree (27)) was performed. Nucleotide-based intergenomic similarities were estimated with VIRIDIC (28) for Pa-X (MN871472.1), Pa-V (MN871471.1), and Pa-oumiga (MN871474.1) and all identified phages genomes with ≥ 5% query coverage (n = 25) to phage iggy. Whole genome and deazaguanine gene cluster comparisons (amino acid-based) were made with clinker v0.0.23 (29). The whole genome nucleotide alignment with iggy and PBPA162 (MK816297.1) was made with Muscle 3.8.425 in Geneious Prime 2020.1.1. Figures are compiled with Biorender.com.

Quantification of modified 2′-deoxynucleosides by LC–MS/MS

DNA analysis followed our previous publications (6,7). Briefly, quantification of the modified 2′-deoxynucleosides (dADG, dQ, $dPreQ_0$, $dPreQ_1$, dCDG and dG^+) and the four canonical 2′-deoxyribonucleosides (dA, dT, dG and dC) in hydrolyzed DNA was achieved by chromatography-coupled tandem mass spectrometric analysis (LC–MS/MS) and an in-line diode array detector (DAD), respectively. Aliquots of hydrolyzed DNA were injected onto a C18 column coupled to an Agilent 1290 Infinity DAD and an Agilent 6490 triple quadruple mass spectrometer (Agilent, Santa Clara, CA). The UV 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 3000 V and nozzle voltage 800 V. Compounds were quantified in multiple reaction monitoring (MRM) mode with the following m/z transitions: 310.1 → 194.1, 310.1 → 177.1, 310.1 → 293.1 for dADG, 394.1 → 163.1, 394.1 → 146.1, 394.1 → 121.1 for dQ, 292.1 → 176.1 for $dPreQ_0$, 296.1 → 163.1, 296.1 → 121.1, 296.1 → 279.1 for $dPreQ_1$, 309.1 → 193.1, 309.1 → 176.1, 309.1 → 159.1 for dG^+, 311.1 → 177.1, 311.1 → 78.9 for dCDG and 266.1 → 150.1, 266.1 → 108.1 for m6dA. Data acquisition and processing were performed using MassHunter software (Agilent, Santa Clara, CA).

Identification of modified 2′-deoxynucleosides by nanopore

A whole genome amplification (WGA) of iggy was constructed using the Illustra Ready-To-Go GenomiPhi V3 DNA amplification kit (GE Healthcare, Pittsburgh PA US) and debranched with S1 nuclease (Thermo Fisher Scientific, Waltham, MA, USA), as described previously (6). Wild-type (WT) and WGA DNA were sequenced on a Nanopore MinION R9.4.1 flowcell as described above and the detection of modified DNA bases, was performed as described earlier (6) in the Tombo suite (v1.5) (30). Briefly, basecalled fast5 files were 'resquiggled' against the iggy reference genome. Modified bases were detected by comparing the WT with unmodified WGA DNA using the Tombo 'model_sample_compare' function and the fraction of modified reads (ModFrac) were reported per nucleotide position. For motif discovery of DNA recognition sites for the modifying enzyme, 5128 genomic positions with ModFrac values higher than 0.7 were used as input for the MEME software (v5.1.0) (31) with the 'zoops' approach (Zero Or One Occurrence Per Sequence). Five nucleotide positions in either direction of the modified bases were included for motif discovery (total of 11 bases per modified region). With Tombo, a receiver operating characteristic (ROC) curve and a precision-recall curve were plotted with the discovered DNA motif, in order to validate the results. ModFrac values per nucleotide position were imported in R (v4.3.0) for further analyses. Here, a ModFrac cut-off value for defining modified bases was determined based on the ModFrac density distribution for each of the four bases. In order to remove noise and false-positive prediction of base modifications stemming from the neighbouring effect, the ModFrac values of two bases up- and downstream of the identified modification motif (GA) were set to 0 unless adjacent bases were also GA, as described in a previous study (6).

The Modfrac values on a single strand of the iggy genome were visualized in CLC Genomics Workbench 22.0.2.

DNA restriction endonuclease digest analysis

Phage iggy DNA restriction was performed according to manufacturer's instructions in 50 μl reactions of ∼1 μg genomic DNA with REs with GA in target site: *Eco*RI, *Bam*HI, *Hinf*I, *Eco*RV and *Sal*I and with REs without GA in target site: *Nco*I and *Nde*I (Thermo Scientific, MA, USA). DNA integrity of digested samples was subsequently analysed by gel electrophoresis.

In vitro cleavage of DNA with cas9

For the *in vitro* cleavage of DNA using Cas9 RNP complex, two different DNA templates were used, i.e. an *Nde*I restriction fragment of native phage iggy DNA (4135 bp, 27 134–31 268 in iggy genome) and a synonymous PCR fragment (primers in Supplementary Table S2). Both fragments were separated on an 1% agarose gel and extracted using QIAEX®II Gel Extraction Kit (Qiagen, Germany) according to manufacturer's protocol prior to digestion. The Alt-R CRISPR Cas9 nuclease, tracrRNA, and crRNAs were ordered and synthetized by Integrated DNA Technologies (IDT, Belgium). The digestion reaction was set up according to manufacturer's protocol using 50 ng of either native phage DNA or PCR product. Products of digestion were analyzed on a TapeStation 4200 system (Agilent, USA) using High Sensitivity D5000 kit. Efficiency of the reaction was calculated based on integrated area under the curve between template and digestion products.

Host range

Host range was assessed by efficiency of plating (EOP) (32) with PAO1 (33) as reference strain, by spotting (5 μl) triplicates of serial 10-fold dilutions (0 to -7) of ∼1.4×10^{11} PFU/ml on soft-agar overlays with 200 μl host culture. EOP was determined on PAO1 and 52 other *P. aeruginosa* isolates; the reference strains PDO300 (34), PAK (CP020659.1) (35), PA14 (KE136335.1) (36), two transposon generated knockout mutants PAO1 ΔpilA and PA14 ΔpilA, 20 clinical isolates from seven Danish cystic fibrosis (CF) patients collected over a 30-year period (1978–2008) (37), and 27 environmental (sewage) isolates from the Lindberg *P. aeruginosa* typing collection (38). The EOP results are presented as log reduction compared to infection efficiency (PFU/ml) on the reference PAO1, that is log(PFU/ml on PAO1) – log(PFU/ml on isolate). On hosts where no individual plaques were observed but growth was inhibited, the minimum dilution resulting in clearing zones was calculated based on PAO1 EOP.

RESULTS AND DISCUSSION

Morphology and genome characteristics

The siphovirus morphology of iggy was confirmed by TEM imaging (Figure 1A). Phage iggy has a slightly elongated icosahedral head and a flexible non-contractile tail with a long central tail spike with a globular appendage, but no protruding tail fibres. On PAO1 (LB-Miller 0.4% soft-agar overlay) iggy gives rise to Ø ∼3 mm turbid plaques.

The bidirectional genome of phage iggy is 60626 bp double stranded DNA (56.4% GC, no tRNAs detected) with 85 coding sequences (CDSs) and direct terminal repeats (DTR) of 1091 bp. Phage iggy is closely related (query coverage: 97%, identity: 94.2%, BLASTn) with the unclassified and undescribed Pseudomonas phage PBPA162 (MK816297.1, 61 286 bp, 85 CDSs, 56.5% GC, no tRNAs) isolated from sewage in South Korea on an undefined strain of *P. aeruginosa* (Figure 1B). Phage iggy and PBPA162 genomes are highly similar with gene synteny and 73 common proteins (>75% protein sequence similarity, Supplementary Table S3). The largest discrepancies are seen in the peptide chain release factor 1 *prfA* gene (gp34), a small hypothetical gene (gp40) not present in PBPA162, *queC* (gp45) and the hypothetical genes gp46, gp52 and gp77-80 (Figure 1B, Supplementary Table S3). Nucleotide sequence similarity between iggy and all other published phages genomes is minimal (query cover: ≤14%, identity: ≤71%) (Supplementary Table S3). Two of the iggy CDSs encoding hypothetical proteins (gp9 and 78) have no similarity to protein sequences in the NCBI database and seven (gp2, 15, 64, 66, 73, 75 and 85) only share similarity with PBPA162 CDSs (Supplementary Table S3). Putative functions or domains can be assigned to less than half of the CDSs ($n = 36$). These represent six categories (1) DNA replication and host take over, (2) host lysis, (3) DNA packaging, (4) morphogenesis, (5) deazaguanine DNA modification, and (6) hypothetical proteins (Figure 1B, Supplementary Table S3). Both iggy and PBPA162 genomes encode all four genes of the 7-cyano-7-deazaguanine ($preQ_0$) biosynthesis pathway: *queE*, *queC, queD* and *folE,* as well as *dpdA* (Supplementary Table S3).

