

Integrative omics analysis of *Pseudomonas aeruginosa* virus PA5oct highlights the molecular complexity of jumbo phages

Cédric Lood^{#1, 4}, Katarzyna Danis-Wlodarczyk^{#1,2}, Bob G. Blasdel¹, Ho Bin Jang¹, Dieter Vandenheuvel¹, Yves Briers¹, Jean-Paul Noben³, Vera van Noort^{4,5}, Zuzanna Drulis-Kawa², Rob Lavigne^{*1}

¹ Laboratory of Gene Technology, KU Leuven, Leuven, Belgium

² Department of Pathogen Biology and Immunology, Institute of Genetics and Microbiology, University of Wroclaw, Wroclaw, Poland

³ Biomedical Research Institute and Transnational University Limburg, Hasselt University, Diepenbeek, Belgium

⁴ Laboratory of Computational Systems Biology, KU Leuven, Leuven, Belgium

⁵ Institute of Biology, Leiden University, Leiden, The Netherlands

Contributed equally

* For correspondence: rob.lavigne@kuleuven.be

Abstract

Pseudomonas virus vB_PaeM_PA5oct is a compelling model phage to investigate phage-bacteria interactions and is a potent candidate for phage therapy. Combining hybrid genome sequencing, RNA-Seq, and mass spectrometry of virion-associated proteins allowed us to accurately assemble and annotate PA5oct's 286,783 bp genome. It encodes 461 coding regions, of which 93 were identified as virion-associated proteins using ESI-MS/MS. RNA-seq revealed twelve tRNA genes, four ncRNA regions, and a gradual take-over of the total cell transcriptome from 21% in early infection to 93% in late infection. Although PA5oct is not organized into strictly contiguous regions of temporal transcription, defined genomic regions are expressed in early, middle, and late phases of infection. Interestingly, we observed distinct regions showing limited transcription activity throughout the infection cycle. The infection of *Pseudomonas aeruginosa* by PA5oct triggers the upregulation of bacterial operons related to NAD biosynthesis, Psl exopolysaccharide, and periplasmic nitrate reductase production and downregulates Type IV pili gene products, suggesting mechanisms of superinfection exclusion. We used the proteome of PA5oct to position this isolate amongst other phages using a protein-sharing network. This study expands our understanding of the molecular diversity of jumbo viruses and opens new questions towards cellular regulation and phage-encoded hijacking mechanisms.

Introduction

The discovery of very large microbial viruses infecting bacteria and protozoa has dramatically expanded the range of known viral genome sizes, starting around 2 kb and ending upward of 2 Mb. These so-called jumbo viruses can have genomes that are larger than those of parasitic bacteria and archaea, thereby bridging the former classical divide between cells and viruses in terms of genome size (Claverie *et al.*, 2009; Koonin *et al.*, 2015). Importantly, these uncharacterized viral genomes are diverse, and most of their genes lack nucleotide similarity to other functionally characterized or indeed even known sequences that can be accessed in public databases. As such, they have been aptly described as “Viral Dark Matter” and there is room for the exploration of this untapped sequence space (Hatfull, 2015).

Bacteriophages with genome sizes beyond 200 kbp are known as jumbo phages, sometimes also referred to as giant phages. They have been isolated from a multitude of environments, including water, soil, plants, and animal tissues (Yuan *et al.*, 2017). A subset of 120 of them have been sequenced to date (May 2019, Supplementary Table S1), including 109 myoviruses, 10 siphoviruses, and 1 unclassified phage. Most infect Gram-negative bacteria (95.4%), with only a few (11 jumbo phages) infecting Gram-positive bacteria, mostly *Bacillus* strains. The tiny plaques typically formed by these large-sized phages are smaller than 0.5 mm on 0.7% soft agar and are easily overlooked during classical propagation procedures. Their large virion size can also lead to their loss during filtration procedures that are standard in many phage isolation protocols (Sewer *et al.*, 2007), suggesting biases against their isolation.

Although jumbo phages have proven to be interesting for structural analysis and have provided insights into infection processes (Fokine *et al.*, 2007; 2005; Wu *et al.*, 2012), the understanding of their functional genetics and genome organization has often remained limited to the results of *in silico* predictions of genomes assembled from short-reads sequencers such as Illumina machines. However, investigation into their basic (molecular) microbiology, including elucidating their transcriptional schemes, or the mechanisms of their co-evolution with their host must be pursued further if we are to acquire new insights (Hendrix, 2009). RNA-sequencing has proven to be an important tool in that respect, allowing a detailed molecular elucidation of phage transcriptional schemes, and the discovery of regulatory elements and sRNAs, but also providing

clues towards the host response during phage infection and its evolutionary implications (Ceyssens *et al.*, 2014; Chevallereau *et al.*, 2016; Blasdel *et al.*, 2017b and 2018).

Pseudomonas virus phiKZ, the first sequenced jumbo phage (Mesyanzhinov *et al.*, 2002) serves as a model phage in this regard. In particular, its transcription scheme is independent from the host transcriptional machinery and is governed by the consecutive action of phage-encoded, non-canonical multisubunit RNAPs (virion associated, vRNAP), distantly related to bacterial β and β' subunits, that are co-packaged alongside the genome and co-injected upon infection to bootstrap the transcription of other RNAP genes present in the phage genome (Ceyssens *et al.*, 2014).

Here we present a detailed molecular analysis of the jumbo *Pseudomonas* virus vB_PaeM_PA5oct, a myovirus infecting *Pseudomonas aeruginosa* (Drulis-Kawa *et al.*, 2014) that lacks a phage-encoded RNAP and thus represents a fundamentally different transcription mode from phage phiKZ. Our investigation was driven by combining short and long read sequencing, temporal transcriptomics, and structural proteomics. Together, these molecular techniques allowed to elucidate the genome organization of this phage species, that currently has no close representatives in the public databases in terms of DNA homology, and the response it elicits during infection of its host *P. aeruginosa* PAO1. Furthermore, we establish the evolutionary relationships between PA5oct and other jumbo bacteriophages using a protein-sharing network analysis.

Material & Methods

Bacteriophage propagation, purification, and morphology

In all experiments, phage PA5oct was propagated as previously described (Danis-Wlodarczyk *et al.*, 2015). Phage lysate was purified by 0.45 and 0.22 μm filtration and incubation with 10% polyethylene glycol 8000 (PEG 8000)–1 M NaCl according to standard procedures (Ceyssens *et al.*, 2006). Finally, CsCl-gradient ultracentrifugation was applied (Ceyssens *et al.*, 2008) and the resulting phage preparation dialyzed three times for 30 min against 250 volumes of phage buffer using Slide-A-Lyzer Dialysis Cassettes G2 (Thermo Fisher Scientific Inc, MA, USA). The phage titer

was assessed using the double-agar layer technique (Adams, 1959) and purified samples were stored at 4°C.

The rifampicin growth assay

One-step growth curves were established according to the method of Pajunen *et al.* (2000), with modifications. An equal volume of bacterial culture at OD₆₀₀ of 0.4 was mixed with phage suspension (10⁶ pfu/ml) to obtain a MOI of 0.01. Phages could adsorb for 8 min at 37°C, after which the mixture was diluted to 10⁻⁴. Triplicate samples were taken every 5 min during 1-1.5 h for titration, (Danis-Włodarczyk *et al.*, 2015). Where appropriate, rifampicin (Sigma) was added to a final concentration of 400 µg/ml.

DNA isolation and hybrid sequencing

Phage genomic DNA was prepared with the use of a modified protocol for lambda DNA isolation according to Ceysens (2009) and sequenced using the Illumina MiSeq platform available at the Nucleomics Core (VIB, Belgium) as previously described by Danis-Włodarczyk *et al.* (2015). The phage genomic DNA was also prepared for long reads sequencing using Oxford Nanopore Technology (ONT) 1D library protocol as described in Kakabadze *et al.* (2018). The library was sequenced in-house on a MinION equipped with a R9.4.1 flowcell (ONT).

RNA extraction and sequencing

P. aeruginosa strain PAO1 was grown overnight in 5 ml LB medium at 37 °C. Next, cells were diluted 1:100 in 50 ml fresh medium and further grown at 37 °C until an OD₆₀₀ of 0.3 was reached (around 1.2 x 10⁸ CFU/mL, early exponential phase). The culture was infected with PA5oct at a MOI of 50. To control for a synchronous infection, we ensured that less than 5% of bacterial survivors remained after five minutes post infection. Biologically independent samples were collected in triplicates at 5 minutes (early), 15 minutes (middle), and 25 minutes (late) into infection and processed as described previously (Blasdel *et al.*, 2017a; Chevallereau *et al.*, 2016).

Genome and transcriptome analysis

The genome of PA5oct was reconstructed using both Nanopore and Illumina reads with the integrated hybrid assembler Unicycler v0.4.7 (Wick *et al.*, 2017). After assembling in a single circular contig, the ends of the genome were determined by mapping the sequencing reads to

the genome using bwa-mem v0.7.17 (Li H, 2013), and by analysis of Illumina reads with PhageTerm v0.12 (Garneau *et al.*, 2017). The genome architecture was also confirmed by investigating potential structural variations using the long nanopore reads with a combination of the tools ngmlr and sniffles (Sedlazeck *et al.*, 2018), and the genome browser ribbonviewer (Nattestad *et al.*, 2016).

The genome annotation was performed by integrating the RAST phage annotation tool (McNair *et al.*, 2018) and a manual curation with standard genome annotation methods (Danis-Wlodarczyk *et al.*, 2015; and 2016). The annotation was further inspected visually with UGENE (Okonechnikov *et al.*, 2012) using the stranded RNA-Seq reads, mapped to the phage genome. The RNA-sequencing reads were aligned to both the phage and host genomes using the software TopHat v2 (Kim *et al.*, 2013) and the resulting alignments were then summarized strand-independently into count tables using HTSeq-Count (Anders *et al.*, 2015). Finally, the counts of each features were normalized using TPM (transcripts per million). For the differential expression of the host features, each statistical comparison presented was performed using the DESeq2 (Love *et al.* 2014) R/Bioconductor package to normalize host transcript populations to host transcript populations, or phage to phage, before testing for differential expression as described by Blasdel *et al.* (2018).

The genome of PA5oct was deposited under the NCBI Genbank accession number MK797984, Illumina and Nanopore genomic reads are available in the NCBI SRA database via the bioproject accession number PRJNA516157, and the transcriptomics dataset is available through the NCBI GEO database via the accession number GSE130190.

[ESI-MS/MS analysis on virion particle proteins](#)

Phage proteins were isolated from a purified phage lysate (10^9 pfu/ml) by a single methanol/chloroform extraction (1:1:0.75, v/v/v) (Acros Organics) and precipitated by addition of an equal volume of methanol (14,000 x g, 6 min). The phage proteins were separated on a 12% SDS-PAGE gel and analyzed by ESI-MS/MS as previously described (Ceysens *et al.*, 2014; Van den Bossche *et al.*, 2014).

Protein family clustering and construction of the relationships network

To build a gene-sharing network, we retrieved 231,166 protein sequences representing the genomes of 2,304 bacterial and archaeal viruses from NCBI RefSeq (version 85) and used the network analytics tool, vConTACT (version 2.0; <https://bitbucket.org/MAVERICLab/vcontact2>) (Jang *et al.*, 2019), as an app at iVirus (Bolduc *et al.*, 2017a). Briefly, including protein sequences from PA5oct, a total of 231,627 sequences were subjected to all-to-all BLASTp searches, with an E-value threshold of 10^{-4} , and defined as the homologous protein clusters (PCs) in the same manner as previously described (Bolduc *et al.*, 2017b). Based on the number of shared PCs between the genomes, vConTACT v2.0 calculated the degree of similarity as the negative logarithmic score by multiplying hypergeometric similarity P-value by the total number of pairwise comparisons. Subsequently, pairs of closely related genomes with a similarity score of ≥ 1 were grouped into viral clusters (VCs), with default parameters of vConTACT v2.0 (Jang *et al.*, 2019). The network was visualized with Cytoscape (version 3.5.1; <http://cytoscape.org/>), using an edge-weighted spring embedded model, which places the genomes or fragments sharing more PCs closer to each other. The taxonomic affiliation was taken from the International Committee on Taxonomy of Viruses (ICTV Master Species List v1.3; <http://www.ictvonline.org/virusTaxonomy>) incorporated 2017 ICTV updates (Adams *et al.*, 2017; Adriaenssens *et al.*, 2017). Further, the proportion of shared PCs between two genomes was characterized, as previously described (Bolduc *et al.*, 2017b and Jang *et al.*, 2019).

