environmental microbiology

Environmental Microbiology (2020) **22**(6), 2165–2181 doi:10.1111/1462-2920.14979

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

**Cédric Lood,[1,2†] Katarzyna Danis-Wlodarczyk,[1,3†] Bob G. Blasdel,[1] Ho Bin Jang,[1] Dieter Vandenheuvel,[1] Yves Briers,[1] Jean-Paul Noben,[4] Vera van Noort,[2,5] Zuzanna Drulis-Kawa[3] and Rob Lavigne[1*]**

[1]*Department of Biosystems, Laboratory of Gene Technology, KU Leuven, Leuven, Belgium.*

[2]*Department of Microbial and Molecular Systems, Laboratory of Computational Systems Biology, KU Leuven, Leuven, Belgium.*

[3]*Department of Pathogen Biology and Immunology, Institute of Genetics and Microbiology, University of Wroclaw, Wroclaw, Poland.*

[4]*Biomedical Research Institute and Transnational University Limburg, Hasselt University, Diepenbeek, Belgium.*

[5]*Institute of Biology, Leiden University, Leiden, The Netherlands.*

## Summary

***Pseudomonas* virus vB_PaeM_PA5oct is proposed as a model jumbo bacteriophage to investigate phage-bacteria interactions and is a candidate for phage therapy applications. Combining hybrid sequencing, RNA-Seq and mass spectrometry allowed us to accurately annotate its 286,783 bp genome with 461 coding regions including four non-coding RNAs (ncRNAs) and 93 virion-associated proteins. PA5oct relies on the host RNA polymerase for the infection cycle and RNA-Seq revealed a gradual take-over of the total cell transcriptome from 21% in early infection to 93% in late infection. PA5oct is not organized into strictly contiguous regions of temporal transcription, but some genomic regions transcribed in early, middle and late phases of infection can be discriminated. Interestingly, we observe regions showing limited transcription activity throughout the infection cycle. We show that PA5oct upregulates specific bacterial operons during infection including operons *pncA-pncB1-nadE* involved in NAD biosynthesis, *psl* for exopolysaccharide biosynthesis and *nap* for periplasmic nitrate reductase production. We also observe a downregulation of T4P gene products suggesting mechanisms of superinfection exclusion. We used the proteome of PA5oct to position our isolate amongst other phages using a gene-sharing network. This integrative omics study illustrates the molecular diversity of jumbo viruses and raises new questions towards cellular regulation and phage-encoded hijacking mechanisms.**

Received 23 June, 2019; revised , ; accepted 6 March, 2020. *For correspondence. E-mail rob.lavigne@kuleuven.be; Tel. +32 16 37 95 24; Fax +32 16 3 21965. †These authors contributed equally to this work.

## 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 and Abergel, 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 functional sequence space (Hatfull, 2015).

Bacteriophages with genome sizes beyond 200 kbp are defined 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 and Gao, 2017). A subset of 152 of them have been sequenced to date (February 2020, Supporting Information Table S1), including 141 myoviruses, 10 siphoviruses and 1 unclassified phage. Most infect Gram-negative bacteria with only a few that can infect Gram-positive bacteria (11 jumbo phages), mostly *Bacillus* strains. The tiny plaques typically formed by these

© 2020 The Authors. *Environmental Microbiology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

large-sized phages are smaller than 0.5 mm on 0.7% soft agar and are easily overlooked during classical propagation assays. Their large virion size can also lead to their loss during filtration procedures that are standard in many phage isolation protocols (Serwer *et al*., 2007), suggesting biases against their isolation. Interestingly, other large bacteriophages currently lacking isolates have been reported using culture-free, deep metagenomics sequencing approaches, with assembled genome sizes up to 735 kbp (Al-Shayeb *et al*., 2020).

Although jumbo phages have proven to be interesting for structural analysis and have provided insights into infection processes (Fokine *et al*., 2005; 2007; 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, investigations into their basic (molecular) microbiology, including elucidating their transcriptional schemes, or the mechanisms of their coevolution with their host must be pursued further if we are to acquire new insights (Hendrix, 2009). RNA-Seq 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*., 2017; 2018a).

*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 RNA polymerases (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. PA5oct is close to the recently described *Pseudomonas* phage MIJ3. The genomic proximity between these two isolates is discussed in the paper by Imam and colleagues (2019), in which they highlight a few peculiar genetic features unique to this phage species, including the presence of 2 host-like proteins (FtsH and TilS), as well as the absence of the machinery thought to be conserved across jumbo phages for the construction of a phage nucleus during the replication cycle. PA5oct, along with MIJ3, represents a new species within a new phage genus that appears to be geographically dispersed and genomically stable, and we propose here to further characterize its molecular biology.

Our investigation was driven by combining short and long read sequencing, temporal transcriptomics and structural proteomics. Taken together, these molecular techniques allowed us to elucidate the genome organization of this phage species and the response it elicits during infection of its host *P. aeruginosa* PAO1. Furthermore, we establish the evolutionary relationships between PA5oct and other characterized phages using a gene-sharing network analysis.

## Results

### *Basic genome properties of* Pseudomonas *phage PA5oct*

Phage PA5oct has a linear, A+T-rich (33.3% GC-content), double-stranded DNA genome of 286,783 bp with 461 putative coding DNA sequences (CDSs) as shown in Fig. 1. The analysis of the genome sequence shows that this isolate possesses little sequence homology to other phages in the currently known, that is, sequenced and published, phage universe except with the *Pseudomonas* phage MIJ3 (Imam *et al*., 2019) with which it shares 99% DNA identity and coverage and with which it forms a new viral species.

The genome of PA5oct has long direct terminal repeats (DTRs) of an estimated 39,080 bases (see Supplementary Report 1), 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 (Fig. 2).

Aside from 445 protein-coding sequences, the phage genome also contains 12 transfer RNAs (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)].

Using the alignment of the RNA-Seq data to the PA5oct genome (presented in the next section), we validated the annotation of the tRNA sites and uncovered four actively transcribed regions lacking putative open reading frames (ORFs) or known promoters that have been annotated as non-coding RNA (ncRNA) (Table 1). This alignment of the RNA-Seq data also enabled the manual curation and discovery of 49 additional CDSs on top of the 412 CDSs initially identified by the automated *in silico* RAST annotation pipeline (McNair *et al*., 2018), bringing the total to 461 CDSs.

© 2020 The Authors. *Environmental Microbiology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