Phage iggy encodes no identified bacterial virulence or antibiotic resistance genes and the BACPHLIP analysis suggests a virulent lifestyle (0.7625). However, even though no temperate lifestyle genes were detected by a PhageLeads screen, the turbid plaques produced by iggy may indicate a lysogenic potential. Furthermore, manual inspection of iggy CDS reveals both a recombinase protein RecA (gp53) and a CDS (gp60) with homology (>98% HHpred probability scores) to both the positive regulator PrtN and the control of excision (Cox) protein first identified in the temperate Escherichia phage P2 (39). RecA can induce a SOS response in *P. aeruginosa* by cleaving LexA (to induce DNA repair), and can also cleave PrtR, which then upregulates PrtN to increase production of anti-bacterial pyocins (40–42). Both control of excision and induction of pyocin production suggests a temperate lifestyle potential, though induction of the SOS system may also be a lytic strategy to inhibit cell division in order to achieve higher yields of progeny as observed in Phage-Antibiotic-Synergy (PAS) (43,44).

A new genus within the subfamily *queuovirinae*

The intergenomic similarity analysis (VIRIDIC) confidently separates iggy, PBPA162, Pa-X, Pa-V and Pa-oumiga from the other 23 phages with ≥ 5% nucleotide query coverage to iggy (4.1–12.1% intergenomic similarity) (Supplementary Figure S1). Phage iggy (and PBPA162) also have very few proteins (<2% with > 75% similarity) shared with

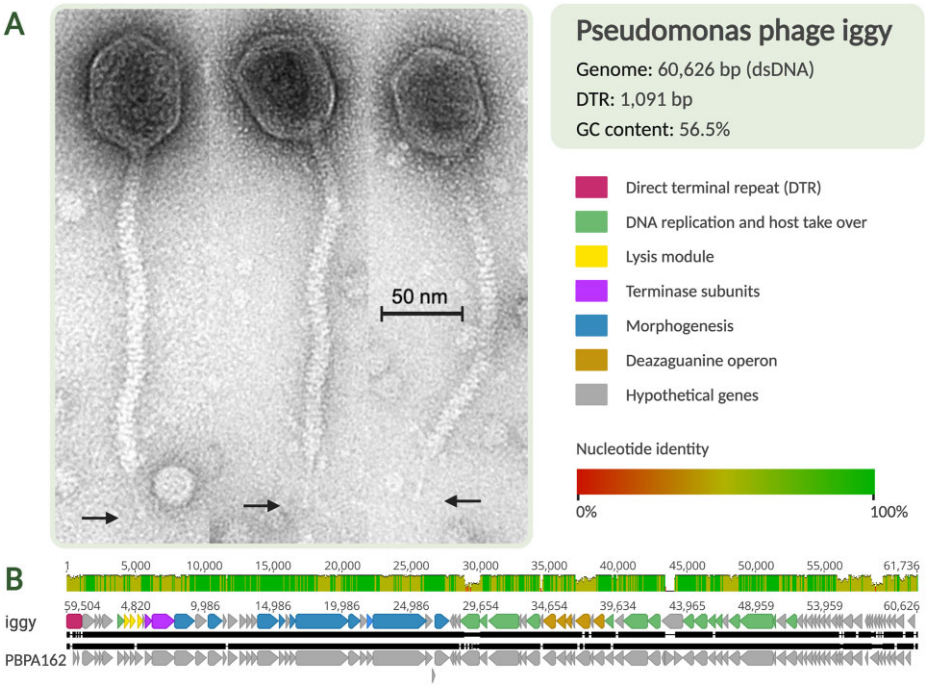


Figure 1. (**A**) Transmission electron micrographs (TEM) of phage iggy with scalebar (50 nm), the globular appendage on the long central tail spike is denoted by arrows. (**B**) Whole genome nucleotide alignment of iggy and PBPA162, the graph (Muscle, Geneious Prime) indicates nucleotide identity (%) by red/green scale bar, gene functions are colour coded according to legend.

all other published phages (Supplementary Table S3). Thus, according to the current ICTV guidelines on genus demarcation (70% nucleotide similarity), iggy, PBPA162, Pa-X, Pa-V and Pa-oumiga constitute a new previously undescribed genus; '*Iggyvirus*'. Despite the very low degree of nucleotide similarity (0.7–9.9%, BLASTn) between iggy and all the phages not belonging to the '*Iggyvirus*' group but included in our analyses (those with ≥ 1% nucleotide query coverage to iggy, n = 68), there are many similarities. Sixty-six of them have (predicted) siphovirus morphology, the majority ($n = 62$) are of comparable genome length (52.4–64.1 kb), though they vary in GC content (41.5–62.1%) and with one exception (Microcystis phage Me-ZS1), they have all been isolated on Alpha-, Beta-, or Gammaproteobacteria (Supplementary Table S4). Interestingly, the sparse nucleotide similarity between iggy and these phages is concentrated in the genes encoding $preQ_0$ biosynthesis and DpdA, the only exception being seven of the eight phages isolated on *Xanthomonadaceae*, Alphaproteobacteria phage PhiJL001 and two partial phage genomes derived from metagenomes (Supplementary Table S4). $PreQ_0$ biosynthesis genes are also a common trait of the recently ratified subfamily *Queuovirinae* comprising ten classified phages in four genera (45). Forty-one of the phages with ≥ 1% nucleotide query coverage to iggy belong to three of the *Queuovirinae* genera (*Nipunavirus, Nonagvirus* and *Seuratvirus*). In a phylogeny (phylogeny.fr), with pa-X, Pa-V, Pa-oumiga and the 50 best amino acid sequence matches (BLASTp) to the iggy *TerL,* the 'iggyviruses' cluster closest with an uncultured *Caudoviricetes* sequenced from a freshwater sample. But they are also in a larger monophyletic clade, with *Queuovirinae* of *Nonagvirus, Seuratvirus* and *Amoyvirus,* as well as the *Vidquintavirus* Pantoea Bacteriophage vB_PagS_Vid5 (NC_042120), which also encode $preQ_0$ biosynthesis and DpdA (Figure 2).

A whole genome amino acid-based phylogeny comprising iggy, the 41 *Queuovirinae* (with ≥ 1% nucleotide query) and the 3433 current phage genomes of the RefSeq database also confidently places iggy (and PBPA162) in a distinct lineage within the *Queuovirinae* subfamily (Supplementary Figure S2). Furthermore, whole genome alignments of iggy and *Queuovirinae* genus representatives reveal analogous bidirectional genome arrangements (Supplementary Figure S3). One strand comprises the lysis module, terminase subunits and morphogenesis genes, and the other strand the host takeover and DNA replication genes, including the deazaguanine modification genes. Consequently, we propose that iggy, PBPA162, Pa-X, Pa-V and Pa-oumiga represent a new genus '*Iggyvirus*' within the *Queuovirinae* subfamily.

The dezaguanine gene cluster composition is conserved across phage genera and host specificity

In previous studies, also the three $dPreQ_0$ derivatives 2′-deoxy-7-aminomethyl-7-deazaguanine ($dPreQ_1$), dADG and dG+ have been detected in phage genomic DNA (2). QueC performs the final two-step conversion (CDG → ADG → $PreQ_0$) in the $preQ_0$ biosynthesis pathway, and is hence highly influential on the type of substitution inserted into DNA by DpdA (Figure 3) (2,7,46). Interestingly, one of the few genes that differ in nucleotide sequence between iggy and PBPA162 is *queC* (57.3% similarity, Supplementary Table S3). The iggy *queC* gene

