



Comprehensive study of new virulent bacteriophages : from transcriptomic and mechanistic characterisations towards evolutionary perspectives

Anne Chevallereau

► To cite this version:

Anne Chevallereau. Comprehensive study of new virulent bacteriophages : from transcriptomic and mechanistic characterisations towards evolutionary perspectives. Microbiology and Parasitology. Université Sorbonne Paris Cité, 2017. English. NNT : 2017USPCC162 . tel-02079406

HAL Id: tel-02079406

<https://theses.hal.science/tel-02079406>

Submitted on 26 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse de doctorat de l'Université Sorbonne Paris Cité

Préparée à l'Université Paris Diderot

Ecole doctorale Bio Sorbonne Paris Cité (ED157)

Unité de Biologie Moléculaire du Gène chez les Extrêmophiles

Groupe Interactions Phages-Bactéries chez l'Animal

Institut Pasteur

Thèse de doctorat de Microbiologie

Par **Anne CHEVALLEREAU**

Comprehensive study of new virulent bacteriophages: From transcriptomic and mechanistic characterisations towards evolutionary perspectives

Présentée et soutenue publiquement à Paris le 19 Mai 2017

Pr. Isabelle Martin-Verstraete (Université Paris Diderot)

Dr. Mireille Ansaldi (CNRS / Université Aix Marseille)

Dr. Darren Smith (University of Northumbria)

Pr. Olga Soutourina (Université Paris Sud)

Dr. Romé Voulhoux (CNRS / Université Aix-Marseille)

Dr. Laurent Derbarbieux (Insitut Pasteur)

Présidente

Rapportrice

Rapporteur

Examinatrice

Examineur

Directeur de thèse



Except where otherwise noted, this work is licensed under
<http://creativecommons.org/licenses/by-nc-nd/3.0/>

Thèse de doctorat de l'Université Sorbonne Paris Cité

Préparée à l'Université Paris Diderot

Ecole doctorale Bio Sorbonne Paris Cité (ED157)

Unité de Biologie Moléculaire du Gène chez les Extrêmophiles

Groupe Interactions Phages-Bactéries chez l'Animal

Institut Pasteur

Thèse de doctorat de Microbiologie

Par **Anne CHEVALLEREAU**

Comprehensive study of new virulent bacteriophages: From transcriptomic and mechanistic characterisations towards evolutionary perspectives

Présentée et soutenue publiquement à Paris le 19 Mai 2017

Pr. Isabelle Martin-Verstraete (Université Paris Diderot)
Dr. Mireille Ansaldi (CNRS / Université Aix Marseille)
Dr. Darren Smith (University of Northumbria)
Pr. Olga Soutourina (Université Paris Sud)
Dr. Romé Voulhoux (CNRS / Université Aix-Marseille)
Dr. Laurent Derbarbieux (Insitut Pasteur)

Présidente
Rapportrice
Rapporteur
Examinatrice
Examineur
Directeur de thèse

**Etude globale de deux nouveaux bactériophages virulents :
Caractérisations transcriptomique, mécanistique et perspectives évolutives**

Résumé:

Soutenue par le renouveau de la phagothérapie, la découverte de nouveaux bactériophages (phages) nous a permis de définir deux nouveaux genres de virus dénommés *Kpp10virus* et *Pakpunavirus* dont les mécanismes infectieux sont inconnus. Il est admis que le succès d'un cycle infectieux est notamment assuré par une réappropriation efficace des ressources de la cellule hôte, conduisant à sa transformation en « virocellule », c'est-à-dire, un organisme cellulaire exclusivement dédié à la production de particules virales. Ce travail de thèse a pour objectif d'apporter une vision globale des stratégies moléculaires utilisées par les virus appartenant aux genres *Kpp10virus* et *Pakpunavirus* (respectivement représentés par les phages PAK_P3 et PAK_P4) pour infecter le pathogène opportuniste *Pseudomonas aeruginosa*.

Dans un premier temps, nous avons évalué leurs propriétés intrinsèques en analysant le contenu de leurs génomes, leurs spectres d'hôtes, leurs paramètres de croissance ainsi qu'en identifiant leur récepteur bactérien. Dans un second temps, une combinaison d'approches transcriptomiques et métabolomiques a permis de montrer que ces deux virus ont des programmes transcriptionnels similaires, incluant notamment une régulation temporelle de leur expression génétique et la production de transcrits antisens. De plus, ils provoquent tous deux la dégradation rapide de 90% des ARNm de l'hôte, qui sont alors remplacés par des ARNm viraux. Malgré cette dégradation, nous avons constaté que ces deux phages redirigent les voies de biosynthèse bactériennes plutôt que de provoquer une extinction totale du métabolisme cellulaire, en utilisant cependant des mécanismes différents. De plus, nous avons détecté l'activation, par l'hôte, d'une réponse commune suite à une infection par PAK_P3 ou PAK_P4 et avons émis l'hypothèse qu'il s'agit d'une tentative de réparation des importants dommages ARN induits par l'infection virale.

Enfin, nous avons étudié les fonctions d'une protéine virale (Gp92), largement conservée chez les virus appartenant à ces deux genres et qui est produite au stade précoce du cycle infectieux. Lorsqu'elle est produite seule chez l'hôte, cette protéine altère la morphologie cellulaire et interagit avec un complexe de régulation bactérien de type facteur sigma/anti-sigma impliqué dans la réponse au stress (appelé AlgU-MucA). Notre étude suggère un rôle potentiel de Gp92 dans l'atténuation du stress provoqué par l'infection virale.

Ce manuscrit fournit un modèle de transformation d'une cellule de *P. aeruginosa* en « virocellule » au cours de l'infection par PAK_P3 ou PAK_P4. De plus, la comparaison des stratégies de ces deux virus, vraisemblablement issus d'un ancêtre commun, nous a permis de discuter l'évolution des mécanismes infectieux chez les phages virulents.

Mots clefs : *Kpp10virus*, *Pakpunavirus*, « virocellule », *P. aeruginosa*, étude transcriptomique.

Comprehensive study of new virulent bacteriophages: From transcriptomic and mechanistic characterisations towards evolutionary perspectives

Abstract:

Previous investigations in the field of phage therapy led to the discovery of two new genera of bacteriophages (phages), namely *Kpp10virus* and *Pakpunavirus* whose infection mechanisms are unknown. It is acknowledged that a successful infection is notably ensured by an effective takeover of host cell resources, leading to its transformation into a *virocell*, a cellular organism exclusively dedicated to the production of progeny phages.

This PhD work aims to provide a comprehensive view of molecular strategies set up by *Kpp10virus* and *Pakpunavirus* (represented by phages PAK_P3 and PAK_P4, respectively) to infect the opportunist pathogen *Pseudomonas aeruginosa*.

First, we assessed phage intrinsic properties by analyzing their genomic content, evaluating their host range and growth parameters and identifying their bacterial receptor.

Then, by coupling transcriptomics and metabolomics approaches, we found that both viruses have similar transcriptional programs, with a temporal regulation of their gene expression and production of antisense transcripts. They both strikingly prompt a rapid degradation of 90% of host mRNAs, which are eventually replaced by viral RNAs. Despite this extensive degradation, we found that both phages do not shutoff host metabolism but redirect biosynthesis pathways, however through different mechanisms. In addition, we found that a common host response is elicited upon both PAK_P3 and PAK_P4 infections and hypothesized it represents an attempt of the host to repair extensive RNA damage.

Finally, we investigated the functions of an early produced phage protein (Gp92), broadly conserved in both phage genera, in order to identify particular mechanisms of host subversion used by these phages. When expressed alone in the host, Gp92 alters cell morphology and interacts with the bacterial regulatory complex sigma/anti-sigma involved in stress response (namely AlgU- MucA). Our study suggests a potential role of Gp92 in alleviating the stress caused by phage infection.

This manuscript provides a model of *virocell* transformation upon infection of *P. aeruginosa* by PAK_P3 or PAK_P4. In addition, by comparing their reproductive strategies, it addresses the evolution of infection mechanisms in virulent phages deriving from a common ancestor.

Keywords: *Kpp10virus*, *Pakpunavirus*, *virocell*-conversion, *P. aeruginosa*, transcriptomic study

A Kadoc

[...]

A l'aube de jours nouveaux

Acknowledgments / Remerciements

First, I would like to thank all the members of the jury for their time and consideration. I thank Isabelle Martin-Verstraete for chairing this committee. Thanks to Mireille Ansaldi and Darren Smith for accepting to review this manuscript and travelling all the way from the very south of France, or from the very north of England. Thanks to Olga Soutourina and Romé Voulhoux for accepting to examine this work.

Je tiens à remercier Patrick Forterre, responsable de l'unité Biologie Moléculaire du Gène chez les Extrêmophiles, pour m'avoir permis de réaliser ma thèse au sein de son laboratoire.

J'adresse mes sincères remerciements à mon directeur de thèse, Laurent Debarbieux, pour m'avoir offert la possibilité, tant convoitée, de travailler sur les phages. Entre l'absence de place disponible au labo et l'incertitude de pouvoir participer au concours de l'école doctorale, tu as quand même rendu cela possible, et je t'en suis reconnaissante. Merci également pour ta disponibilité, ton soutien dans l'adversité et surtout pour ta confiance et la liberté que tu m'as laissée pour mener ce projet.

I express my deep gratitude to Rob Lavigne who made this project possible in the first place, for welcoming me in his lab several times and for the fruitful collaboration all along this PhD. Thanks also to Bob Blasdel for working with me on the transcriptomic project and to Anne-Sophie Delattre who mentored me during my Master 2 project in Leuven.

Je souhaite également adresser mes remerciements à mes collaborateurs de l'Institut Pasteur pour leur aide dans ces différents projets : à Gouzel Karimova pour m'avoir initiée à la technique du double hybride et accompagnée dans les manips, à Marc Monot pour avoir rendues plus digestes les analyses des données RNA-Seq, à Elisa Brambilla pour l'acquisition des images de microscopie time lapse et pour son implication intellectuelle. Les moments passés à travailler avec vous ont été de véritables bouffées d'air frais lorsque j'ai eu besoin de lâcher ma paillasse.

Mille mercis à mes collègues de l'unité BMGE pour leur bonne humeur, les moments passés ensemble, les discussions, les déjeuners, l'aide intellectuelle, matérielle, administrative, j'en passe et des meilleurs. En particulier, merci à Ana Cova Rodrigues pour son aide indispensable, sa gentillesse, son efficacité et tant de choses encore. Merci à Chantal Archambaut et Estelle Atse pour avoir rendues les tâches quotidiennes bien plus faciles et les couloirs plus animés ! Merci à Elena Rensen pour toutes nos discussions, les compatissantes ou les plus décontractées, pour le partage des recettes

végétariennes, et surtout pour m'avoir fait rire si souvent. Merci à Pierre Béguin pour les discussions, son humour et pour avoir partagé la célébration de l'Escalade tous les ans avec nous tous.

Un grand merci aux membres du groupe Phage qui sont devenus un peu plus que des collègues :

A Marta Lourenço pour sa gentillesse et ses imitations remarquables des cris d'animaux, obrigada ;)

A Luisa De Sordi, avec qui j'ai partagé tellement de choses ; joies, peines, crises de nerfs, anecdotes quotidiennes, interesting facts, des conseils scientifiques et moins scientifiques, un voyage à Seattle et le très particulier Evergreen phage meeting, des expressions françaises... en somme des tas de bons souvenirs. I will miss you !

A Nicolas Dufour avec qui j'ai partagé tout autant de choses. Des histoires de violoneux à la fine fleur de la chanson française, en passant par des réflexions philosophiques (et déprimantes) à propos de l'avenir... Ta présence au labo était d'un réconfort indispensable et ton départ a presque été un traumatisme... (Papuche ?)

Merci aux copains de l'AC15 pour m'avoir permis de me défouler sur le terrain et de me vider l'esprit le vendredi soir chez Momo. Merci à mes amis thésards pour leur soutien, nos échanges pour se remonter le moral, les soirées/week ends (beaucoup trop peu nombreux) qu'on a passé ensemble : Jerzy, Mélanie, Julie, Ana, David. Merci à Gwen d'avoir toujours pris le temps d'envoyer des petits messages quand bien même elle se trouvait à l'autre bout de la planète.

Je souhaite également exprimer ma profonde gratitude à Zhou, sans qui je n'aurais jamais pu faire cette thèse. Tu m'as apporté une aide ô combien précieuse, à la fois intellectuelle et morale, pour laquelle je te remercie sincèrement.

Mes remerciements vont à ma famille, mon frangin pour avoir le courage de venir assister à ma soutenance et parce que je sais que je peux compter sur vous tous. En particulier, je remercie mes parents, pour m'avoir soutenue inconditionnellement toutes ces années. J'espère que vous savez à quel point je vous suis reconnaissante...

Olivier, pour ton intérêt et ta curiosité, ton soutien sans faille, ton extrême patience, ta vision optimiste des choses (que j'essaye de me forcer à partager, promis !), pour t'occuper des tâches quotidiennes seul pendant que j'écris ce manuscrit, pour être mon partenaire à la maison et sur le terrain, je ne te dirai jamais assez merci... Tu as été présent dans tous ces moments si pénibles...Je n'en serais pas venue à bout sans toi (et Plume !). Merci d'être là, tout simplement.

(Mai 2017 sera le mois n°28, je dis ça...)