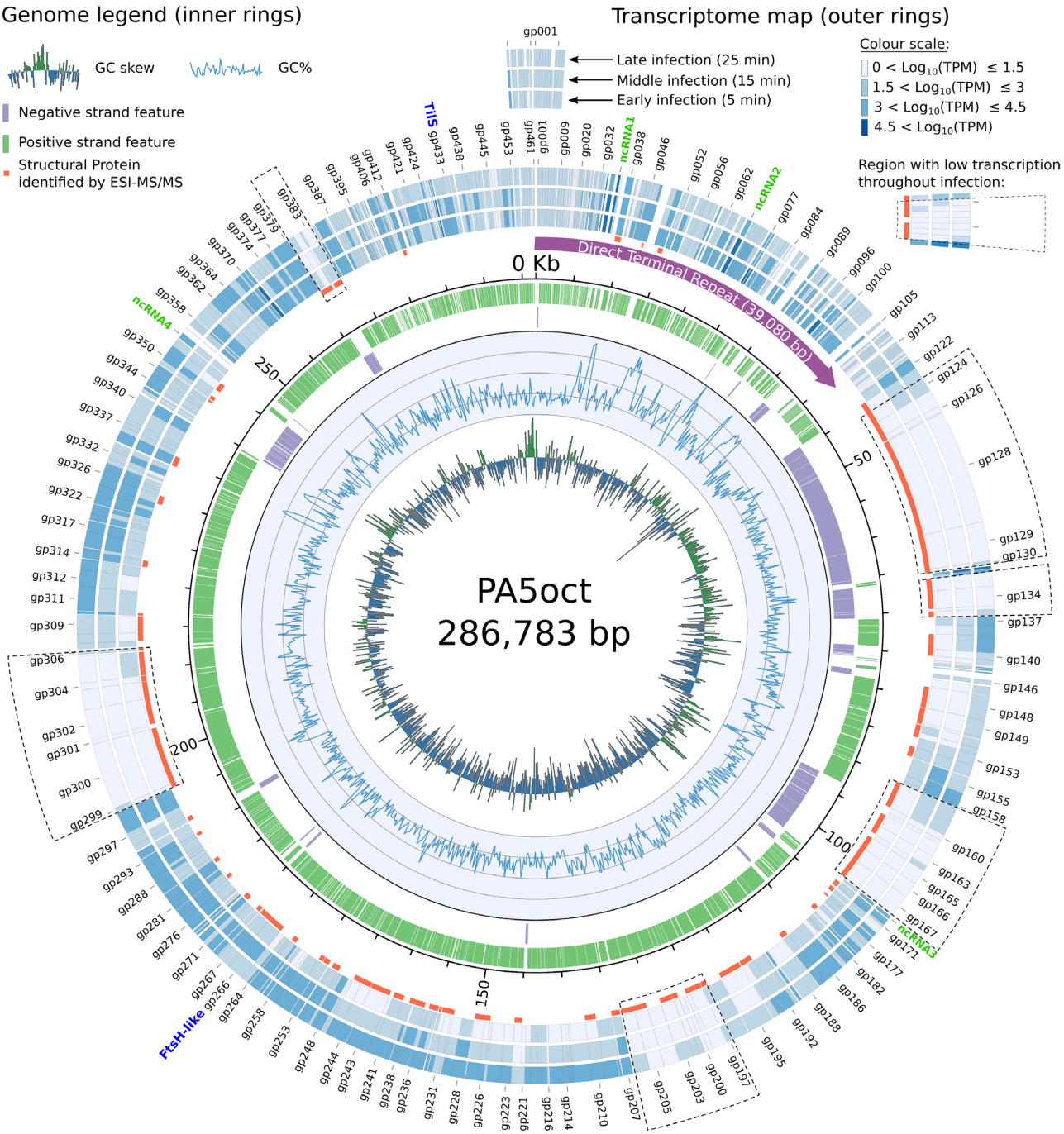



**Fig. 1.** Circular representation of PA5oct genome and its transcriptome. The two inner circles display a histogram of the GC-skew followed by a graph of the GC content along the genome of PA5oct. It is followed by a stranded depiction of the CDSs and ncRNAs indicating that most features are encoded 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 marked with red dots. The three outer rings depict normalized RNA read counts for each of the annotated feature. Each annotated feature is represented to size, with for example gp128 and gp300 being the largest proteins encoded by PA5oct. Contiguous regions with limited amounts of transcripts detected across multiple timepoints are encased in dotted lines. The two host-like proteins FtshH and TilS reported in Imam and colleagues (2019) are annotated in blue-colored text. The figure was created using the software Circos (Krzywinski *et al*., 2009). [Color figure can be viewed at wileyonlinelibrary.com]

### *Transcription of the PA5oct genome and its temporal regulation*

In accordance with the results from Drulis-Kawa and colleagues (2014), we calculated that PA5oct has a latent period just over 40 min, with a burst size of about 30–40 phage particles per infected bacterial cell (Supporting Information Fig. S1). Based on these parameters, we selected timepoints 5, 15 and 25 min as proxies for early, middle and late infection respectively.

© 2020 The Authors. *Environmental Microbiology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

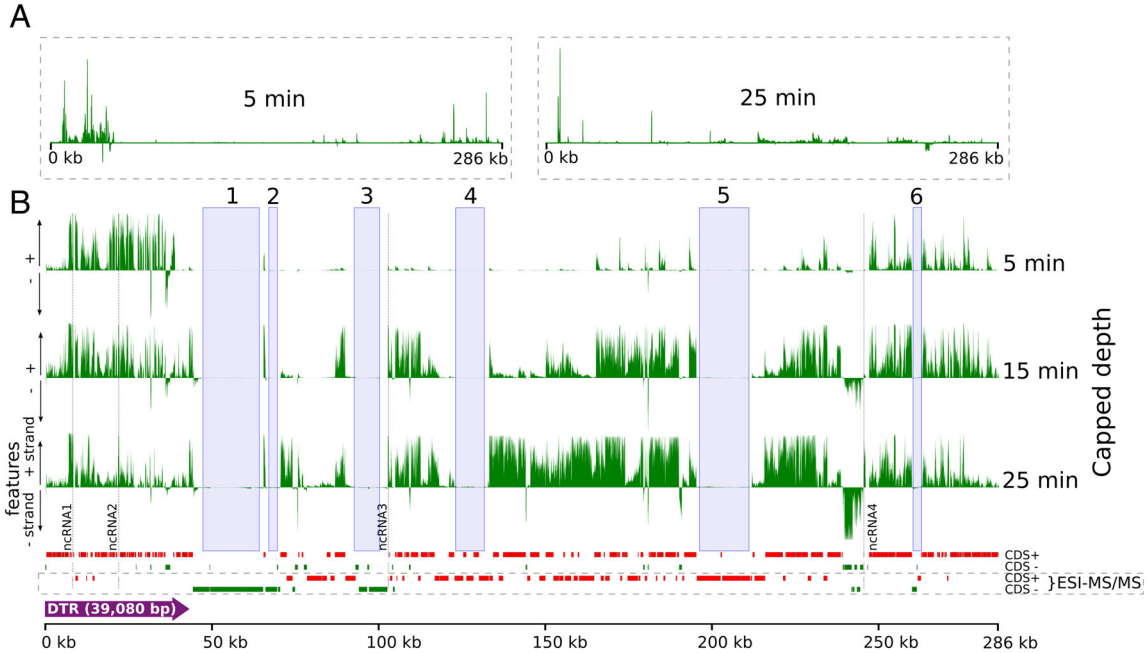



**Fig. 2.** RNA-Seq analysis of PA5oct transcriptome after *P. aeruginosa* PAO1 infection. Genome-wide overview of reads mapped to the PA5oct genome for samples taken 5, 15 and 25 min post-infection.
A. Coverage of samples 5 min and 25 min (15 min omitted) show a limited number of regions which dominate 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 was normalized to reflect the depth across the triplicates. The four tracks at the bottom of the graph (B) indicate the features annotated on both the positive strand (in red) and negative strand (in green), these features have been further subdivided in two sets of tracks in order to highlight the structural features identified by the ESI-MS/MS technique. The figure was created using the python software *pyGenomeTracks*. [Color figure can be viewed at wileyonlinelibrary.com]

**Table 1.** Identified non-coding RNA species.

| | | | Percentage of phage reads | | |
|---|---|---|---|---|---|
| Name | Location | Strand | 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 |

Four non-coding RNAs have been identified based on RNA-Seq data from transcripts aligning to genomic loci that lack plausible ORFs.

Importantly, when 400 μg ml$^{-1}$ of rifampicin, a host RNAP transcription inhibitor, is added prior PA5oct infection, the progeny phage production is completely suppressed (Supporting Information Fig. S1).

Throughout the infection, it appears that the 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) (Supporting Information - Table S2). The strand specificity of the results revealed that early transcription is almost exclusively occurring from the Watson strand. We also observe localized section of the genome with high transcription levels that dominate the transcription landscape of PA5oct (Fig. 2A).

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, were 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 min onwards (Fig. 2B). Many of the (functionally annotated) gene features transcribed in those middle and late time points were predicted to be responsible for structural proteins as well as DNA metabolism and replication. Two of the four non-coding RNAs located in the DTR region (ncRNA1 and ncRNA2), and which lack similarity to any known primary DNA sequence in the GenBank database (release 235)

© 2020 The Authors. *Environmental Microbiology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

with the exception of the isolate MIJ3, were highly transcribed throughout the infection cycle (Figs 1 and 2).

Remarkably, six genomic regions, including two of the three major regions transcribed on the crick strand (Fig. 2B, boxes 1–6), showed 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 (Fig. 1, gene products marked in red) and which have been detected by electrospray ionization tandem mass spectrometry (ESI-MS/MS) (described in the *Transcriptome and proteome-based functional annotations* section).

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 (Supporting Information Table S3) whereas 36 putative factor-independent terminators with conserved stem loops could be identified around transcription stop sites (Supporting Information Table S4).

### *Transcriptome and proteome-based functional annotations*

The ESI-MS/MS technique allowed us to identify a total of 93 gene products linked to the phage virion particles, comprising 7 virion-unrelated enzymes, 13 virion-associated proteins and 73 other gene products (Supporting Information Fig. S2 and Table S5). Combined with the bioinformatics analysis of the genome and the transcriptome, we could functionally annotate a third of the CDSs 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 fibre 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*., 1983).

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

**Table 2.** Putative peptidoglycan degradation proteins.

| Gene product | ESI-MS/MS | Functional analysis (bioinformatics) |
|---|---|---|
| gp133 | Yes | Baseplate hub subunit with tail-associated lysozyme homologous to the gp5 of *Escherichia coli* bacteriophage T4 which has a needle-like structure attached to the end of the tail tube serving in the injection process (Kanamaru *et al*., 2005). |
| gp173 | Yes | Glycosyl hydrolase, lysozyme-like protein with a 159 amino acids long domain (11–170 aa). |
| gp245 | Yes | Baseplate hub and tail muramyl-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). |

Five proteins are predicted to be associated with the degradation of the peptidoglycan layer. For each protein we indicate (i) the related CDSs by its gene product (gp) number, (ii) whether they were detected during the ESI-MS/MS analysis and (iii) their bioinformatically predicted function.