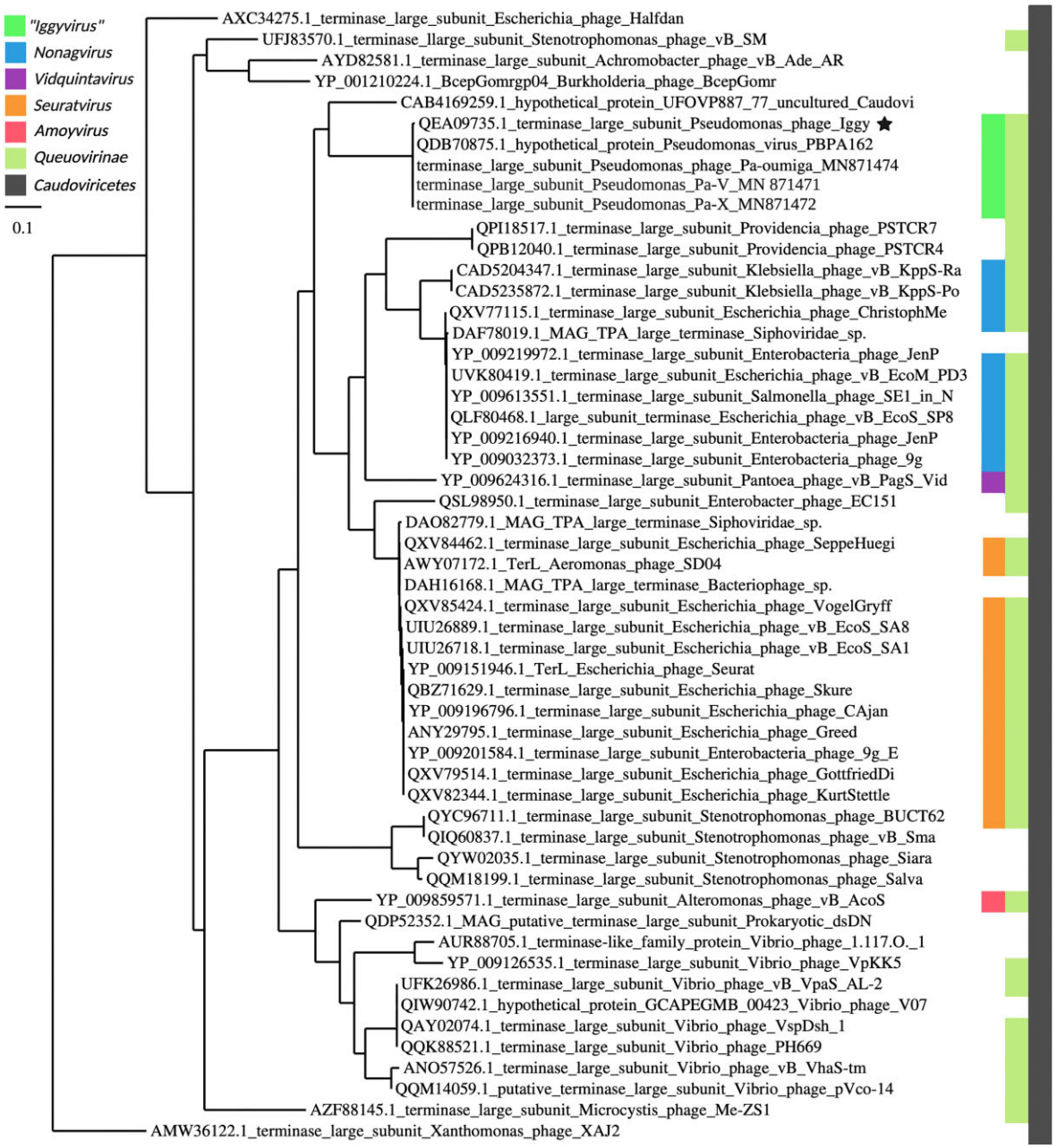


Figure 2. Phylogenomic tree based on the amino acid sequence of the terminase large subunit (*TerL*) of phage iggy (denoted by a star), Pa-X, Pa-V, Pa-oumiga and the top 50 best matches in BLASTp, excluding Bacteria taxid:2. Xanthomonas phage XAJ2 is included as outgroup. The tree was made with phylogeny.fr, alignment by MUSCLE 3.8.31, refined by GBlocks 0.91b, phylogeny PhyML 3.1/4.0 aLRT. Colours denote taxonomy. Legend for branch length in upper left corner.

(1062 bp) is longer than the usually observed phage *queC* genes (Figure 3 & Supplementary Figure S4). The first ∼200 amino acids are conserved (99.91 probability score, HHpred), but the last ∼140 amino acids have no clear match. Phages with a glutamine amidotransferase class II domain fused to QueC (Gat-QueC) have dG^+ modifications as opposed to $dPreQ_0$ in phages encoding conventional QueC (2,46). Phages with conventional QueC, may also encode stand-alone amidinotransferases (QueF-L), which can prompt substitution with $dPreQ_1$ and/or dG^+ and/or dADG, whereas encoding the archaeosine synthase (ArcS) can instigate substitution with dG^+ and/or dADG (2). Despite of the prolonged QueC, iggy strictly substitutes dG with $dPreQ_0$ (Supplementary Figure S4).

The organization of the genes encoding the deazaguanine modification in the large group of phages, spanning a wide variety of hosts, and with ≥ 1% query cover (BLASTn) to iggy, including representatives from all four *Queuovirinae* genera, is highly conserved (Figure 3). The genes follow the order *queE, queC, queD, folE*, and *dpdA*, though some

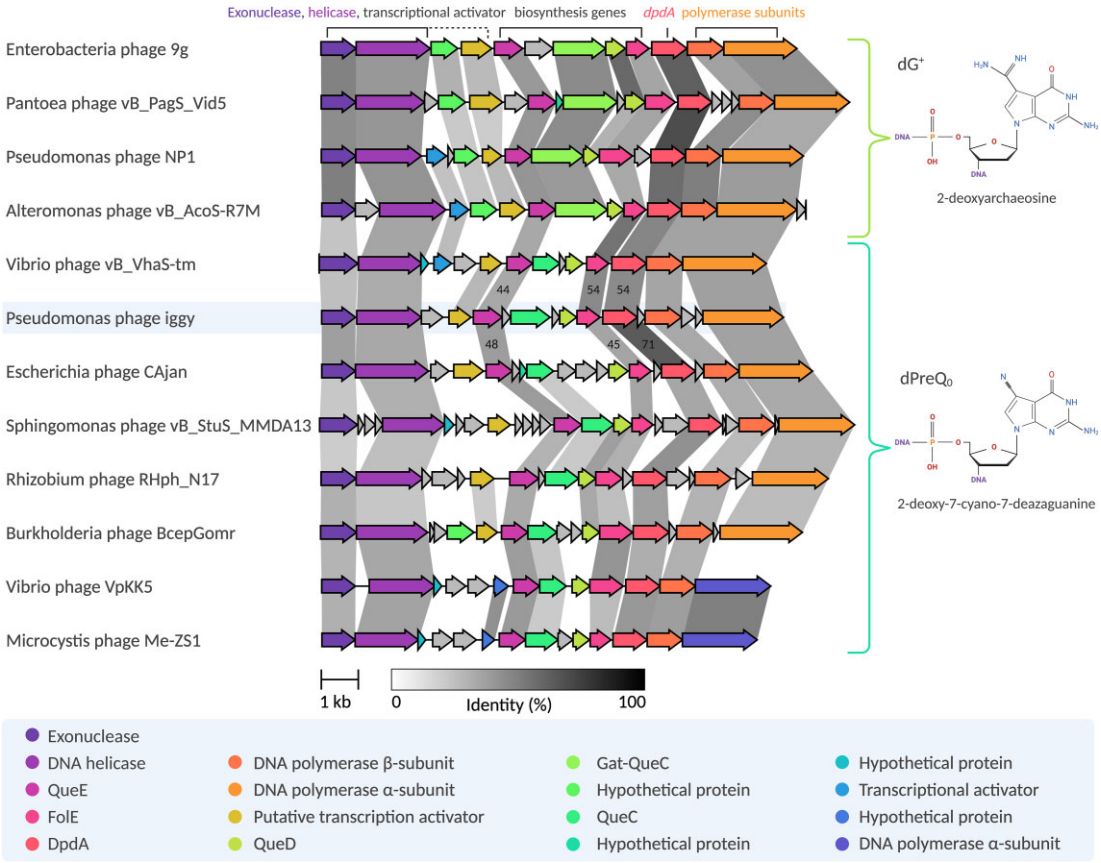


Figure 3. Gene cluster alignments (amino acid-based) of phage genome regions encoding deazaguanine modifications (clinker v0.0.23). Gene clusters are colour coded according to homology (legend with gene products in blue box), grey genes have no similarity to other genes in the alignment but may have a determined or putative function. Similarities (%) of iggy (highlighted in blue) QueE, QueC, QueD, FolE and DpdA to those of VhaS-tm and CAjan are written in links. Predicted resulting deazaguanine modifications and chemical structure hereof is shown on the right side.

phages including iggy have hypothetical genes inserted in between, and Cajan also encodes *yhhQ* (a putative $preQ_0$ transporter) directly downstream of *queC* (Figure 3). The *queE, folE, queD* and *dpdA* genes are conserved in all assessed phages irrespectively of the resulting deazaguanine modification, with 44–71% amino acid similarity between the corresponding iggy genes *queE*, *folE*, *queD* and *dpdA* and those of vB_VhaS-tm and CAjan (Figure 3). But the *queC* genes differ highly; no amino acid similarity exists between iggy and vB_VhaS-tm and CAjan *queC* (identity threshold 0.3). The *Mycobacterium* phages of the genus *Rosebushvirus*, subfamily *Bclasvirinae*, also encode a complete set of $preQ_0$ biosynthesis genes and DpdA, but in the order *dpdA*, *queC*, *queD*, *queE*, *folE* (2). No nucleotide similarity was detected between Mycobacterium phage Rosebush (AY129334.1) deazaguanine modification genes and those with similarity to the iggy deazaguanine modification genes, and only the Rosebush FolE has amino acid similarity (35%) (Supplementary Figure S4). This suggests separate origins of the deazaguanine modification gene clusters in these two distinct groups of phages. Likewise, the genomic organization of the rosebushviruses also differ, their deazaguanine modification genes are located directly upstream of the terminase genes and not next to the polymerase genes as the phages in this study (Supplementary Figure S5).