Index

ACKNOWLEDGMENTS / REMERCIEMENTS	6
INDEX	8
FIGURE LIST	12
TABLE LIST	14
ABBREVIATION LIST.....	15
FOREWORD	17
PART 1 - LITERATURE REVIEW	18
CHAPTER 1 - WHAT IS A VIRUS?	19
1. <i>History</i>	19
2. <i>Definitions and current debate</i>	20
CHAPTER 2 - NATURE AND USE OF BACTERIOPHAGES.....	22
1. <i>Classification and genomic organisation</i>	22
The predominance of Caudovirales	22
Others families of bacteriophages.....	23
Limitations of a genomic-based taxonomy of phages: content and architecture of phage genomes.....	25
2. <i>Lytic Infection cycle</i>	27
Adsorption	27
Entry	27
Intracellular lifecycle: phage temporal transcriptional scheme	28
Assembly of new virions	29
Lysis	29
Lysogeny, Pseudolysogeny and Chronic infections	32
3. <i>Phage-Host interactions and ecological importance</i>	34
Resistance mechanisms: phage-bacteria arms race	34
Co-evolution between phage and bacteria	36
Ecological perspective	37
4. <i>Medical and biotechnological application of phages</i>	39
From molecular biology origins to nanotechnologies	39
Pathogen detection	40
Agriculture and food industry.....	40
Phage therapy	41
Source of novel antimicrobials	43
CHAPTER 3 - TURNING THE RIBOCELL INTO A VIROCELL: KEYS FOR SUCCESS.....	46
1. <i>Sabotage and enslavement of cell machineries</i>	46
2. <i>Global impact of phage infection on cellular transcriptional dynamic</i>	50

Complete shutoff of host transcription during phage infection: a mistaken view	50
Reprogramming of cell transcriptome during phage infection revealed by whole-genome transcriptional analyses	50
3. <i>Impact of phage infection on host cellular metabolism</i>	53
CHAPTER 4 - THE VERSATILE OPPORTUNISTIC PATHOGEN <i>PSEUDOMONAS AERUGINOSA</i>	55
1. <i>A threatening opportunistic pathogen</i>	55
2. <i>A master of adaptation</i>	56
Detection and signalling systems	56
Complex transcriptional regulation network.....	58
Post transcriptional regulation: prevalent role of small non coding RNAs	60
Two examples of adaptation mechanisms characteristic of <i>P. aeruginosa</i>	61
3. <i>Diversity and genome plasticity</i>	64
Core genome, accessory genome and pangenome.....	64
Different types of accessory genetic elements and role of HGT	64
CHAPTER 5 - MODEL OF STUDY AND RESEARCH GOALS.....	66
1. <i>Isolation of new phages with therapeutic potential</i>	66
2. <i>Definition of two new phage genera</i>	68
3. <i>Research goals, scientific questions and manuscript organisation</i>	70
PART 2 - RESULTS AND DISCUSSION	72
CHAPTER 1 - GENERAL DESCRIPTION OF PHAGES PAK_P3 AND PAK_P4	73
1. <i>Comparative genomic: PAK_P3 and PAK_P4 share common life-history traits</i>	73
Definition of core genes	73
Predicted functions	75
2. <i>Lifecycle parameters</i>	77
3. <i>Host range</i>	78
Evaluation of strain collection diversity.....	78
PAK_P3 and PAK_P4 have similar host ranges	79
4. <i>Phage receptor characterization</i>	81
Estimation of spontaneous mutation frequencies leading to the emergence of phage resistant variants	81
Identification of Tn insertion mutants resistant to phage infection.....	82
CHAPTER 2 - PHAGE TRANSCRIPTIONAL PROGRAMS.....	87
1. <i>Temporal regulation of phage gene expression</i>	87
2. <i>Production of non-coding RNAs</i>	89
Antisense transcription of the structural region during early infection	89
PAK_P3 displays strong expression of temporally regulated small non-coding RNAs.....	90
3. <i>Core genes are similarly expressed in both infections but display phage-specific adjustments</i>	92
CHAPTER 3 - EXAMPLE OF A VIRAL STRATEGY TO CO-OPT THE BACTERIAL SYSTEM DURING EARLY INFECTION	94
1. <i>Identification of a small phage protein inhibiting <i>P. aeruginosa</i> growth</i>	94
2. <i>Gp92 causes cell morphology modification</i>	96
3. <i>Gp92 transmembrane location is crucial for its activity</i>	96
4. <i>Gp92 interacts with the stress response-associated transcriptional regulators AlgU and MucA</i>	98

5. Gp92 leads to cell death in strains with impaired stress response	99
6. Gp92 homologous protein encoded by PAK_P4 is not toxic for <i>P. aeruginosa</i> cells	103
7. Proposed model.....	104
CHAPTER 4 - IMPACT OF PHAGE INFECTION ON HOST METABOLISM	107
1. Dramatic redirection of RNA metabolism by PAK_P3 and PAK_P4	107
2. PAK_P3 induces changes in host metabolome	109
Global variation of metabolite content	109
Drastic changes in amino acid, nucleotide-sugar and pyrimidine pathways	110
Variations in host metabolome composition are not otherwise mediated through differential expression of host genes	114
3. Modulation of the expression of iron uptake-related host genes upon PAK_P4 infection	116
CHAPTER 5 - IDENTIFICATION OF A COMMON HOST TRANSCRIPTIONAL RESPONSE TO PHAGE INFECTION	119
1. Prerequisite: reannotation of strain PAK genome using transcriptomic data	119
2. A common host response is elicited during infection by PAK_P3 and PAK_P4	121
PART 3 - CONCLUSIONS AND PERSPECTIVES	125
1. OVERVIEW OF PAK_P3 AND PAK_P4 INFECTION CYCLES	126
2. EVOLUTIONARY PERSPECTIVE EMERGING FROM PAK_P3 AND PAK_P4 COMPARISON	131
AFTERWORD	133
PART 4 - METHODS.....	134
STRAINS AND GROWTH CONDITIONS.....	135
PHAGE BIOLOGY	136
<i>Host range</i>	136
<i>Receptor identification</i>	137
<i>Adsorption assay and one-step growth experiment</i>	138
<i>Dependence to host RNA polymerase assay</i>	138
“-OMICS” EXPERIMENTS.....	139
<i>Strains and growth conditions</i>	139
<i>Whole transcriptome sequencing</i>	139
<i>Statistical analysis</i>	140
<i>Strain PAK reannotation and prediction of ncRNAs with RNA-Seq data</i>	140
<i>High coverage metabolomics analysis</i>	140
MOLECULAR AND CELLULAR MICROBIOLOGY	142
<i>General and basic molecular biology</i>	142
<i>Generation of knock-out mutants</i>	142
<i>Expression of phage predicted genes in <i>Pseudomonas aeruginosa</i></i>	142
<i>Toxicity assay</i>	142
<i>Bacterial growth curves</i>	143
<i>Time Lapse Microscopy</i>	143

<i>Bacterial Adenylate Cyclase Two-Hybrid assays</i>	143
<i>Gp92 topology analysis</i>	144
REFERENCES	145
ANNEX - LISTS	157
STRAIN LIST	158
PRIMER LIST	160
PLASMID LIST.....	163
ARTICLE 1 - The search for therapeutic bacteriophages uncovers one new subfamily and two new genera of <i>pseudomonas</i>-infecting <i>myoviridae</i>	165
ARTICLE 2 - Next-generation “-omics” approaches reveal a massive alteration of host rna metabolism during bacteriophage infection of <i>pseudomonas aeruginosa</i>	166
ARTICLE 3 - Comparative transcriptomics analyses reveal the conservation of an ancestral infectious strategy in two bacteriophage genera	167
POPULARISATION OF SCIENCE - Les bactériophages: comment ces virus alliés fonctionnent-ils?	168

Figure List

Figure 1: Morphologies of Caudovirales	23
Figure 2: Overview of virion morphotypes of prokaryotic viruses	24
Figure 3: Phage Proteomic Tree	25
Figure 4: Example of modular organisation of phage genomes	26
Figure 5: Mechanisms of cell lysis	30
Figure 6: Lytic infection cycle	31
Figure 7: Graphic representation of alternative infection cycles	33
Figure 8: Allele frequencies over time in directional and negative frequency-dependent selection	37
Figure 9: Antimicrobial drug discovery strategy based on phage genomics	45
Figure 10: Summary of cell-to-cell signalling systems in <i>P. aeruginosa</i>	58
Figure 11: <i>P. aeruginosa</i> virulence regulatory network	60
Figure 12: Model of activation of AlgU (σ_{22}) regulon.....	62
Figure 13: Phage treatment of pulmonary infection.....	67
Figure 14: PAK_Px phylogenetic tree	68
Figure 15: Genomic comparison of phages PAK_P3 and PAK_P4	74
Figure 16: Proteomic analysis of PAK_P3 virions	76
Figure 17: PAK_P3 and PAK_P4 rapidly adsorb to their host and efficiently produce new progenies	77
Figure 18: Number of phages leading to a productive infection on a given isolate	78
Figure 19: Extract of PAK_P3 and PAK_P4 host range	80
Figure 20: <i>P. aeruginosa</i> PAK strain spontaneous mutants resistant to PAK_P3 infection	82
Figure 21: Adsorption assay of PAK_P3 on LPS- and flagellin-mutant strains	84
Figure 22: Strain PAK Δ 4980 is resistant to phage infection	85
Figure 23: Phage transcription is temporally regulated.....	88
Figure 24: Example of PAK_P3 antisense transcripts	90
Figure 25: PAK_P3 small non coding RNAs	91
Figure 26: Comparison of PAK_P3 and PAK_P4 homologous genes expression	92
Figure 27: Expression of phage gene gp92 affects growth of <i>P. aeruginosa</i> and <i>E. coli</i>	95
Figure 28: Gp92 affects bacterial cell morphology	95
Figure 29: Topology analysis of Gp92	97
Figure 30: In vivo interaction of Gp92 and anti-sigma factors (RseA/MucA).....	99
Figure 31: Effect of gp92 expression in mucA22 strain	100
Figure 32: Effect of gp92 expression in Δ algU strain.....	101
Figure 33: Effect of gp92 expression in Δ algD strain.....	102
Figure 34 Alignment of Gp92 and Gp109 protein sequences.....	103
Figure 35: gp109 expression does not lead to cell toxicity.....	103
Figure 36: Effect of D-cycloserine treatment on the growth of cells expressing gp92	106

<i>Figure 37: PAK_P4 takes over the host cell transcription faster than PAK_P3</i>	108
<i>Figure 38: PAK_P3 and PAK_P4 amplification are dependent of host RNAP</i>	108
<i>Figure 39: PAK_P3 alters P. aeruginosa metabolite content over the course of infection</i>	109
<i>Figure 40: Pathway enrichment analysis during PAK_P3 infection revealed its requirement on pyrimidine metabolism</i>	111
<i>Figure 41: Comparison of significant changes in transcriptomics and metabolomics data from PAK_P3 infected cells reveals no direct correlation</i>	115
<i>Figure 42: Strain PAK displays high number of non-coding RNAs</i>	120
<i>Figure 43: Overview of differentially expressed genes upon PAK_P3 and PAK_P4 infections</i>	121
<i>Figure 44: Host strain specifically up-regulates the expression of one operon involved in RNA processing in response to both phage infections.</i>	123
<i>Figure 45: Phage-induced alteration of host gene expression during late infection</i>	124
<i>Figure 46: Schematic progression of PAK_P3 and PAK_P4 infection cycles</i>	130
<i>Figure 47: Hypothetical evolution of a phage ancestral program</i>	132

Table list

<i>Table 1: Overview of bacterial virus families</i>	24
<i>Table 2: Mechanisms of host subversion mediated by phage early proteins</i>	48
<i>Table 3: Examples of analyses of transcriptional dynamics of phage infected cell</i>	52
<i>Table 4: ECF sigma factors identified in P. aeruginosa PAO1</i>	59
<i>Table 5: Proteomic analysis of PAK_P3 virions</i>	75
<i>Table 6: Genes identified with a Tn insertion</i>	83
<i>Table 7: Gp92-binding proteins identified with bacterial two-hybrid screen</i>	98
<i>Table 8: Differential expression of AlgU regulon upon PAK_P4 infection</i>	105
<i>Table 9: PAK_P3 infection causes a dramatic increase in free deoxynucleotides</i>	112
<i>Table 10: PAK_P3 infection influences amino sugar metabolism</i>	113
<i>Table 11: Differential expression of genes involved in iron uptake upon phage infection</i>	118

Abbreviation list

AMG	Auxiliary Metabolic Gene
Abi	Abortive infection
AHL	N acyl Homoserine Lactone
ARD	Arms race dynamics
asRNA	antisense RNA
ATP	Adenosine TriPhosphate
BACTH	Bacterial Adenylate Cyclase Two Hybrid
BLAST	Basic Local Alignment Search Tool
bp	base pair
Cas	CRISPR associated
CDS	Coding Sequence
CF	Cystic Fibrosis
cfu	Colony forming unit
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DNA	Desoxyribonucleic Acid
DSF	Diffusible signal factor
ECF	Extracytoplasmic function
EOP	Efficiency of plaquing
ESI-MS/MS	Electrospray tandem mass spectrometry
FC	Fold Change
FSD	Fluctuating Selection Dynamic
Fur	Ferric uptake regulator
GAC	Global activator of Antibiotic and Cynaide synthesis
Gp	Gene product
HGT	Horizontal Gene Transfer
HMC	hydroxymethylcytosine
ICTV	International Committee of Taxonomy of Viruses
LB	Lysogeny Broth
LPS	Lipopolysaccaride
MOI	Multiplicity Of Infection
NCBI	National Center for Biotechnology Information
OD	Optical density
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
pfu	plaque forming unit
PQS	<i>Pseudomonas</i> quinolone signal
QS	Quorum sensing
RGP	Region of genomic plasticity
Rif	Rifampicin
R-M	Restriction Modification
RNA	Ribonucleic Acid
RNAP	RNA polymerase
Rnase	Ribonuclease
RNA-Seq	RNA Sequencing
SAR	Signal Anchor Release

Sie	Superinfection exclusion
SNP	Single Nucleotide Polymorphism
SP	Signal Peptide
sRNA	small RNA
Tn	Transposon

Foreword

As a tribute to Dr. Luisa De Sordi, I provide here my last list of (self-centred) so-called interesting facts.

This research project is the result of 4 years and a half spent under the supervision of Laurent Debarbieux, being about 1015 working days, including 150 days as a Master 2 student. I spent 10% of my Master-PhD in Leuven (Belgium) in the Laboratory of Gene Technology headed by Rob Lavigne as the result of a key collaboration. Three other collaborations have contributed to produce the data presented here: with Gouzel Karimova for bacterial two-hybrid assays (D. Ladant's lab), Marc Monot for analysing transcriptomic data (B.Dupuy's lab) and Elisa Brambilla for time lapse microscopy (S. van Teeffelen's lab).

I am grateful to my PhD supervisor for giving me the opportunity to participate to 8 scientific meetings in 4 countries (France, England, Switzerland and the United States) during which I gave 5 talks and presented 4 posters, one of which was awarded with a prize from the Microbiology Society.

For 2 years, I had spent 120h teaching at the University Paris Diderot and mentored 3 students in the lab.

This PhD is also characterized by intense bench work. For instance, 5200 (underestimated) PCRs were performed, 367 strains were constructed and stocked, about 1000 DNA samples were Sanger sequenced and more than 200 primers were ordered.

To continue with figures, we can count 3054 uses of swearwords (likely underestimated), a dozen of "bitchy-notes-with-little-heart", 5 voodoo dolls, 28 sessions of hysterical laugh, 35h of nice fun time with labmates (Nicolas, Luisa, Elena...), twentyish expressions that became cult (e.g. Sonia killer, couscoussière, Bam bam bam ok girl, don't gratte your croute...), 3h spent shaving mice.

More seriously, I would like to emphasize that this molecular biology-oriented project was initiated in a phage therapy lab (originally). Consequently, tools to (genetically) manipulate *Pseudomonas* needed to be gathered or constructed. I have also established several protocols, notably for phage adsorption assays and one-step growth experiments, which I hope will be useful for future students.

I initiated several side projects ranging from the construction of a Cas9-based editing tool to generate recombinant phages, to the production of fluorescent virulent phages and also the isolation and characterization of phages infecting *Rouxiella chamberiensis*.

In summary, these four years have been quite intense and have resulted in the publication of 2 (hopefully 4) peer-reviewed articles, one article in a French popular science journal and the present manuscript.

PART 1

-

LITERATURE REVIEW

CHAPTER 1 - WHAT IS A VIRUS?

1. History

The discovery of viruses was made possible with the invention of Chamberland filters by the eponymous French microbiologist in 1884. This porcelain filter, with pore size of 0.1 – 1 μM , allows to remove any cell (bacteria included) contained in a solution. This technological advance enabled the establishment of a whole new scientific discipline called «Virology», as a result of a succession of discoveries that are summarized below.

In the late 19th century, the acknowledged “*germ theory of disease*” postulated that any disease agent could be retained on filters. The Russian biologist Ivanovsky is generally given credit for being the first scientist suggesting, in 1892, that the causal agent of the tobacco mosaic disease is a filterable agent¹. The conceptual leap was made few years later (1898) when the Dutch microbiologist Beijerinck demonstrated that this filterable agent was able to reproduce and multiply only within living plants. He proposed it was a new form of infectious agent and called it «*virus*» to indicate its non-bacterial nature². This virus, now known as tobacco mosaic virus, was the first to be crystallized in 1935, proving that viruses were particles and not liquid as Beijerinck thought³.