Results

Basic microbiological and genome properties of *Pseudomonas* phage PA5oct

Under standard conditions, one-step growth experiments indicate a latent period of approximately 40 min and a burst size of about 30 - 40 phage particles per infected bacterial cell (Drulis-Kawa *et al.*, 2014). When 400 $\mu\text{g/ml}$ of rifampicin, a host RNA polymerase (RNAP) transcription inhibitor, is added prior PA5oct infection, the progeny phage production is completely suppressed. This suggests that PA5oct, in contrast to another jumbo phage *Pseudomonas* virus phiKZ, relies on the host RNAP for phage transcription.

Phage PA5oct has a linear, A+T-rich (33.3% GC-content), double-stranded DNA genome of 286,783 bp with 461 putative CDS that could be validated by inspection with the RNA-seq data (Figure 1). In comparison, the *in silico* RAST annotation pipeline (McNair *et al.*, 2018) allowed the identification of 412 CDS. The genome has long direct terminal repeats (DTR) of 39,080 bases, a characteristic also found in previously published jumbo phages (Yoshikawa *et al.*, 2018). Interestingly, the DTR region includes 105 genes, many of which are highly transcribed in the early phase of the infection (Figure 2). Aside from 449 protein coding sequences, the phage genome also contains twelve tRNAs [Met (CAT), Leu (TAA), Asn (GTT), Thr (TGT), Met (CAT), Leu (TAA), Arg (TCT), Met (CAT), Pro (TGG), Gly (TCC), Ser (TGA) and Ser (GCT)] that were identified by tRNAscan-SE (Lowe *et al.*, 2016) and confirmed in the RNA-Seq data. Transcriptomics data also revealed four actively transcribed regions lacking putative ORFs or known promoters that have been annotated as ncRNA (see Table 1).

Table 1. Identified non-coding RNA species. Four non-coding RNAs have been identified based on RNA-Seq data from transcripts aligning to genomic loci lacking plausible ORFs or known promoters.

Name	Location	Strand	Percentage of phage reads		
			Early	Middle	Late
ncRNA001	8263..8498	+	2.40%	5.07%	4.68%
ncRNA002	21951..22401	+	4.65%	5.04%	1.23%
ncRNA003	103065..103368	+	0.01%	0.05%	0.03%
ncRNA004	246116..246734	+	0.08%	0.03%	0.02%

Temporal regulation of PA5oct genome expression

PA5oct transcripts progressively come to dominate the host transcription. They represent 21% of total non-rRNA transcripts after the early infection timepoint (5 min), increase to 69% by middle infection (15 min), and culminate to 93% during late infection (25 min). The strand specificity of the results revealed that early transcription is almost exclusively occurring from the Watson strand.

Most of the genes contained within the long DTR between gp34 and gp105 (uncharacterized proteins, except gp39, 48, and 51, which are structural proteins), but also gp450 (uncharacterized protein), are each found to be highly transcribed at this stage and can be defined as early genes. Phage PA5oct appears to have few genomic regions that can be distinguished between the middle and late phase of transcription as captured at 15 minutes onwards (Figure 2-B). Many of the (functionally annotated) gene features transcribed in those middle and late timepoints are predicted to be responsible for structural proteins as well as DNA metabolism and replication. Two of the four non-coding RNAs found in the DTR region (ncRNA1 and ncRNA2), and which lack similarity to any known primary DNA sequence in the Genbank database (release 231), are highly transcribed throughout the infection cycle (Figure 1, 2 and Supplementary table S2).

Remarkably, six genomic regions, including two of the three major regions transcribed on the Crick strand (Figure 2-B, boxes 1-6), were found to show limited transcription under our laboratory growth conditions, despite a high sequencing depth (between 11M and 21M reads mapping for each individual sample). These six regions correspond primarily to genes that encode structural proteins (Figure 1, gene products marked in red) and which have been detected by ESI-MS/MS.

A total of 32 putative promoters with highly conserved, AT-rich intergenic motifs (5'-TATAATA-3') and (5'-TTGAC-3') were identified around transcription start sites (Supplementary table S2) whereas 36 putative factor-independent terminators with conserved stem loops could be identified around transcription stop sites (Supplementary table S3).

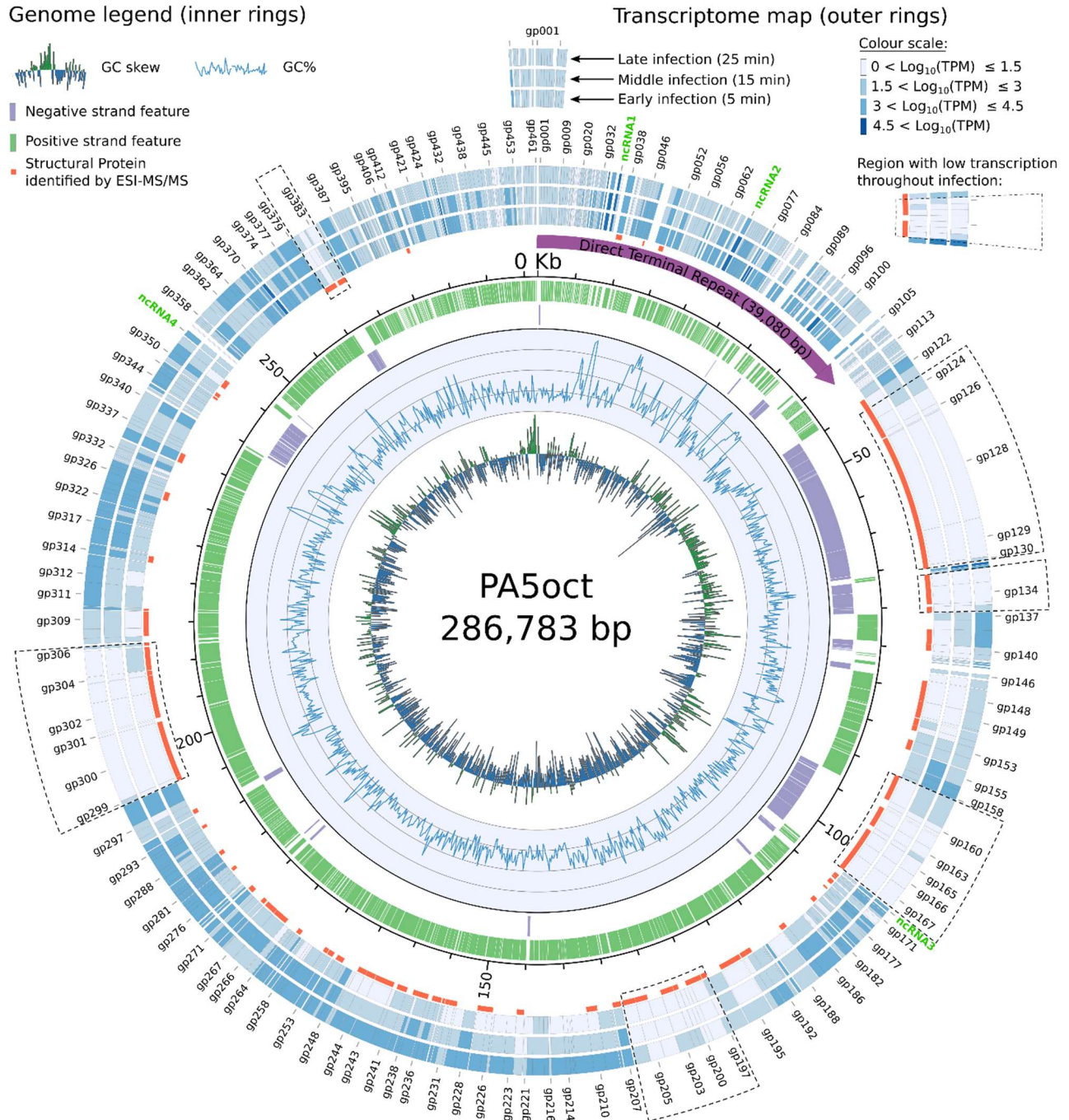


Figure 1. Circular representation of PA5oct genome and its transcriptome. The two inner circles display a histogram of the GC-skew followed by a graph of the GC content along the genome of PA5oct. It is followed by a stranded depiction of the CDS and ncRNA indicating that most features are encoded on the positive strand (in green) with intermittent regions encoding on the negative strand (in purple). Structural proteins confirmed by ESI-MS/MS analysis are highlighted in red color. The three outer rings depict normalized RNA read counts for each of the annotated feature

(note that some gp numbers are not shown on the graph to ease the visualization). Contiguous regions with limited amounts of transcription activity detected across multiple timepoints are encased in dotted lines. The figure was created with the software *Circos* (Krzywinski *et al.*, 2009)

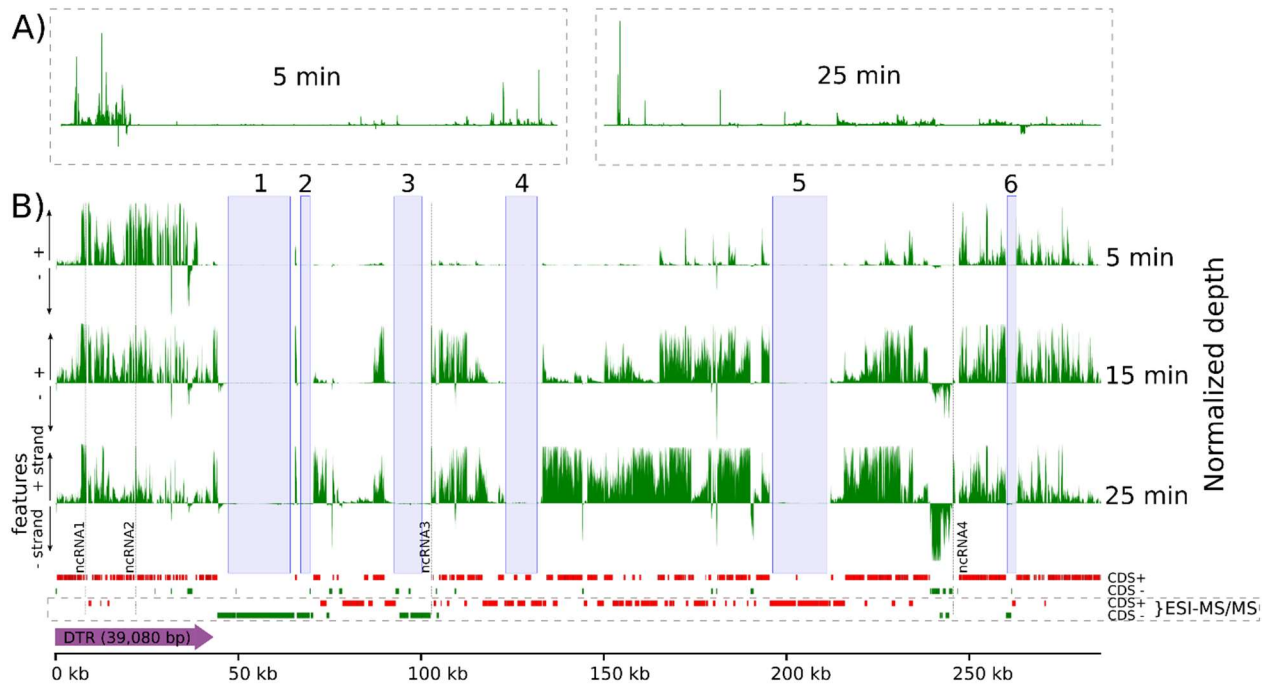


Figure 2. RNA-Seq analysis of PA5oct transcriptome after *P. aeruginosa* PAO1 infection. Genome-wide overview of reads mapped to the PA5oct genome for samples taken 5, 15, and 25 minutes post infection. A) Coverage of samples 5 min and 25 min (15 min omitted) show a limited number of regions which dominate the transcription landscape. B) Zooming in on the samples by capping the strongly transcribed regions allows to delineate regions differentially transcribed throughout the infection. Six marginally transcribed regions are highlighted with blue bars. Each of the three timepoints were normalized to reflect the depth across the triplicates. The figure was created using the python software *pyGenomeTracks*.