Pseudomonas phage iggy genomic DNA is modified with $dPreQ_0$

LC–MS/MS analysis identified $dPreQ_0$ as the only deazaguanine substitution in phage iggy DNA. The frequency is estimated by LC–MS/MS to be 16.90% of dG (Table 1). However, a very small fraction of iggy genomic A is also modified. Less than one dA per iggy genome is substituted with N6-methyl-2′-deoxyadenosine (m^6dA), equivalent to mere 0.006% of all dAs per genome (Figure 4, Table 1).

The $dPreQ_0$ modification occurs in a single GA motif

The Nanopore sequencing coverage of WT iggy (295 × cov) and iggy WGA (1000 × cov) (Supplementary Table S5), was sufficient for Tombo DNA modification analysis. The mean read length of the WGA sample was shorter than that of WT, a probable consequence of the WGA treatment. The WT DNA had a lower mean Phred quality score (9.2) compared to WGA DNA (10.7) and the yield from WT was likewise only one third of that of WGA DNA. Likely, the DNA modifications in the WT DNA present a challenge for the pores in the Nanopore flowcell which leads to lower quality and throughput, although the difference in yield may also be a consequence of uneven library pooling (47). The same trend of lower Phred scores for DNA with

Table 1. Frequency of modified nucleotides in iggy genomic DNA as determined by Nanopore sequencing (ModFrac) and LC–MS/MS (per 10^6 2′-deoxynucleosides)

		Nanopore				LC–MS/MS	
Nucleotide	Total count	Modified frequency	Modified count	Corrected modified frequency	Corrected modified count	m^6dA frequency	$dPreQ_0$ frequency
A	12 987	15.7	2046	0.05	7	0.12	–
C	17 141	8.5	1452	0.02	4	–	–
G	17 120	23.3	3990	16.11	2758	–	16.90
T	13 378	10.6	1417	0.02	2	–	–

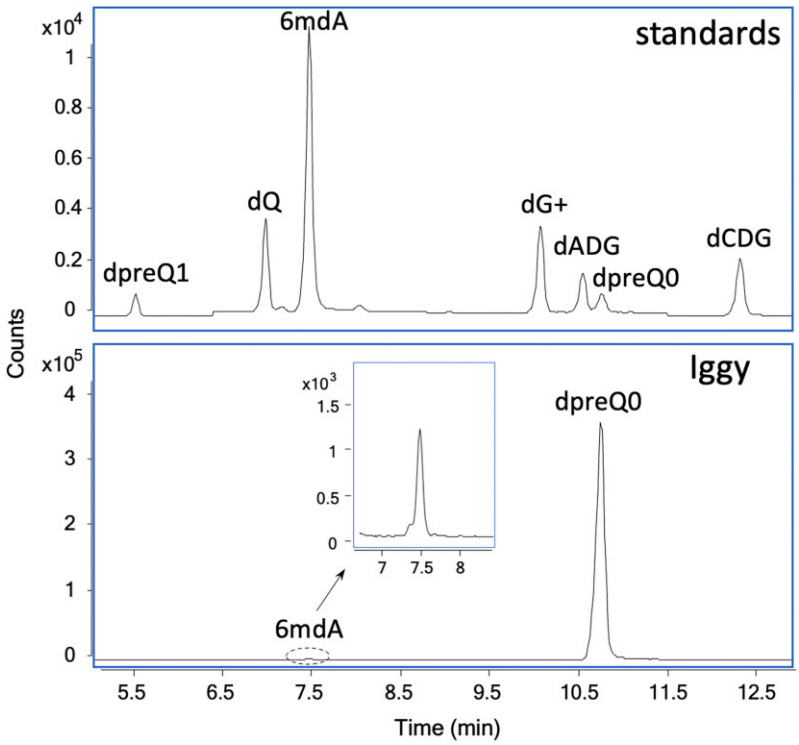


Figure 4. LC–MS/MS chromatograms. Top: standards for dpreQ1, dQ, 6mdA (m6dA), dG+, dADG, dPreQ0 and dCDG. Bottom: hydrolyzed phage iggy genomic DNA.

deazaguanine modifications was also observed in a previous study on phage CAjan (6). A single, clear 'GA' DNA motif was identified from 5128 positions in the iggy genome with ModFrac values higher than 0.7, indicating that at least 70% of aligned Nanopore reads show that a position is modified (Figure 5). No less than 97.14% of all GA sites have ModFrac values above the cut-off, showing that nearly all GA sites ($n = 2840$) are modified. This is similar to previous findings of $dPreQ_0$ modifications and suggests that the system is saturated with $preQ_0$ (6). A ROC curve is a plot test of sensitivity, it shows the performance of a classification model at all classification thresholds by plotting two parameters: the true and the false positive rate, while a precision-recall curve shows the trade-off between precession and recall for different thresholds. To validate the Tombo results, the iggy GA motif, with the G being the modified base, was used to create ROC- and precision-recall curves, with an area under the curve (AUC) and mean average precision (AP) of 0.95 and 0.66, respectively (Supplementary Figure S6). This shows that almost all GA sites are modified, although some false-positive non-GA modified sites are present. As described below, these are from the neighbouring effect of Nanopore sensing of base modifications and can be corrected. Contrary to the dual GA/GGC motif in phage CAjan (6), there is clearly only one GA motif in phage iggy (Figure 5A, Supplementary Figure S6).

The clear-cut GA motif was used to correct false-positive high ModFrac values of the two bases up- and downstream of the GA motif, as these bases are affected by the previously described neighbouring effect (Figure 5B). The neighbouring effects occur because the signal of each sequenced base is affected by multiple adjacent bases within the nanopores (48). The ModFrac values of the two bases adjacent to GA motifs was set to '0', unless they themselves were GA. Density estimates of corrected ModFrac values for each base (Supplementary Figure S7), were evaluated to determine a ModFrac cut-off of 0.75, used to distinguish whether a base is modified or not. With this cut-off, we find that 16.11% of Gs are modified in iggy (Table 1). Correspondingly, 99.75% of all modified Gs ($n = 2758$) are found

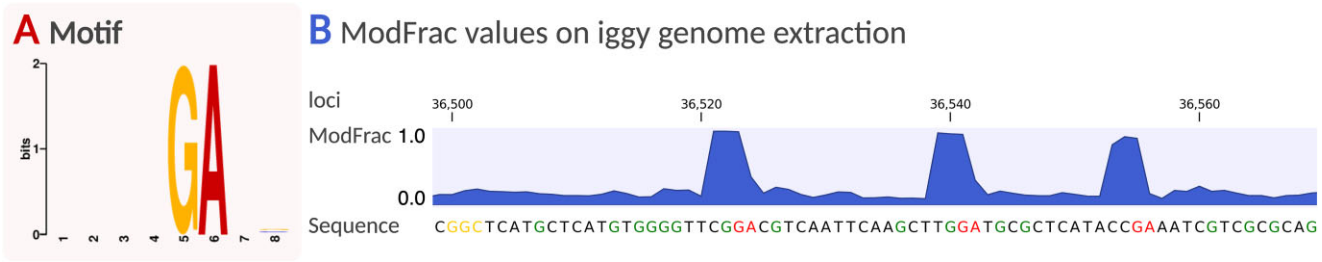


Figure 5. (**A**) MEME (no SSC) generated motif for the dPreQ$_0$ modification in iggy genomic DNA. (**B**) Unadjusted ModFrac values on a representative extraction of the forward strand of the iggy genome (CLC Genomics workbench), the GA motifs are red, GGC are yellow and Gs not followed by an A or GC are green.

in a GA context, strongly supporting the GA motif identified from un-corrected ModFrac values and the associated AUC value of 0.95 (Supplementary Figure S7). In comparison, the applied correction for the neighbouring effect efficiently adjusts the number of modifications for A, C and T to 7, 4 and 2, respectively. These low values may be attributed to unknown low frequency modifications or simply be false positives.

Approximately 16.5% of iggy genomic G is substituted by dPreQ0

The slight discrepancy (<1 percentage point) between the Nanopore (16.11%) and LC–MS/MS (16.90%) substitution estimates is even lower than what was observed in our previous study of phage CAjan (3.6 percentage points) (6). The estimated degree of dPreQ$_0$ modification of dG in iggy is considering both estimates approximately 16.5%, this is substantially lower (36–43%) than what has previously been observed in phage genomes with comparable deazaguanine gene clusters, like CAjan, 9g and Rosebush (∼25–28% of dG) (2). However, the difference is not a consequence of the efficiency of the iggy deazaguanine modification system, but due to the lower count of GA sites in the iggy genome. Indeed, the Tombo analysis estimates an almost complete (97.14% of total iggy GA sites) substitution of the DpdA motif comparable to what was observed in phage CAjan (6). But only 2840 (16.59%) of iggy dG are in a GA motif, while 3213 (23.96%) of CAjan dG are in a GA motif, and the CAjan DpdA is also active on GGC (4.28% of dG) (6).