The early 20th century has seen successive discoveries of new viruses and notably viruses that infect bacteria. This discovery is attributed to two independent researchers: Frederick Twort and Felix d’Hérelle. In 1915, the British bacteriologist Twort discovered a filterable agent able to lyse a bacterial culture and reported it as “*a bacteriolytic agent*”, as he had alternative hypotheses on its true nature. Independently, two years later, the French-Canadian microbiologist d’Hérelle discovered “*an invisible, antagonistic microbe of the dysentery bacillus*” that he claimed was a “*virus parasitic on bacteria*”⁴. He called these new viruses «*bacteriophages*» (from the Greek *phagein* – to eat) and rapidly proposed and promoted their use as therapeutic agents. However, the question of the viral nature of bacteriophages met with rapid scepticism and controversy and had to wait until 1940 to be partly solved with the first transmission electron micrographs. From this point, studies on bacteriophages multiplied and led to major discoveries in fundamental molecular biology [reviewed in ⁵].

2. Definitions and current debate

Beijerinck originally defined viruses as “*contagium vivum fluidum*” (contagious living fluid), *i.e.* filterable agents able to cause diseases.

This definition has evolved over the 20th century and differed from one author to another. A common and modern view of viruses is the following: they are small particles (typically <0.2 µm) consisting of proteins and nucleic acids (with or without a lipid envelope), obligate intracellular parasites able to infect cells from all three domains of life (*Eukarya*, *Bacteria* and *Archaea*) and even to parasitize other viruses.

However, this definition has been challenged few years ago with the discovery of giant viruses (so called “girus”) which refute the original, and universally adopted, property of viruses as filterable agents. Indeed, these viruses (*i.e.* *Mimiviridae*⁶, *Megaviridae*⁷ and *Pandoraviridae*⁸) are larger than some cellular organisms in terms of both particle and genome sizes (encoding >1000 genes) and possess many gene features usually associated with cellular life.

Consequently, the nature of viruses and the inherent question of whether or not “*viruses are alive*” are still actively debated nowadays. In 2008, Raoult and Forterre⁹ proposed a dichotomous classification system of the living world that would be divided between the capsid-encoding organisms (*i.e.* viruses) and the ribosome-encoding organisms (*i.e.* eukaryotic, archaeal and bacterial organisms), based on the observation that the only universal feature distinguishing viruses and cells is the presence of ribosomes. According to this proposed classification, they suggested to define viruses as “*capsid-encoding organisms that are composed of proteins and nucleic acids, self-assemble in a nucleocapsid and use a ribosome-encoding organism for the completion of their life cycle*”⁹. This definition implicitly considers viruses as living organisms and assumes a clear distinction between a virus and a virion. According to this definition, the living form of the virus is the infected cell, or *virocell*, whose main function is to produce infectious particles (virions)¹⁰. The metabolism of the *virocell* is therefore radically different than the metabolism of the original non-infected cell (*ribocell*)¹⁰, but nonetheless very active as exemplified by the induction of specific pathways, such as glycolysis or fatty acid synthesis, by Hepatitis C virus¹¹. This concept is best illustrated by the formation of viral factories induced by most eukaryotic viruses, giving a physical reality to the *virocell*. In *Bacteria* and *Archaea*, growing evidence supports that viruses also redirect their hosts’ metabolisms toward their own viral metabolisms¹²⁻¹⁴ and formation of a nucleus-like structure during bacteriophage infection has been reported for the first time earlier this year¹⁵.

By focusing on virus-host interactions, in particular on the intracellular stage of the viral lifecycle, this manuscript will attempt to provide additional arguments in favour of the virocell concept and therefore considers viruses as capsid-encoding organisms.

CHAPTER 2 - NATURE AND USE OF BACTERIOPHAGES

As mentioned in Chapter 1, d'Hérelle's bacteriophages (or phages) are viruses that specifically infect bacteria. They are ubiquitous and extremely numerous as we estimate their number reaches 10^{31} particles on Earth. After having been used mostly as convenient experimental tools for more than half a century, they are now being rediscovered, notably for their ecological importance and their medical applications.

1. Classification and genomic organisation

Several classifications have been proposed since first observations of phages under an electron microscope in the 1940's. They were based on various criteria such as virion morphology, type of genetic material, presence of an envelope or even host range. The International Committee of Taxonomy of Viruses (ICTV), founded in 1966 adopted a classification system based on virion morphology and nature of the nucleic acid¹⁶. In this system, the virus species is the lowest taxon in a branching hierarchy of viral taxa, which is organised as follows: Order > Family > (Sub-family) > Genus > Species. The ICTV defined a virus species as "*a polythetic class of viruses that constitute a replicating lineage and occupy a particular ecological niche*", therefore assuming that all members within one species "*have several properties in common but do not necessarily all share a single common defining property*", in other words, a virus species is defined by "*a consensus group of properties*" (<http://www.ictvonline.org/virusTaxInfo.asp>).

The predominance of Caudovirales

The vast majority of phages observed under an electron microscope (96%) possesses double-stranded DNA (dsDNA) contained in a tailed, non-enveloped virion. These properties define the order of *Caudovirales*. Tailed phages share a similar organisation comprising an icosahedral head, containing the genome, attached to a helical tail with variable length generally provided with structures allowing the fixation of the viral particle on a host surface receptor (baseplate and fibers). The *Caudovirales* order is then subdivided in three families defined by the properties of the tails (Figure 1):

- i. *Podoviridae*: These small viruses have short and non-contractile tails. They usually contain a rather small genome (40 to 50 kb). They represent about 10% of tailed phages with 83 species listed by the ICTV (January 2017). A well-studied podovirus is phage T7 infecting *Escherichia coli*.
- ii. *Siphoviridae*: Siphoviruses are the most represented within Caudovirales (65%) with 374 species inventoried (ICTV, January 2017). They have a long flexible and non-contractile tail. The lambda (λ) phage (infecting *E. coli*) is the most well characterized siphovirus.

- iii. *Myoviridae*: The most well-known myovirus is phage T4 (infecting *E. coli*). It is characterized by a tail whose contraction is necessary to inject the viral genome within the host cell. Twenty five percent of tailed phages belong to the *Myoviridae* family (205 species defined by the ICTV January, 2017).

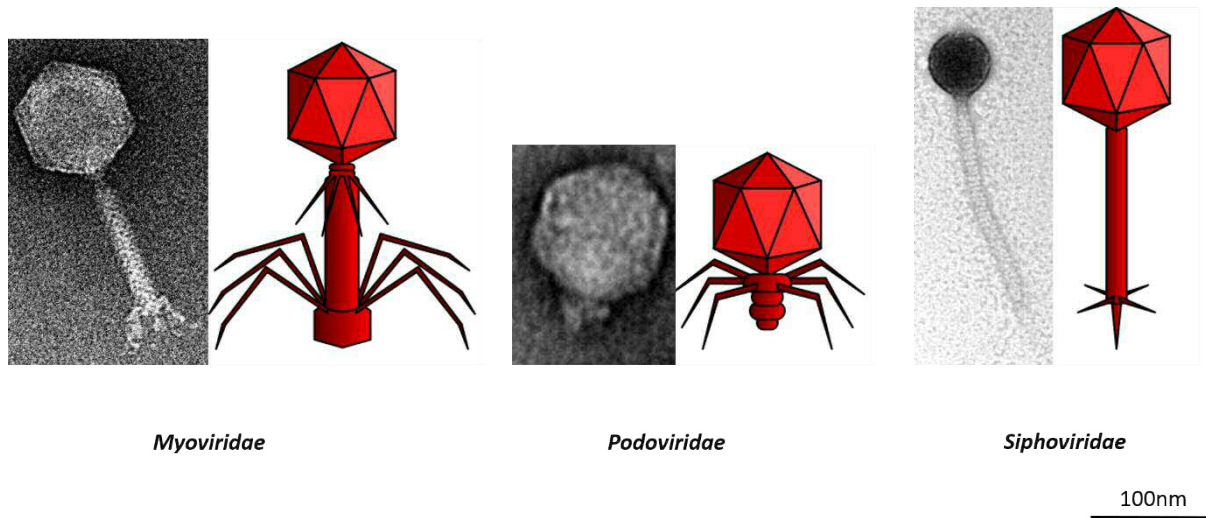


Figure 1: Morphologies of Caudovirales

Electron micrographs and associated schematic representations and of *Myoviridae*, *Podoviridae* and *Siphoviridae*. (Cartoons extracted from <https://commons.wikimedia.org/w/index.php?curid=35782604>, by Ninjatacoshell - Own work, CC BY-SA 4.0 and electron microscopy images from laboratory images collection)

Others families of bacteriophages

Even if most known bacteria-eating viruses belong to the order of *Caudovirales*, having a tail is not an absolute rule. In fact, the 4% non-tailed viruses display a greater diversity than tailed viruses in terms of morphology (Figure 2) and genetic material nature (double or single-stranded, DNA or RNA, segmented or not). Among viruses that infect *Bacteria*, nine additional families have been reported¹⁷, as indicated in Table 1.

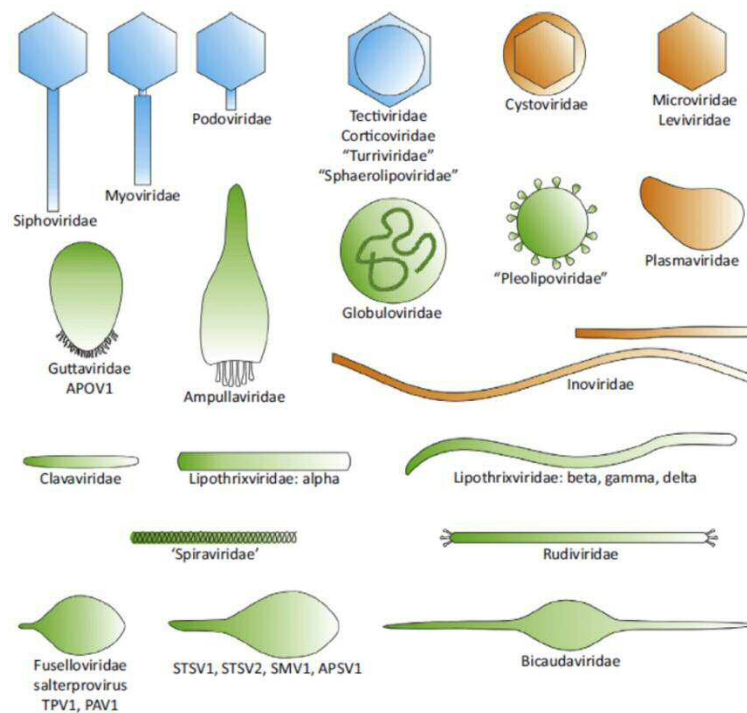
Although they are largely underrepresented, some of these viral ‘outsiders’ have been extensively studied, like the first sequenced virus phiX174 (*Microviridae*)¹⁸ or the filamentous phage M13 (*Inoviridae*) which allowed the development of the phage display technique¹⁹.

Henceforth, as this manuscript focuses on tailed phages, the generic term “phage” refers to *Caudovirales*.

Table 1: Overview of bacterial virus families

^aThe number of referenced species was obtained from the ICTV Taxonomy (January 2017). L, linear; C, circular; S, segmented. Modified from Krupovic *et al.* (2011)¹⁷

Family	Capsid morphology	Additional feature(s)	Genome type	No. of referenced species ^a	Example
<i>Myoviridae</i>	Icosahedral	Tail (contractile)	dsDNA, L	205	T4
<i>Siphoviridae</i>	Icosahedral	Tail (long non-contractile)	dsDNA, L	374	λ
<i>Podoviridae</i>	Icosahedral	Tail (short non-contractile)	dsDNA, L	83	T7
<i>Tectiviridae</i>	Icosahedral	Internal membrane	dsDNA, L	5	PRD1
<i>Corticoviridae</i>	Icosahedral	Internal membrane	dsDNA, C	1	PM2
<i>Plasmaviridae</i>	Pleomorphic	Enveloped	dsDNA, C	1	L2
<i>Microviridae</i>	Icosahedral	Non-enveloped	ssDNA, C	21	Φ X174
<i>Inoviridae</i>	Filamentous	Long flexible or short rigid	ssDNA, C	43	M13
<i>Cystoviridae</i>	Icosahedral	Enveloped, multilayered	dsRNA, L, S	1	Φ 6
<i>Leviviridae</i>	Icosahedral	Non-enveloped	ssRNA, L	4	MS2
<i>Sphaerolipoviridae</i>	Icosahedral	Internal membrane	dsDNA	6	P23-77

**Figure 2: Overview of virion morphotypes of prokaryotic viruses**

In blue are represented virion morphotypes found in viruses infecting both *Bacteria* and *Archaea*. Morphotypes found only in viruses infecting *Archaea* or only in bacterial viruses are coloured in green and brown, respectively. Extracted from Pietilä *et al.* (2014)²⁰. Even if it is off topic in the present manuscript, it should be noticed that the most impressive morphological diversity is awarded to viruses infecting *Archaea*.

Limitations of a genomic-based taxonomy of phages: content and architecture of phage genomes

With the tremendous progress of sequencing technologies and the resulting massive increase of sequenced whole genomes available, one could have expected that phage classification and phylogeny would have been greatly refined based on comparative genomic analyses.

Instead, phage genomes revealed an extreme diversity and divergence. Most of their annotated genes has no relative sequences detectable in databases and pairwise comparisons show that two genes coding for the same predicted function (e.g. major capsid protein) usually have very low sequence similarity, even in two related phages. Consequently, there is no strictly conserved single genetic, or protein, marker that could allow the construction of a universal tree of phages, following the example of Woese's ribosomal DNA-based taxonomy that led to the revolutionized view of a tripartite division of the living world²¹.

Despite these difficulties, efforts have been made to bring phage taxonomy to the postgenomic era and a new taxonomic system based on predicted phage proteome has been proposed²² (Figure 3). The originally proposed "Phage Proteomic Tree" was based on 105 predicted proteomes deduced from complete phage genomes. The resulting taxonomy was reported to be compatible with the ICTV system and was supported by phage biology knowledge.

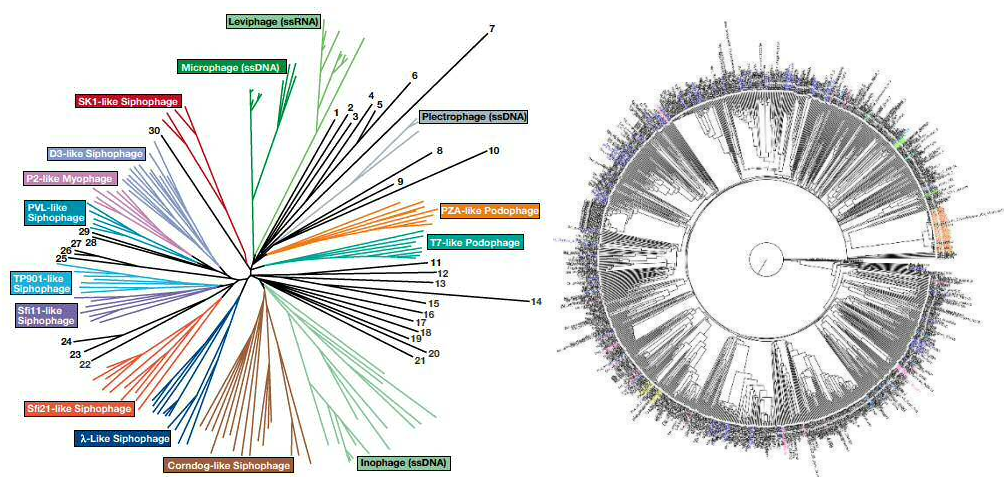


Figure 3: Phage Proteomic Tree

Phage Proteomic Tree is a whole-genome-based taxonomy system that can be used to identify similarities between complete phage genomes and metagenomic sequences. Left: version of the tree containing 167 phage genomes (2005). Phages in black cannot be classified into any clade. Extracted from Edwards and Rohwer, 2005²³. Right: Latest version of the tree containing 1184 genomes (April 2013). Available on www.phantome.org

This taxonomic system is potentially valuable for classifying uncultured phages identified through metagenomic data, since it is based on degrees of divergence between genome sequences. Moreover, the multiplication of large scale metagenomic projects on environmental samples reveals that viral diversity far exceeds that of experimentally characterized virus isolates. Indeed, a striking disparity between the number of potential new taxa and the number currently defined by the ICTV has been estimated. A robust sequence-based taxonomy of viruses is needed to provide a comprehensive view of the global virome and is currently considered by the executive committee of the ICTV²⁴.