Transcriptome and proteome-based functional annotations

The ESI-MS/MS technique allowed to delineate a total of 93 gene products linked to the phage particle, comprising experimentally confirmed seven virion-unrelated enzymes, 13 virion-associated proteins, and 73 other gene products (Figure 1S, supplementary table 2S). Combined

with the bioinformatics analysis of the genome and the transcriptome, we could functionally annotate of a third of the CDS found in PA5oct.

Three capsid proteins were predicted, including the portal vertex protein of the head (gp149), the prohead core scaffolding protein and protease (gp152) and the precursor to the major head subunit (gp154). An additional protein, assigned as a putative head completion protein (gp200), was annotated based on sequence similarity to *Klebsiella* phage vB_KleM_RaK2 head completion protein. Moreover, several proteins associated with the tail apparatus were annotated, including a neck protein (gp193) and four tail proteins, a phage tail fiber protein H (gp159), a tail sheath stabilizer and completion protein (gp162), another tail fiber protein (gp165) and a tail sheath monomer (gp195) as well as three baseplate proteins, gp129, gp130 and gp133 (baseplate wedge, baseplate protein, baseplate hub subunit and tail lysozyme, respectively). Gp159, the putative phage tail fiber protein H, has a putative endo-N-acetylneuraminidase region on its C-terminus (868 – 1008 aa), suggesting a (2->8)-alpha-sialosyl linkage hydrolase function of oligo- or poly(sialic) acids, activity associated with the tail spikes and exopolysaccharide depolymerases (Kwiatkowski *et al.*, 1982).

PA5oct also encodes several genes predicted to be associated with peptidoglycan layer degradation, including three ESI-MS/MS detected product gp133, gp173, gp245, and two that were not detected gp368, and gp441. Their putative function was analyzed using BLASTP, HHpred, HMMER, InterPro, and Phyre2 (Table 2).

Table 2. Putative peptidoglycan degradation proteins. Five genes are predicted to be associated with the degradation of the peptidoglycan layer.

Gp	ESI-MS/MS	Functional analysis
gp133	yes	Baseplate hub subunit with tail-associated lysozyme homologous to the gp5 of <i>Escherichia coli</i> bacteriophage T4 which has a needle-like structure attached to the end of the tail tube serving in the injection process (Kanamaru <i>et al.</i> , 2005).
gp173	yes	Glycosyl hydrolase, lysozyme-like protein with a 159 amino acids long domain (11 - 170 aa).

gp245	yes	Baseplate hub and tail muramoyl-pentapeptidase
gp368	no	Cell-wall hydrolase domain protein with an N-acetylmuramyl-L-alanine amidase SleB domain cleaving the bond between N-acetylmuramoyl and L-amino acid residues. It is potentially an endolysin given that it was not recognized during ESI-MS/MS analysis
gp441	no	Murein hydrolase, peptidase M23, with a 103 amino acids long domain (30 - 133 aa).

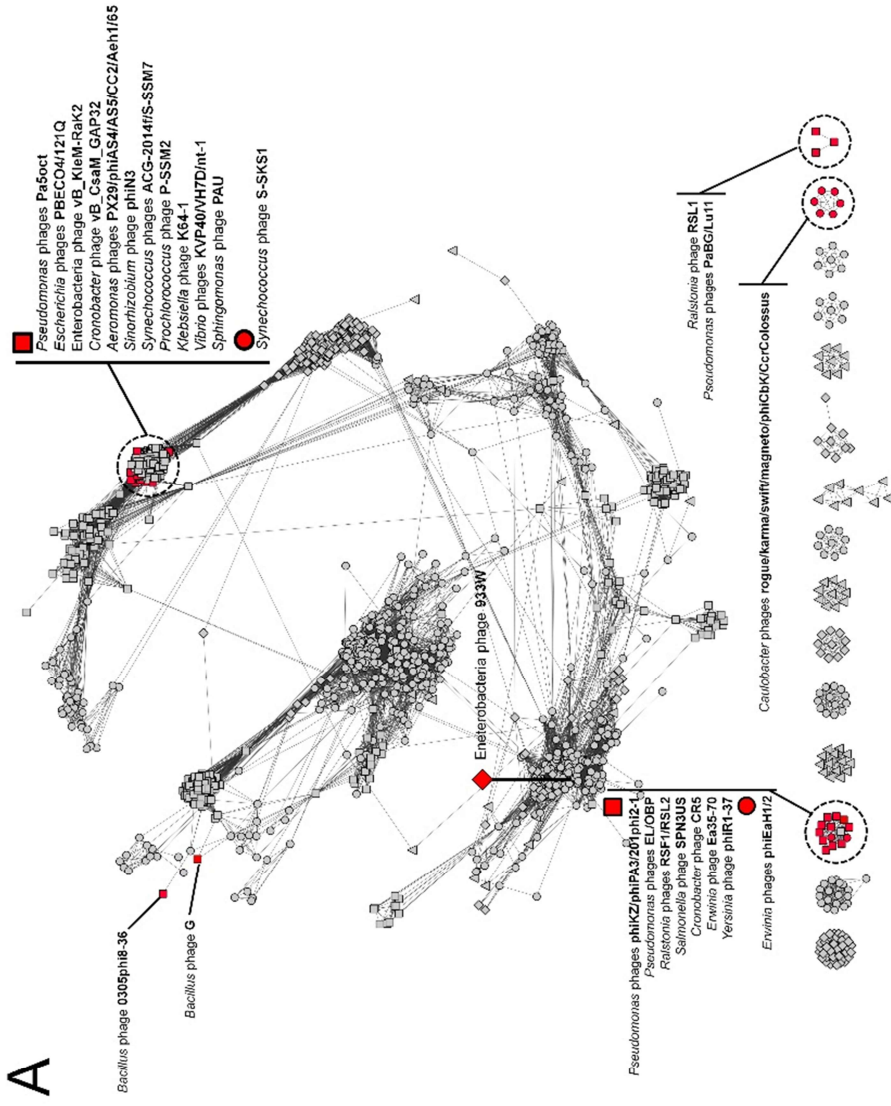
Gene-sharing relationships with other jumbo phages including PA5oct to other viruses

To represent genetic relationships between *Pseudomonas* phage PA5oct and other bacteriophages with genomes exceeding 200 kbp (Hendrix, 2009), a gene-sharing network was built. Viruses that do not show significant similarity to the jumbo were excluded for clarity. The resulting network was composed of 1,355 viral genomes (nodes) belonging to the *Myoviridae*, *Siphoviridae*, *Podoviridae*, or uncharacterized phages and 37,955 relationships (edges) between them.

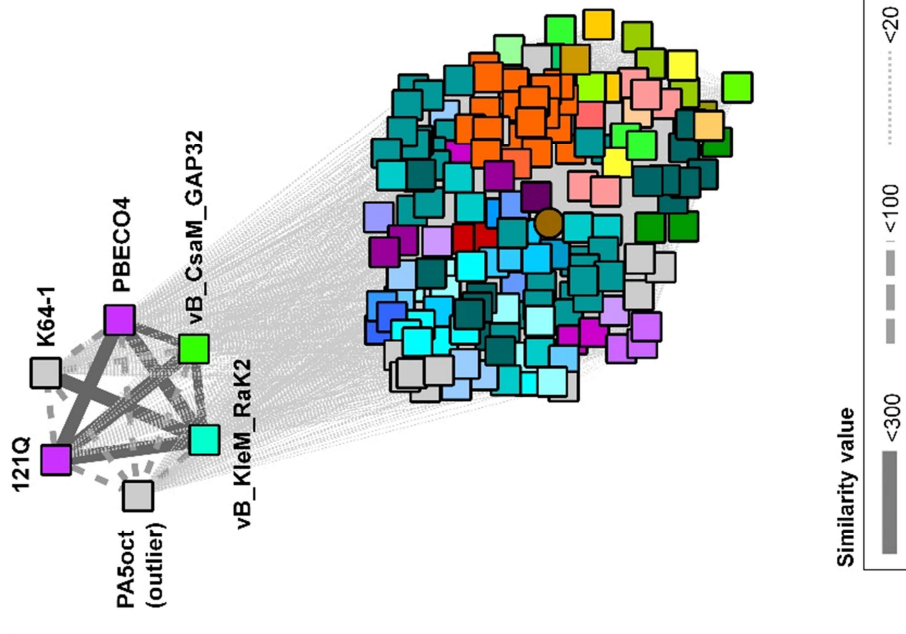
As shown in Figure 3A, 24 jumbo phages were placed into the largest connected component (LCC) that predominates the double-stranded DNA phages, whereas 22 phages were placed into three isolated components. A majority (19) of the 24 jumbo phages in the LCC, which includes PA5oct, fall within a highly interconnected region (subnetwork) that is rich in the *Tevenvirinae* viruses (Upper right in Figure 3A; Table 5S). A subsequent analysis found 60 coherent groups of viral genomes (i.e., viral clusters, VCs) embedded in this subnetwork (Figure 3B; Table S5), of which six jumbo phages including PA5oct and *Escherichia* phages 121Q/PBECO 4, *Klebsiella* virus RaK2, *Klebsiella* virus K64-1, and *Cronobacter* virus GAP32 showed stronger connections to each other but weaker connections outside the group due to their more common genes (i.e., homologous protein clusters, PCs) than the rest of the network (Figure 3B). Notably, a further inspection of their relationships (phage K64-1 was excluded as it has an incomplete protein annotation profile in NCBI) uncovered its sub-structure where PA5oct shares less common PCs (~11-15%) to the remaining members (Figure 3C), indicating its distant relationships as a outlier genome (Jang et al., 2019).

With respect to other jumbo phages, the network identifies informative connections. For example, the *Phicbkvirus*, and phages RSL1, PaBG, and Lu11 form two isolated components (Fig. 3A). This discontinuous structure of two viral groups, due to their distinct gene pools, can reflect their evolutionary relationships as separate viral lineages (Gill *et al.*, 2012; Yamada *et al.*, 2010); whereas the connection of jumbo phages to their relatives having smaller genomes (<200kbp) such as *Bacillus* phage G, phages of the *Phikzvirus*, *Elvirus*, *Agrican357virus*, and *Rsl2virus*, and *Bacillus* phages 0305phi8-36 (Figure 3A and Table S5) can support evolutionary links between jumbo phages and smaller-genome phages (Hendrix, 2009; Adriaenssens *et al.*, 2012; Jang *et al.*, 2013). Together, the visualization and analysis of gene content relationships of the jumbo phages as a network revealed the global distribution of genetically diverse groups of jumbo phages across 2,259 viruses, in which most of them together with smaller-genome phage(s) are placed into two major groups comprising viruses belonging to the *Tevenvirinae* as well as the *Phikzvirus/Elvirus/Agrican357virus/Rsl2virus*, respectively, and others form evolutionary distinct clades (Gill *et al.*, 2012; Yamada *et al.*, 2010) (Table S5). In particular, despite the similarities of PA5oct to the recently characterized jumbo phages vB_CsaM_GAP32, 121Q, PBECO 4, vB_KleM-RaK2, and K64-1 (Abbasifar *et al.*, 2014; Ackermann *et al.*, 1983; Kim *et al.*, 2013; Šimoliūnas *et al.*, 2013; Pan *et al.*, 2015), follow-up analyses indicate that PA5oct is a more distant member of that group as evidenced by the fewer number of proteins shared with the group (Figure 3B and 3C).