From the LC–MS/MS analysis it is observed that a small fraction of iggy genomic A is also modified (0.006% of all dAs per genome). This is also somewhat indicated in the Tombo analysis which detected a slightly elevated count of modified A nucleotides (7 dA per iggy genome) compared to the C and T nucleotides (4,2). However, this degree of modification is very close to the limit of detection and cannot confidently be separated from background noise or false positives in the Tombo analysis. Phage iggy does not encode any DNA adenine methylases (Dam), but the propagation host *P. aeruginosa* PAO1 encodes HsdM (PA2735) as part of a type 1 R-M system. The sequence motif of HsdM is GATC(N)$_6$GTC (the underlined base is modified on the opposite strand), in which > 80% of m^{6}dA are found and with a methylation level varying from 65% to 85% depending on growth conditions (49). The iggy genome has 5 HsdM target sites corresponding to 10 m6dA per genome, but they are all partially shared with the iggy DpdA target site (GA). Hence, the incomplete saturation (<1 out of 5 sites per strand) of m^{6}dA in the iggy genome is likely caused by a combination of the HsdM not initially being 100% efficient, the fact that it is competing with the iggy DpdA for binding sites and the potential inhibition of HsdM caused by dPreQ$_0$ (Table 1). It has previously been proposed that dPreQ$_0$, like glc-HMC in T4, at least partially, inhibits adenine methylation during phage infection. In phage Cajan, the Δ*queC* mutants have a 10-fold higher degree of m^{6}dA (0.3% of dA), than the WT (0.03% of dA), supporting this notion (6). Notably, phage 9g, which has dG$^+$ substitutions (25–27%) of dG instead of dPreQ$_0$ in an unknown motif, can be further modified by non-specific methylation (*Eco*GII) to a much higher degree (45% of dA) (50).

The partially shared target site of m^{6}dA and dPreQ$_0$ prevents the distinction between the two modifications by the Tombo software, although a higher degree of Dam modifications would have been detectable from the MEME-based motif discovery. The low degree of dC and dT modification indicated by Nanopore sequencing was not supported by the LC–MS/MS analysis and is thus characterised as representing false positives.

The iggy genomic DNA is resistant towards selected REs

The GA motif was further examined by RE analysis (Supplementary Figure S8). The iggy genomic DNA is, as expected, sensitive towards *Nco*I (C|CATGG) and *Nde*I (CA|TATG) while resistant towards *Eco*RV (GAT|ATC) and *Sal*I (G|TCGAC). However, as also observed for CAjan (dPreQ$_0$, ∼25% of G) (2), it is sensitive towards *Eco*RI and *Bam*HI targeting G|AATTC and G|GATCC, respectively. Conversely, dG$^+$ substitutions seem to protect phage DNA against cleavage from both *Eco*RI and *Bam*HI. Pantoea phage vB_PagS_Vid5, which is predicted to have dG$^+$ substitutions is resistant towards *Bam*HI and phage 9g (dG$^+$, 25–27% of G) is resistant towards ∼100 REs including *Eco*RI and *Bam*HI (7,50,51). Interestingly, Mycobacterium phage Rosebush DNA (dPreQ$_0$, ∼28% of dG) is also resistant towards *Bam*HI (no *Eco*RI sites in genome). However, Rosebush also has a miniscule amount of dADG substitutions (∼0.003% of G) or perhaps other supplementary undetected modifications (2). It has previously been shown that *Hinf*I has a complex pattern of interaction with DNA and is highly susceptible to minor perturbations in secondary structure (52). Correspondingly, the dPreQ$_0$ in iggy DNA

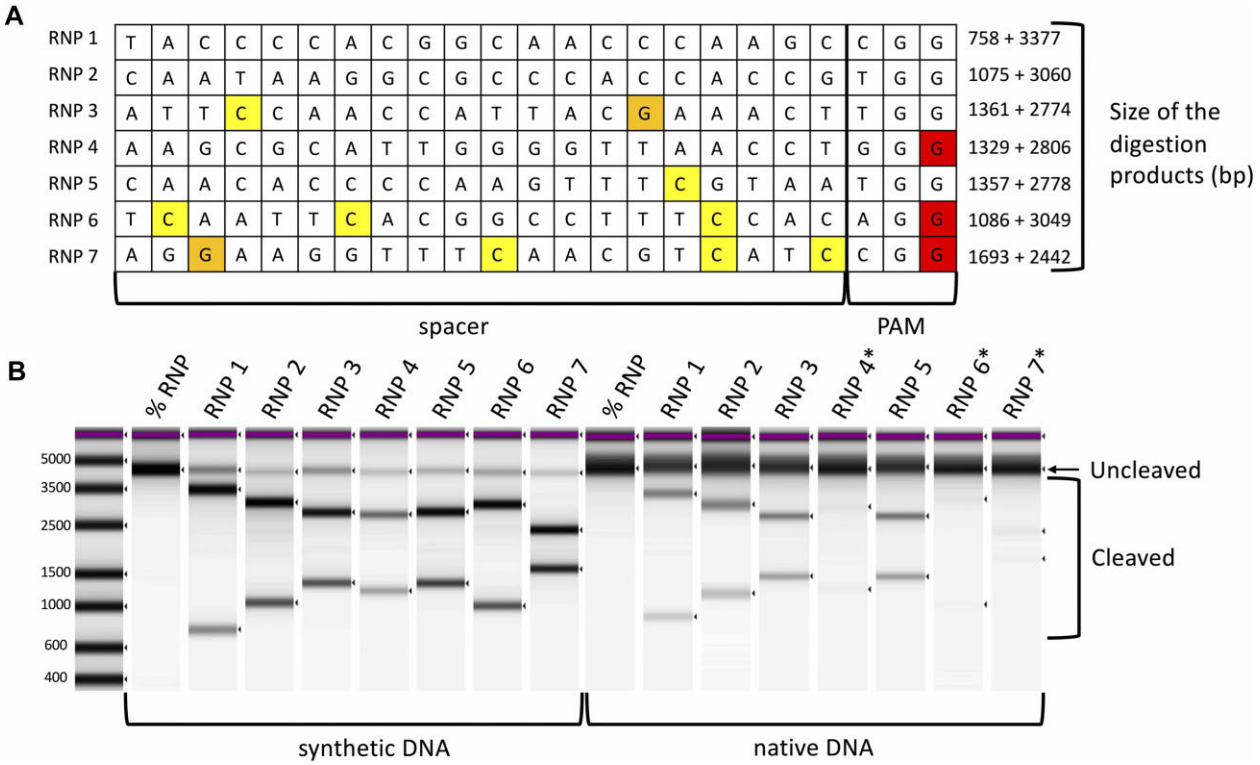


Figure 6. Effect of PAM region modification on RNP complex activity. (**A**) Graphic showing the protospacer sequence of seven individual RNPs used in the experiment. Modifications on the target strand, non-target strand, or PAM region are highlighted in yellow, orange, and red, respectively. Digestion product sizes are indicated on the right side. (**B**) Electropherograms of the seven individual RNPs targeting synthetic DNA with canonical bases only (lanes 2–9) or synonymous, native DNA obtained from phage iggy DNA (lanes 10–17). The first lane represents the size ladder, while lanes 2 and 10 serve as controls without RNPs. The positions of uncleaved and cleaved fragments are indicated on the right side. RNPs 4, 6 and 7, which recognize a modified PAM and have 75% reduction in cleavage efficiency are marked with an asterisk (*).

appears to interfere with effectivity of *Hinf*I. Even though *Hinf*I is clearly able to restrict some sites, it is far from being active at all 164 sites in the iggy genome.

The dPreQ0 modification provides protection towards *in vitro* digestion with cas9

To further investigate the defensive nature of $dPreQ_0$ and whether it also protects against adaptive immune systems such as CRISPR-Cas, we identified seven protospacers based on the Nanopore motif analysis for testing with Cas9 RNP complexes. Reactions were made on both native iggy genomic DNA with modifications ($dG \rightarrow dPreQ_0$) in predicted sites and a synthetised version of the same fragment containing only canonical bases. The crRNAs were selected to target genomic sites with different context of modifications; in the protospacer adjacent motif (PAM), PAM-proximal and distal region, and lack of modifications. The positions of modifications are summarised in Figure 6. For digestions made on fragments with canonical bases, the average efficiency of reaction was 84.5 +/- 5.5% regardless of the crRNA used in the RNP. In all cases examined, the efficiency of DNA digestion was observed to be lower for the native iggy DNA. This decrease in efficiency is likely attributable to the co-purification of additional *NdeI* restriction fragments. Nevertheless, the reaction remained specific, and no extra products resulting from digestion were detected. The proportion of the 4135 bp fragment of interest within the native DNA template was determined to be 32.8%. Intriguingly, the efficiency of the digestion process in native DNA exhibited a significant discrepancy between sites with or without a modification in the PAM sequence (NGG for cas9). In instances where a modification in PAM was present, the calculated efficiency experienced a reduction of 75% (from $27.7 \pm 0.5\%$ to $7.1 \pm 0.9\%$, see also Figure 6).