### *Gene-sharing relationships to other bacteriophages*

Viruses lack a universal marker gene that would enable an integrated taxonomy. In order to represent genetic relationships between *Pseudomonas* phage PA5oct, the jumbo phages and the other viral genomes available in RefSeq, a gene-sharing network was built using the approach described in Jang and colleagues (2019). In such networks, nodes are virus genomes, edges between nodes indicate the gene content similarities between genomes and viruses sharing a high number of genes [i.e., homologous protein clusters (PCs)] localize into viral clusters (VCs, approximating the genus-rank viral groups) (Jang *et al*., 2019).

Using the Cytoscape software, we removed disconnected components (or subnetworks that did not contain jumbo phages) from our gene-sharing network to simplify its visualization on Fig. 3 [see, e.g., fig. 2C of Jang *et al*., 2019 for a full gene-sharing network of all known phages]. The resulting network was composed of 1355

© 2020 The Authors. *Environmental Microbiology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

viral genomes (nodes) belonging to the *Myoviridae*, *Siphoviridae* and *Podoviridae* families, or uncharacterized phages and 37,955 relationships (edges) between them.

A subset of 23 jumbo phages were placed into the largest connected component (LCC) that predominates the double-stranded DNA phages, whereas 22 others were placed into three isolated components (Fig. 3). Note that phage K64-1 was excluded from our gene content-based analyses (Figs 3 and 4A and B) since its incomplete protein annotation (accession No. NC_027399.1; only includes 64 proteins from a total of 300 predicted proteins) can lead to underestimating its genetic relationship to the other phages.

We observe that a majority (18) of the 23 jumbo phages in the LCC, which includes PA5oct, are related to the T4-like viruses belonging to the *Tevenvirinae* (upper right in Fig. 3; Supporting Information Table S6). Of these viruses, five jumbo phages including *Pseudomonas* phage PA5oct, *Escherichia* phages 121Q/PBECO 4, *Klebsiella* phage RaK2 and *Cronobacter* phage GAP32 showed stronger connections to each other but weaker connections outside the group due to their more common genes (PCs) than the rest of the network (Fig. 4A and B, Supporting Information Table S6). Notably, further inspection of their relationships uncovered that PA5oct shares less common PCs (~11–15%) to the remaining members (Fig. 4B), indicating its distant relationships as an outlier genome (Jang *et al*., 2019).

The visualization of the gene-sharing network and the analysis of gene content relationships of the jumbo phages shows us that most of the jumbo phages are related to two major groups including the T4-like viruses belonging to the *Tevenvirinae* and the PhiKZ-like viruses to the *Phikzvirus*, *Elvirus*, *Agrican357virus or Rsl2virus*, respectively, while the others form evolutionary distinct clades (Yamada *et al*., 2010; Gill *et al*., 2012) (Figs 3 and 4A and B; Supporting Information Table S6).

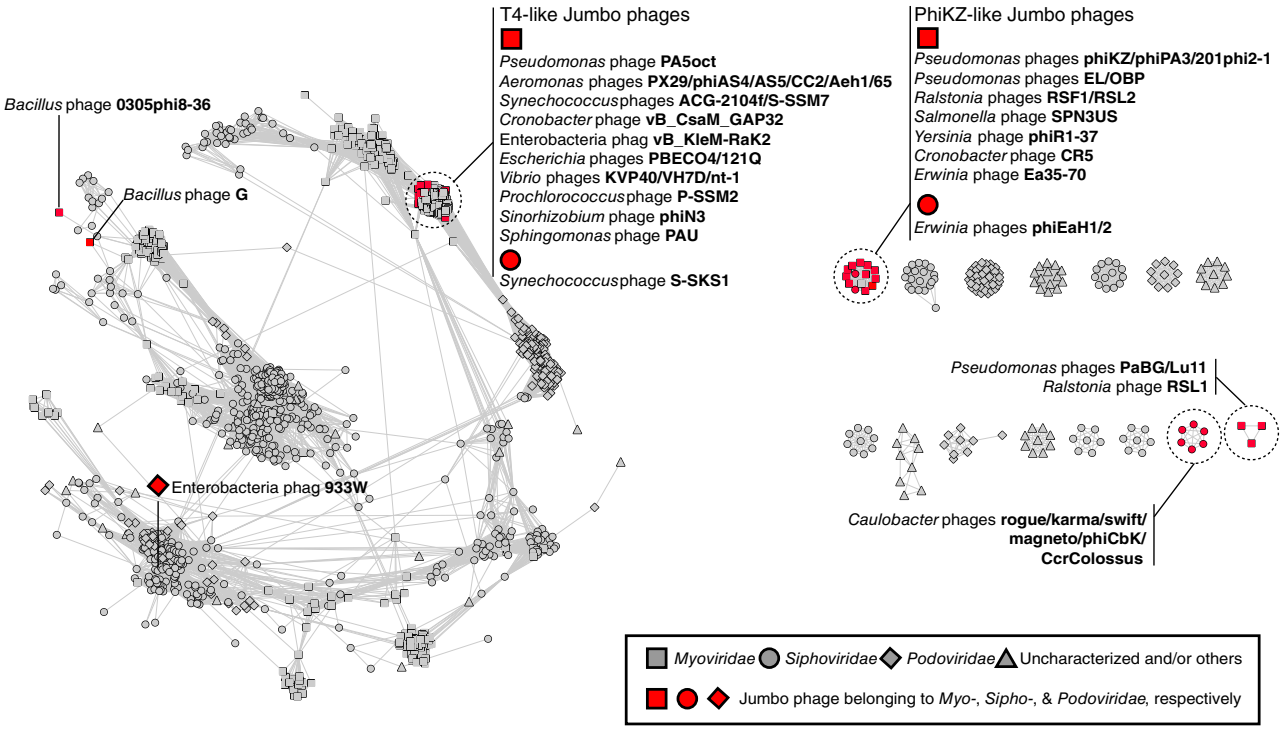
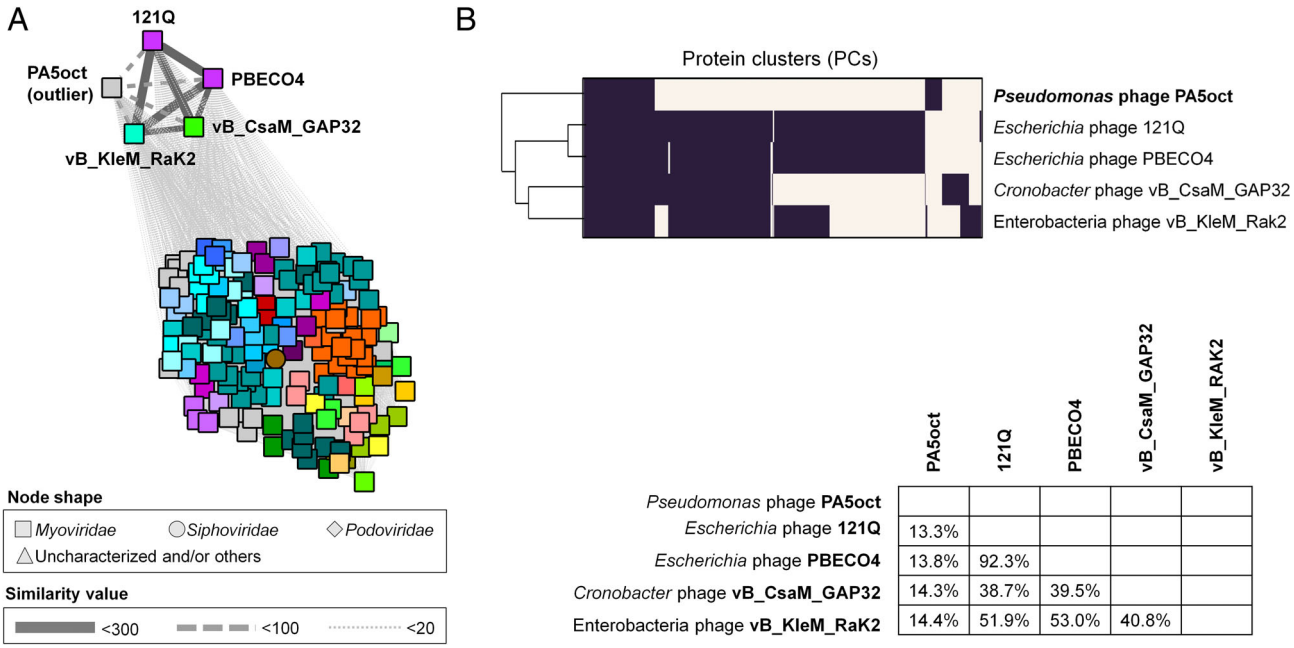