A



B



C

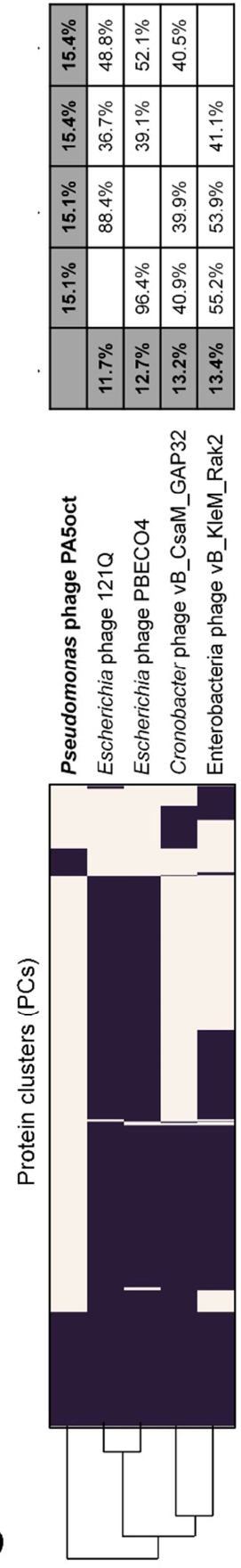


Figure 3. Protein-sharing network for PA5oct and other jumbo phages. (A) A network representation was produced using the edge-weighted spring embedded layout of Cytoscape version 3.5.1. Each node is depicted as a different shape, representing bacteriophages belonging to the *Myoviridae* (rectangle), *Podoviridae* (diamond), *Siphoviridae* (circle), or uncharacterized phages (triangle). Selected jumbo phages are shown in red. Edges between two nodes indicate statistically weighted pairwise similarities with similarity scores of ≥ 2 . (B) An enlarged view of the subnetwork comprising PA5oct and its relatives. Edge thickness is proportional to similarity values estimated with the hypergeometric equation (Materials and methods) and viral clusters are denoted as different colors. Note that due to the lower number of common genes, *Pseudomonas* virus PA5oct was identified as a distant relative (outlier; Jang *et al.*, 2019) to five phages including 121Q, K64-1, PBECO4, GAP32, and Rak2 (C) Module profiles showing the presence (dark) and absence (light) of homologous protein clusters (PCs) across genomes. Each row represents a virus and each column represents a PC. The genomes were hierarchically grouped based on hierarchical patterns of gene sharing (Left). A matrix showed the percentage of shared PCs between five-member viruses (Right). From these proteome-based analyses, phage K64-1 was excluded due to its incomplete protein annotation in NCBI.

PA5oct-specific differential gene expression of the host genome

The conserved PAO1 host-mediated transcriptional response to phage infection present during the infection by phages 14-1, PEV2, YuA, and LUZ19 (Blasdel *et al.*, 2018) is also found during the lytic infection cycle of PA5oct (Figure 4). However, PA5oct also appears to elicit several phage-specific impacts on transcript abundance, which result in a unique differential expression pattern of specific host genes. PA5oct progressively dominates the non-ribosomal RNA environment of the cell and this shift from host transcripts to phage transcripts (Figure 4) depletes all host transcripts 0.08-fold relative to the total in the cell. Note that this global depletion of host transcripts relative to the total phage/host transcript population is not accounted for in the Log_2 (Fold Change) values we report (Figure 4, Supplementary table S6), as described previously (Blasdel *et al.*, 2018). Doing this allows us to illustrate the impact of PA5oct on host transcripts relative to other host transcripts, rather than relative to the total.

As shown on Figure 4, the strongest upregulation is that of the PA4918-PA4920 operon, which contains an enzyme involved in NAD biosynthesis (Okon et al., 2016). Another operon specifically upregulated during PA5oct infection encodes a part of the *psl A-L* gene cluster (*pslE-J*), which is involved in the production of Psl exopolysaccharide (Jackson et al., 2004). PA5oct also uniquely manipulates the host into upregulating the transcription of the *napABCDEF* operon, which encodes a periplasmic nitrate reductase. Interestingly, several genes associated with the biosynthesis of Type IV pili are downregulated relative to other host transcripts. As appears to be common in different infection settings with other virulent phages (Blasdel et al., 2018), PA5oct appears to mobilize the transcription of genes of a potentially cryptic Pf1 prophage (family *Inoviridae*) encoded from ORF PA0618 to PA0642.

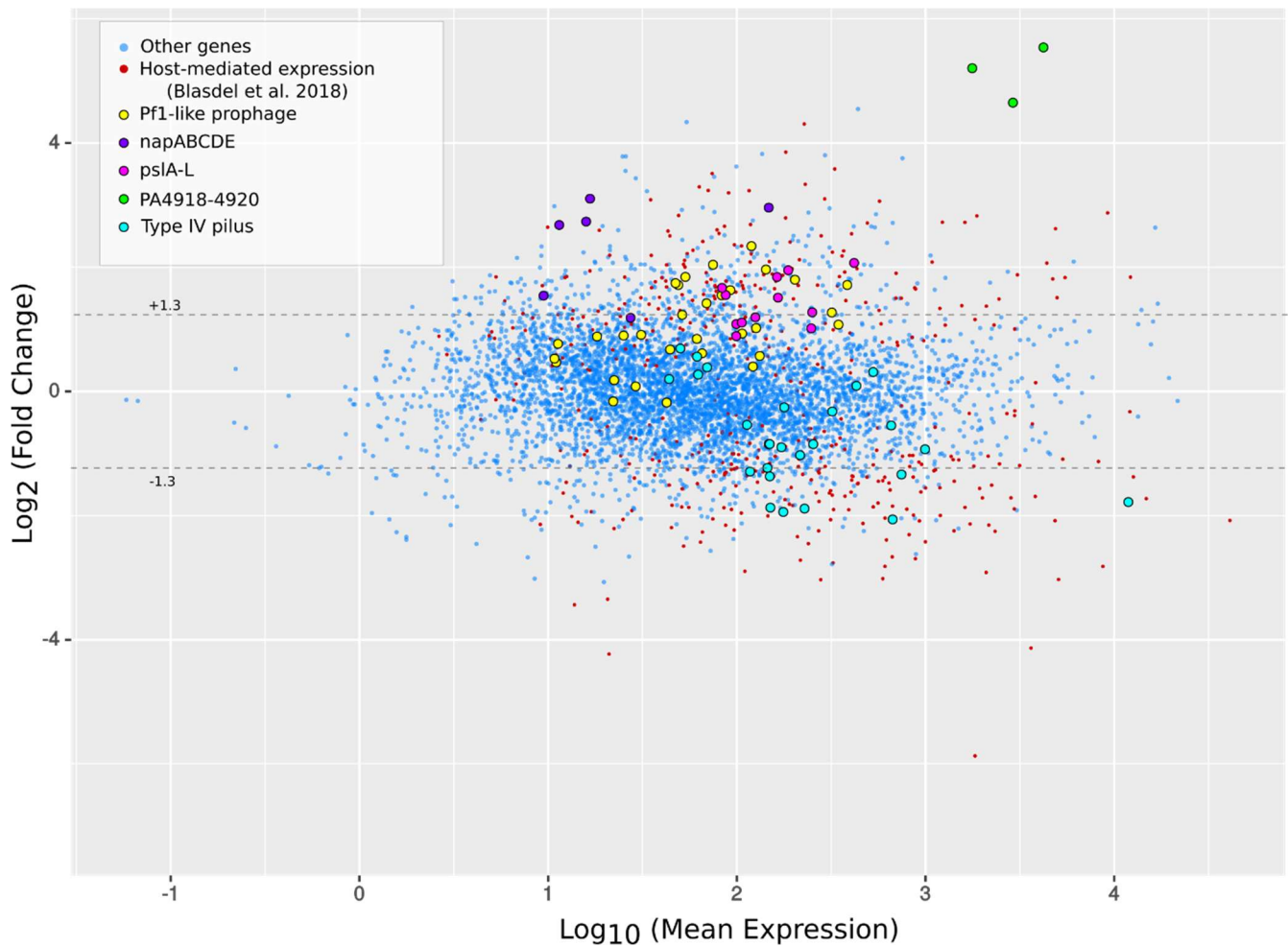


Figure 4. MA Plot highlighting the differential expression of host genes during PA5oct infection.

In addition to the host-mediated stress response to phage infection identified by Blasdel *et al.* (2018), we observe PA5oct-specific differential expression of host genes during PA5oct infection. Specifically, we see a PA5oct-specific upregulation of the *pslE-J* part of the *pslA-L* exopolysaccharide operon, the *napABCDE* operon which encodes the periplasmic nitrate reductase, parts of a Pf1 prophage, and the operon PA4918-4920 involved in NAD biosynthesis. We also observe a downregulation of Type IV pili genes which is a necessary receptor to initiate PA5oct infection.

Discussion

Insights from the integration of multiple omics analyses

Phage PA5oct has currently the third largest genome amongst the sequenced *Pseudomonas* jumbo phages in available databases, after 201phi2-1 (316.674 kbp, Thomas *et al.*, 2008) and phiPA3 (309.208 kbp, Monson *et al.*, 2011). To date, the largest myovirus known is *Bacillus megaterium* phage G (497,513 bp genome, Table 1S). It is important to keep in mind that some jumbo phages, especially those reported prior to the 1990s, were identified only by electron microscopy (e.g. *Gluconobacter* phage GW6210) (Abbasifar *et al.*, 2014) and that ongoing sequencing efforts of these isolates, or metagenomics efforts will likely disrupt the current rankings of jumbo phage genome sizes.

In terms of annotation, functions could be attributed for a third of PA5oct proteins and a large portion of predicted CDS are unique, lacking sequence-based similarity to the genomes available in the public database Genbank (r231). For this reason, the functional characterization of jumbo phage orphan genes is an important endeavor to undertake in order to elucidate basic phage biology as well as phage application safety when thinking of therapeutic settings (Yuan *et al.*, 2017). Leveraging RNA-seq for a whole genome analysis is a powerful way to elucidate differential expression of gene features across different conditions but also delineate possible phage transcriptional modules. By experimentally defining the timing and expression levels of transcripts in both phage and host, directional RNA-Seq helps to discover and accurately annotate novel sequences, particularly for ncRNAs that are difficult to predict using genomics

data only, but also for small phage peptides that fall below gene prediction thresholds of annotation pipelines (Ceyssens *et al.*, 2014). In this manner, we could refine annotations of existing coding sequences, correcting and adding about 10% of the total predicted ORFs from those predicted *in silico* as well as annotating four ncRNA regions of high interest for future investigation. Interestingly, we also discovered a ribonuclease H-like protein on the genome of PA5oct (gp325), putatively involved in the processing of the host transcripts during the infection cycle.

ESI-MS/MS allowed us to identify 93 structural proteins of PA5oct, including four structural head related proteins, one neck protein and four tail related proteins. This is a similar result to that seen with giant *Pseudomonas* 201phi2-1 (89 structural proteins, Thomas *et al.*, 2010). However, other jumbo phages can possess a much smaller number of structural proteins, such as *Pseudomonas* virus phiKZ (62), *Aeromonas* virus phiAS5 (26) or *Ralstonia* virus phiRSL1 (25) (Lecoutere *et al.*, 2009; Yamada *et al.*, 2010; Kim *et al.*, 2012). Furthermore, the capsid of jumbo phages can have very complex structures, as we could observe in the case of phage phiKZ (30 head structural proteins), where in the middle it has an “inner body”, a spool-like protein structure that plays an important role in DNA packaging and genome ejection during, respectively, phage virion assembly and subsequent virion adsorption (Wu *et al.*, 2012).

Importantly, the large capsids seen in jumbo phages can accommodate large genomes allowing these phages to possess many accessory genes comparing to small phages (Ceyssens *et al.*, 2014; Yuan & Gao, 2017; Mesyanzhinov *et al.*, 2002; Hertveldt *et al.*, 2005; Kiljunen *et al.*, 2005; Thomas *et al.*, 2007; Gill *et al.*, 2012). For example, PA5oct encodes five different genes encoding peptidoglycan-degrading enzymes and 12 tRNA genes to increase the translation efficiency (Yuan & Gao, 2016). Note that other jumbo phages may contain even more tRNAs genes, such as *Vibrio* virus phi-pp2 and KVP40 (30 tRNA); *Caulobacter* virus CcrColossus and *Vibrio* virus nt-1 (28 tRNA); *Aeromonas* virus Aeh1, *Caulobacter* virus Swift and Magneto (27 tRNA); *Cronobacter* virus GAP32, *Caulobacter* virus Karma and phiCbK (26 tRNA); *Aeromonas* virus PX29 (25 codons) (Supplementary table 1S).