It is known that PAM is crucial for DNA interrogation (53) and kinetic analyses have shown that 7-deaza guanine modifications in PAM destabilize target binding in the type I-F CRISPR system of *P. aeruginosa* (54). Furthermore, it was recently reported that deaza-modified nucleobases have altered pK_a values which can alter interactions with nucleic acids and proteins (55). The results obtained in our experiment suggest that defence against the Cas9 RNP complex is provided already in the PAM scanning phase, possibly before the DNA is unwound and checked for the complementarity with crRNA (RNP 4, 6 and 7, Figure 6) (56). Furthermore, the presence of modified bases does not seem to affect the DNA cleavage by the Cas9 RNP complex regardless of the modification being present in the PAM-proximal seed sequence (RNP 5, Figure 6) or PAM-distal region (RNP 3). Strains of *P. aeruginosa* have substantial genetic diversity and clinical strains harbour a range of CRISPR-Cas systems such as type I-F, I-E and I-C (57). The many types

Table 2. Host range assessed by efficiency of plating (EOP), presented as log(10) reduction in infectivity (PFU/ml) of averages of triplicate PFU counts with PAO1 as reference (1.4×10^{11} PFU/ml). A plus (+) denotes the detection of clearing zones where no individual plaques were observed in any dilutions. Clinical strain denotation: CF patient/isolate/clone

	Number	Strain	Isolation (year)	Mucoid[a]	EOP (log reduction)
Reference	R0	PAO1	-	NM	Reference
	R1	PDO300[a]	-	M	0.4
	R2	PAK	-	-	4.9
	R3	PA14	-	-	5.7
Mutant	M1	PAO1 ΔpilA	2021	-	-
	M2	PA14 ΔpilA	-	-	-
Clinical isolates	C1	CF66/20538	1992	-	-
	C2	CF66/4085a	2002	-	-
	C3	CF66/5561a	2008	-	-
	C4	CF66/5561b	2008	M	-
	C5	CF104/89378	2004	-	+
	C6	CF128/1398a/DK-1	1992	M	3.5
	C7	CF128/1398b/DK-1	1992	NM	3.1
	C8	CF128/8503a/DK-1	2002	M	3.2
	C9	CF128/8503b/DK-1	2002	NM	3.0
	C10	CF89/15061	1978	NM	4.4
	C11	CF238/82847b/DK-1	2001	NM	3.2
	C12	CF46/474a	1988	-	3.4
	C13	CF46/474b	1988	-	4.0
	C14	CF46/52329a	1997	-	3.7
	C15	CF46/52329b/DK-1	1997	NM	3.8
	C16	CF46/1395a	2003	-	5.4
	C17	CF30/19731b/DK-1	1992	NM	+
	C18	CF30/75887a	2001	-	-
	C19	CF30/75887b/DK-1	2001	NM	3.5
	C20	CF/NA	-		3.9
Environmental isolates	E1	PsF7	-	-	-
	E2	Ps16A	-	-	+
	E3	Ps16B	-	-	2.6
	E4	Ps21	-	NM	0.6
	E5	Ps24A	-	-	-
	E6	Ps24B	-	-	-
	E7	Ps24C	-	-	5.1
	E8	Ps31	-	-	4.9
	E9	Ps44	-	-	3.8
	E10	PsF7	-	-	-
	E11	PsF8	-	NM	-
	E12	Ps10A	-	-	-
	E13	PsF10B	-	-	0.6
	E14	Ps109	-	NM	-
	E15	Ps352	-	-	+
	E16	Ps119x	-	NM	+
	E17	PsM4A	-	-	2.8
	E18	PsM4B	-	-	0.3
	E19	PsM6	-	NM	-
	E20	Ms1214A	-	-	4.3
	E21	Ps1214B	-	-	+
	E22	PsCol2	-	NM	-
	E23	PsCol11	-	-	+
	E24	PsCol18	-	NM	1.5
	E25	PsCol21	-	NM	+
	E26	PsCol188A	-	-	+
	E27	PsCol188B	-	-	-

[a]M: mucoid, NM: non-mucoid (37).

and sub-types of CRISPR-Cas systems have evolved diverse PAM sequences and readout mechanisms to counter escape strategies, they have assorted stringency and some are able to recognize multiple PAM sequences (56,58). Future studies are needed to determine if phage-encoded $dPreQ_0$ does indeed provide protection against host Cas cleavage during infection. But these results suggests that $dPreQ_0$ in iggy and other phage genomes are indeed a part of their counter-defence strategy against CRISPR-Cas.

Pseudomonas phage iggy has a broad host range on clinical and environmental *P. aeruginosa* isolates

Phage iggy successfully infects and forms plaques on its isolation host PAO1, the reference strains PDO300, PAK and PA14, on 13 out of 20 clinical isolates and on 10 out of 27 environmental *P. aeruginosa* strains. In addition, clearing zones without visible plaque formation were observed on two additional clinical strains and seven of the environmental strains (Table 2). The infectivity (here defined

as PFU/ml) of iggy on PDO300 is in the same range as on PAO1, but is ∼5 log lower on PAK and ∼6 log lower on PA14. The main cell surface receptors on *P. aeruginosa* used by phages are lipopolysaccharides (LPS) and type IV pili (T4P) (59). T4P are virulence factors used for adherence, biofilm formation, DNA uptake, and twitching motility composed of single type IVa pilin protein encoded by the *pilA* gene which occurs as at least five distinct alleles associated with distinct accessory genes (59,60). PAO1, PDO300 and PAK all encode group II *pilA*, whereas PA14 encodes group III *pilA* (60). PAK is a virulent lab-strain characterised by high expression of pili and flagella (35). To assess whether iggy utilises pili as phage receptor, infectivity tests were performed with both a PAO1 Δ*pilA* and a PA14 Δ*pilA* transposon knockout mutant. The loss of pili caused prevention of infection and thus strongly suggests that iggy uses pili as primary receptors for initial attachment to the host, though the secondary receptors for irreversible binding remain unknown.

The 10 environmental *P. aeruginosa* strains sensitive to iggy represent a broad range of *P. aeruginosa* phage types (38). The infectivity of iggy on four of them (E4, E13, E18 and E24) is comparable (0.3–1.5 log reduction) to infectivity on PAO1, while it is ∼3–5 log lower on the remaining phage types (Table 2). Phage iggy also infects the clinical isolates with relatively high infectivity, only ∼3–4 log lower than on PAO1. The sensitive strains originate from six different CF patients and the isolates retrieved from patients over time have remained sensitive over decades (Table 2). Though there is a slight decrease in infectivity in the isolates collected from a single patient since 1988 (C12-C16), for which iggy infects the 2003 isolate with ∼6 log lower efficiency than PAO1. Mucoid conversion of *P. aeruginosa* isolates from CF lungs are frequently observed (34,61) and iggy infects both non-mucoid and mucoid clinical isolates as well as PDO300. The inactivation (ΔG430) of the *mucA* gene in PDO300, an isogenic derivative of PAO1, causes a constitutive expression of alginate biosynthesis genes leading to a stable mucoid phenotype (34). The ability of iggy to infect this phenotype increases its potential for use as a phage therapy agent, though further studies are required to fully assess the suitability of iggy including biofilm assays, co-treatment with antibiotics, *in vivo* studies of pharmacokinetics, and -dynamics and verification of a strictly lytic lifestyle.

The occasionally observed phage propagation inhibition devoid of plaque formation suggests that the iggy infection cycles was incomplete on these isolates ($n = 9$) (Table 2). In most instances ($n = 7$) clearing zones were only observed when high phage concentrations ($\geq 1.4 \times 10^7$ PFU/ml) were spotted. It cannot be ruled out that these clearing zones are caused by lysis from without (62), but they may also be a consequence of host encoded defence systems (including the altruistic abortive infection (Abi) systems (63)), which result in cell death but prevent propagation of phage progeny and hence do not lead to plaque formation. On two isolates (C5 and E25), turbid lysis was observed when adding as few as ∼700 to ∼7000 PFU, rendering lysis from without highly unlikely. Bacteria carry a multitude of anti-phage defence systems, oftentimes organized in genomic defence-islands (64). The $preQ_0$ modification in phage iggy DNA may counter host defence-systems targeting non-self-DNA, such as RM-systems and CRISPR-Cas, by preventing recognition or binding. Through this, the $preQ_0$ modification may contribute to the relatively broad host range of phage iggy, observed on both the reference, clinical and environmental isolates. $PreQ_0$ modifications have previously been shown to protect against a broad range of restriction enzymes (6). The theory, that preQ0 modifications are part of the phage anti-defence system, is supported in our study by RE digestion and Cas9 RNP assays of iggy DNA showing *in vitro* resistance towards selected restriction enzymes with GA in their target and CRISPR Cas9 when dG in PAM is modified.