**Fig. 3.** Protein-sharing network of PA5oct and other jumbo phages. A network was built for 2304 bacterial and archaeal virus genomes retrieved from ViralRefSeq v.85 (see *Experimental procedures* section). We removed several disconnected components or subnetworks that do not contain the jumbo phages or their relatives for clarity, resulting in 1355 viral genomes (nodes) and 37,955 relationships (edges) between genomes. Each node is depicted as a different shape, representing bacteriophages belonging to the *Myoviridae* (rectangle), *Podoviridae* (diamond), *Siphoviridae* (circle) or uncharacterized phages (triangle). Edges between two nodes indicate statistically weighted pairwise similarities with similarity scores of ≥ 2. The positions of 45 selected jumbo phages (red node) and two major viral groups they belong to are indicated for illustrative purposes. The network consists of multiple disconnected components. On the left-hand side is the component that contains the largest number of nodes, including with the phages from the RaK2-like clade in which we find PA5oct (highlighted and annotated with the isolates names). The other components on the right-hand side of the figure are smaller clusters of related phages that could not be connected are disconnected from to the rest of the network. Amongst those, three separate clusters comprise jumbo phages. The network representation was produced using the edge-weighted spring embedded layout of Cytoscape version 3.5.1. The Supporting Information Table S6 contains the list of phages and assigned VCs. [Color figure can be viewed at wileyonlinelibrary.com]

© 2020 The Authors. *Environmental Microbiology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.,




| | PA5oct | 121Q | PBECO4 | vB_CsaM_GAP32 | vB_KleM_RAK2 |
|---|---|---|---|---|---|
| *Pseudomonas* phage **PA5oct** | | | | | |
| *Escherichia* phage **121Q** | 13.3% | | | | |
| *Escherichia* phage **PBECO4** | 13.8% | 92.3% | | | |
| *Cronobacter* phage **vB_CsaM_GAP32** | 14.3% | 38.7% | 39.5% | | |
| Enterobacteria phage **vB_KleM_RaK2** | 14.4% | 51.9% | 53.0% | 40.8% | |

**Fig. 4.** Gene content-based analysis of phage PA5oct and its relatives.
A. Enlarged view of the *Tevenvirinae*-related subnetwork (Fig. 3) comprising phage PA5oct and its relatives. Each node indicates an individual viral genome. Nodes are colored according to the VC to which they belong. Viruses belonging to this section of the network are overwhelmingly of the *Myoviridae* family (legendary box). Edge thickness is proportional to similarity values estimated with the hypergeometric equation (see *Experimental procedures* section; Bolduc *et al*., 2017a,b). Note that due to the lower number of common genes, Pseudomonas phage PA5oct was identified as a distant relative (outlier; Jang *et al*., 2019) to five phages including 121Q, K64-1, PBECO4, GAP32 and Rak2.
B. Module profiles showing the presence (dark) and absence (light) of homologous PCs across genomes. Each row represents a given virus isolate and each column represents a PC. The genomes were hierarchically grouped based on hierarchical patterns of gene sharing (Upper). A table showed the average shared PC percentage (proteome similarity; see *Experimental procedures* section) between five viruses (Bottom). Note that phage K64-1 was excluded from our gene content-based analyses (Figs 3 and 4A and B) since its incomplete protein annotation in NCBI (i.e., only includes 64 proteins from a total of 300 predicted proteins) can lead to incorrect estimation of genetic relationships of K64-1 to others. [Color figure can be viewed at wileyonlinelibrary.com]

### *PA5oct-specific differential gene expression of the host genome*

In order to visually summarize the results of the differential gene expression, we created a MA plot in which we compare the late infection conditions ($t$ = 25 min) to the uninfected cell (Fig. 5). In total, 698 genes were found to be differentially expressed (Supporting Information - Table S7). In this work, we highlight the host-mediated transcriptional response described in the next paragraph, and a few genomic systems of interest that are part of the phage-specific transcriptional response and are of relevance to phage microbiology or phage therapeutic settings. Further dissection of the results is possible, and we have made the data publicly available for that purpose (see *Data availability statement* section).

The conserved PAO1 host-mediated transcriptional response to phage infection present during the infection by phages 14-1, PEV2, YuA and LUZ19 and described by Blasdel and colleagues (2018a) is also found in the infection with PA5oct and we used it to annotate our results on Fig. 5. Briefly, this conserved response was uncovered by infecting *P. aeruginosa* PAO1 with strictly lytic phages from divergent clades and studying the commonalities in the transcriptional response of the host to these infections using RNA-Seq. Using this technique, Blasdel and colleagues (2018a) showed that around 5.5% of the CDSs of *P. aeruginosa* PAO1, including genetic systems that reduce the efficiency of the lytic phage cycle, were differentially transcribed and could be defined as part of a host-mediated stress response.

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 (Fig. 5) 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 (Fig. 5, Supporting Information Table S7), as described previously (Blasdel *et al*., 2017; 2018a). Doing this allows us to illustrate the impact of PA5oct on host transcripts relative to other host transcripts, rather than relative to the total.

© 2020 The Authors. *Environmental Microbiology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

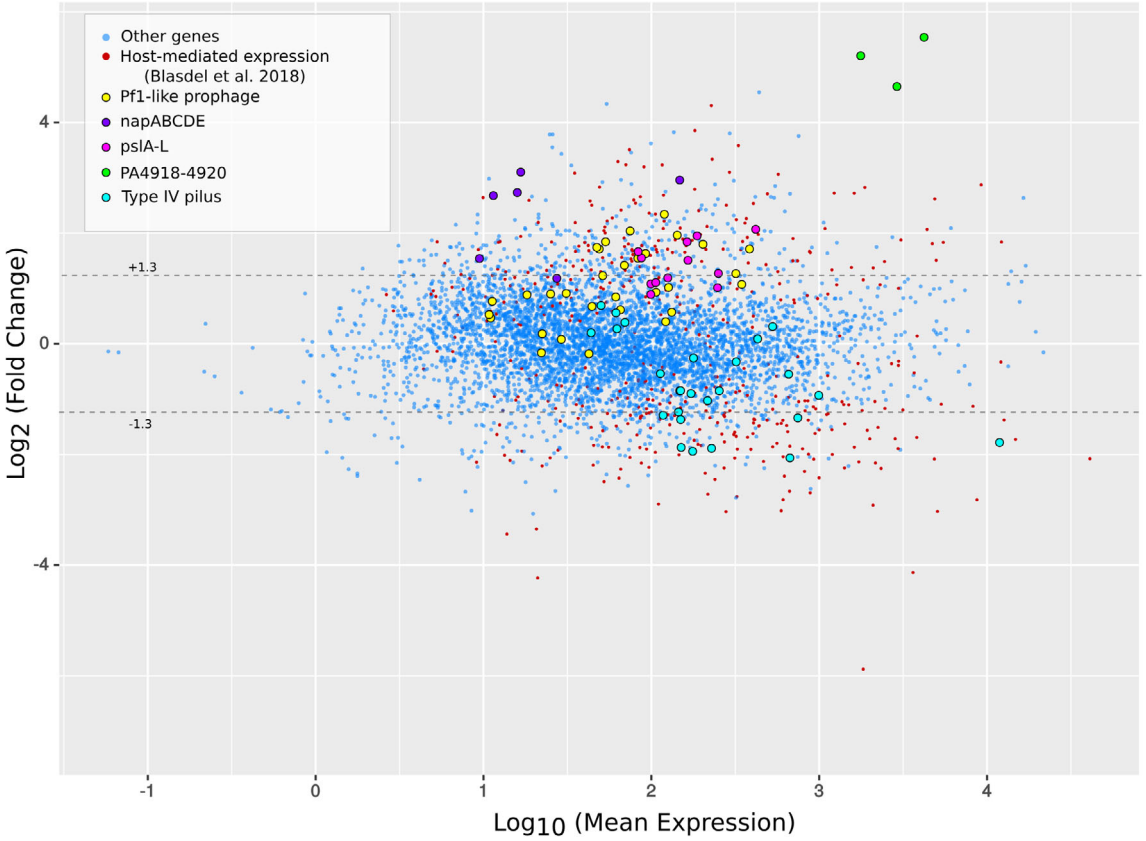



**Fig. 5.** MA plot highlighting the differential expression of host genes from *P. aeruginosa* PAO1 during PA5oct infection. We compared the gene expression between the late infection timepoint ($t$ = 25 min) and the uninfected or control timepoint ($t$ = 0 min) to identify host features that are differentially transcribed. We found that 698 genes from PAO1 were differentially transcribed, of which 281 were downregulated, and 417 were upregulated. In addition to the host-mediated stress response to phage infection identified by Blasdel and colleagues (2018a), we observe a 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 have been shown to be a necessary receptor to initiate PA5oct infection (Olszak *et al*., 2019). The exhaustive list of differentially expressed genes can be found in the Supporting Information Table S7. [Color figure can be viewed at wileyonlinelibrary.com]

As shown on Fig. 5, the strongest upregulation is that of the PA4918-PA4920 operon, which contains an enzyme involved in NAD biosynthesis (Okon *et al*., 2017). 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 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*., 2018a), PA5oct appears to mobilize the transcription of genes of a potentially cryptic Pf1 prophage (family *Inoviridae*) encoded from ORF PA0618 to PA0642.