Hijacking the transcriptional environment of the host

The strategy of phage transcriptional mechanisms depends primarily on the presence or absence of a phage RNA polymerase (Yang *et al.*, 2014). After careful analysis of PA5oct genes, no RNAP-like proteins could be identified. Taken together with the absence of phage production in the presence of rifampicin and the host's ability to mount a transcription-level defense response, this suggests that this phage may depend on host transcriptional mechanism, though we cannot rule out the possibility that an unknown but essential phage protein is also sensitive to rifampicin. The main role in phage gene transcription is thus likely played by the large DNA-dependent RNAP of bacterial host, a 400-kDa protein complex composed of five subunits (α_2 , β , β' , ω), that transcribes mRNA following binding to the DNA template (Murakami, 2015; Tagami *et al.*, 2014, Davis *et al.*, 2017). This process is often supported by phage-encoded σ factors forming a holoenzyme with RNAP (α_2 , β , β' , ω , σ) (Williams *et al.*, 1987; Pavlova *et al.*, 2012).

Some phages can also implement other transcriptional strategies, e.g. viral gene transcription may rely on both the host RNAP along with a single-subunit phage RNAP (*Enterobacteria* virus T7, T3, *Pseudomonas* virus phiKMV and *Xanthomonas oryzae* virus Xp10). The single-subunit RNAP is then responsible for transcription of phage genes in the middle and late stage of infection, whereas early genes are transcribed by the host RNAP (Savalia *et al.*, 2010; Semenova *et al.*, 2005, Yang *et al.*, 2014). A similar situation is observed in the case of *Enterobacteria* virus N4, where early and middle stage of transcription depends on two phage-encoded RNAPs and the late genes are transcribed by host RNAP (Haynes *et al.*, 1985; Willis *et al.*, 2002). Another example is the giant *Pseudomonas* virus phiKZ that can infect its host independently of the host RNAP by using a virion-associated phage RNAP active in early transcription, and a non-virion-associated phage RNAP, important for late and possibly middle transcripts (Ceyssens *et al.*, 2014; Yakunina *et al.*, 2015).

Phage transcription patterns

PA5oct progressively dominates host transcription. After 5 min, PA5oct transcripts represented 21% of total non-rRNA transcripts, eventually proceeding to 69% and then 92% by middle and late infection. This possibly reflects a globally accelerated degradation of RNA in the cell akin to what was described by Chevallereau *et al.* (2016). In a similar fashion to giant *Pseudomonas* virus

phiKZ (Ceyssens *et al.*, 2014), giant Yersinia virus phiR1-37 (Leskinen *et al.*, 2016) and giant *Bacillus* virus AR9 (Lavysh *et al.*, 2017), PA5oct is not organized into strictly contiguous regions of temporal transcription. In the early timepoint, we observe a lot of transcription in the DTR region (Figure 2). These early coding sequences are classically understood to be involved in the transitioning from host to phage metabolism. Other regions of early, middle, and late expression are scattered throughout the genomes, a feature that increasingly appears to be characteristic of phages with large genomes, including the 168 kbp *Escherichia* virus T4.

We also found six regions with low levels of transcription that encode for unknown structural genes correlated to baseplate and tail that were confirmed to be translated and incorporated into phage particles by ESI-MS/MS (a very sensitive technique), suggesting either a notably efficient translation or a novel and unknown bias during RNA-seq. It is tempting to speculate that the expression of these structural proteins may be conditional, for example based on environmental cues which result in altered tail-receptor proteins to be displayed. Indeed, these specific structural proteins are not conserved between phages within the protein-sharing network, hinting at functions associated with specificity. These genes could have been captured from other phages and be part of the accessory genetic content of these jumbo phages.

Comparative genome analysis and protein-sharing network

To examine the genetic relationships between PA5oct and other jumbo bacteriophages as well as *Myoviridae*, *Siphoviridae*, *Podoviridae* viruses, a protein-sharing network was constructed. The PA5oct, *Escherichia* virus 121Q/PBECO4, *Klebsiella* virus vB_KleM-RaK2, *Klebsiella* virus K64-1, and *Cronobacter* virus vB_CsaM_GAP32 presented closer relationships based on shared conserved core genes. Furthermore, PA5oct, PBECO4, GAP32, T5, FelixO1 appear to be distantly diverged members of *Tevenvirinae*. The T5 and FelixO1 phages appear also to act as bridge for T4-related components due to their links to other phage groups. Unlike the marker genes of bacteria (i.e., 16S rRNA genes), which can be used for their taxonomic classification, viruses lack universal genes. Thus, traditional single-gene-based phylogenies, such as those based on major capsid proteins, will always be limited to related subsets of the total virus universe. Protein-sharing networks, as presented here, introduce an alternative approach, which provides a more global view that is also consistent with current taxonomic groupings. Using this approach, the

possible genetic relationships for PA5oct, but also all giant virus among the whole virus population can be observed.

Host stress response during PA5oct virus infection

Phages have a substantial impact on bacterial cellular systems, including transport/export, energy production/conversion, ribosomal proteins, cell wall modification, conversion of ribonucleotides to deoxyribonucleotides, cold shock and osmotic stress response (Fallico *et al.*, 2011; Ainsworth *et al.*, 2013; Ravantti *et al.*, 2008; Leskinen *et al.*, 2016). In this study, we observed that PA5oct upregulates the operon encoding *pslE-J*, responsible for production of Psl exopolysaccharide. Psl is essential for initiating and maintaining biofilm architecture in both mucoid and non-mucoid strains. It is responsible for the formation of a fabric-like matrix that holds cells closely together, thereby impeding migration within the biofilm (Ma *et al.*, 2009). This may lead to increased slime production, reducing the relative abundance of LPS phage receptors, hence secondary phage infections.

The *napABCDEF* operon, involved in the production of a periplasmic nitrate reductase, is upregulated during PA5oct infection. This protein is required for anaerobic growth in cystic fibrosis (CF) sputum *in vitro*. In a CF environment, *P. aeruginosa* prefers anaerobic growth that significantly enhances its biofilm formation and antibiotic resistance (Yoon *et al.*, 2002). Under these anaerobic conditions, *P. aeruginosa* modifies the structure of its LPS from a highly electronegative surface to a neutral surface (Sabra *et al.*, 2003, Palmer *et al.*, 2007). This suggests that phage PA5oct may influence *Pseudomonas* biofilm formation. While the purpose and mechanistic origin of this upregulation is unclear, it would seem to suggest that PA5oct may be well adapted to the environment of human-associated *Pseudomonas* biofilms, an interesting prospect for phage therapy given this phage's broad host-range.

The upregulation of the operon PA4918-4920 can be linked to NAD biosynthesis, an important molecule that mediates the transfer of energy in the cell, potentially part of the phage mediated conversion of the bacterial cell into an optimized phage production system.

We observed here again an upregulation of prophage pf1, a key filamentous prophage that is linked to the pathogenicity of *P. aeruginosa* in lung infections (Burgener *et al.*, 2019), which appears to be ubiquitous during phage infection. Indeed, this has been observed in other

transcriptomics assays of phage infection with phiKZ (Ceyssens *et al.*, 2014), as well as with phages LUZ19, PEV2, PakP3, and PakP4 (Blasdel *et al.*, 2018). It can reasonably be proposed to be the result of a general stress response of phage Pf1 that is unrelated to superinfection exclusion mechanisms.

Since PA5oct downregulates biosynthesis of Type IV pili and indirectly modifies bacterial LPS, the signaling networks might also influence the overproduction of nitrate reductase. As it was proven in our related manuscript, Type IV pili and LPS structures serve as phage PA5oct receptors (Olszak *et al.*, 2019). Receptor downregulation possibly functions as a tactic to reduce losses of “sister” phages and progeny adsorption (i.e., secondary adsorption) to already phage-infected bacteria (Abedon, 2017).

Funding

This study was supported by research grant no. 2012/04/M/NZ6/00335 of the National Science Centre, Poland. KDW was co-financed by the European Union as part of the European Social Fund and a postdoctoral fellowship at KU Leuven, Belgium. CL is supported by an SB PhD fellowship from FWO Vlaanderen (1S64718N). RL, YB, and ZDK are members of the “PhageBiotics” research community, supported by the FWO Vlaanderen. ZDK was a members of the EU COST BM1003 Action in Microbial cell surface determinants of virulence (http://www.cost.eu/COST_Actions/bmbs/BM1003). JPN is supported by a grant from the Hercules Foundation (project R-3986).

Acknowledgements

We are grateful to Steve Abedon and Harald Brüßow for critical comments and helpful suggestions on the manuscript.

Authors contributions

KDW performed one-step growth experiments and analysis of ESI-MS/MS results. CL, KDW performed DNA isolation, and annotation of the genome. HBJ generated protein family clustering, network construction and analyzed the dataset. CL, DV helped with promoter identification. CL, KDW, YB performed the genome sequencing and assembly. JPN performed ESI-

MS/MS analysis. BB and KDW prepared samples for RNA-Seq. BB, CL, KDW analyzed the transcriptomics results. BB, CL, HBJ, KDW, RL, ZDK designed experiments and wrote the manuscript.

Supporting Information

Supplementary table 1S. List of jumbo phages found in the literature.

Supplementary table 2S. Alignments of PA5oct promoters. Conserved AT-rich intergenic motifs: -10 box (5'- TATAATA - 3') and -35 box (5' - TTGAC - 3'). The corresponding sequence logos from MEME/MAST are depicted below the alignments. ORFs in front of which promoters are located are listed on the left side, whereas the corresponding sequence locus in the phage genome is indicated on the right side.

Supplementary table 3S. Phage PA5oct predicted terminators with palindromes marked blue.

From the left site: list of ORFs after which terminators are located by ARNold with strandedness, location in phage genome, motif of terminator, terminator sequence with conserved stem-loop structures marked blue.

Supplementary table 4S. The ESI-MS/MS analysis of denatured phage PA5oct particles after fractionation on SDS-PAGE gel.

Supplementary table 5S. List of jumbo phages (yellow color) and relevant phages that belong to the viral clusters (VCs) and their ICTV taxonomy.

Supplementary table 6S. The differential expression of non-rRNA gene feature within *Pseudomonas* PAO1 genome by late PA5oct infection.

Supplementary figure 1S. The SDS-PAGE pattern of PA5oct structural proteome against LMW Ladder (Thermo Scientific) in the first lane. The corresponding molecular weight is mentioned left. The numbered fractions on the right, correspond to gel slices analyzed individually by ESI-

MS/MS. The proteins are mentioned in the slice in which they were most abundantly present (see Band no. in Supplementary Table S4).

References

- Abbasifar R, Griffiths MW, Sabour PM, Ackermann HW, Vandersteegen K, Lavigne R, Noben JP, Alanis Villa A, Abbasifar A, Nash JH, Kropinski AM. (2014). Supersize me: *Sakazakii* phage GAP32. *Virology*; 460, 138–146. doi: 10.1016/j.virol.2014.05.003.
- Abedon ST. 2017. Phage “delay” towards enhancing bacterial escape from biofilms: a more comprehensive way of viewing resistance to bacteriophages. *AIMS Microbiology*, 3(2): 186-226. doi: 10.3934/microbiol.2017.2.186.
- Abedon ST. 2015. Bacteriophage secondary infection. *Viol Sin30*: 3-10. doi: 10.1007/s12250-014-3547-2.
- Ackermann HW, Nguyen TM. 1983. Sewage coliphages studied by electron-microscopy. *Appl. Environ. Microbiol.* 45, 1049–1059.
- Adams MH. 1959. Bacteriophages. In: Adams MH, editor. *Bacteriophages*. New York: Interscience Publishers; 27–30.
- Adams MJ, Lefkowitz EJ, King AMQ, Harrach B, Harrison RL, Knowles NJ, Kropinski AM, Krupovic M, Kuhn JH, Mushegian AR, Nibert M, Sabanadzovic S, Sanfaçon H, Siddell SG, Simmonds P, Varsani A, Zerbini FM, Gorbalenya AE, Davison AJ. 2017. Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses. *Arch Virol.* 162:2505-2538. doi: 10.1007/s00705-017-3358-5.
- Adriaenssens EM, Krupovic M, Knezevic P, Ackermann HW, Barylski J, Brister JR, Clokie MR, Duffy S, Dutilh BE, Edwards RA, Enault F, Jang HB, Klumpp J, Kropinski AM, Lavigne R, Poranen MM, Prangishvili D, Rumnieks J, Sullivan MB, Wittmann J, Oksanen HM, Gillis A, Kuhn JH. 2017. Taxonomy of prokaryotic viruses: 2016 update from the ICTV bacterial and archaeal viruses subcommittee. *Arch Virol.* 162:1153-1157. doi: 10.1007/s00705-016-3173-4.