In conclusion, the iggy DpdA is unique in having a single GA motif resulting in ∼16.5% substitution of dG with $dPreQ_0$ in the iggy genome. We found that at least 97% of GA sites were modified, indicating an efficient DpdA modifying enzyme that saturates the modification sites. We show that DpdAs can have distinct recognition sites, as the CAjan DpdA binds to both GA, GGC and GGT whereas the iggy DpdA only has affinity for GA. Diversity in the DpdA DNA and protein sequences in distinct phage groups indicates that the true range of DpdA motifs is even broader. Furthermore, we show that $preQ_0$ modifications in PAM sites significantly lower the DNA cleavage activity of Cas9 RNP complex *in vitro*. This expands the potential range of interactions between deazaguanine modifications and host defence systems across phage-host couples and have interesting implications in terms of potential involvement of deazaguanine modifications in epigenetic regulation. Based on the separate clustering in phylogenomic analyses, the low degree of nucleotide sequence similarity (<10%) and few common proteins (<2% with > 75% similarity) shared between iggy (and PBPA162, Pa-X, Pa-V, Pa-oumiga) and all other published phages, we propose that these five phages represent a new genus. This new genus has analogous genome composition and characteristics with phages within the subfamily *Queuovirinae*, class *Caudoviricetes*, and phylogenomic analyses confirm this taxonomic assignment. Phage iggy efficiently infects a broad range of *P. aeruginosa* isolates under the given conditions, possibly aided by the protective nature of the deazaguanine modification, but additional studies are required to validate the suitability of iggy for phage therapy.

DATA AVAILABILITY

The commands for modified base detection using Tombo and R are provided in Supplementary Data, Supplementary Table S6.

The phage iggy annotated genome has been deposited in NCBI under accession number MN029011.1.

The Nanopore sequencing data is available in the NCBI Sequence Read Archive (SRA) with the following accession numbers SRR22407576 for WGA and SRR22405777 for WT.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Associate Professor Oana Ciofu (University of Copenhagen) and Professor Thomas Bjarnsholt (University of Copenhagen, Rigshospitalet) for supplying us with *P. aeruginosa* isolates, and the BIOFOS wastewater treatment plant Damhusåen, a member of the Danish Water and Wastewater Association (DANVA) for supplying us with inlet wastewater samples. Angela Back (Max Rubner-Institut, Kiel) is acknowledged for her technical assistance with the electron microscopic analysis.

FUNDING

Human Frontier Science Program Grant [RGP0024/2018]; Villum Experiment Grant [17595]; National Research Foundation of Singapore through the Singapore-MIT Alliance for Research and Technology Antimicrobial Resistance Interdisciplinary Research Group (to L.C. and P.). Funding for open access charge: Human Frontier Science Program Grant [RGP0024/2018].

Conflict of interest statement. None declared.

REFERENCES

1. Lee,Y.J. and Weigele,P.R. (2021) Detection of modified bases in bacteriophage genomic DNA. *Methods Mol. Biol.*, **2198**, 53–66.
2. Hutinet,G., Kot,W., Cui,L., Hillebrand,R., Balamkundu,S., Gnanakalai,S., Neelakandan,R., Carstens,A.B., Lui,C.F., Tremblay,D. *et al.* (2019) 7-Deazaguanine modifications protect phage DNA from host restriction systems. *Nat. Commun.*, **10**, 5442.
3. Crippen,C.S., Lee,Y.-J., Hutinet,G., Shajahan,A., Sacher,J.C., Azadi,P., de Crécy-Lagard,V., Weigele,P.R. and Szymanski,C.M. (2019) Deoxyinosine and 7-deaza-2-deoxyguanosine as carriers of genetic information in the DNA of Campylobacter viruses. *J. Virol.*, **93**, e01111-19.
4. Hutinet,G., Lee,Y.-J., de Crécy-Lagard,V. and Weigele,P.R. (2021) Hypermodified DNA in viruses of *E.coli* and Salmonella. *EcoSal Plus.*, **9**, 2.
5. Lin,W., Li,D., Sun,Z., Tong,Y., Yan,X., Wang,C., Zhang,X. and Pei,G. (2020) A novel freshwater cyanophage vB_MelS-me-ZS1 infecting bloom-forming cyanobacterium microcystis elabens. *Mol. Biol. Rep.*, **47**, 7979–7989.
6. 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 preQ0deazaguanine modifications in bacteriophage CAjan DNA using Nanopore sequencing reveals same hypermodification at two distinct DNA motifs. *Nucleic Acids Res.*, **48**, 10383–10396.
7. 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 SciUS A*, **113**, e1452–e1459.
8. Turner,D., Kropinski,A.M. and Adriaenssens,E.M. (2021) A roadmap for genome-based phage taxonomy. *Viruses*, **13**, 506.
9. Santamaría,R.I., Bustos,P., Sepúlveda-Robles,O., Lozano,L., Rodríguez,C., Fernández,J.L., Juárez,S., Kameyama,L., Guarneros,G., Dávila,G. *et al.* (2014) Narrow-host-range bacteriophages that infect rhizobium etli associate with distinct genomic types. *Appl Environ Microbiol.*, **80**, 446–454.
10. Olsen,N.S., Hendriksen,N.B., Hansen,L.H. and Kot,W. (2020) A new high-throughput screening (HiTS) method for phages – Enabling crude isolation and fast identification of diverse phages with therapeutic potential. *PHAGE*, **1**, 137–148.
11. Olsen,N.S., Lametsch,R., Wagner,N., Hansen,L.H. and Kot,W. (2022) Salmonella phage akira, infecting selected *Salmonella enterica* Enteritidis and Typhimurium strains, represents a new lineage of bacteriophages. *Arch Virol*, **167**, 2049–2056.
12. Wood,E. (1983) Molecular cloning. A laboratory manual. *Biochem Educ.*, **11**, 82.
13. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) In: *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press.
14. Bankevich,A., Nurk,S., Antipov,D., Gurevich,A.A., Dvorkin,M., Kulikov,A.S., Lesin,V.M., Nikolenko,S.I., Pham,S., Prjibelski,A.D. *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.*, **19**, 455–477.
15. Brettin,T., Davis,J.J., Disz,T., Edwards,R.A., Gerdes,S., Olsen,G.J., Olson,R., Overbeek,R., Parrello,B., Pusch,G.D. *et al.* (2015) RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep.*, **5**, 8365.
16. Besemer,J. and Borodovsky,M. (2005) GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res.*, **33**, W451–W454.
17. 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.
18. Soding,J., Biegert,A. and Lupas,A.N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.*, **33**, W244–W248.
19. Zankari,E., Hasman,H., Cosentino,S., Vestergaard,M., Rasmussen,S., Lund,O., Aarestrup,F.M. and Larsen,M.V. (2012) Identification of acquired antimicrobial resistance genes. *J Antimicrob. Chemother.*, **67**, 2640–2644.
20. Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2015) REBASE-a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.*, **43**, D298–D299.
21. Yukgehnaish,K., Rajandas,H., Parimannan,S., Manickam,R., Marimuthu,K., Petersen,B., Clokie,M.R.J., Millard,A. and Sicheritz-Pontén,T. (2022) PhageLeads: rapid assessment of phage therapeutic suitability using an ensemble machine learning approach. *Viruses*, **14**, 342.
22. Hockenberry,A.J. and Wilke,C.O. (2021) BACPHLIP: predicting bacteriophage lifestyle from conserved protein domains. *PeerJ*, **9**, e11396.
23. Benson,D.A., Cavanaugh,M., Clark,K., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Sayers,E.W. (2017) GenBank. *Nucleic Acids Res.*, **45**, D37–D42.
24. Dereeper,A., Guignon,V., Blanc,G., Audic,S., Buffet,S., Chevenet,F., Dufayard,J.F., Guindon,S., Lefort,V., Lescot,M. *et al.* (2008) Phylogeny.Fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.*, **36**, 465–469.
25. 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,***2**, 214–223.
26. Ondov,B.D., Treangen,T.J., Melsted,P., Mallonee,A.B., Bergman,N.H., Koren,S. and Phillippy,A.M. (2016) Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.*, **17**, 132.
27. Nishimura,Y., Yoshida,T., Kuronishi,M., Uehara,H., Ogata,H. and Goto,S. (2017) ViPTree: the viral proteomic tree server. *Bioinformatics*, **33**, 2379–2380.
28. Moraru,C., Varsani,A. and Kropinski,A.M. (2020) VIRIDIC—A novel tool to calculate the intergenomic similarities of prokaryote-infecting viruses. *Viruses*, **12**, 1268.
29. Gilchrist,C.L.M. and Chooi,Y.-H. (2021) clinker & clustermap.Js: automatic generation of gene cluster comparison figures. *Bioinformatics*, **37**, 2473–2475.
30. Stoiber,M.H., Quick,J., Egan,R., Lee,J.E., Celniker,S.E., Neely,R., Loman,N., Pennacchio,L. and Brown,J.B. (2017) *De novo* identification of DNA modifications enabled by genome-guided nanopore signal processing. bioRxiv doi: https://doi.org/10.1101/094672, 10 April 2017, preprint: not peer reviewed.
31. Bailey,T.L. and Elkan,C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **2**, 28–36.
32. Letarov,A.V. and Kulikov,E.E. (2018) Determination of the bacteriophage host range: culture-based approach. *Methods Mol. Biol.*, **1693**, 75–84.