## 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 201ϕ2-1 1 (316.674 kbp, Thomas *et al*., 2008) and phiPA3 (309.208 kbp, Monson *et al*., 2011). To date, the largest myovirus isolate known is *Bacillus megaterium* phage G (497,513 bp genome, Supporting Information Table S1). 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) (Ackermann and Nguyen, 1983; Abbasifar *et al*., 2014) and that ongoing sequencing efforts of these isolates, or metagenomics (Al-Shayeb *et al*., 2020) efforts will likely disrupt the rankings of jumbo phage genome sizes.

© 2020 The Authors. *Environmental Microbiology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

In terms of annotation, functions could be attributed for a third of PA5oct proteins and a large portion of predicted CDSs are unique, lacking sequence-based similarity to the genomes available in the public database GenBank (r235) with the notable exception of the same species isolate, *Pseudomonas* phage MIJ3 (Imam *et al*., 2019). For this reason, the functional characterization of phage orphan genes is an important endeavour to undertake in order to elucidate basic phage biology as well as phage application safety when thinking of therapeutic settings (Yuan and Gao, 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) that is putatively involved in the transitioning of the host by processing the transcripts during the infection cycle.

ESI-MS/MS allowed us to identify 93 structural proteins of PA5oct, including 4 structural head-related proteins, 1 neck protein, and 4 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). Interestingly, it is known that the capsid of jumbo phages can have very complex structures, as we can observe in the case of phage phiKZ (30 head structural proteins), where in the middle of the capsid there is 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 hold sizeable genomes allowing these phages to possess many accessory genes comparing to small phages (Mesyanzhinov *et al*., 2002; Hertveldt *et al*., 2005; Kiljunen *et al*., 2005; Thomas *et al*., 2007; Gill *et al*., 2012; Ceyssens *et al*., 2014; Yuan and Gao, 2017). As such, the capsid can accommodate a genome with DTRs that are longer (39,080 bp) than entire phage genomes, and which encodes a large set of virion-associated proteins, some of which appear to be barely transcribed. PA5oct's genome also contains 12 tRNA genes that likely increase the translation efficiency (Bailly-Bechet *et al*., 2007). This is not uncommon, as we show in the Supporting Information Table S1 with other jumbo phages that are known to possess numerous tRNA genes, such as *Xanthomonas* virus XacN1 with its record 56 tRNAs. As expected, the difference in GC content between *P. aeruginosa* PAO1 (66%) and PA5oct (33%) is reflected in alternate codon usage (see Supporting Information Table S8). However, this codon bias is not fully explained by the phage-encoded tRNAs.

### *Hijacking the transcriptional environment of the host*

The strategy of phage transcriptional mechanisms depends primarily on the presence or absence of a phage RNAP (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 apparent ability to mount a conserved host-mediated transcription-level defence response, this suggests that PA5oct depends 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 ($\alpha 2$, $\beta$, $\beta'$,$\omega$), that transcribes mRNA following binding to the DNA template (Tagami *et al*., 2014; Murakami, 2002). This process is often supported by phage-encoded $\sigma$ factors forming a holoenzyme with RNAP ($\alpha 2$, $\beta$, $\beta'$,$\omega$, $\sigma$) (Williams *et al*., 1987; Pavlova *et al*., 2012).

Some phages can also implement other transcriptional strategies, for example, 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 (Semenova *et al*., 2005; Savalia *et al*., 2010; 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 and Rothman-Denes, 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,

© 2020 The Authors. *Environmental Microbiology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

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 as described by Chevallereau and colleagues (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 time point, we observe a lot of transcription in the DTR region (Fig. 2). These early coding sequences are classically understood to be involved in the transitioning from host to phage metabolism. Distinctive 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 report 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.

The unique host-like proteins FtsH and TilS discovered previously in the jumbo phage MIJ3 and hypothesized to be involved in the maintenance of host viability during the infection (Imam *et al*., 2019) can also be found in PA5oct, and are both shown to be transcribed throughout the infection cycle (Fig. 1).

### *Comparative genome analysis and protein-sharing network*

To examine the genetic relationships between PA5oct and other *Myoviridae*, *Siphoviridae* and *Podoviridae* viruses, a protein-sharing network was constructed. This technique allowed us to locate our isolate within the phylogenetic clade of Rak2-like viruses, named after the Klebsiella phage vB_KleM-RaK2 (Šimoliūnas *et al*., 2013), and which also includes *Escherichia* viruses 121Q/PBECO4, *Klebsiella* virus K64-1 and *Cronobacter* virus vB_CsaM_GAP32 and has since been shown to include other jumbo phages (Yoshikawa *et al*., 2018). This clustering is consistent with the more targeted phylogeny of *Pseudomonas* phage MIJ3 from Imam *et al*. (2019) in which they used a multiple sequence alignment of the amino acid sequences from the major capsid and terminase proteins to build a more phylogeny using a neighbour-joining tree method.

With respect to other jumbo phages, the overall gene-sharing network (Fig. 3) identifies informative connections. For example, the *Phicbkvirus*, and phages RSL1, PaBG and Lu11 form two isolated components (Fig. 3). This discontinuous structure of two viral groups, due to their distinct gene pools, is indicative of their evolutionary relationships as separate viral lineages (Yamada *et al*., 2010; Gill *et al*., 2012) whereas the connection of jumbo phages to their relatives having smaller genomes (< 200 kbp) such as Bacillus phage G, phages of the *Phikzvirus*, *Elvirus*, *Agrican357virus* and *Rsl2virus*, and *Bacillus* phages 0305phi8-36 (Fig. 3A and Supporting Information Table S6) supports evolutionary links between jumbo phages and smaller-genome phages (Hendrix, 2009; Adriaenssens *et al*., 2012; Jang *et al*., 2013). Furthermore, PA5oct, PBECO4, GAP32, T5, FelixO1 appear to be distantly diverged members of *Tevenvirinae*. The T5 and FelixO1 phages appear 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 informs about the possible genetic relationships for PA5oct and includes all the giant viruses within the entire population of viruses.

### *Host stress response during PA5oct phage 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 (Ravantti *et al*., 2008; Fallico *et al*., 2011; Ainsworth *et al*., 2013; Leskinen *et al*., 2016). In total, 698 genes of the PAO1 host were found to be differentially expressed.

© 2020 The Authors. *Environmental Microbiology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

We observed that PA5oct upregulates the operon encoding *psIE-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 matrix production, reducing the relative abundance of lipopolysaccharide (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 and likely 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). This upregulation appears to be ubiquitous during phage infection and 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*., 2018a). 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 signalling 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 clonal phages through secondary adsorption to already phage-infected bacteria (Abedon, 2017).

## Experimental procedures

### *Bacteriophage propagation, purification and morphology*

In all experiments, phage PA5oct was propagated using the *P. aeruginosa* strain PAO1 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.

### *One-step growth curve and rifampicin growth assay*

One-step growth curves were established according to the method of Pajunen and colleagues (2000), with modifications. An equal volume of *P. aeruginosa* PAO1 bacterial culture in LB medium at $OD_{600}$ of 0.3 was infected with phage PA5oct ($10^6$ pfu $ml^{-1}$) to obtain a multiplicity of infection (MOI) of 0.01. Phages could adsorb for 8 min at 37°C, after which the mixture was diluted to $10^{-4}$ to limit post-adsorption to uninfected bacteria. Triplicate samples were taken every 5 min during 1–1.5 h for titration, (Danis-Wlodarczyk *et al*., 2015). The burst size was calculated as the ratio of the final amount of liberated phage particles to the initial number of phages in the infected culture. Where appropriate, rifampicin (Sigma) was added to a final concentration of 400 μg $ml^{-1}$.

### *DNA isolation and hybrid sequencing*

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

### *RNA extraction and sequencing*

*Pseudomonas 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

© 2020 The Authors. *Environmental Microbiology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

until an $OD_{600}$ of 0.3 was reached (around $1.2 \times 10^8$ CFU $ml^{-1}$, early exponential phase). The culture was infected with PA5oct at a MOI of 50. To establish a synchronous infection, we ensured that less than 5% of bacterial survivors remained after 5 min post-infection. Biologically independent samples were collected in triplicates at 5 min (early), 15 min (middle) and 25 min (late) into infection and processed as described previously (Chevallereau *et al*., 2016; Blasdel *et al*., 2017).