- Adriaenssens EM, Mattheus W, Cornelissen A, Shaburova O, Krylov VN, Kropinski AM, Lavigne R. 2012. Complete genome sequence of the giant phage Lu11. *J Virol.*;86(11):6369-70. doi: 10.1128/JVI.00641-12.
- Ainsworth S, Zomer A, Mahony J, van Sinderen D. 2013. Lytic infection of *Lactococcus lactis* by bacteriophages tuc2009 and c2 triggers alternative transcriptional host responses. *Applied and environmental microbiology*, 79 (16), 4786–4798. pages 29, 30, 177. doi: 10.1128/AEM.01197-13.
- Anders S, Pyl PT and Huber W, 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2), pp.166-169.
- Arens DK, Brady TS, Carter JL, Pape JA, Robinson DM, Russell KA, Staley LA, Stettler JM, Tateoka OB, Townsend MH and Whitley KV, 2018. Characterization of two related *Erwinia myovirus*s that are distant relatives of the phiKZ-like jumbo phages. *PloS one*, 13(7), p.e0200202.
- Bae B, Davis E, Brown D, Campbell EA, Wigneshweraraj S, Darst SA. 2013. Phage T7 Gp2 inhibition of *E. coli* RNA polymerase involves misappropriation of $\sigma 70$ domain 1.1. *PNAS*, 110(49), 19772-19777. doi: 10.1073/pnas.1314576110.
- Baxter K, Lee J, Minakhin L, Severinov K, Hinton DM. 2006. Mutational analysis of s70 region 4 needed for appropriation by the bacteriophage T4 transcription factors AsiA and MotA. *J Mol Biol* 363: 931–944. doi: 10.1016/j.jmb.2006.08.074.
- Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* 29, 2607–2618.
- Bhunchoth A, Blanc-Mathieu R, Mihara T, Nishimura Y, Askora A, Phionrit N, Leksomboon C, Chatchawankanphanich O, Kawasaki T, Nakano M, Fujie M, Ogata H, Yamada T. 2016. Two asian jumbo phages, phiRSL2 and phiRSF1, infect *Ralstonia solanacearum* and show common features of phiKZ-related phages. *Virology.*;494:56-66. doi: 10.1016/j.virol.2016.03.028.

- Blasdel BG, Ceyssens PJ, Chevallereau A, Debarbieux L, Lavigne R. 2018. Comparative transcriptomics reveals a conserved Bacterial Adaptive Phage Response (BAPR) to viral predation. bioRxiv 248849. doi: <https://doi.org/10.1101/248849>.
- Blasdel BG, Ceyssens PJ, Lavigne R. 2017a. Preparing cDNA libraries from lytic phage infected cells for whole transcriptome analysis by RNA-Seq. In Clokie MRJ and Kropinski A. (ed), Bacteriophages: Methods and Protocols. Vol. 3, Humana Press. doi: 10.1007/978-1-4939-7343-9_14.
- Blasdel BG, Chevallereau A, Monot M, Lavigne R, Debarbieux L. 2017b. Comparative transcriptomics analyses reveal the conservation of an ancestral infectious strategy in two bacteriophage genera. ISME J.;11(9):1988-1996. doi: 10.1038/ismej.2017.63.
- Bolduc B, Jang HB, Doucier G, You ZQ, Roux S, Sullivan MB. 2017a. vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. PeerJ. 2017;5:e3243. doi: 10.7717/peerj.3243.
- Bolduc B, Youens-Clark K, Roux S, Hurwitz BL, Sullivan MB. 2017b. iVirus: facilitating new insights in viral ecology with software and community data sets imbedded in a cyberinfrastructure. ISME J. 11:7-14. doi: 10.1038/ismej.2016.89.
- Burgener EB, Sweere JM, Bach MS, Secor PR, Haddock N, Jennings LK *et al.* (2019). Filamentous bacteriophages are associated with chronic Pseudomonas lung infections and antibiotic resistance in cystic fibrosis. Science translational medicine, 11(488), eaau9748.
- Buttimer C, O'Sullivan L, Elbreki M, Neve H, McAuliffe O, Ross RP, Hill C, O'Mahony J, Coffey A. 2016. Genome sequence of jumbo phage vB_AbaM_ME3 of *Acinetobacter baumannii*. Genome Announc. 4:e00431-16. doi: 10.1128/genomeA.00431-16.
- Buttimer C, Born Y, Lucid A, Loessner MJ, Fieseler L and Coffey A, 2018. Erwinia amylovora phage vB_EamM_Y3 represents another lineage of hairy Myoviridae. *Research in microbiology*, 169(9), pp.505-514.
- Camara B, Liu M, Reynolds J, Shadrin A, Liu B, Kwok K, Simpson P, Weinzierl R, Severinov K, Cota E, Matthews S, Wigneshweraraj SR. 2010. T7 phage protein Gp2 inhibits the *Escherichia coli* RNA polymerase by antagonizing stable DNA strand separation near the transcription start site. Proc Natl Acad Sci 107: 2247–2252. doi: 10.1073/pnas.0907908107.

- Ceyssens PJ, Lavigne R, Mattheus W, Chibeu A, Hertveldt K, Mast J, Robben J, Volckaert G. 2006. Genomic analysis of *Pseudomonas aeruginosa* phages LKD16 and LKA1: establishment of the KMV subgroup within the T7 supergroup. *J. Bacteriol.* 188:6924 –6931. doi: 10.1128/JB.00831-06.
- Ceyssens PJ, Mesyanzhinov V, Sykilinda N, Briers Y, Roucourt B, Lavigne R, Robben J, Domashin A, Miroshnikov K, Volckaert G, Hertveldt K. 2008. The genome and structural proteome of YuA, a new *Pseudomonas aeruginosa* phage resembling M6. *J Bacteriol.*; 190: 1429–1435. doi: 10.1128/JB.01441-07.
- Ceyssens PJ, Minakhin L, Van den Bossche A, Yakunina M, Klimuk E, Blasdel B, De Smet J, Noben JP, Bläsi U, Severinov K, Lavigne R. 2014. Development of giant bacteriophage ϕ KZ is independent of the host transcription apparatus. *J Virol.*;88(18):10501-10. doi: 10.1128/JVI.01347-14.
- Ceyssens PJ, Miroshnikov K, Mattheus W, Krylov V, Robben J, Noben JP, Vanderschraeghe S, Sykilinda N, Kropinski AM, Volckaert G, Mesyanzhinov V, Lavigne R. 2009. Comparative analysis of the widespread and conserved PB1-like viruses infecting *Pseudomonas aeruginosa*. *Environ Microbiol.*; 11: 2874–2883. doi: 10.1111/j.1462-2920.2009.02030.x.
- Chevallereau A, Blasdel BG, De Smet J, Monot M, Zimmermann M, Kogadeeva M, Sauer U, Jorth P, Whiteley M, Debarbieux L, Lavigne R. 2016. Next-Generation "-omics" Approaches Reveal a Massive Alteration of Host RNA Metabolism during Bacteriophage Infection of *Pseudomonas aeruginosa*. *PLoS Genet* 12: e1006134. doi: 10.1371/journal.pgen.1006134.
- Chow MS, Rouf MA. 1983. Isolation and Partial Characterization of Two *Aeromonas hydrophila* Bacteriophages. *Appl Environ Microbiol.* ;45(5):1670-6.
- Claverie J M ,Abergel C. 2009. Mimivirus and its virophage. *Annu.Rev.Genet.*43, 49–66. doi: 10.1146/annurev-genet-102108-134255.
- Comeau AM, Arbiol C, Krisch HM. 2014. Composite conserved promoter-terminator motifs (PeSLs) that mediate modular shuffling in the diverse T4-Like Myoviruses. *Genome Biol. Evol.* 6, 1611–1619. doi: 10.1093/gbe/evu129

- Cornelissen A, Ceyssens PJ, Krylov VN, Noben J-P, Volckaert G, Lavigne R. 2012a. Identification of EPS-degrading activity within the tail spikes of the novel *Pseudomonas putida* phage AF. *Virology* 434:251–256. doi: 10.1016/j.virol.2012.09.030.
- Cornelissen A, Hardies SC, Shaburova OV, Krylov VN, Mattheus W, Kropinski AM, Lavigne R. 2012b. Complete genome sequence of the giant virus OBP and comparative genome analysis of the diverse phiKZ-related phages. *J.Virol.*86, 1844–1852. doi: 10.1128/JVI.06330-11.
- Danis-Wlodarczyk K, Olszak T, Arabski M, Wasik S, Majkowska-Skrobek G, Augustyniak D, Gula G, Briers Y, Jang HB, Vandenheuvel D, Duda KA, Lavigne R, Drulis-Kawa Z. 2015. Characterization of the newly isolated lytic bacteriophages KTN6 and KT28 and their efficacy against *Pseudomonas aeruginosa* biofilm. *PLoS One.*;10(5):e0127603. doi: 10.1371/journal.pone.0127603.
- Danis-Wlodarczyk K, Vandenheuvel D, Jang HB, Briers Y, Olszak T, Arabski M, Wasik S, Drabik M, Higgins G, Tyrrell J, Harvey BJ, Noben JP, Lavigne R, Drulis-Kawa Z. 2016. A proposed integrated approach for the preclinical evaluation of phage therapy in infections. *Sci Rep.*, 6:28115. doi: 10.1038/srep28115.
- Day A, Ahn J and Salmond GP, 2018. Jumbo Bacteriophages Are Represented Within an Increasing Diversity of Environmental Viruses Infecting the Emerging Phytopathogen, *Dickeya solani*. *Frontiers in microbiology*, 9.
- Donelli G, Dore E, Frontali C, Grandolfo ME. 1975. Structure and physico-chemical properties of bacteriophage G. III. A homogeneous DNA of molecular weight 5 times 10^8 . *J. Mol. Biol.* 94, 555–565.
- Dömötör D, Becsagh P, Rakhely G, Schneider G, Kovacs T. 2012. Complete genomic sequence of *Erwinia amylovora* phage PhiEaH2. *J. Virol.* 86,10899–10899. doi: 10.1128/JVI.01870-12.
- Drulis-Kawa Z, Olszak T, Danis K, Majkowska-Skrobek G, Ackermann HW. 2014. A giant phage from Poland. *Arch Virol.*; 159(3): 567–572. doi: 10.1007/s00705-013-1844-y.
- Esplin IND, Berg JA, Sharma R, Allen RC, Arens DK, Ashcroft CR, Bairett SR, Beatty NJ, Bickmore M, Bloomfield TJ, Brady TS, Bybee RN, Carter JL, Choi MC, Duncan S, Fajardo CP, Foy BB, Fuhrman DA, Gibby PD, Grossarth SE, Harbaugh K, Harris N, Hilton JA, Hurst E, Hyde JR,

- Ingersoll K, Jacobson CM, James BD, Jarvis TM, Jaen-Anievas D, Jensen GL, Knabe BK, Kruger JL, Merrill BD, Pape JA, Payne Anderson AM, Payne DE, Peck MD, Pollock SV, Putnam MJ, Ransom EK, Ririe DB, Robinson DM, Rogers SL, Russell KA, Schoenhals JE, Shurtleff CA, Simister AR, Smith HG, Stephenson MB, Staley LA, Stettler JM, Stratton ML, Tateoka OB, Tatlow PJ, Taylor AS, Thompson SE, Townsend MH, Thurgood TL, Usher BK, Whitley KV, Ward AT, Ward MEH, Webb CJ, Wienclaw TM, Williamson TL, Wells MJ, Wright CK, Breakwell DP, Hope S, Grose JH. 2017. Genome Sequences of 19 Novel *Erwinia amylovora* Bacteriophages. *Genome Announc.*;5(46). pii: e00931-17. doi: 10.1128/genomeA.00931-17.
- Fallico V, Ross RP, Fitzgerald GF, McAuliffe O. 2011. Genetic response to bacteriophage infection in *Lactococcus lactis* reveals a fourstrand approach involving induction of membrane stress proteins, d- alanylation of the cell wall, maintenance of proton motive force, and energy conservation. *Journal of virology*, 85 (22), 12032–12042. pages 29, 30, 177. doi: 10.1128/JVI.00275-11.
- Fokine A, Battisti AJ, Bowman VD, Efimov AV, Kurochkina LP, Chipman PR, Mesyanzhinov VV, Rossmann MG. 2007. Cryo-EM study of the bacteriophage phiKZ. *Structure.*;15(9):1099-104. <https://doi.org/10.1016/j.str.2007.07.008>.
- Fokine A, Kostyuchenko VA, Efimov AV, Kurochkina LP, Sykilinda NN, Robben J, Volckaert G, Hoenger A, Chipman PR, Battisti AJ, Rossmann MG, Mesyanzhinov VV. 2005. A three-dimensional cryo-electron microscopy structure of the bacteriophage phiKZ head. *J Mol Biol.*;352(1):117-24. <https://doi.org/10.1016/j.jmb.2005.07.018>.
- Garneau JR, Depardieu F, Fortier LC, Bikard D and Monot M, 2017. PhageTerm: a tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. *Scientific reports*, 7(1), p.8292.
- Gill JJ, Berry JD, Russell WK, Lessor L, Escobar-Garcia DA, Hernandez D, Kane A, Keene J, Maddox M, Martin R, Mohan S, Thorn AM, Russell DH, Young R. 2012. The *Caulobacter crescentus* phage ϕ CbK: genomics of a canonical phage. *BMC Genomics*. 13:542. doi: 10.1186/1471-2164-13-542.