However, such comparative genomic-based methods are limited by frequent horizontal gene transfers and recombination events that are acknowledged to occur in viral genomes, thus making it difficult to reconstruct their evolutionary histories. Indeed, phages usually have a particular genome architecture organized in modules. Typically, genes encoding proteins involved in the same processes are grouped in functional modules (e.g. structural, related to nucleic acid metabolism, lysis cassette, *etc.*) (Figure 4). The gene order (synteny) within these modules is usually well conserved in related phages, despite their low sequence similarity (even at the amino acid level). Moreover, each phage genome can be seen as a combination of modules originating from different phage lineages (Figure 4). The remarkable extent of this mosaicism in phage genomes explains why it is difficult to infer phylogenetic relationships from genomic comparisons²⁵.

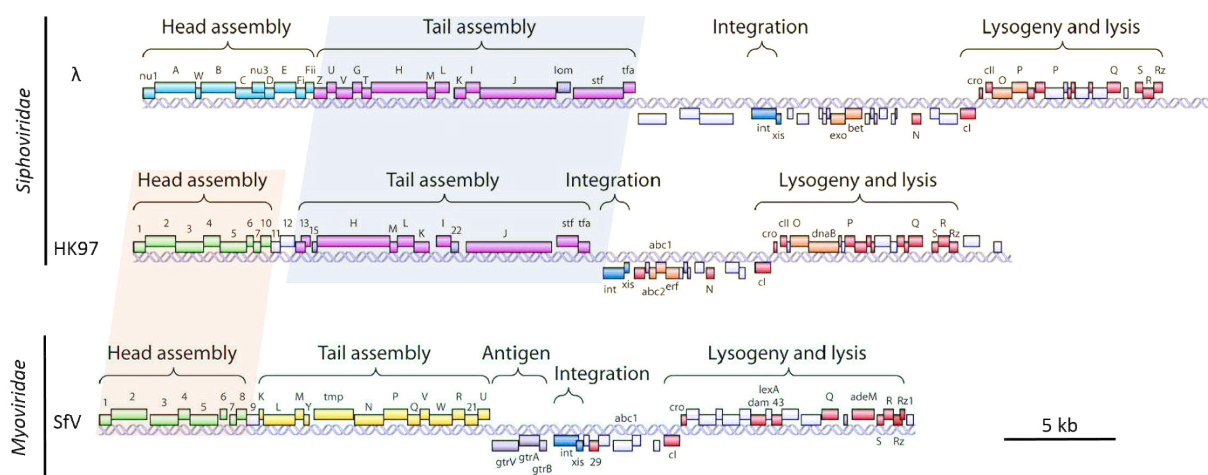


Figure 4: Example of modular organisation of phage genomes

Homologous genes in cassettes are colored similarly; analogous but not homologous genes are shown in orange. Modified from Krupovic *et al.* (2011)¹⁷

2. Lytic Infection cycle

Depending on their proteomic contents and interactions with the host cell, phages can display different types of infectious cycles. Four have been described and are known as chronic, lysogenic, pseudolysogenic and lytic. For the purpose of this manuscript, the following paragraphs will describe the successive steps occurring during a lytic cycle. Some of these steps are common with the other types of lifecycles that will be briefly introduced at the end of this section. Phages that replicate exclusively using lytic cycles are called virulent phages (or strictly lytic phages). This cycle systematically results in the burst of the host cell (or lysis).

Adsorption

The first step of the infection cycle consists in the adsorption of the virion on the surface of the targeted cell through interactions between phage tail proteins and bacterial receptors (e.g. lipopolysaccharides, proteins, *etc.*). The specificities of these interactions are such that one given phage is able to infect only a limited number of strains, often belonging to the same bacterial species.

Entry

Once attached to the host cell, the entry of the virus is a two-step process including (i) the degradation of the host cell envelope allowing the virion to reach the plasma membrane followed by (ii) the ejection of the genomic DNA into the cytoplasm of the infected cell.

The first step depends on the nature and complexity of the host cell envelope which comprises the cell inner membrane and either (i) a thin peptidoglycan cell wall covered with an outer membrane containing lipopolysaccharides (LPS) in Gram-negative bacteria or (ii) a thick peptidoglycan in Gram-positive bacteria. Moreover, all of these structures can be surrounded by other protective structures such as capsules (composed of polysaccharides), proteinaceous layers (S-layer) or mucolic acids. Consequently, phages carry enzymatic equipment (usually associated to tail fibers) to penetrate through these structures such as exopolysaccharides depolymerases (e.g. endosialidases²⁶ or alginate lyase²⁷) and exolysins (murein hydrolases²⁸).

The mechanism of DNA delivery depends on the nature of the tail. For instance, the tail sheath of myoviruses contracts, allowing the internal tail tube to go through the envelope. Then, viral DNA is ejected from the capsid head, where it is tightly packed, into the cytoplasm of the infected cell via a combination of pressure-driven (osmotic force imbalance between virus capsid and host cytoplasm) and/or protein-driven processes^{29,30}.

Intracellular lifecycle: phage temporal transcriptional scheme

Once the invading DNA has penetrated the cell, phage intracellular lifecycle begins. It is commonly acknowledged that most viral intracellular cycles are temporally regulated and comprise at least two steps related to the progression of the transcription of phage genes.

During the early stage of infection, a first set of genes (so called early genes) is transcribed either by the host RNA polymerase (RNAP) or by a viral RNA polymerase (vRNAP). When using the host RNAP, promoters of early genes compete with host promoters for binding the transcription machinery. Effector proteins can also be co-injected from the viral particle with phage DNA to favour the recruitment of the RNAP on phage promoters (e.g. protein Alt encoded by T4 phage)³¹. Alternatively, a vRNAP can initiate the transcription of early genes as in phage N4 lifecycle where a vRNAP is packaged into the capsid and injected into *E. coli* cytoplasm along with phage DNA³².

Early gene products are mainly involved in redirecting host metabolism towards the production of new phage particles. Through protein-protein interactions, these early proteins redirect or activate host proteins that are needed for the completion of the viral cycle. They also tend to inactivate potential bacterial defence mechanisms. Consequently, they target a wide range of key metabolic processes such as transcription, translation, and replication. Several examples of such phage hijacking strategies will be expanded in Chapter 3. Additionally, some early gene products are also involved in the regulation of subsequent phage processes. For example, early transcription of phage T7 is mediated by the host RNAP and allows the expression of T7 gene 1, encoding T7 RNAP. The vRNAP takes over the transcription of the rest of phage genome during a so called middle infection step.

Middle infection is acknowledged to be dedicated to phage DNA replication. Most phages carry replication genes arranged in modules (e.g. DNA polymerase, helicase, primase, *etc.*). Virtually all known types of DNA replication mechanisms have been reported to be used by phages, whether it be rolling circle-type (*E. coli* phage P2), theta-type (*E. coli* phage λ), protein-primed (*Bacillus subtilis* phage Φ 29) or transcription-initiated (*E. coli* phage T7)³³. Phage genomes are usually replicated as concatemers of individual genomes flanked by terminally redundant sequences. Each new copy of the genome is eventually incorporated in new capsids that are assembled once genes encoding structural proteins, DNA packaging and assembly systems have been massively transcribed during the late stage of infection (further described in the following paragraph).

It has to be noted that the scheme described above is an informative overview of phage temporal transcription and that each virus possesses its own specificities. Notably, the dependency of the host RNAP can be total (phage T4³⁴), partial (phage T7³⁴) or inexistent (*Pseudomonas aeruginosa* phage PhiKZ³⁵) and some phages do not display a middle stage of transcription. In conclusion, each phage possesses its proper strategy for regulating gene expression.

Assembly of new virions

Once individual virion components have been produced, they need to be assembled into functional particles. The first step is the genome packaging (or encapsidation) into a procapsid composed of the major (and potentially minor) capsid protein(s) whose assembly may require the help of scaffolding proteins. Genomic DNA translocates into the procapsid through a channel protein called portal protein. This involves the activity of an ATP-dependent motor transiently associated to the portal vertex known as terminase. This terminase enzyme has multiple roles: on top of catalysing DNA translocation, it is also involved in resolving concatemers of individual genomes. This is performed by endonucleolytic cleavages either on specific sites (phage T7) or independently of the sequence in phages using a “headful mechanism” of DNA packaging^{36,37} (phage T4). Concomitant to encapsidation, the procapsid undergoes a maturation process consisting in scaffolding proteins cleavage and shell rearrangement (*i.e.* usually expansion) and ends with the substitution of the terminase by neck proteins that make the link between the mature head and the tail. Tail proteins are either sequentially attached to the capsid (podoviruses) or pre-assembled in a separate pathway (sipho- and myoviruses). For example, tail assembly of *Myoviridae* follows a sequential order: the baseplate is first assembled and forms the base for tail tube polymerization which length is controlled by the tape measure protein. Preassembled fibers then attach to the distal end of the tail. Finally the mature tail and head are spontaneously joined via the neck proteins³⁷.

Lysis

Lysis step is a tightly regulated and temporally scheduled process. Two main pathways for releasing progeny phage from the host cell have been described in lytic phages. Although underlying mechanisms of these two pathways are different, they both include three steps (in Gram negative bacteria) corresponding to the disruption of the inner membrane, peptidoglycan and outer membrane. This tripartite mechanism allows a better regulation of the lysis process as it offers just as many regulatory levels³⁸.

The **canonical holin/endolysin pathway** has long been considered as a universal lysis mechanism (Figure 5). It involves the joint activity of a membrane protein (holin) with a muralytic enzyme degrading the peptidoglycan (endolysin). During the late stage of the infection cycle, holins accumulate in the membrane of the infected cell until reaching a critical concentration that triggers their aggregation, resulting in the formation of large pores in the inner membrane thus exposing the peptidoglycan to cytosolic accumulated endolysins. As a result, holins are responsible for the regulation of lysis timing, which is thought to be an allele-specific process. Indeed, it has been shown that missense mutations could affect lysis timing by altering holin structures resulting in a change in the critical concentration required for aggregation. Moreover, it has also been demonstrated that holins can be specifically inhibited by antiholins. Once accessible, endolysins escape the cytoplasm to

reach the peptidoglycan and degrade it. Lysis of Gram negative bacteria requires an additional step completed by proteins spanning both inner and outer membranes, either as a unique (u-spanin) or a heterodimer protein (i-spanin and o-spanin). Once the peptidoglycan is degraded, the outer-membrane lipoprotein part and the transmembrane part of the spanin complex can freely diffuse through the periplasm and aggregate, resulting in the fusion of the inner and outer membrane allowing the lytic burst to occur³⁸.

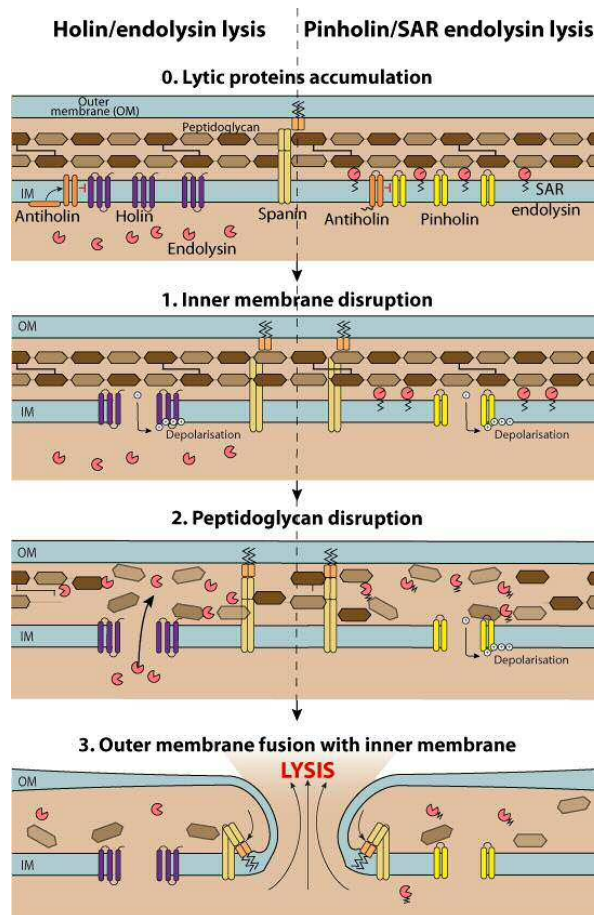


Figure 5: Mechanisms of cell lysis

Holin/endolysin (left) and Pinholin/SAR-endolysin (right) mechanisms of lysis are represented. Description of the different steps are given in the text. Extracted from <http://viralzone.expasy.org/>

Some lytic phages have been found to use an alternative molecular mechanism of lysis fulfilling the same overall strategy (Figure 5). In the **SAR-endolysin/pinholin system**, a different type of holin is triggered to form small heptameric channels leading to cell membrane depolarization that activates SAR endolysins. In contrast to canonical endolysins, SAR (for Signal Anchor Release) endolysins are exported in the periplasm through the host *sec* system that recognizes the N-terminal SAR domain without cleaving it. As a result, they accumulate into the periplasm in an inactive, membrane-anchored form. Upon pinholin-induced membrane depolarization, SAR-endolysins are released and refolded into their active form which ultimately degrades the peptidoglycan enabling the final spanin-mediated step to take place³⁸.

To summarize, the main steps of lytic infection cycle are summarized in Figure 6.

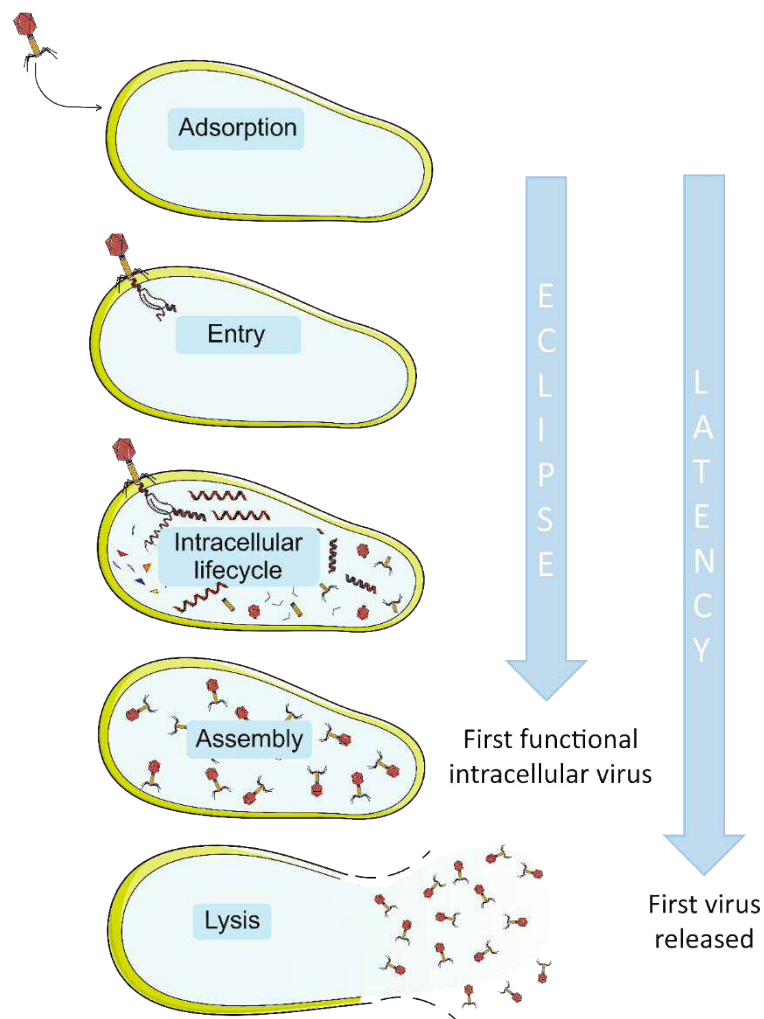


Figure 6: Lytic infection cycle

The main steps of the lytic cycle are represented. The eclipse period is defined as the time necessary to produce the first functional viral particle within the cell. It is distinct from the latency period that ends when the first virion is released upon cell lysis.