### *Genome 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, 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 search for potentially related genome sequences was performed using the blastn tool against the nr/nt nucleotide database (NCBI Resources Coordinators, 2016), and further inspected as described in *Protein family clustering and construction of the relationships network* section.

The genome annotation was performed by integrating the results of the RAST phage annotation tool (McNair *et al*., 2018) with a manual curation using standard genome annotation methods (Danis-Wlodarczyk *et al*., 2015; 2016). tRNAs were identified using tRNAscan-SE (Lowe and Chan, 2016). The annotation of the CDSs was further inspected visually with UGENE (Okonechnikov *et al*., 2012) using the stranded RNA-Seq reads, mapped to the phage genome (see *Transcriptome analysis* section).

The detection of regulatory sequences extracted using a custom python script in the region up to 100 nt upstream of the CDSs was done using the MEME software v5.0.5 with defaults parameters (Bailey *et al*., 2015), and we used the software ARNold to identify putative factor-independent terminators with default parameters (Naville *et al*., 2011). The analysis of the codon usage was performed by concatenating the CDSs of PAO1 and separately of PA5oct, before using the Sequence Manipulation Suite (SMS) tool Codon Usage (Stothard, 2000).

### *Transcriptome analysis*

The RNA-Seq reads were aligned to both the phage and host genomes using the software TopHat v2 (Kim *et al*., 2013a,b) 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 transcripts per million. The alignments were parsed in order to count the number of reads mapping to the chromosome of PAO1 and the chromosome of PA5oct respectively using the idxstats function within the software package Samtools v1.9 (Li *et al*., 2009).

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 and colleagues (2018b).

### *ESI-MS/MS analysis on virion particle proteins*

Phage proteins were isolated from a purified phage lysate ($10^9$ pfu $ml^{-1}$) by a single methanol/chloroform extraction (1:1:0.75, vol./vol./vol.) (Acros Organics) and precipitated by addition of an equal volume of methanol ($14{,}000 \times g$, 6 min). The phage proteins were separated on a 12% SDS-PAGE gel and analysed by ESI-MS/MS as previously described (Ceyssens *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 2304 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*., 2017b). 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 PCs in the same manner as previously described (Bolduc *et al*., 2017a). 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 $\geq 1$ were grouped into VCs, with default parameters of vConTACT v2.0 (Jang *et al*., 2019). The network was

© 2020 The Authors. *Environmental Microbiology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

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). As previously described (Bolduc *et al*., 2017a; Jang *et al*., 2019), the fraction of PCs between two genomes (i.e., proteome similarity) was computed using the geometric index (*G*). The proteome similarity was estimated as

$$G_{AB} = \lfloor N(A) \cap N(B) \rfloor / (\lfloor N(A) \rfloor \times \lfloor N(B) \rfloor)$$

where $N(A)$ and $N(B)$ indicate the numbers of PCs and singletons in the genomes of A and B respectively.

## Acknowledgements

We are grateful to Steve Abedon, Harald Brüssow and the reviewers for their critical comments and helpful suggestions on the manuscript. 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 post-doctoral 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).

## Author contributions

KDW performed one-step growth experiments and analysis of ESI-MS/MS results. KDW and CL performed DNA isolation and annotation of the genome. HBJ generated protein family clustering, network construction and analysed the data set. DV and CL helped with promoter identification. CL, KDW and YB performed the genome sequencing and assembly. JPN performed ESI-MS/MS analysis. KDW and BB prepared samples for RNA-Seq. CL, KDW and BGB analysed the transcriptomics results. KDW, CL, BGB, HBJ, ZDK and RL designed experiments and wrote the manuscript.

## Data availability statement

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 data set is available through the NCBI GEO database via the accession number GSE130190.

## References

Abbasifar, R., Griffiths, M.W., Sabour, P.M., Ackermann, H. W., Vandersteegen, K., Lavigne, R., *et al*. (2014) Supersize me: *Cronobacter sakazakii* phage GAP32. *Virology* **460**: 138–146.

Abedon, S.T. (2017) Phage "delay" towards enhancing bacterial escape from biofilms: a more comprehensive way of viewing resistance to bacteriophages. *AIMS Microbiol* **3**: 186–226.

Ackermann, H.W., and Nguyen, T.M. (1983) Sewage coliphages studied by electron microscopy. *Appl Environ Microbiol* **45**: 1049–1059.

Adams, M.H. (1959) *Bacteriophages*. New York, NY, USA: Interscience Publishers, Inc.

Adams, M.J., Lefkowitz, E.J., King, A.M., Harrach, B., Harrison, R.L., Knowles, N.J., *et al*. (2017) Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2017). *Arch Virol* **162**: 2505–2538.

Adriaenssens, E.M., Mattheus, W., Cornelissen, A., Shaburova, O., Krylov, V.N., Kropinski, A.M., and Lavigne, R. (2012) Complete genome sequence of the giant *Pseudomonas* phage Lu11 (2012). *J Virol* **86**: 6369–6370.

Adriaenssens, E.M., Krupovic, M., Knezevic, P., Ackermann, H.W., Barylski, J., Brister, J.R., *et al*. (2017) Taxonomy of prokaryotic viruses: 2016 update from the ICTV bacterial and archaeal viruses subcommittee. *Arch Virol* **162**: 1153–1157.

Ainsworth, S., Zomer, A., Mahony, J., and van Sinderen, D. (2013) Lytic infection of *Lactococcus lactis* by bacteriophages Tuc2009 and c2 triggers alternative transcriptional host responses. *Appl Environ Microbiol* **79**: 4786–4798.

Al-Shayeb, B., Sachdeva, R., Chen, L.X., Ward, F., Munk, P., Devoto, A., *et al*. (2020) Clades of huge phages from across Earth's ecosystems. *Nature* **578**: 425–431.

Anders, S., Pyl, P.T., and Huber, W. (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169.

Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015) The MEME suite. *Nucleic Acids Res* **43**: W39–W49.

Bailly-Bechet, M., Vergassola, M., and Rocha, E. (2007) Causes for the intriguing presence of tRNAs in phages. *Genome Res* **17**: 1486–1495.

Blasdel, B.G., Chevallereau, A., Monot, M., Lavigne, R., and Debarbieux, L. (2017) Comparative transcriptomics analyses reveal the conservation of an ancestral infectious strategy in two bacteriophage genera. *ISME J* **11**: 1988.

Blasdel, B.G., Ceyssens, P.J., Chevallereau, A., Debarbieux, L., and Lavigne, R. (2018a) Comparative transcriptomics reveals a conserved Bacterial Adaptive

© 2020 The Authors. *Environmental Microbiology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

Phage Response (BAPR) to viral predation. BioRxiv. 248849.

Blasdel, B.G., Ceyssens, P.J., and Lavigne, R. (2018b) Preparing cDNA libraries from lytic phage-infected cells for whole transcriptome analysis by RNA-Seq. In *Bacteriophages*. Clokie M., Kropinski A., Lavigne R. (eds). New York, NY, USA: Humana Press, pp. 185–194.

Bolduc, B., Jang, H.B., Doulcier, G., You, Z.Q., Roux, S., and Sullivan, M.B. (2017a) vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. *PeerJ* **5**: e3243.

Bolduc, B., Youens-Clark, K., Roux, S., Hurwitz, B.L., and Sullivan, M.B. (2017b) iVirus: facilitating new insights in viral ecology with software and community data sets imbedded in a cyberinfrastructure. *ISME J* **11**: 7.

Burgener, E.B., Sweere, J.M., Bach, M.S., Secor, P.R., Haddock, N., Jennings, L.K., *et al*. (2019) Filamentous bacteriophages are associated with chronic *Pseudomonas* lung infections and antibiotic resistance in cystic fibrosis. *Sci Transl Med* **11**: eaau9748.

Ceyssens, P.J., Lavigne, R., Mattheus, W., Chibeu, A., Hertveldt, K., Mast, J., *et al*. (2006) Genomic analysis of *Pseudomonas aeruginosa* phages LKD16 and LKA1: establishment of the φKMV subgroup within the T7 supergroup. *J Bacteriol* **188**: 6924–6931.

Ceyssens, P.J., Mesyanzhinov, V., Sykilinda, N., Briers, Y., Roucourt, B., Lavigne, R., *et al*. (2008) The genome and structural proteome of YuA, a new *Pseudomonas aeruginosa* phage resembling M6. *J Bacteriol* **190**: 1429–1435.

Ceyssens, P.J., Miroshnikov, K., Mattheus, W., Krylov, V., Robben, J., Noben, J.P., *et al*. (2009) Comparative analysis of the widespread and conserved PB1-like viruses infecting *Pseudomonas aeruginosa*. *Environ Microbiol* **11**: 2874–2883.