- Gregory AC, Solonenko SA, Ignacio-Espinoza JC, LaButti K, Copeland A, Sudek S, Maitland A, Chittick L, Dos Santos F, Weitz JS, Worden AZ, Woyke T, Sullivan MB. 2016. Genomic differentiation among wild cyanophages despite widespread horizontal gene transfer. *BMC Genomics.*;17(1):930. doi: 10.1186/s12864-016-3286-x.
- Hatfull GF. 2015. Dark matter of the biosphere: the amazing world of bacteriophage diversity. *J. Virol.* 89, 8107–8110. doi: 10.1128/JVI.01340-15.
- Hendrix RW. 2009. Jumbo bacteriophages. *Curr Top Microbiol Immunol.*;328:229-40.
- Hertveldt K, Lavigne R, Pleteneva E, Sernova N, Kurochkina L, Korchevskii R, Robben J, Mesyanzhinov V, Krylov VN, Volckaert G. 2005. Genome comparison of *Pseudomonas aeruginosa* large phages. *J. Mol. Biol.* 354, 536–545. <https://doi.org/10.1016/j.jmb.2005.08.075>.
- Høyland-Kroghsbo NM, Mærkedahl RB, Svenningsen SL. 2013. A Quorum-Sensing-Induced Bacteriophage Defense Mechanism. *mBio.*; 4(1): e00362-12. doi: 10.1128/mBio.00362-12.
- Jackson KD, Starkey M, Kremer S, Parsek MR, Wozniak DJ. 2004. Identification of *psl*, a locus encoding a potential exopolysaccharide that is essential for *Pseudomonas aeruginosa* PAO1 biofilm formation. *Journal of bacteriology*, 186(14), 4466-4475. doi: 10.1128/JB.186.14.4466-4475.2004.
- Jacquemot L, Bettarel Y, Monjol J, Corre E, Halary S, Desnues C, Bouvier T, Ferrier-Pagès C and Baudoux AC, 2018. Therapeutic potential of a new jumbo phage that infects *Vibrio coralliilyticus*, a widespread coral pathogen. *Frontiers in microbiology*, 9.
- James E, Liu M, Sheppard C, Mekler V, Cámara B, Liu B, Simpson P, Cota E, Severinov K, Matthews S, Wigneshweraraj S. 2012. Structural and mechanistic basis for the inhibition of *Escherichia coli* RNA polymerase by T7 Gp2. *Mol Cell.*;47(5):755-66. doi: 10.1016/j.molcel.2012.06.013.
- James TD, Cardozo T, Abell LE, Hsieh ML, Jenkins LM, Jha SS, Hinton DM. 2016. Visualizing the phage T4 activated transcription complex of DNA and *Escherichia coli* RNA polymerase. *Nucleic Acids Res.* ;44(16):7974-88. doi: 10.1093/nar/gkw656.

- Jang HB, Fagutao FF, Nho SW, Park SB, Cha IS, Yu JE, Lee JS, Im SP, Aoki T, Jung TS. 2013. Phylogenomic network and comparative genomics reveal a diverged member of the phiKZ-related group, marine *Vibrio* phage Φ JM-2012. *J. Virol.* 87:12866-12878. doi: 10.1128/JVI.02656-13.
- Jang HB, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, Brister JR, Kropinski AM, Krupovic M, Lavigne R, Turner D, Sullivan MB. 2019. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat Biotechnol.* doi: 10.1038/s41587-019-0100-8. [Epub ahead of print]
- Kakabadze E, Makalatia K, Grdzlishvili N, Bakuradze N, Goderdzishvili M, Kusradze I, Phoba MF, Lunguya O, Lood C, Lavigne R and Jacobs J, 2018. Selection of Potential Therapeutic Bacteriophages that Lyse a CTX-M-15 Extended Spectrum β -Lactamase Producing *Salmonella enterica* Serovar Typhi Strain from the Democratic Republic of the Congo. *Viruses*, 10(4), p.172.
- Kanamaru S, Ishiwata Y, Suzuki T, Rossmann MG, Arisaka F. 2005. Control of bacteriophage T4 tail lysozyme activity during the infection process. *J. Mol. Biol.* 346 p.1013. <https://doi.org/10.1016/j.jmb.2004.12.042>.
- Kiljunen S, Hakala K, Pinta E, Huttunen S, Pluta P, Gador A, Lönnberg H, Skurnik M. 2005. Yersiniophage ϕ R1-37 is a tailed bacteriophage having a 270 kb DNA genome with thymidine replaced by deoxyuridine. *Microbiology-SGM* 151, 4093–4102. doi: 10.1099/mic.0.28265-0
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L., 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology*, 14(4), p.R36.
- Kim JH, Son JS, Choi YJ, Choresca CH, Shin SP, Han JE, Jun JW, Park SC. 2012. Complete genome sequence and characterization of a broad-host range T4-like bacteriophage ϕ AS5 infecting *Aeromonas salmonicida* subsp *salmonicida*. *Vet. Microbiol.* 157, 164–171. doi: 10.1016/j.vetmic.2011.12.016.
- Kim MS, Hong SS, Park K, Myung H. 2013. Genomic analysis of bacteriophage PBECO4 infecting *Escherichia coli* O157:H7. *Arch. Virol.* 158, 2399–2403. doi: 10.1007/s00705-013-1718-3.

- Klimuk E, Akulenko N, Makarova K S, Ceysens P J, Volchenkov I, Lavigne R, Severinov K 2013. Host RNA polymerase inhibitors encoded by phiKMV-like phages of . *Virology*, 436(1), 67-74. doi: 10.1016/j.virol.2012.10.021.
- Koonin EV, Dolja VV, Krupovic M. 2015. Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology*.;479-480:2-25. doi: 10.1016/j.virol.2015.02.039.
- Krylov V, Pleteneva E, Bourkaltseva M, Shaburova O, Volckaert G, Sykilinda N, Kurochkina L, Mesyanzhinov V. 2003. bacteriophages of *Pseudomonas aeruginosa*: a long and complex evolutionary pathway. *Res Microbiol.*;154(4):269-75. [https://doi.org/10.1016/S0923-2508\(03\)00070-6](https://doi.org/10.1016/S0923-2508(03)00070-6).
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ and Marra MA, 2009. Circos: an information aesthetic for comparative genomics. *Genome research*, 19(9), pp.1639-1645.
- Kwan T, Liu J, Dubow M, Gros P, Pelletier J. 2006. Comparative genomic analysis of 18 *Pseudomonas aeruginosa* bacteriophages. *J Bacteriol.*; 188(3):1184-7. doi: 10.1128/JB.188.3.1184-1187.2006.
- Kwiatkowski B, Boschek B, Thiele H, Stirm S. 1983. Substrate specificity of two bacteriophage-associated endo-N-acetylneuraminidases. *J Virol.*;45(1):367-74.
- Lal TM, Sano M, Hatai K, Ransangan J. 2016. Complete genome sequence of a giant *Vibrio* phage ValKK3 infecting *Vibrio alginolyticus*. *Genom Data.*;8:37-8. doi: 10.1016/j.gdata.2016.03.002.
- Lavysh D, Sokolova M, Minakhin L, Yakunina M, Artamonova T, Kozyavkin S, Makarova KS, Koonin EV, Severinov K. 2016. The genome of AR9, a giant transducing *Bacillus* phage encoding two multisubunit RNA polymerases. *Virology.*;495:185-96. doi: 10.1016/j.virol.2016.04.030.
- Lavysh D, Sokolova M, Slashcheva M, Förstner KU, Severinov K. 2017. Transcription Profiling of *Bacillus subtilis* Cells Infected with AR9, a Giant Phage Encoding Two Multisubunit RNA Polymerases. *MBio.*;8(1). pii: e02041-16. doi: 10.1128/mBio.02041-16.

- Lee JH, Bai J, Shin H, Kim Y, Park B, Heu S, Ryu S. 2015. A Novel Bacteriophage Targeting *sakazakii* Is a Potential Biocontrol Agent in Foods. *Appl Environ Microbiol.*;82(1):192-201. doi: 10.1128/AEM.01827-15.
- Lee JH, Shin H, Kim H, Ryu S. 2011. Complete Genome sequence of *Salmonella* bacteriophage SPN3US. *J. Virol.* 85, 13470–13471. doi: 10.1128/JVI.06344-11.
- Leskinen K, Blasdel BG, Lavigne R, Skurnik M. 2016. RNA sequencing reveals the progression of phage-host interactions between ϕ r1-37 and *Yersinia enterocolitica*. *Viruses*, 8 (4), 111. doi: 10.3390/v8040111.
- Li H, 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.
- Lin YR, Lin CS. 2012. Genome-wide characterization of *Vibrio* phage ϕ pp2 with unique arrangements of the mob-like genes. *BMC Genomics.* ;13:224. doi: 10.1186/1471-2164-13-224.
- Liu B, Shadrin A, Sheppard C, Mekler V, Xu Y, Severinov K, Matthews S, Wigneshweraraj S. 2014. A bacteriophage transcription regulator inhibits bacterial transcription initiation by σ -factor displacement. *Nucleic Acids Res.*;42(7):4294-305. doi: 10.1093/nar/gku080.
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, 15, 550. doi: 10.1186/s13059-014-0550-8
- Lowe, T.M. and Chan, P.P., 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic acids research*, 44(W1), pp.W54-W57.
- Lukashin AV, Borodovsky M. 1998. GeneMark.hmm: New solutions for gene finding. *Nucleic Acids Res.* 26, 1107–1115.
- Luo ZH, Yu YP, Jost G, Xu W, Huang XL. 2015. Complete genome sequence of a giant *Vibrio* bacteriophage VH7D. *Mar Genomics.*;24 Pt 3:293-5. doi: 10.1016/j.margen.2015.
- Ma L, Conover M, Lu H, Parsek MR, Bayles K, Wozniak DJ. 2009. Assembly and development of *Pseudomonas aeruginosa* biofilm matrix. *PLoS Pathog*;5(3):1000354. doi: 10.1371/journal.ppat.1000354.