Lysogeny, Pseudolysogeny and Chronic infections

As previously mentioned, phages can display alternative lifecycles (represented in Figure 7).

The **lysogenic cycle**, undertaken by so called temperate phages (such as phage λ), is characterized by a latency state during which the phage genome is integrated into the host chromosome. The following description of lysogenic cycle hallmarks will be based on phage λ lifecycle, the principal model temperate phage^{39,40}. Once the viral genome has penetrated into the host cytoplasm, an intermediate early transcription occurs, allowing the expression of regulatory genes (constituting a genetic switch system) that control the decision to enter a lytic or a lysogenic pathway. This decision depends on environmental signals, notably the nutritional state of the cell and the number of co-infecting phages and involves a complex regulatory network. It relies on the concentration ratio of the regulatory proteins Cro and CII. If the concentration of the latter is high, the lytic pathway is blocked and the lysogenic switch is initiated. CII indirectly promotes the integration of the phage genome into the chromosome (then called prophage) as well as expression of the CI regulator which represses expression of lytic genes. The prophage is then replicated as part of the bacterial genome and transmitted to daughter cells (so called lysogens) through cell divisions. During this very stable state, a latent viral transcription is still present, notably allowing the production of the main regulator CI. Upon cellular stress (e.g. DNA damage inducing SOS-response), early lytic genes are expressed and phage genome is excised^{39,40}. The virus then follows a lytic cycle scheme, as described in the above paragraph.

Besides these probabilistic lysis-lysogeny decisions mainly based on the metabolic state of infected cells, Erez and coworkers have discovered earlier this year that phages can use small communication molecules to coordinate lysis-lysogeny decisions⁴¹. They demonstrated that phage-specific peptides are released upon *Bacillus* host cell infection and govern the decision of progeny phages to employ lytic or lysogenic cycle in subsequent infections, according to their concentration in the medium. This *arbitrium* system, as it has been called, is composed of three genes; one encodes the communication peptide, the second encodes the intracellular peptide receptor and the last is responsible for the production of a negative regulator of lysogeny.

In addition to lytic and lysogenic routes, another phage propagation mode has been proposed and termed **pseudolysogeny**. It has been defined as a phage carrier state where the phage is under a non-replicative episomal form inside the infected cell *i.e.* neither engaged in a lytic cycle nor in lysogenization, and assymetrically segregated upon bacterial divisions⁴². Pseudolysogenic interactions have been commonly thought to be associated with starved cells, thus representing a phage survival strategy as virus development can subsequently restarts when the host encounters more favourable growth conditions. Recently, Cenens and colleagues showed that pseudolysogeny also exists in actively

growing phage P22 infected *Salmonella* cells⁴³. Although molecular mechanisms and genetic regulations are still poorly understood, this phenomenon adds a level of complexity in our view of phage-host interactions and potentially has a significant impact on phage-bacteria coevolution as it was suggested to favour the selection of phage-resistant variants⁴⁴.

A fourth infection cycle type is characterized by the continuous release of low amounts of phage particles, without cell lysis. In this **chronic infection cycle**, progeny phages exit the cell through a concerted mechanism of assembly and secretion called extrusion. Such cycles are typically performed by filamentous phages such as M13, famous for its application in the protein-protein interaction technique known as phage display.

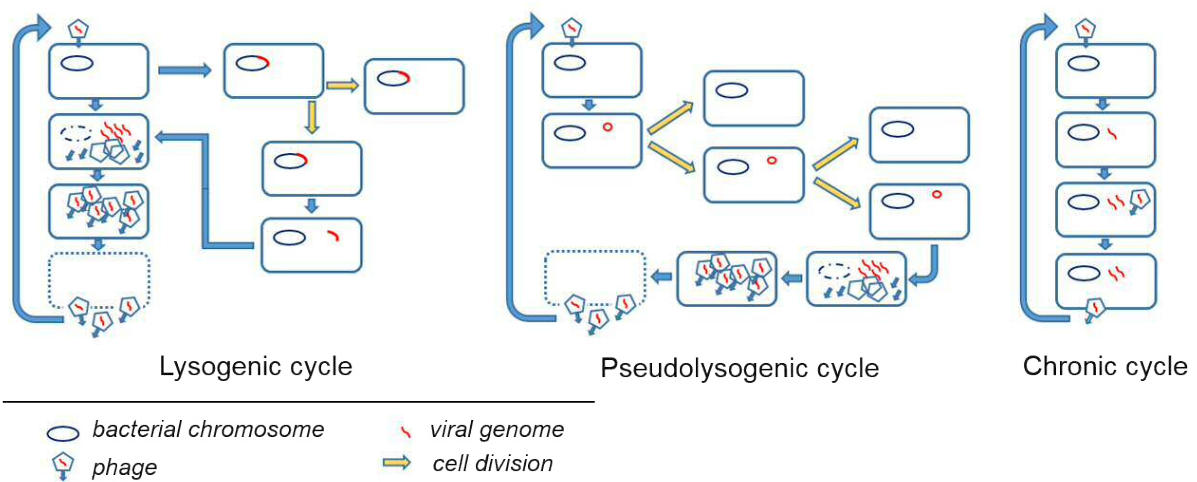


Figure 7: Graphic representation of alternative infection cycles

A brief description of the different cycles is given in the text. Extracted from 45

3. Phage-Host interactions and ecological importance

Resistance mechanisms: phage-bacteria arms race

To fight back against phage invaders, bacteria have evolved various mechanisms counteracting the infection steps described above. By giving a brief overview of bacterial defences, I will demonstrate that phages evolved to counterattack and thereby illustrate the concept of phage-bacteria arms race.

- Surface modification

The most straightforward solution to prevent phage infection consists in blocking phage adsorption, by modifying the receptor or making it inaccessible through production of extracellular matrix or competitive inhibitors. Mutations affecting receptor recognition are the most frequently encountered when looking for phage insensitive bacterial mutants, as exemplified by the experiment reported in the results section of the current manuscript (p.81). To circumvent this, phages can adapt to recognize modified receptors or by producing enzymes that cleave the extracellular matrix⁵.

- Superinfection exclusion systems blocking viral DNA translocation

Membrane-associated proteins have been reported to block entry of phage DNA into the host cell. As they are often encoded by phages themselves (mostly prophages), they protect the bacterial host from a secondary infection by a closely related (or identical) phage, hence the appellation “superinfection exclusion (Sie) systems”. One example of the few well-documented Sie system is encoded by the virulent phage T4 and acts by inhibiting the transfer of DNA into the cytoplasm by changing the conformation of the injection site, preventing subsequent infection by other T-even-like phages⁵.

- Degradation of invading DNA: bacterial immune system

Intracellular defences that prevent phage replication by degrading viral DNA have been extensively characterized and often compared to eukaryotic immune system with both its innate and adaptive aspects.

Analogous to the innate immunity is the restriction-modification (R-M) system able to discriminate the *self* (bacterial DNA) from the *non-self* (foreign DNA). It involves the activity of a methyltransferase that methylates *self* DNA on specific sites and the endonucleolytic activity of a restriction endonuclease that cleaves the nonmethylated invading DNA at the same recognition site. To evade the R-M system, some phage genomes contain modified nucleotides that can be methylated (e.g. phage T4 contains hydroxymethylcystosine (HMC)). In retaliation, some bacteria have acquired systems that can cleave HMC-containing DNA⁵.

The CRISPR-Cas system (Clustered Regularly Interspaced Short Palindromic Repeats – CRISPR associated genes) is considered as a bacterial adaptive immune system. It integrates short sequences (spacers) of about 20 nucleotides, derived from foreign DNA on the host genome, in between short

palindromic repeated sequences. This dedicated locus, the CRISPR array, thus constitutes a genomic memory of previously encountered parasites. A set of *cas* genes encodes the protein machinery that carries out the immune response. The CRISPR locus is transcribed in a long precursor RNA (pre-crRNA) which is subsequently processed into small crRNAs composed of one spacer unit and one repeat. These crRNAs guide a complex of Cas proteins, including an endonuclease, to detect and cleave complementary DNA (protospacers) on parasitic genomes⁵. Phages can escape this surveillance complex by acquiring point mutations either in the targeted protospacers or in a short conserved motif flanking the protospacer in the phage genome known as the Protospacer Adjacent Motif (PAM). Once again, bacteria can restore phage resistance by acquiring additional spacers targeting the escaping mutant⁵. In addition, phages can carry a set of anti-CRISPR proteins that inhibit the CRISPR-associated endonuclease activity⁴⁶.

- Abortive infection

The abortive infection (Abi) strategy may be considered as an altruistic bacterial behaviour as it prevents the release of new phage progeny by triggering the premature death of the infected cell, thereby limiting phage dissemination in bacterial populations. Abi mechanisms are diverse but generally rely on a phage-induced activation of cellular proteins interfering with key metabolic processes. For instance, a well characterized Abi system is the RexAB system encoded by phage λ . A protein-DNA complex formed during the replication of a secondary infecting phage (e.g. phage T4) activates the intracellular sensor RexA. The subsequent activation of the membrane protein RexB, an ion channel, results in the depolarization of the bacterial membrane ultimately leading to cell death and thus limiting the propagation of the secondary virus within the lysogenized population. Phage T4 *rII* locus encoding RIIA and RIIB has been demonstrated to inhibit RexAB-mediated abortive infection, again exemplifying a phage evasion mechanism⁵. More recently, an additional Abi system has been discovered in *Staphylococci* infected cells. The presence of a specific phage protein in the cell activates a eukaryotic-like serine/threonine kinase (Stk2) which then phosphorylates a variety of proteins involved in a wide range of essential processes (e.g. translation, transcription). This results in cell death before new infectious phage particles can be released, thereby preventing the propagation of phage infection to neighbouring bacteria⁴⁷.

Co-evolution between phage and bacteria

“Now, here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!”

The Red Queen Lewis Carroll - *Through the Looking-Glass* (1871).

Named after the above Lewis Carroll’s quotation, the Red Queen evolutionary hypothesis posits that, within a pressuring and changing environment, organisms must “*run faster*”, that is, they must constantly adapt to maintain their fitness while competing with opposing organisms. Reciprocally, to avoid being outcompeted and driven to extinction, the opposing organism has to counter-adapt (“*running even faster*”). As it has been exemplified in the section above, the Red Queen hypothesis applies to phage-bacteria interactions. This extensive “adaptation-counter adaptation” leads to the generation of substantial diversity within both populations since co-evolution is based on reciprocal genetic changes and subsequent modification of allele frequencies within populations.

Two main types of selection dynamics can be distinguished: The **arms race dynamic** (ARD) takes place when an allele providing a fitness advantage rapidly increases in frequency and eventually reaches fixation within a population (Figure 8). In the case of phage-bacteria antagonistic co-evolution, this would be illustrated by a phage with an extended infectivity range (generalist) whose frequency increases within phage population (as it is able to infect more hosts, thus having a greater reproductive success) and eventually becoming the only phage genotype present (strong positive selection, also called selective sweep). Alongside, same phenomenon applies to the host where the sub-population carrying an allele conferring resistance to this generalist phage will be strongly selected. Consequently, both species continually accumulate adaptive mutations rendering them more and more infective/resistant⁴⁸.

However, in a given environment, generalism may be associated with evolutionary costs and higher levels of specificity in host-parasite interactions may be favoured. In these conditions, a negative frequency-dependent selection (where a rare allelic variant has a selective advantage on a relatively short timescale) is likely to occur and results in the establishment of a **fluctuating selection dynamic** (FSD) (Figure 8). For instance, a rare phage mutant adapted to the most common host genotype will increase in frequency. The resulting selective pressure will favour a rare host genotype that will become more common in turn and so on. It is well established that FSD, implying specificity between host and phage genotypes, allows for the maintenance of diversity and the persistence of coevolution⁴⁸.

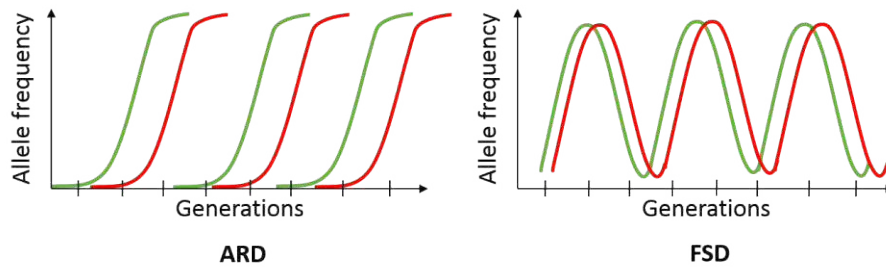


Figure 8: Allele frequencies over time in directional and negative frequency-dependent selection

Green and red curves correspond to host and phage, respectively. Directional selection (left) is characterized by recurrent selective sweeps and leads to the establishment of an arms race dynamic while in fluctuating selection dynamics, negative frequency-dependent selection (right) results in oscillation of host and parasite genotypes over time.

It clearly appears that environmental factors such as nutrient availability, host population density and composition or temperature have an impact on the probability of ARD versus FSD to occur, but their influence is not fully understood⁴⁹. These different types of coevolution potentially have different consequences on global interactions of phages and bacteria (in terms of genetic diversity and dynamics of populations) which have been proved to have a profound impact on ecological systems, biogeochemical cycles and also on the evolution of bacterial pathogenicity in human as the paragraph below illustrates.

Ecological perspective

Initially based on electron microscopy observations, and later refined by other measurements in various environments, the current estimated number of tailed virus particles in the biosphere reaches 10^{31} . This impressive number is actually underestimated as it does not include non-tailed viruses or viral genomes present in cells as prophages¹⁷. Globally, we acknowledge that phages outnumber bacteria by tenfold.

Moreover, tailed viruses are ubiquitous as they have been isolated in a great variety of environments, whether it be in soil, water or more extreme environments such as Sahara desert⁵⁰. They are also present on human skin and digestive tract as exemplified by the widespread so called *crAssphage* that is thought to be carried by half of the world population⁵¹. On top of being numerous and ubiquitous, phages are also extremely diverse. Although phage diversity still remains poorly described, it has been estimated that 100 millions of different phage genomes would be present in nature⁵².

It is therefore reasonable to concede that phages must play a major role in dynamics of ecological systems and this through, at least, two mechanisms: (i) First, by affecting bacterial abundance (and thus composition of communities) in local environments as a result of lytic infections (killing) and (ii) by promoting the acquisition of new genetic traits in bacteria through co-evolutionary relationships (as previously explained) and horizontal gene transfer (HGT).