Ceyssens, P.J., Minakhin, L., Van den Bossche, A., Yakunina, M., Klimuk, E., Blasdel, B., *et al*. (2014) Development of giant bacteriophage ϕKZ is independent of the host transcription apparatus. *J Virol* **88**: 10501–10510.

Chevallereau, A., Blasdel, B.G., De Smet, J., Monot, M., Zimmermann, M., Kogadeeva, M., *et al*. (2016) Next-generation "-omics" approaches reveal a massive alteration of host RNA metabolism during bacteriophage infection of *Pseudomonas aeruginosa*. *PLoS Genet* **12**: e1006134.

Claverie, J.M., and Abergel, C. (2009) Mimivirus and its virophage. *Annu Rev Genet* **43**: 49–66.

Danis-Wlodarczyk, K., Olszak, T., Arabski, M., Wasik, S., Majkowska-Skrobek, G., Augustyniak, D., *et al*. (2015) Characterization of the newly isolated lytic bacteriophages KTN6 and KT28 and their efficacy against *Pseudomonas aeruginosa* biofilm. *PLoS One* **10**: e0127603.

Danis-Wlodarczyk, K., Vandenheuvel, D., Jang, H.B., Briers, Y., Olszak, T., Arabski, M., *et al*. (2016) A proposed integrated approach for the preclinical evaluation of phage therapy in *Pseudomonas* infections. *Sci Rep* **6**: 28115.

Drulis-Kawa, Z., Olszak, T., Danis, K., Majkowska-Skrobek, G., and Ackermann, H.W. (2014) A giant *Pseudomonas* phage from Poland. *Arch Virol* **159**: 567–572.

Fallico, V., Ross, R.P., Fitzgerald, G.F., and McAuliffe, O. (2011) Genetic response to bacteriophage infection in *Lactococcus lactis* reveals a four-strand approach involving induction of membrane stress proteins, D-alanylation of the cell wall, maintenance of proton motive force, and energy conservation. *J Virol* **85**: 12032–12042.

Fokine, A., Kostyuchenko, V.A., Efimov, A.V., Kurochkina, L. P., Sykilinda, N.N., Robben, J., *et al*. (2005) A three-dimensional cryo-electron microscopy structure of the bacteriophage ϕKZ head. *J Mol Biol* **352**: 117–124.

Fokine, A., Battisti, A.J., Bowman, V.D., Efimov, A.V., Kurochkina, L.P., Chipman, P.R., *et al*. (2007) Cryo-EM study of the *Pseudomonas* bacteriophage φKZ. *Structure* **15**: 1099–1104.

Garneau, J.R., Depardieu, F., Fortier, L.C., 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. *Sci Rep* **7**: 8292.

Gill, J.J., Berry, J.D., Russell, W.K., Lessor, L., Escobar-Garcia, D.A., Hernandez, D., *et al*. (2012) The *Caulobacter crescentus* phage phiCbK: genomics of a canonical phage. *BMC Genomics* **13**: 542.

Hatfull, G.F. (2015) Dark matter of the biosphere: the amazing world of bacteriophage diversity. *J Virol* **89**: 8107–8110.

Haynes, L.L., and Rothman-Denes, L.B. (1985) N4 virion RNA polymerase sites of transcription initiation. *Cell* **41**: 597–605.

Hendrix, R.W. (2009) Jumbo bacteriophages. In *Lesser Known Large dsDNA Viruses*. Van Etten, J.L. (ed). Berlin, Germany: Springer, pp. 229–240.

Hertveldt, K., Lavigne, R., Pleteneva, E., Sernova, N., Kurochkina, L., Korchevskii, R., *et al*. (2005) Genome comparison of *Pseudomonas aeruginosa* large phages. *J Mol Biol* **354**: 536–545.

Imam, M., Alrashid, B., Patel, F., Dowah, A.S., Brown, N., Millard, A.D., *et al*. (2019) vB_PaeM_MIJ3, a novel jumbo phage infecting *Pseudomonas aeruginosa*, possesses unusual genomic features. *Front Microbiol* **10**: 2772.

Jackson, K.D., Starkey, M., Kremer, S., Parsek, M.R., and Wozniak, D.J. (2004) Identification of psl, a locus encoding a potential exopolysaccharide that is essential for *Pseudomonas aeruginosa* PAO1 biofilm formation. *J Bacteriol* **186**: 4466–4475.

Jang, H.B., Fagutao, F.F., Nho, S.W., Park, S.B., Cha, I.S., Yu, J.E., *et al*. (2013) Phylogenomic network and comparative genomics reveal a diverged member of the ϕkz-related group, marine *Vibrio* phage ϕJM-2012. *J Virol* **87**: 12866–12878.

Jang, H.B., Bolduc, B., Zablocki, O., Kuhn, J.H., Roux, S., Adriaenssens, E.M., *et al*. (2019) Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat Biotechnol* **37**: 632.

Kakabadze, E., Makalatia, K., Grdzelishvili, N., Bakuradze, N., Goderdzishvili, M., Kusradze, I., *et al*. (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**: 172.

Kanamaru, S., Ishiwata, Y., Suzuki, T., Rossmann, M.G., and Arisaka, F. (2005) Control of bacteriophage T4 tail

© 2020 The Authors. *Environmental Microbiology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

lysozyme activity during the infection process. *J Mol Biol* **346**: 1013–1020.

Kiljunen, S., Hakala, K., Pinta, E., Huttunen, S., Pluta, P., Gador, A., *et al*. (2005) Yersiniophage ϕR1-37 is a tailed bacteriophage having a 270 kb DNA genome with thymidine replaced by deoxyuridine. *Microbiology* **151**: 4093–4102.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013a) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**: R36.

Kim, J.H., Son, J.S., Choi, Y.J., Choresca, C.H., Jr., Shin, S. P., Han, J.E., *et al*. (2012) Complete genome sequence and characterization of a broad-host range T4-like bacteriophage phiAS5 infecting *Aeromonas salmonicida subsp*. *salmonicida*. *Vet Microbiol* **157**: 164–171.

Kim, M.S., Hong, S.S., Park, K., and Myung, H. (2013b) Genomic analysis of bacteriophage PBECO4 infecting *Escherichia coli* O157: H7. *Arch Virol* **158**: 2399–2403.

Koonin, E.V., Dolja, V.V., and Krupovic, M. (2015) Origins and evolution of viruses of eukaryotes: the ultimate modularity. *Virology* **479**: 2–25.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., *et al*. (2009) Circos: an information aesthetic for comparative genomics. *Genome Res* **19**: 1639–1645.

Kwiatkowski, B., Boschek, B., Thiele, H., and Stirm, S. (1983) Substrate specificity of two bacteriophage-associated endo-N-acetylneuraminidases. *J Virol* **45**: 367–374.

Lavysh, D., Sokolova, M., Slashcheva, M., Förstner, K.U., and Severinov, K. (2017) Transcription profiling of *Bacillus subtilis* cells infected with AR9, a giant phage encoding two multisubunit RNA polymerases. *MBio* **8**: e02041–e02016.

Lecoutere, E., Ceyssens, P.J., Miroshnikov, K.A., Mesyanzhinov, V.V., Krylov, V.N., Noben, J.P., *et al*. (2009) Identification and comparative analysis of the structural proteomes of ϕKZ and EL, two giant *Pseudomonas aeruginosa* bacteriophages. *Proteomics* **9**: 3215–3219.

Leskinen, K., Blasdel, B., Lavigne, R., and Skurnik, M. (2016) RNA-sequencing reveals the progression of phage-host interactions between φR1-37 and *Yersinia enterocolitica*. *Viruses* **8**: 111.

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv: 1303.3997.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al*. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Love, M.I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

Lowe, T.M., and Chan, P.P. (2016) tRNAscan-SE on-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res* **44**: W54–W57.

Ma, L., Conover, M., Lu, H., Parsek, M.R., Bayles, K., and Wozniak, D.J. (2009) Assembly and development of the *Pseudomonas aeruginosa* biofilm matrix. *PLoS Pathog* **5**: e1000354.

McNair, K., Aziz, R.K., Pusch, G.D., Overbeek, R., Dutilh, B. E., and Edwards, R. (2018) Phage genome annotation using the RAST pipeline. In *Bacteriophages*. Clokie M., Kropinski A., Lavigne R. (eds). New York, NY, USA: Humana Press, pp. 231–238.

Mesyanzhinov, V.V., Robben, J., Grymonprez, B., Kostyuchenko, V.A., Bourkaltseva, M.V., Sykilinda, N.N., *et al*. (2002) The genome of bacteriophage φKZ of *Pseudomonas aeruginosa*. *J Mol Biol* **317**: 1–19.

Monson, R., Foulds, I., Foweraker, J., Welch, M., and Salmond, G.P. (2011) The *Pseudomonas aeruginosa* generalized transducing phage φPA3 is a new member of the φKZ-like group of 'jumbo' phages, and infects model laboratory strains and clinical isolates from cystic fibrosis patients. *Microbiology* **157**: 859–867.

Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O., and Darst, S.A. (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* **296**: 1285–1290.

Nattestad, M., Chin, C.S., and Schatz, M.C. (2016) Ribbon: visualizing complex genome alignments and structural variation. BioRxiv; 082123.

Naville, M., Ghuillot-Gaudeffroy, A., Marchais, A., and Gautheret, D. (2011) ARNold: a web tool for the prediction of Rho-independent transcription terminators. *RNA Biol* **8**: 11–13.

NCBI Resources Coordinators. (2016) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **44**: D7.

Okon, E., Dethlefsen, S., Pelnikevich, 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*. *Int J Med Microbiol* **307**: 83–94.

Okonechnikov, K., Golosova, O., Fursov, M., and Ugene Team. (2012) Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28**: 1166–1167.

Olszak, T., Danis-Wlodarczyk, K., Arabski, M., Gula, G., Maciejewska, B., Wasik, S., *et al*. (2019) *Pseudomonas aeruginosa* PA5oct jumbo phage impacts planktonic and biofilm population and reduces its host virulence. *Viruses* **11**: 1089.

Pajunen, M., Kiljunen, S., and Skurnik, M. (2000) Bacteriophage φYeO3-12, specific for *Yersinia enterocolitica* serotype O: 3, is related to coliphages T3 and T7. *J Bacteriol* **182**: 5114–5120.

Palmer, K.L., Brown, S.A., and Whiteley, M. (2007) Membrane-bound nitrate reductase is required for anaerobic growth in cystic fibrosis sputum. *J Bacteriol* **189**: 4449–4455.

Pavlova, O., Lavysh, D., Klimuk, E., Djordjevic, M., Ravcheev, D.A., Gelfand, M.S., *et al*. (2012) Temporal regulation of gene expression of the *Escherichia coli* bacteriophage phiEco32. *J Mol Biol* **416**: 389–399.

Ravantti, J.J., Ruokoranta, T.M., Alapuranen, A.M., and Bamford, D.H. (2008) Global transcriptional responses of *Pseudomonas aeruginosa* to phage PRR1 infection. *J Virol* **82**: 2324–2329.

Sabra, W., Lünsdorf, H., and Zeng, A.P. (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.

© 2020 The Authors. *Environmental Microbiology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

Savalia, D., Robins, W., Nechaev, S., Molineux, I., and Severinov, K. (2010) The role of the T7 Gp2 inhibitor of host RNA polymerase in phage development. *J Mol Biol* **402**: 118–126.

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**: 461.

Semenova, E., Djordjevic, M., Shraiman, B., and Severinov, K. (2005) The tale of two RNA polymerases: transcription profiling and gene expression strategy of bacteriophage Xp10. *Mol Microbiol* **55**: 764–777.

Serwer, P., Hayes, S.J., Thomas, J.A., and Hardies, S.C. (2007) Propagating the missing bacteriophages: a large bacteriophage in a new class. *Virol J* **4**: 21.

Šimoliūnas, E., Kaliniene, L., Truncaitė, L., Zajančkauskaitė, A., Staniulis, J., Kaupinis, A., *et al*. (2013) *Klebsiella* phage vB_KleM-RaK2—a giant singleton virus of the family *Myoviridae*. *PLoS One* **8**: e60717.

Stothard, P. (2000) The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **28**: 1102–1104.

Tagami, S., Sekine, S.I., Minakhin, L., Esyunina, D., Akasaka, R., Shirouzu, M., *et al*. (2014) Structural basis for promoter specificity switching of RNA polymerase by a phage factor. *Genes Dev* **28**: 521–531.

Thomas, J.A., Hardies, S.C., Rolando, M., Hayes, S.J., Lieman, K., Carroll, C.A., *et al*. (2007) Complete genomic sequence and mass spectrometric analysis of highly diverse, atypical *Bacillus thuringiensis* phage 0305ϕ8–36. *Virology* **368**: 405–421.

Thomas, J.A., Rolando, M.R., Carroll, C.A., Shen, P.S., Belnap, D.M., Weintraub, S.T., *et al*. (2008) Characterization of *Pseudomonas chlororaphis* myovirus 201ϕ2-1 via genomic sequencing, mass spectrometry, and electron microscopy. *Virology* **376**: 330–338.

Thomas, J.A., Weintraub, S.T., Hakala, K., Serwer, P., and Hardies, S.C. (2010) Proteome of the large *Pseudomonas* myovirus 201φ2-1: delineation of proteolytically processed virion proteins. *Mol Cell Proteomics* **9**: 940–951.

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 Comput Biol* **13**: e1005595.

Williams, K.P., Kassavetis, G.A., and Geiduschek, E.P. (1987) Interactions of the bacteriophage T4 gene 55 product with *Escherichia coli* RNA polymerase. Competition with *Escherichia coli* sigma 70 and release from late T4 transcription complexes following initiation. *J Biol Chem* **262**: 12365–12371.

Willis, S.H., Kazmierczak, K.M., Carter, R.H., and Rothman-Denes, L.B. (2002) N4 RNA polymerase II, a heterodimeric RNA polymerase with homology to the single-subunit family of RNA polymerases. *J Bacteriol* **184**: 4952–4961.

Wu, W., Steven, A.C., Cheng, N., Thomas, J., and Black, L. (2012) Bubblegrams reveal the inner body of bacteriophage φKZ. *Microsc Microanal* **18**: 112–113.

Yakunina, M., Artamonova, T., Borukhov, S., Makarova, K. S., Severinov, K., and Minakhin, L. (2015) A non-canonical multisubunit RNA polymerase encoded by a giant bacteriophage. *Nucleic Acids Res* **43**: 10411–10420.

Yamada, T., Satoh, S., Ishikawa, H., Fujiwara, A., Kawasaki, T., Fujie, M., and Ogata, H. (2010) A jumbo phage infecting the phytopathogen *Ralstonia solanacearum* defines a new lineage of the *Myoviridae* family. *Virology* **398**: 135–147.

Yang, H., Ma, Y., Wang, Y., Yang, H., Shen, W., and Chen, X. (2014) Transcription regulation mechanisms of bacteriophages: recent advances and future prospects. *Bioengineered* **5**: 300–304.

Yoon, S.S., Hennigan, R.F., Hilliard, G.M., Ochsner, U.A., Parvatiyar, K., Kamani, M.C., *et al*. (2002) *Pseudomonas aeruginosa* anaerobic respiration in biofilms: relationships to cystic fibrosis pathogenesis. *Dev Cell* **3**: 593–603.

Yoshikawa, G., Askora, A., Blanc-Mathieu, R., Kawasaki, T., Li, Y., Nakano, M., *et al*. (2018) *Xanthomonas citri* jumbo phage XacN1 exhibits a wide host range and high complement of tRNA genes. *Sci Rep* **8**: 4486.

Yuan, Y., and Gao, M. (2017) Jumbo bacteriophages: an overview. *Front Microbiol* **8**: 403.

## Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

**Supplementary Table S1** List of jumbo phages reported in the literature.

**Supplementary Table S2**. The RNA-seq alignments were parsed in order to count the reads mapping to either the chromosome of the *Pseudomonas aeruginosa* PAO1 host, and the chromosome of PA5oct. For each timepoint (5 min, 15 min, and 25 min), we averaged the % of reads mapping to PA5oct compared to the total amount of reads mapping.

**Supplementary Table S3**. 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 S4**. Phage PA5oct predicted terminators with palindromes marked blue. From the left-hand side: 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 S5**. List of the detected proteins by ESI-MS/MS, based on the analysis of denatured phage PA5oct particles after fractionation on SDS-PAGE gel.

**Supplementary Table S6**. List of jumbo phages (yellow color) and relevant phages that belong to the viral clusters (VCs) and their ICTV taxonomy.

**Supplementary Table S7**. The differential expression of non-rRNA gene features within *Pseudomonas aeruginosa* PAO1 genome by late PA5oct infection.

© 2020 The Authors. *Environmental Microbiology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

**Supplementary Table S8.** The codon usage of PAO1 and PA5oct are compared side by side, together with the specific codons encoded on PA5oct highlighted.

**Supplementary Figure S1** Left-hand figure: PA5oct growth curve at MOI 0.01 on *P. aeruginosa* PAO1. The numbers of PFU $ml^{-1}$ in chloroform-treated cultures shown in presence and absence of rifampin (RIF) at 400 $\mu g\ ml^{-1}$. Right-hand side figure: PA5oct growth curve in the presence of RIF.

**Supplementary Figure S2** 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 S5).

**Supplementary Report S1**. Report of the termini of PA5oct identified by the software PhageTerm (Garneau *et al*., 2017)

© 2020 The Authors. *Environmental Microbiology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.,