33. Klockgether,J., Munder,A., Neugebauer,J., Davenport,C.F., Stanke,F., Larbig,K.D., Heeb,S., Schöck,U., Pohl,T.M., Wiehlmann,L. *et al.* (2010) Genome diversity of *Pseudomonas aeruginosa* PAO1 laboratory strains. *J Bacteriol.*, **192**, 1113–1121.
34. Bragonzi,A., Wiehlmann,L., Klockgether,J., Cramer,N., Worlitzsch,D., Döning,G. and Tümmler,B. (2006) Sequence diversity of the mucABD locus in *Pseudomonas aeruginosa* isolates from patients with cystic fibrosis. *Microbiology*, **152**, 3261–3269.
35. Cain,A.K., Nolan,L.M., Sullivan,G.J., Whitchurch,C.B., Filloux,A. and Parkhill,J. (2019) Complete genome sequence of *Pseudomonas aeruginosa* reference strain PAK. *Microbiol. Resour. Announc.*, **8**, e00865-19.
36. Mikkelsen,H., McMullan,R. and Filloux,A. (2011) The *Pseudomonas aeruginosa* reference strain PA14 displays increased virulence due to a mutation in ladS. *PLoS One*, **6**, e29113.
37. Henriksen,K., Rørbo,N., Rybtke,M.L., Martinet,M.G., Tolker-Nielsen,T., Høiby,N., Middelboe,M. and Ciofu,O. (2019) *P. Aeruginosa* flow-cell biofilms are enhanced by repeated phage treatments but can be eradicated by phage–ciprofloxacin combination —monitoring the phage – *P. aeruginosa* biofilms interactions. *Pathog Dis.*, **77**, ftz011.
38. Lindberg,R.B. and Latta,R.L. (1974) Phage typing of *Pseudomonas aeruginosa* : clinical and epidemiologic considerations. *J Infect Dis.*, **130**, S33–S42.
39. Yu,A. and Haggard-Ljungquist,E. (1993) The Cox protein is a modulator of directionality in bacteriophage P2 site- specific recombination. *J Bacteriol.*, **175**, 7848–7855.
40. Dingemans,J., Ghequire,M.G.K., Craggs,M., De Mot,R. and Cornelis,P. (2016) Identification and functional analysis of a bacteriocin, pyocin S6, with ribonuclease activity from a *Pseudomonas aeruginosa* cystic fibrosis clinical isolate. *Microbiology open,*, **5**, 413–423.
41. Wu,W. and Jin,S. (2005) PtrB of *Pseudomonas aeruginosa* suppresses the type III secretion system under the stress of DNA damage. *J. Bacteriol.*, **187**, 6058–6068.
42. Penterman,J., Singh,P.K. and Walker,G.C. (2014) Biological cost of pyocin production during the SOS response in *Pseudomonas aeruginosa*. *J. Bacteriol.*, **196**, 3351–3359.
43. Comeau,A.M., Tétart,F., Trojet,S.N., Prère,M.F. and Krisch,H.M. (2007) Phage-antibiotic synergy (PAS): β-lactam and quinolone antibiotics stimulate virulent phage growth. *PLoS One*, **2**, e799.
44. Kamal,F. and Dennis,J.J. (2015) *Burkholderia cepacia* complex phage-antibiotic synergy (PAS): antibiotics stimulate lytic phage activity. *Appl Environ Microbiol.*, **81**, 1132–1138.
45. International Committee on Taxonomy of Viruses (2020) *Create one new subfamily (psQueuovirinae) including one new genus (Amoyvirus) and three existing genera (Caudovirales: Siphoviridae). International Committee on Taxonomy of Viruses (ICTV).* https://talk.ictvonline.org/taxonomy/p/taxonomy-history?taxnode_id=202011640, (10 July 2023, date last accessed).
46. Hutinet,G., Swarjo,M.A. and de Crécy-Lagard,V. (2017) Deazaguanine derivatives, examples of crosstalk between RNA and DNA modification pathways. *RNA Biol.*, **14**, 1175–1184.
47. Nielsen,T.K., Forero-Junco,L.M., Kot,W., Moineau,S., Hansen,L.H. and Riber,L. (2023) Detection of nucleotide modifications in bacteria and bacteriophages: strengths and limitations of current technologies and software. *Mol Ecol.*, **32**, 1236–1247.
48. Rang,F.J., Kloosterman,W.P. and de Ridder,J. (2018) From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol.*, **19**, 90.
49. Doberenz,S., Eckweiler,D., Reichert,O., Jensen,V., Bunk,B., Sproer,C., Kordes,A., Frangipani,E., Luong,K., Korlach,J. *et al.* (2017) Identification of a *Pseudomonas aeruginosa* PAO1 DNA methyltransferase, its targets, and physiological roles. *MBio*, **8**, e02312-16.
50. Tsai,R., Corrêa,I.R., Xu,M.Y. and Xu,S.Y. (2017) Restriction and modification of deoxyarchaeosine (dG+)-containing phage 9 g DNA. *Sci Rep.*, **7**, 8348.
51. Šimoliūnas,E., Šimoliūnienė,M., Kaliniene,L., Zajančkauskaitė,A., Skapas,M., Meškys,R., Kaupinis,A., Valius,M. and Truncaitė,L. (2018) Pantoea bacteriophage vB_PagS_Vid5: a low-temperature siphovirus that harbors a cluster of genes involved in the biosynthesis of archaeosine. *Viruses*, **10**, 583.
52. Bodnar,J.W., Zempsky,W., Warder,D., Bergson,C. and Ward,D.C. (1983) Effect of nucleotide analogs on the cleavage of DNA by the restriction enzymes AluI, DdeI, HinfI, RsaI, and TaqI. *J Biol Chem.*, **258**, 15206–15213.
53. Sternberg,S.H., Redding,S., Jinek,M., Greene,E.C. and Doudna,J.A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, **507**, 62–67.
54. 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.
55. Mlotkowski,A.J., Schlegel,H.B. and Chow,C.S. (2023) Calculated pKa values for a series of aza- and Deaza-modified nucleobases. *J. Phys. Chem. A.*, **27**, 3526–3534.
56. Gleditzsch,D., Pausch,P., Müller-Esparza,H., Özcan,A., Guo,X., Bange,G. and Randau,L. (2019) PAM identification by CRISPR-cas effector complexes: diversified mechanisms and structures. *RNA Biol.*, **16**, 504–517.
57. Belkum,A.V., Soriaga,L.B., Lafave,M.C., Akella,S., Veyrieras,J., Barbu,E.M., Brami,D., Schicklin,S., Felderman,M., Schwartz,A.S. *et al.* (2015) Phylogenetic distribution of CRISPR-Cas systems in antibiotic-resistant *Pseudomonas aeruginosa*. *MBio*, **6**, e01796-15.
58. Vink,J.N.A., Baijens,J.H.L. and Brouns,S.J.J. (2021) PAM-repeat associations and spacer selection preferences in single and co-occurring CRISPR-Cas systems. *Genome Biol.*, **22**, 281.
59. Harvey,H., Bondy-Denomy,J., Marquis,H., Sztanko,K.M., Davidson,A.R. and Burrows,L.L. (2018) *Pseudomonas aeruginosa* defends against phages through type IV pilus glycosylation. *Nat Microbiol.*, **3**, 47–52.
60. Asikyan,M.L., Kus,J.V. and Burrows,L.L. (2008) Novel proteins that modulate type IV pilus retraction dynamics in *Pseudomonas aeruginosa*. *J Bacteriol.*, **190**, 7022–7034.
61. Mathee,K., Ciofu,O., Sternberg,C., Lindum,P.W., Campbell,J.I.A., Jensen,P., Johnsen,A.H., Givskov,M., Ohman,D.E., Søren,M. *et al.* (1999) Mucoid conversion of *Pseudomonas aeruginosa* by hydrogen peroxide: a mechanism for virulence activation in the cystic fibrosis lung. *Microbiology*, **145**, 1349–1357.
62. Abedon,S.T. (2011) Lysis from without. *Bacteriophage*, **1**, 46–49.
63. Hampton,H.G., Watson,B.N.J. and Fineran,P.C. (2020) The arms race between bacteria and their phage foes. *Nature*, **577**, 327–336.
64. Doron,S., Melamed,S., Ofir,G., Leavitt,A., Lopatina,A., Keren,M., Amitai,G. and Sorek,R. (2018) Systematic discovery of antiphage defense systems in the microbial pangenome. *Science*, **359**, eaar4120.