The best illustration of the first point is given by marine ecosystems where it has been roughly estimated that phages are responsible for 10-50% of the daily microbial mortality⁵³. Then we can easily understand they indirectly affect biogeochemical cycling of major elements (carbon, nitrogen, phosphorus) by shaping the communities of autotrophic organisms (primary producers) as exemplified by cyanophage-cyanobacteria interactions⁵⁴ (major producer of oxygen) or, more directly, by catalysing the transformation of prokaryotic cells towards dissolved organic matter (so-called viral shunt)⁵⁵.

Phages also influence the metabolism of their host as they encode “auxiliary metabolic genes” (AMGs), *i.e.* genes involved in supplying additional resources within the infected cell, to ensure the functioning of metabolic pathways of significance to the phage, thereby fostering reproductive success. Phage AMGs involved in Calvin cycle¹³, photosynthesis, phosphorus and sulphur metabolisms have already been identified⁵⁶. Moreover, a recent metagenomic analysis of the global ocean virome identified more than 230 new AMGs, suggesting that viruses may directly manipulate sulfur and nitrogen cycling⁵⁷. However, integrating quantitative effects of viral infections into large-scale models of ecological processes is challenging. Therefore, the fundamental question of the role of viruses in global earth systems remains largely unresolved.

Additionally, phages also influence the evolution of bacterial genomes by promoting the acquisition of new genes. As a consequence, these new phage-derived genes may confer selective advantages to bacteria and favour the colonization of new ecological niches (e.g. human hosts). Therefore, phages are often associated with the evolution of bacterial pathogenicity. Numerous mechanisms of phage-mediated evolution of bacterial genomes have been reported and few examples are given below.

One of the major mechanisms generating diversity in bacterial genomes is **lysogenic conversion**⁵⁸. Temperate phages can carry accessory genes encoding proteins beneficial for the lysogen and often appear to be virulence factors relative to a human host. For instance, the *Vibrio cholera* toxin is encoded by the transferable phage CTXφ. Prophages are very abundant in bacterial genomes and are sometimes responsible for the conversion of a non-pathogenic strain into a pathogen (e.g. Shiga toxin producing *E. coli* O157:H7)⁵⁹. Phages can also be responsible for HGT between bacteria through various mechanisms such as **transformation** (release of bacterial DNA upon lysis, subsequently acquired by neighbouring competent cells) or transduction (most frequent). **Generalized transduction** is a process during which random host DNA is accidentally packaged within newly formed virions and subsequently delivered to cells upon secondary infection whereas in specialized transduction, it is the host DNA adjacent to the prophage integration site that is packaged after imprecise excision⁵.

4. Medical and biotechnological application of phages

Since their discovery, phages have had a major impact on fundamental and applied biology. While Félix d'Hérelle straightaway investigated their potential use as therapeutic agents, other scientists known as 'the phage group' (including Nobel awarded Max Delbrück, Salvador Luria, Alfred Hershey, James Watson, among others) tackled fundamental biological questions by working with phages as models.

From molecular biology origins to nanotechnologies

Breakthrough discoveries such as the pre-existence of mutations in absence of selection, the demonstration that DNA is the support of genetic information, the triplet nature of the genetic code as well as the development of molecular biology tools (e.g. restriction enzymes, cloning vector, sequencing...) have been based on phage research. Today, phages still have a major importance in biotechnologies, notably through technologies based on phage display technique. This technique, elaborated more than 30 years ago¹⁹, relies on modifying filamentous phage genomes to display polypeptides as fusion to coat proteins at the virion surface and has been widely used in protein-protein interaction studies. Numerous applications have subsequently derived from phage display such as vaccine design, where the recombinant phage is exploited for its immunogenic property and used as a carrier to elicit the production of antibodies against the displayed epitope. Such attempts have been made in animal models to develop immunotherapies against chronic diseases and cancer⁶⁰. Another application of phage display in nanomedicine is the use of recombinant phages to target specific locations (e.g. tumour site) where they can subsequently deliver conjugated cytotoxic drug or recruit immune effector cells.

An additional example of a major contribution of fundamental phage research in the elaboration of sophisticated biotechnological tools is given by the "CRISPR craze". The discovery of the bacterial adaptive immune system led to the development of a powerful genome editing tool based on type II CRISPR-Cas system (better known as CRISPR-Cas9). It relies on the ability of Cas9 endonuclease to generate double-strand breaks in DNA sequences complementary to a Cas9-associated single guide RNA (sgRNA). This sgRNA can be engineered to direct Cas9 towards a chosen DNA sequence, allowing then to select for spontaneous or specific mutations of this sequence. This technology has been tested in various organisms (phages, bacteria, plants, animals and even humans with the approval of the first clinical trial consisting in reinjecting T cells back to a patient upon Cas9 editing to improve cancer therapy⁶¹). These past few years have known an explosion in the number of CRISPR-based applications, in a wide range of domains, from fundamental research to agriculture and biomedicine.

Pathogen detection

The specificity with which phages recognize their host has motivated their use as identification tools in order to detect particular bacterial species (or strains) among complex communities colonizing various environments (such as plants, food products or human host).

Detections methods can be based on phage-induced lysis, where the release of bacterial cytoplasmic markers (such as ATP) upon phage lysis of the targeted pathogen is monitored. Alternatively, the presence of one particular strain can be revealed by measuring phage multiplication. In the so-called phage amplification assay, phages are mixed with the tested sample, allowed sufficient time to infect the targeted pathogen and then titered on a propagating strain⁶². Besides using unmodified phages to detect bacteria, other techniques based on genetically engineered phages have been developed. One example is provided by reporter phages carrying genes that are expressed upon infection of targeted cells and allow the detection of their product by measuring luminescence, fluorescence or production of a chromogenic molecule⁶². Among the reporter genes that have been tested, the most frequently used are derived from luciferase genes, whether it be the full bacterial *luxCDABE* operon (with the limitation associated with such a large DNA fragment to be incorporated in phage size-constrained genomes) or eukaryotic version such as the 516bp *nluc* (NanoLuc)⁶³. This approach makes it possible to rapidly detect viable bacterial cells within a sample but also suffers from disadvantages since the construction of reporter phages requires a deep characterisation of phage genomes, and most phages are not genetically modifiable.

Agriculture and food industry

Similar to their applications in medicine, phages can also be used in agriculture and food industry, whether it be for detection (as explained above) or for the eradication of plant and foodborne pathogens.

This is well illustrated by the development and the commercialization of phage cocktails targeting *Listeria monocytogenes*. The so-called ListShield™ (Intralytix) and Listex P100™ (EBI Food Safety)⁶⁴ products have been approved by American and Canadian regulatory agencies and are used to reduce potential contamination of 'ready-to-eat' food products. Another example is the use of phages as biopesticides to control populations of plant pathogens (e.g. *Erwinia amylovora*). One product, known as AgriPhage™ (OmniLytics), is sold for environmental applications in Canada. However, some raise concerns regarding such large-scale spreading of phages in the environment since, similarly to the overuse of antibiotics, it may have harmful effects on the spread of antibiotic resistance/virulence genes and on the composition of microbial communities within the plants and soil⁶⁵.

Phage therapy

Following its first successful application to treat dysentery, the therapy envisioned by d'Hérelle in the early 20th century entered an 'early enthusiastic' period during which phage therapy trials multiplied. However, associated publications reported inconsistent results (partly due to inappropriate applications and lack of rigorous study designs) which did not meet with the excessive expectations regarding phage therapy. At the same time, the poor knowledge of phages' nature represented an additional point of scepticism and criticism. The discovery of antibiotics, associated to the complex geopolitical context of the Cold War, precipitated the decline of phage therapy in Western countries while the Soviet Union kept developing the use of therapeutic phages⁶⁶.

The current threat of increasing antibiotic resistance, coupled with the paucity of new efficient molecules have motivated the resurgence of phage therapy worldwide for the last couple of decades. Considerable efforts have been made to understand better phage-bacteria interactions (through genetic, functional, ecological and co-evolutionary approaches) but also their interplay with a human host (particularly, its immune system).

Phages' intrinsic properties make them particularly interesting for therapeutic purposes:

- i. Their specificity (ranging from infecting only few strains within a species to many strains even across different genera⁶⁷) is valuable to eradicate harmful bacteria only. Maintaining intact the community of commensal bacteria has been demonstrated to be crucial: broad-spectrum antibiotics alter gut microbiota balance resulting in dysbiosis, largely associated to inflammatory bowel diseases⁶⁸. Although this 'narrow' spectrum can also be seen as a hurdle (e.g. when the pathogen responsible for the infection is not precisely identified), it can be circumvented by using cocktails of phages with distinct host ranges.
- ii. They represent local self-replicating antibacterial agents since they multiply as long as their bacterial host is present resulting in increased phage titers on the infection site. Therefore, phages could be administrated as single doses, in theory.
- iii. Their unprecedented abundance and diversity constitute a large reservoir of alternative antimicrobial solutions and thus facilitate the rapid discovery of phages efficient against one particular problematic strain.
- iv. Their fast bacteriolytic activity is not hampered by antibiotic resistance mechanisms characterizing the most problematic bacterial pathogens.

However, some concerns have been raised regarding their generalised utilisation in humans. For example, their propensity to mediate HGT may favour the spreading of resistance and virulence genes

(through generalized transduction). This is the main reason why temperate phages are not recommended for phage therapy purposes.

A second example is given by infections caused by intracellular bacteria (exclusively or partially) such as *Rickettsia spp.*, *Legionella pneumophila* or *Helicobacter pylori*. They can hardly be treated with phages because they cannot infect (nor penetrate in whatever way) eukaryotic cells.

A main concern is related to the emergence of bacterial resistance to phages, a lesson taken from overuse of antibiotics. However, phage resistance is thought to be more easily circumvented than antibiotic resistance because new phages efficient against resistant bacteria are expected to be found in nature. Even if this is not the case, several studies have reported that acquisition of phage resistance is associated with trade-offs (e.g. reduced virulence, decreased mobility, lower growth rate) rendering resistant bacteria more susceptible to conventional treatments⁶⁹, although this is not a general rule.

Another fundamental question regards innocuousness of phages. A century of phage studies have reported no evidence of phage-induced disorders, despite their omnipresence in the environment (food, water, human organism). This would tend to prove that phages are rather well tolerated. As long as phage preparations are sufficiently purified, that is, cleared from bacterial degradation products, it seems that the activation of the first line of immune defences (innate immunity) is weak. On the other hand, evidence of adaptive immune response to phage administration have been reported (e.g. production of antibodies against phages) but not associated to unsuccessful treatment⁷⁰. Moreover, the potential action of immune memory, that would prevent the reuse of a particular phage, may be avoided by limiting the capture of phages by antibodies through improved phage formulations⁷¹. Overall, these three-player interactions are far from being fully understood and further studies are required.

Some of the drawbacks of phage therapy are being tackled by synthetic biology, enabled by continual progress in sequencing technologies and genetic engineering. For example, recombinant phages with modified tail fibers that allow them to access a wider range of hosts have been produced^{72,73}. As an alternative approach, research groups have proposed to use phage virions to deliver a synthetic plasmid carrying antimicrobial genes leading to rapid and non-lytic bacterial death (thereby avoiding the release of bacterial toxic components or endotoxins). Benefits of using these so-called phagemids are threefold: (i) the specificity of cell targeting can be even more refined; on top of the intrinsic virion-based strain specificity (phage receptor), the delivered plasmid can encode a CRISPR-Cas system directed against a specific genetic signature (e.g. antibiotic resistant marker)⁷⁴. (ii) Their non-replicative nature ensure no possible HGT (but reciprocally, the benefit associated to phage replication is lost). (iii) The delivered genetic material is fully understood (unlike native phages genomes that mostly encode unknown functions).

From a clinical point of view, even if eastern European countries (e.g. Georgia, Poland and Russia) have used phage therapy for decades, their clinical experience of phage administration is bound to ancient clinical trials that do not meet with current standards. As the renewed interest in phage therapy is recent, safety standards and legal frameworks are not adapted (even inexistent) to phage products. As a result, current development of phase II clinical trials is slowed down. So far, two have been conducted. The first one aimed to treat chronic otitis caused by antibiotic resistant *P. aeruginosa* and resulted in significant reduction of bacterial load in treated patients⁷⁵. The second reported no benefit of oral phage administration to treat acute *E. coli* diarrhea which might be explained by the fact that targeted bacterial strain was actually at low concentration while increased levels of *Streptococcus* were found⁷⁶. Moreover, two phase I-II clinical trials are currently in progress: the PhagoBurn trial (<http://www.phagoburn.eu>) for treating cutaneous infections on burn wounds and the PHOSA trial (<http://www.phosa.eu>) combatting osteoarticular infections due to *Staphylococcus epidermidis* and *S. aureus*.

Phage therapy, as one of the most promising alternative to antibiotics, needs further rigorous studies that must define its frame of application in terms of safety and efficacy. However, one of the main impediment to the development of phage therapy is the lack of adapted regulatory frameworks. A more straightforward approach to cope with the urgent need of new antimicrobial weapons commercially available may still be the design of new antibacterial molecules. Once again, phages are valuable allies for this purpose.

Source of novel antimicrobials

Over billion years of coevolution, phages have developed strategies to optimally inhibit or co-opt key bacterial processes in order to ensure efficient infection cycles. These strategies have therefore inspired the design of new antimicrobial molecules. Many efforts have been made to develop the use of phage antibacterial enzymes as therapeutic agents and among those, endolysins are the most popular. Their exogenous application to lyse Gram positive bacteria has proven to be very efficient both *in vitro* and in animal models (mainly infections due to *S. aureus* and *Streptococcus spp*)⁷⁷. Efforts have also been made to engineer these lytic enzymes, increasing their activity or broadening their lytic spectrum. Although endolysin-based products are still not approved for therapeutic use in humans, two phase I clinical trials targeting *S. aureus* infections are being performed⁷⁷. In contrast, endolysin application is considerably more challenging in the case of Gram negative bacteria due to their impermeable outer membrane. Consequently, strategies to enhance their access to peptidoglycan (e.g. combination with cell wall disrupting agents, example of Artilysin[®]S⁷⁸) have been studied. Few have been demonstrated natural bacteriolytic activity against Gram negative bacteria but have not been tested in animal models yet⁷⁷.

Although resistance mechanisms against endolysins have still not been reported yet, resistance to such 'lysis from without' mediated by external application of peptidoglycan hydrolases have been documented (e.g. resistance to the bacteriocin called lysostaphin). However, fitness costs associated to a modification of endolysins' targets (*i.e.* peptidoglycan bounds) are expected to be important and may come with reduced viability and/or virulence.

It has to be noted that the therapeutic use of other envelope-targeting phage proteins (*i.e.* virion-associated peptidoglycan hydrolases, phage-associated depolymerases and holins) are also being investigated⁷⁷.

On top of cell wall degrading molecules, current commercially available antimicrobials mostly interfere with DNA biosynthesis, transcription and translation and globally target fewer than 20 families of proteins while there are hundreds of essential bacterial proteins⁷⁹. Therefore, identifying additional bacterial targets and their inhibitors is essential to widen our therapeutic arsenal. Deep studies of phage-host interactions suggested that phage non-structural metabolic proteins could be utilized in drug discovery. Such proteins, which are responsible for the transition of cellular metabolism from a host-oriented state to a phage-centred state (virocell), are mostly produced during the early stage of infection. As some of these early genes are lethal to the host, deciphering their mode of action represents a powerful way to discover new antibacterial strategies. Numerous mechanisms have already been described (most of them in *E. coli* model phages) and will be discussed in the following chapter. High-throughput screening strategy to identify such phage-encoded antimicrobial candidates has been proposed by Liu and co-workers⁸⁰ and is summarized in Figure 9.