- McNair K, Aziz RK, Pusch GD, Overbeek R, Dutilh BE and Edwards, R., 2018. Phage genome annotation using the RAST pipeline. In *Bacteriophages* (pp. 231-238). Humana Press, New York, NY.
- Matsui T, Yoshikawa G, Mihara T, Chatchawankanphanich O, Kawasaki T, Nakano M, Fujie M, Ogata H and Yamada T, 2017. Replications of two closely related groups of jumbo phages show different level of dependence on host-encoded RNA polymerase. *Frontiers in microbiology*, 8, p.1010.
- Meczker K, Dömötör D, Vass J, Rakhely G, Schneider G, Kovacs T. 2014. The genome of the *Erwinia amylovora* phage Φ EaH1 reveals greater diversity and broadens the applicability of phages for the treatment of fire blight. *FEMS Microbiol. Lett.* 350, 25–27. <https://doi.org/10.1111/1574-6968.12319>
- Mesyanzhinov VV, Robben J, Grymonprez B, Kostyuchenko VA, Bourkaltseva MV, Sykilinda NN, Krylov VN, Volckaert G. 2002. The genome of bacteriophage ϕ KZ of *Pseudomonas aeruginosa*. *J. Mol. Biol.* 317, 1–19. <https://doi.org/10.1006/jmbi.2001.5396>.
- Miller ES, Heidelberg JF, Eisen JA, Nelson WC, Durkin AS, Ciecko A, Feldblyum TV, White O, Paulsen IT, Nierman WC, Lee J, Szczypinski B, Fraser CM. 2003. Complete genome sequence of the broad-host-range *Vibrio* phage KVP40: comparative genomics of a T4-related bacteriophage. *J. Bacteriol.* 185, 5220–5233. doi: 10.1128/JB.185.17.5220-5233.2003.
- Minakhin L, Severinov K. 2005. Transcription regulation by bacteriophage T4 AsiA. *Protein Expr Purif.*; 41(1):1-8. <https://doi.org/10.1016/j.pep.2004.09.019>.
- Monson R, Foulds I, Foweraker J, Welch M, Salmond GP. 2011. The *Pseudomonas aeruginosa* generalized transducing phage ϕ PA3 is a new member of the ϕ KZ-like group of 'jumbo' phages, and infects model laboratory strains and clinical isolates from cystic fibrosis patients. *Microbiology.*; 157(Pt 3):859-67. doi: 10.1099/mic.0.044701-0.
- Montero-Diez C, Deighan P, Osmundson J, Darst SA, Hochschild A. 2013. Phage-encoded inhibitor of *Staphylococcus aureus* transcription exerts context-dependent effects on promoter function in a modified *Escherichia coli*-based transcription system. *J Bacteriol.*; 195(16):3621-8. doi: 10.1128/JB.00499-13.

- Nattestad M, Chin CS and Schatz MC, 2016. Ribbon: visualizing complex genome alignments and structural variation. *BioRxiv*, p.082123.
- Okon E, Dethlefsen S, Pelnikovich A, van Barneveld A, Munder A, and Tümmler B (2017). Key role of an ADP-ribose-dependent transcriptional regulator of NAD metabolism for fitness and virulence of *Pseudomonas aeruginosa*. *International Journal of Medical Microbiology*, 307(1), 83-94.
- Okonechnikov K, Golosova O, Fursov M and Ugene Team, 2012. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, 28(8), pp.1166-1167.
- Olszak T, Danis-Wlodarczyk K, Arabski M, Gula G, Maciejewska B, Wasik S, Lood C, Higgins G, Harvey BJ, Lavigne R, Drulis-Kawa Z. 2019. *Pseudomonas aeruginosa* PA5oct jumbo phage impact planktonic and biofilm population and reduces host virulence. *bioRxiv* <https://doi.org/10.1101/405027>.
- Pajunen M, Kiljunen S, Skurnik M. 2000. Bacteriophage ϕ YeO3-12, specific for *Yersinia enterocolitica* serotype O:3, is related to coliphages T3 and T7. *J Bacteriol.*;182:5114–5120.
- Palmer KL, Brown SA, Whiteley M. 2007. Membrane-Bound Nitrate Reductase Is Required for Anaerobic Growth in Cystic Fibrosis Sputum. *J Bacteriol.*; 189(12): 4449–4455. doi: 10.1128/JB.00162-07.
- Pan YJ, Lin TL, Lin YT, Su PA, Chen CT, Hsieh PF, Hsu CR, Chen CC, Hsieh YC, Wang JT. 2015. Identification of capsular types in carbapenem-resistant *Klebsiella pneumoniae* strains by wzc sequencing and implications for capsule depolymerase treatment. *Antimicrob. Agents Chemother.* 59, 1038–1047. doi:10.1128/AAC.03560-14
- Pavlova O, Lavysh D, Klimuk E, Djordjevic M, Ravcheev DA, Gelfand MS, Severinov K, Akulenko N. 2012. Temporal regulation of gene expression of the *Escherichia coli* bacteriophage ϕ Eco32. *J Mol Biol.*;416(3):389-99. doi: 10.1016/j.jmb.2012.01.002.
- Petrov VM, Ratnayaka S, Nolan JM, Miller ES, Karam JD. 2010. Genomes of the T4-related bacteriophages as windows on microbial genome evolution. *Virology*. 7:292. doi:10.1186/1743-422X-7-292.

- Ravantti JJ, Ruokoranta TM, Alapuranen AM, Bamford DH. 2008. Global transcriptional responses of *Pseudomonas aeruginosa* to phage prr1 infection. *Journal of virology*, 82 (5), 2324–2329. pages 28, 177, 193. doi: 10.1128/JVI.01930-07.
- Sabra W, Lunsdorf H., Zeng AP. 2003. Alterations in the formation of lipopolysaccharide and membrane vesicles on the surface of *Pseudomonas aeruginosa* PAO1 under oxygen stress conditions. *Microbiology* 149:2789-2795. doi: 10.1099/mic.0.26443-0.
- Sayers EW, Barrett T, Benson D, Bolton E, Bryant S, Canese K, Chetvernin V, Church D, DiCuccio M, Federhen S, Feolo M, Fingerman I, Geer L, Helmberg W, Kapustin Y, Landsman D, Lipman D, Lu Z, Madden T, Madej T, Maglott D, Marchler-Bauer A, Miller V, Mizrahi I, Ostell J, Panchenko A, Phan L, Pruitt K, Schuler G, Sequeira E, Sherry S, Shumway M, Sirotkin K, Slotta D, Souvorov A, G S, Tatusova T, Wagner L, Wang Y, Wilbur W, Yaschenko E, Ye J. 2011. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 39:D38–51. doi: 10.1093/nar/gkq1172.
- Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A. and Schatz, M.C., 2018. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods*, 15(6), pp.461-468.
- Serwer P, Hayes SJ, Thomas JA, Hardies SC. 2007. Propagating the missing bacteriophages: a large bacteriophage in a new class. *Virology*; 4:21. doi: 10.1186/1743-422X-4-21.
- Shen CJ, Liu YJ, Lu CP. 2012. Complete genome sequence of *Aeromonas hydrophila* Phage CC2. *J. Virol.* 86, 10900–10900. doi: 10.1128/JVI.01882-12.
- Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW. 2005. Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *Plos Biol.* 3:e144. doi: 10.1371/journal.pbio.0030144.
- Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR, DeFrancesco AS, Kern SE, Thompson LR, Young S, Yandava C, Fu R, Krastins B, Chase M, Sarracino D, Osburne MS, Henn MR, Chisholm SW. 2010. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ. Microbiol.* 12, 3035–3056. doi: 10.1111/j.1462-2920.2010.02280.x.

- Sykilinda NN, Bondar AA, Gorshkova AS, Kurochkina LP, Kulikov EE, Shneider MM, Kadykov VA, Solovjeva NV, Kabilov MR, Mesyanzhinov VV, Vlassov VV, Drukker VV, Miroshnikov KA. 2014. Complete genome sequence of the novel giant phage PaBG. *Genome Announc.*;2(1). pii: e00929-13. doi: 10.1128/genomeA.00929-13.
- Šimoliūnas E, Kaliniene L, Truncaitė L, Zajančauskaitė A, Staniulis J, Kaupinis A, Ger M, Valius M, Meškys R. 2013. phage vB_KleM-RaK2 – a giant singleton virus of the family. *PLoS ONE* 8:e60717. doi: 10.1371/journal.pone.0060717.
- Tagami S, Sekine S, Minakhin L, Esyunina D, Akasaka R, Shirouzu M, Kulbachinskiy A, Severinov K, Yokoyama S. 2014. Structural basis for promoter specificity switching of RNA polymerase by a phage factor. *Genes Dev.*;28(5):521-31. doi: 10.1101/gad.233916.113.
- Taylor MJ, Thorne CB. 1963. Transduction of *Bacillus licheniformis* and *Bacillus subtilis* by each of two phages. *J Bacteriol.*;86:452-61.
- Thomas JA, Hardies SC, Rolando M, Hayes SJ, Lieman K, Carroll CA, Weintraub ST, Serwer P. 2007. Complete genomic sequence and mass spectrometric analysis of highly diverse, atypical *Bacillus thuringiensis* phage 0305 ϕ 8-36. *Virology* 368, 405–421. doi: 10.1016/j.virol.2007.06.043.
- Thomas JA, Rolando MR, Carroll CA, Shen PS, Belnap DM, Weintraub ST, Serwer P, Hardies SC. 2008. Characterization of *chlororaphis* myovirus 2012-1 via genomic sequencing, mass spectrometry, and electron microscopy. *Virology* 376:330 –338. doi: 10.1016/j.virol.2008.04.004.
- Thomas JA, Weintraub ST, Hakala K, Serwer P, Hardies SC. 2010. Proteome of the large myovirus 201 phi 2-1: delineation of proteolytically processed virion proteins. *Mol. Cell. Proteomics* 9, 940–951. doi: 10.1074/mcp.M900488-MCP200.
- White RA 3rd, Suttle CA. 2013. The Draft Genome Sequence of *Sphngomonas paucimobilis* Strain HER1398 (Proteobacteria), Host to the Giant PAU Phage, Indicates That It Is a Member of the Genus *Sphingobacterium* (*Bacteroidetes*). *Genome Announc.*;1(4). pii: e00598-13. doi: 10.1128/genomeA.00598-13.

- Wick, R.R., Judd, L.M., Gorrie, C.L. and Holt, K.E., 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS computational biology*, 13(6), p.e1005595.
- Wu W, Thomas JA, Cheng N, Black LW, Steven AC. 2012. Bubblegrams reveal the inner body of bacteriophage phiKZ. *Science*.;335(6065):182. doi: 10.1126/science.1214120.
- Yagubi AI, Castle AJ, Kropinski AM, Banks TW, Svircev AM. 2014. Complete genome sequence of *Erwinia amylovora* bacteriophage vB_EamM_Ea35-70. *Genome Announc.* 2:e00413-14. doi: 10.1128/genomeA.00413-14.
- Yamada T, Satoh S, Ishikawa H, Fujiwara A, Kawasaki T, Fujie M, Ogata H. 2010. A jumbo phage infecting the phytopathogen *Ralstonia solanacearum* defines a new lineage of the family. *Virology* 398, 135–147. doi: 10.1016/j.virol.2009.11.043.
- Yang H, Ma Y, Wang Y, Yang H, Shen W, Chen X. 2014. Transcription regulation mechanisms of bacteriophages: recent advances and future prospects. *Bioengineered*;5(5):300-4. doi: 10.4161/bioe.32110.
- Yoon SS, Hennigan RF, Hilliard GM, Ochsner UA, Parvatiyar K, Kamani MC, Allen HL, DeKievit TR, Gardner PR, Schwab U, Rowe JJ, Iglewski BH, McDermott TR, Mason RP, Wozniak DJ, Hancock RE, Parsek MR, Noah TL, Boucher RC, Hassett DJ. 2002. *Pseudomonas aeruginosa* anaerobic respiration in biofilms: relationships to cystic fibrosis pathogenesis. *Dev. Cell* 3:593-603. [https://doi.org/10.1016/S1534-5807\(02\)00295-2](https://doi.org/10.1016/S1534-5807(02)00295-2).
- Yoshikawa, G., Askora, A., Blanc-Mathieu, R., Kawasaki, T., Li, Y., Nakano, M., Ogata, H. and Yamada, T., 2018. *Xanthomonas citri* jumbo phage XacN1 exhibits a wide host range and high complement of tRNA genes. *Scientific reports*, 8(1), p.4486.
- Yuan Y, Gao M. 2016a. Proteomic Analysis of a Novel *Bacillus* Jumbo Phage Revealing Glycoside Hydrolase As Structural Component. *Front Microbiol.*; 7:745. doi: 10.3389/fmicb.2016.00745.
- Yuan Y, Gao M. 2016b. Characteristics and complete genome analysis of a novel jumbo phage infecting pathogenic *Bacillus pumilus* causing ginger rhizome rot disease. *Arch Virol.*;161(12):3597-3600.

Yuan Y, Gao M. 2017. Jumbo Bacteriophages: An Overview. *Front Microbiol.*; 8: 403. doi: 10.3389/fmicb.2017.00403.