Although this approach has been proposed since 2004, implementations by industry and pharmaceutical companies are limited. This is probably due to the fact that fundamental research steps required upstream product development are numerous and involve technical challenges. However, the urge for new antibacterial treatments may tip the scale in favour of such proteomic and metabolomics-based research.

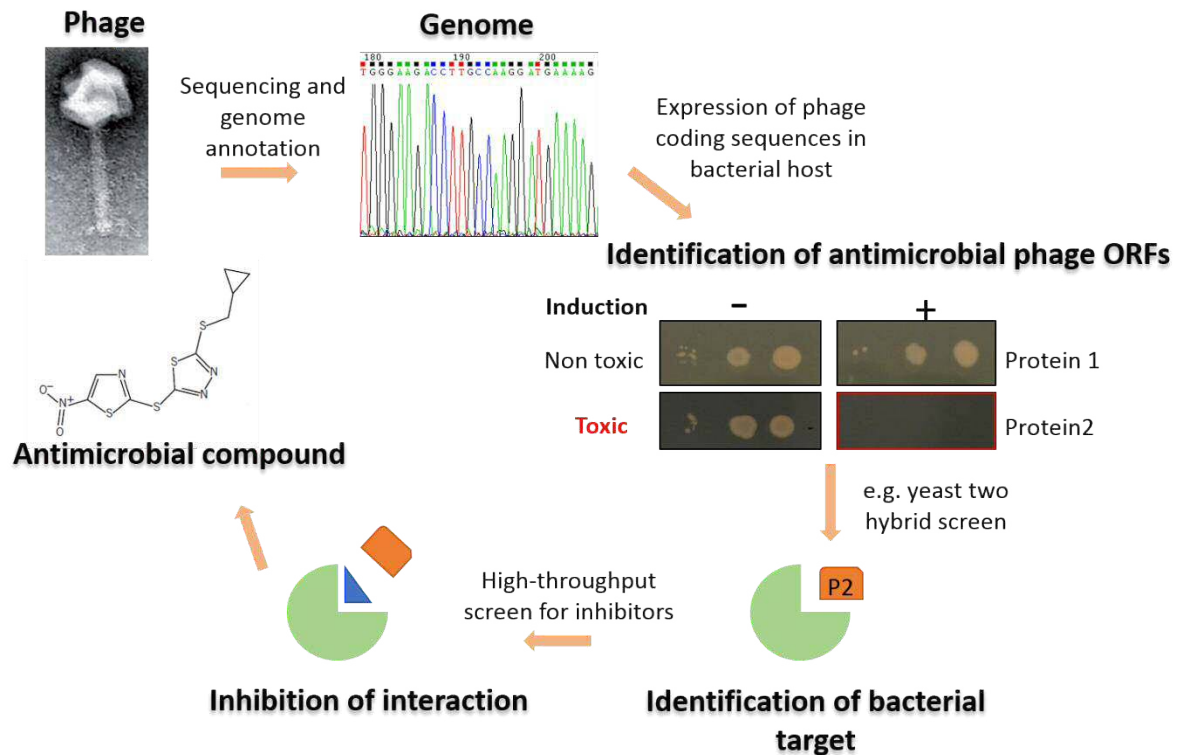


Figure 9: Antimicrobial drug discovery strategy based on phage genomics

Genome of a newly isolated phage is sequence and annotated. Identified candidate ORFs are then cloned and expressed in bacterial host to assess their antimicrobial activity. If the gene product reveals toxicity, its targeted interaction partner is identified. Finally, a library of compound is screened to identify a molecule inhibiting this interaction and thereby mimicking the effect of phage protein.

CHAPTER 3 - TURNING THE RIBOCELL INTO A VIROCELL: KEYS FOR SUCCESS

As previously mentioned, phage early proteins are globally responsible for shutting down host ‘energy-wasting processes’ (from a viral perspective) and defences as well as recruiting cell machineries for the production of new phage particles. In addition to deciphering molecular interactions between phage early proteins and their host targets, a number of studies have recently assessed the impact of phage infection on the global host metabolism, with an emphasis on transcriptional effects, by taking advantage of “-omics” approaches.

1. Sabotage and enslavement of cell machineries

Numerous examples of phage proteins interacting with (and thereby inhibiting, stimulating or redirecting) host proteins involved in crucial cellular functions have been reported for decades, notably with extensive studies of model phages like T7 and T4 in the 1970s-80s. The more targeted machineries are replication, transcription, translation and cell division systems (Table 2). In a large majority, these phage early proteins are small (<250 amino acids), non-essential proteins that usually display little sequence similarity to any other proteins in databanks and that are responsible for about 60% of phage-host protein-protein interactions according to an estimation made by Roucourt and Lavigne⁸¹.

Depending on the level of dependence regarding host machineries (*i.e.* whether they encode their own machinery or not), phages either shut off or redirect related host processes. As most lytic phages encode their own DNA replication enzymes, they usually inhibit host replication through direct interactions with host apparatus as exemplified by Gp104 from *S. aureus* phage 77 which targets DnaI (helicase loader) or Gp8 from *E. coli* phage N4 which directly targets the DNA polymerase (Table 2). Furthermore, to rapidly synthesize new viral DNA, lytic phages need to access a large pool of available nucleotides. Instead of relying on *de novo* nucleoside production, many phages were shown to induce host chromosome degradation and recycle degradation products by incorporating them into their new genomic copies. For instance, phage T4 encodes endonucleases (DenA and DenB) involved in the degradation of host chromosome specifically, as T4 DNA contains modified cytosine residues protecting it against self-cleavage (and against foreign restriction enzymes)³⁴. Degradation of host DNA provides an additional advantage by eliminating competing DNA and RNA synthesis from the host or from other competing genetic elements (such as prophages).

Similarly, phages extensively interact with bacterial transcriptional machinery by completely or partially hijacking host RNAP thereby favouring their own transcription while reducing the burden of host RNA synthesis. Consequently, phages have developed a large variety of mechanisms to impact

bacterial transcriptional systems, depending on whether they encode a viral RNAP. If not, phages encode proteins allowing the appropriation of the host RNAP through its recruitment by viral sigma factors (e.g. phiEco32 Gp36) which direct the RNAP towards phage middle and late genes. On the other hand, phage early promoters are usually similar to bacterial promoters and thus transcribed by the host RNAP, associated to its canonical sigma70 factor. Alternatively, phage proteins can also induce modifications to RNAP subunits to increase its affinity for phage-specific promoters (e.g. ADP-ribosylation catalysed by T4 Alt and ModA proteins)³⁴. In contrast, phage-encoded inhibitors can also shut off host transcription, thereby favouring the transcription of viral genes by a phage-encoded RNAP (see examples listed in Table 2).

Subsequent to this massive transcription, the cellular translation apparatus needs to be efficiently redirected to the overflowing phage mRNAs. For example, *T. thermophilus* phage ϕ YS40 produces middle and late transcripts devoid of ribosome-binding sites, thereby differentiating from bacterial transcripts. To favour the translation of phage leaderless mRNAs over host mRNAs, this virus is thought to induce the upregulation of a translation initiation factor implicated in translation of leaderless RNAs⁸².

Besides these three main cellular machineries, phage proteins have been reported to interact with various other bacterial proteins involved in cell division or various metabolic pathways (e.g. glycolysis, nitrogen metabolism⁷⁹). These numerous interactions suggest that phage impact on host cell must be more complex and subtle than a rough wide inhibition of bacterial processes. Recent studies investigating impacts of phage infection on global dynamics of host transcription or metabolism have demonstrated that phages manipulate the *ribocell* to various extents: from little modification of specific pathways to massive replacement of cell content by phage content.

Table 2: Mechanisms of host subversion mediated by phage early proteinsCompleted from Drulis-Kawa *et al.*⁸³ ND: Non determined, su: subunit

Host mechanism	Phage	Host	Phage protein	Host protein	Stage	Effect	Ref
Replication	λ	<i>E.coli</i>	CIII	FtsH	Early	Competitive inhibition of FtsH	84
			Gam	RecBCD	delayed early	Inhibition of nuclease and helicase activities	85
	77	<i>S. aureus</i>	gp104	DnaI	Early	Inhibition of host replication	80
	Twort	<i>S.aureus</i>	gp168	DnaN	ND	Shutoff of host replication	80
	G1	<i>S.aureus</i>	gp240	DnaN	ND	Shutoff of host replication	80
	N4	<i>E.coli</i>	gp8	HolA (DNaPol III δ su)	ND	Shutoff of host replication	86
Transcription	ϕ Eco32	<i>E.coli</i> 55	gp36 (α factor)	RNAP complex	ND	Recognition of phage promoters	87
	ϕ Eco32	<i>E.coli</i> 55	gp79	RNAP complex	Early	Inhibition of host transcription	87
	T7	<i>E.coli</i>	gp0.7 C-term	RNAP β' su	Early	Host transcription shutoff	88
			gp0.7 N-term	RNAP β and β' sus	ND	Efficiency of termination	89
			gp0.7 N-term	RNase III	ND	Processing of T7 mRNA	90
			gp0.7 N-term	RNase E and RhlB	ND	Protection of T7 mRNA from degradation	91
			gp2	RNAP β' su	ND	Inhibition of transcription initiation	92
	T4	<i>E.coli</i>	Alc	RNAP β su	Early	Host transcription shutoff	93
			ModA	Both RNAP β su	Early	Lower expression of T4 early and host genes	94
			AsiA	σ^{70} factor	Early	Inhibition of σ^{70} promoter recognition	95
			MotA	σ^{70} factor	Early	Recognition of T4 middle promoters	95
			RpbA	RNAP core	Early/middle	Stimulation of T4 late genes expression	96
			Mrh	σ^{32}	Early	Decoy of σ^{32} from RNAP core	97
			Srh	RNAP core enzyme	Early	Decoy of σ^{32} from RNAP core	97
			Srd	RNAP core enzyme	Early	Decoy of σ^{70}/σ^{38} from RNAP core	97
			Srd	Rnase E	Early	Stimulates host mRNA degradation	98
	λ	<i>E.coli</i>	N	RNAP core, NusA, NusG	Early	Antitermination, delayed early transcription	99
			Q	RNAP core	delayed early	Antitermination, late transcription	100
	P2	<i>E.coli</i>	Org	RNA ¹⁰¹ P α su	Early	Inhibition of late transcription	102
	SPO1	<i>B. subtilis</i>	gp44	RNAP β su	Early	Inhibition of host RNAP	103
			gp28	RNAP core	Early	σ factor for middle transcription	104
	Xp10	<i>E.coli</i>	P7	RNAP β' su	Early	Inhibition and antitermination of transcription	105
	P23-45	<i>T.thermophilus</i>	gp39	RNAP holoenzyme	ND	Shutoff of host transcription	105

Host mechanism	Phage	Host	Phage protein	Host protein	Stage	Effect	Ref
Transcription	G1	<i>S. aureus</i>	gp76	RNAP holoenzyme	ND	Shutoff of host transcription	105
			gp67	RNA polymerase σ A factor	ND	Shutoff of host transcription	106
	PaP3	<i>P. aeruginosa</i>	gp70.1	RNA polymerase σ S factor	Early	?	107
	LUZ19		gp25.1	RNAP	ND	Inhibition of transcription initiation	108
	LKA1		gp36	RNAP	ND	Inhibition of transcription initiation	108
	PhiKZ		gp37	RNAse E	ND	Inhibition of RNA degradosome	109
Translation	T4	<i>E.coli</i>	ModB	30S ribosomal su S1,TF	Early	Promotes binding of mRNAs	101
			Alt	EF-Tu, TF	ND	Aminoacyl-tRNAs recruitment	101
	T7	<i>E.coli</i>	gp0.7	IF1,IF2,IF3	ND	Helper function for IF2 and IF3	101
	ϕ YS40	<i>T.thermophilus</i>	ND	IF2,IF3	Early	Translation of phages leaderless mRNAs	110
Cell division	Rac	<i>E.coli</i>	Kil	FtsZ	Early	Cell division arrest	111
	T7	<i>E.coli</i>	gp0.6	MreB	Early	Cell division arrest	112
	T7	<i>E.coli</i>	gp0.4	FtsZ	Early	Cell division arrest	113

2. Global impact of phage infection on cellular transcriptional dynamic

The common view that phage infection necessarily leads to a phage-induced global and total shutoff of host gene expression (both at transcript and protein levels) relies on studies of model phages (e.g. T4 or T5) and on the assumption that it must be beneficial for the phage since it both prevents the establishment of a host stress response and excludes a transcriptional competition with other parasitic genetic elements (such as prophages or plasmids) that might interfere with an efficient phage production.

Conversely, maintaining a host transcriptional system functional, at least partially, may have advantages. This would allow the phage to keep active dedicated host regulatory mechanisms/stress responses enabling either the maintenance of cell homeostasis or the adaptation to changing extracellular conditions. Indeed, recent studies have suggested that total exclusion of host transcription is not a general rule.

Complete shutoff of host transcription during phage infection: a mistaken view

A study of infected *Synechococcus* cells by the Syn9 T4-like cyanophage, based on a combination of microarray and RNA-Sequencing (RNA-Seq), revealed a similar transcriptome dynamic as T4, that is, an almost complete replacement of host transcripts with phage transcripts which constitute 98% of the total cellular mRNA content by late infection¹¹⁴. In contrast, transcripts derived from phage ϕ R1-37 barely reach 16% of the total mRNA content of *Yersinia* infected cells¹¹⁵. In between is the case of *Pseudomonas* phage LUZ19 that ultimately produces 60% of the cell transcripts by the late stage of infection¹¹⁶. These data (and others) illustrate that the extent of host transcripts replacement with phage transcripts depends on the phage and is even variable in alternative hosts. Indeed, Howard-Varona and co-workers showed that the transcriptional takeover induced by phage Phi38:1 (*Bacteroidetes*-virus) was less extensive on an alternative host (<30% of total RNA by late infection) than on its primary host (>90%)¹¹⁷.

Reprogramming of cell transcriptome during phage infection revealed by whole-genome transcriptional analyses

Besides these differences in dynamics of transcripts replacement, manipulations of host gene expression have also been demonstrated to be variable among phages. Some phages barely modify host gene expression patterns while others induce drastic changes.

An example of the former case is given by the giant *P. aeruginosa* phage PhiKZ that triggers the up-regulated expression of a single bacterial operon which happens to belong to prophage Pf1 and may represent the prophage attempt to escape the infected cell. Therefore, it appears that PhiKZ does not reprogram host gene expression³⁵. This result could have been expected as PhiKZ fits the T4-based

model by excluding, almost completely, host transcription over the course of infection (>98% of non-ribosomal RNAs in the cell upon late infection is phage-derived)³⁵. In contrast, the T4-like Syn9 (see above) has been reported to up-regulate the expression of genes involved in various processes (cell envelope, DNA repair, carbon fixation, respiration, tRNA synthesis, nutrient utilization) while depleting almost all bacterial RNAs, like PhiKZ does. Therefore, Syn9 seems to maintain a functional specific host transcription. It has to be noted however that we cannot fully discriminate whether this transcription maintenance is phage driven or simply reflects a general host response to stress¹¹⁴.

Table 3 lists (non-exhaustively) the main results of several works studying transcriptome dynamics of phage-infected cells based on various phage-host models. This illustrates that a wide-range of impacts on host transcription exists (whether it be host-driven or phage-induced) and are specific to each phage-host system. More specifically, these impacts appear to be host-specific as reported Howard-Varona and colleagues when studying phage phi38:1 infections of two alternative, almost identical, *C. baltica* hosts (sharing 93% of their protein-coding genes with an average nucleotide identity of 99.99%). They showed that the two host transcriptomes were largely different upon infection¹¹⁷.

As illustrated by Table 3, many host genes that are differentially expressed upon phage infection are clearly part of bacterial stress response but are also involved in various pathways such as carbon, amino acid, nucleotide metabolisms or export systems. This suggests that phage infection is likely to have a profound impact on cellular metabolism.

Table 3: Examples of analyses of transcriptional dynamics of phage infected cell

Phage	Host	Technique	Consequences of phage infection on host transcription (main results as reported by the authors)	Reference
PRD1	<i>E. coli</i>	microarray	Induction of osmotic stress response, exopolysaccharide synthesis, phage shock proteins	Poranen <i>et al.</i> , 2006 ¹¹⁸
PRR1	<i>P. aeruginosa</i>	microarray	Low impact on cell metabolism Affected genes belong to transport, energy production, and protein synthesis functional classes	Ravantti <i>et al.</i> , 2008 ¹¹⁹
P-SSP7	<i>Prochlorococcus</i>	microarray	Decline of transcript levels for approximately 75% of the host genes 41 up-regulated genes involved in high-light-inducible stress response, carbon metabolism, transcription and ribosome genes. Subsequent up-regulated genes belong to RNA degradation and modification, protein turnover and stress responses pathways	Lindell <i>et al.</i> , 2007 ¹²⁰
Tuc2009	<i>Lactococcus lactis</i>	microarray	Modest impact on host transcription Genes with increased transcription are mainly involved in amino acid metabolism and nucleotide conversion	Ainsworth <i>et al.</i> , 2013 ¹²¹
SWU1	<i>Mycobacterium smegmatis</i>	RNA-Seq	16.9% of <i>M. smegmatis</i> genes were differentially regulated by phage infection Up-regulated pathways include ribosome, protein export, bacterial secretion system, glycerophospholipid metabolism, RNA degradation. Down-regulated pathways include the biosynthesis of siderophore non-ribosomal peptides, nitrotoluene degradation.	Fan <i>et al.</i> , 2016 ¹²²
PhiR1-37	<i>Yersinia enterocolitica</i>	RNA-Seq	Low replacement of host transcripts with phage transcripts (15%) Up-regulation of several specific bacterial gene products known to be involved in stress response and membrane stability, phage- and cold-shock proteins	Leskinen <i>et al.</i> , 2016 ¹¹⁵
Phi29	<i>Bacillus Subtilis</i>	RNA-Seq	Upregulated involved in nucleic acid metabolism (including DNA replication) and protein metabolism Repressed genes involved in the utilization of specific carbon sources such as ribose and inositol	Mojardin and Salas 2016 ¹²³

3. Impact of phage infection on host cellular metabolism

The reprogramming of metabolic genes expression during infection along with their ability to encode AMGs (see *Ecological perspective* section in Chapter 2), supported the hypothesis that phages reprogram their host's metabolic processes to favour their replication. Moreover, efforts in exploring the 'viral dark matter' (*i.e.* early expressed ORFs with unknown function) have shown that early proteins not only interact with RNA, DNA and replication machineries but also with protein complexes involved in metabolic pathways (e.g. Calvin cycle¹³, fatty acid biosynthesis¹²⁴). Despite these strong indications, experimental data investigating global metabolic variations occurring in infected cells remain scarce.

The first study tackling this question investigated the interaction between *Sulfitobacter sp.*, a marine bacterium involved in biogeochemical transformations, and its lytic phage ϕ 2047B¹². They observed a progressive increase in intracellular concentrations of ~70% of the metabolites over the course of infection (e.g. UDP-activated sugars, amino acids). Ultimately, late infected cells display a core metabolome composition radically different compared to uninfected cells. This increase in metabolite concentrations has been demonstrated to be due to an elevated metabolic activity of infected cells rather than a decreased utilisation of these molecules. Therefore, the conclusions of this study supported that phages induce an active reprogramming of bacterial metabolism rather than a rough consumption and degradation of available cellular metabolites¹².

Subsequently, a recent work by De Smet *et al.* (2016)¹⁴ investigated how generalizable these conclusions are, regarding other phage-host systems. They studied the metabolite dynamics in *P. aeruginosa* cells over the course of individual infection by six different phages. By addressing the question of a 'universal' trend of metabolic variations, they showed that phages generally elicit an intensification of pyrimidine and nucleotide sugar metabolism. They also made a positive correlation between increased metabolic pathways and related phage-encoded AMGs. For instance, phages encoding AMGs potentially involved in host genome degradation (e.g. nucleases) have a distinct impact on pyrimidine metabolism compared to phages lacking such AMGs. In addition, they found clear phage-specific and infection stage-dependent variations. Schematically, we can distinguish phages draining all host available resources for viral production (YuA phage) and phages actively reprogramming host metabolism to provide necessary resources to produce new phage building blocks (PhiKZ phage)¹⁴. Overall, finding that a large variety of metabolic pathways are impacted by phage infection, rather than pathways involved in nucleotide and protein synthesis only, confirms that phage-host intracellular interactions are more intertwined and complex than previously assumed and represents an additional argument reinforcing the *virocell* model.

Additional similar studies are still required to evaluate the diversity of impacts of phage infections on cell metabolisms. Notably, additional phage-host systems need to be investigated: a very recent study characterized the alteration of *E. coli* physiology by the shiga-toxin ϕ 24B prophage. Naïve and lysogen strains were reported to display distinct metabolic profiles, demonstrating an impact of the prophage on host metabolism despite the absence of new phage progeny production. In addition, lysogen host has been found to be more tolerant to antimicrobials, even though the prophage does not carry related resistance genes. This further illustrates the importance of studying the metabolic impact of phage infection¹²⁵.

CHAPTER 4 - THE VERSATILE OPPORTUNISTIC PATHOGEN *PSEUDOMONAS AERUGINOSA*

1. A threatening opportunistic pathogen

Pseudomonas aeruginosa is a Gram negative rod-shaped bacterium acknowledged for its ability to colonize diverse ecological niches. This remarkable adaptability makes it an important human opportunistic pathogen. Particularly susceptible are patients with damaged epithelial barriers (burn wounds), a compromised immune system (notably neutrophil depletion) or an altered mucociliary clearance favouring lung colonization (typically in cystic fibrosis (CF) and chronic obstructive pulmonary disease)¹²⁶. *P. aeruginosa* is often associated to hospital acquired infections, through colonization of medical equipment, such as ventilator-associated pneumonia (in mechanically ventilated patients) or urinary tract infections (occurring secondary to catheterization). It is also associated with nosocomial bloodstream and intra-abdominal infections and, most commonly, with otitis externa (so called “swimmer’s ear”) and keratitis (eye infection). This wide range of *P. aeruginosa*-associated infections is the reflection of unique properties: (i) a great capacity to colonize hostile environments (hospital sinks and even disinfectant solutions); (ii) a large arsenal of pathogenicity factors and strategies to evade host defences and (iii) intrinsic and acquired resistance to a large number of antibiotics. Altogether, these properties are key factors to *P. aeruginosa* success and will be further described in the next paragraphs.

P. aeruginosa was given a “threat level of serious” in 2013 CDC’s¹ report: *Antimicrobial resistance threats in the United States*. It mentions that “the most serious gram-negative infections are healthcare-associated, and the most common pathogens are *Enterobacteriaceae*, *Pseudomonas aeruginosa*, and *Acinetobacter*”. In this report, the CDC estimated that multi-drug resistant *P. aeruginosa* cause 51 000 infections per year in the United States and an increasing number of strains have been found to be resistant to all existing antibiotics.

¹ Centers for Disease Control and Prevention (www.cdc.gov)

2. A master of adaptation

Adaptation to changing environments requires efficient stress responses to harmful conditions such as limitation of metabolic resources, host immune defences (oxidative stress) or antibacterial treatment. *Pseudomonads* species are known for their enormous versatile and adaptive physiology as reflected by their high coding potential (genome size from 6 to 7 Mb carrying ~6000 predicted genes). Indeed, they can produce numerous virulence factors (e.g. exotoxin A, type III secretion system, pyocyanin, *etc.*), utilize a broad spectrum of nutrients, form biofilms and have an important metabolic capacity¹²⁷. All these phenotypic changes are tightly controlled by an unusually large repertoire of regulatory pathways (e.g. transcriptional regulators, two-component systems, *etc.*) as illustrated by the identification of more than 500 regulatory genes in the model strain PAO1¹²⁸.

Detection and signalling systems

- Two-component systems

Bacteria have developed molecular sensors allowing them to detect environmental fluctuations and adapt their colonization strategy accordingly. Notably, the so-called two component system is an important regulatory complex consisting of a sensor (histidine kinase) inducing the activation of a response regulator through a phosphorylation cascade. Usually, the activated regulator directly controls gene expression through a DNA binding domain. A large variety of sensors can recognize numerous environmental signals (oxygen, nutrient, osmolarity, *etc.*). In average, bacteria possess 30 two-component systems¹²⁹ but this number is highly variable: ranging from 1 in *Chlamydomophila pneumoniae* to more than 200 predicted in *Myxococcus xanthus*¹³⁰. In accordance with its adaptability, *P. aeruginosa* has been found to encode about 60 two-component systems. One of the most extensively studied is the global activator of antibiotic and cyanide synthesis (GAC) and consists of the transmembrane sensor GacS and its regulator GacA whose activation eventually results in favouring biofilm formation and reducing acute virulence and motility¹³¹. GacA can also be activated by other sensor kinases (such as LadS and RetS) with reciprocal effects, illustrating the complexity of this regulation network.

- Quorum sensing

In addition to two-component systems, *P. aeruginosa* cells are able to adapt their lifestyle coordinately within the bacterial population through cell-to-cell communication by using diffusible signalling molecules. This process, termed quorum sensing (QS), involves the regulation of gene expression in response to critical concentrations of signal molecules reflecting cell density. Currently, three major QS systems are known for *P. aeruginosa*: two utilize N-acyl homoserine lactones (AHLs) as signal molecules while the third recognizes quinolone molecules (so-called *Pseudomonas* quinolone signal or PQS)¹³¹.

Two distinct **AHL systems** coexist in *P. aeruginosa*, namely Las and Rhl systems, and are hierarchically organized in PAO1 (the former controlling the latter). Briefly, system-specific AHLs are synthesized by LasI and sensed by the transcriptional regulator LasR (or RhlI and RhlR, respectively) leading to the control of the production of virulence factors such as elastases, exotoxin A and proteases (or production of rhamnolipids, and repression of the type III secretion system in the case of Rhl system). Obviously, regulation of AHL systems is far more complex and is integrated within a regulatory network involving transcriptional and posttranscriptional control by other regulators¹³¹.

The **PQS molecule** is synthesized by the *pqsABCD*-encoded system. Once it reaches a threshold concentration in extracellular medium, it binds to its receptor and induces its own production as well as the expression of several virulence factors (e.g. *pvd* and *pch* operons responsible for the synthesis of siderophores, see further). Similarly to AHL systems, this quinolone signalling pathway is complex and involves additional structurally-related quinolones (other than PQS) along with other regulatory networks¹³¹.

As illustrated in Figure 10, other signalling pathways involving extracellular molecules exist: (i) those relying on the production of typical pigmented molecules (e.g. pyoverdine, phenazines/pyocyanin); (ii) remarkable diffusible signal factors (DSF) allowing interkingdom communication (e.g. between *P. aeruginosa* and the fungal pathogen *Candida albicans*) and (iii) cyclic dipeptides (diketopiperazine, DKPs) proven to interfere with QS systems although their function is not fully understood¹³¹.

- Nucleotide-based signals

In addition to diffusible (extracellular) communication signals, a large group of nucleotide-based messengers have been demonstrated to play a crucial role in orchestrating intracellular processes.

First discovered, and most well-known, is cAMP. In *P. aeruginosa*, cAMP binds to the regulator Vfr which eventually induce the production of multiple virulence factors. This **cAMP-Vfr signalling** is also a complex signalling cascade that integrates multiple regulatory levels (e.g. it has been demonstrated to be inhibited by the sigma factor AlgU regulon¹³²).

Also largely conserved across many bacterial species, **the messenger cyclic di-GMP** is involved in the control of a multitude of phenotypes (motility, stress adaptation, secondary metabolism, virulence, etc.) and is of key importance for triggering the switch between motile single cell lifestyle to sessile biofilm formation.

Another group of nucleotide-based signalling comprises the **cellular alarmones**, namely ppGpp and pppGpp, which mediate the redirection of cellular metabolism toward survival mode upon amino acid starvation¹³¹.

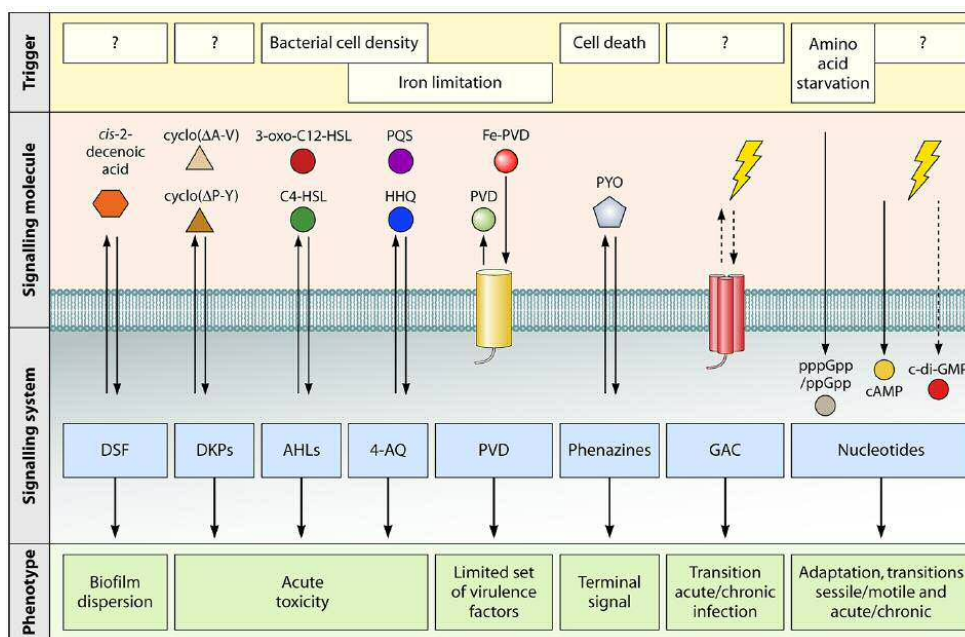


Figure 10: Summary of cell-to-cell signalling systems in *P. aeruginosa*

Extracted from Jimenez *et al.*¹³¹

DSF: Diffusible signal factors, DKP: Diketopiperazines, AHL: Acyl-homoserine lactone, 4-AQ: 4-alkyl quinolone, PVD: pyoverdine, GAC: global activator of antibiotic and cyanide synthesis

Complex transcriptional regulation network

As suggested above, signalling pathways eventually result in the activation of transcription factors that transform the initial external signal into a coordinated transcriptional response. Typically, many of these environmental stress responses are initiated and coordinated by dedicated extracytoplasmic function (ECF) sigma factors, which associate with core RNAP and direct it to a specific set of targeted genes (so called regulon). Most of these ECF sigma factors regulate their own gene expression which are often co-transcribed with genes encoding membrane-anchored inhibitors. These so called anti-sigma factors prevent association of ECF sigma factors to core RNAP by sequestering them at the membrane. Upon environmental signal, the interaction anti sigma/sigma is disrupted and freed ECF sigma factor enables the transcription of corresponding regulon. Each ECF sigma factor is induced by a specific type of stimuli (such as cell envelope stress or iron starvation) as illustrated by Table 4.

Table 4: ECF sigma factors identified in *P. aeruginosa* PAO1As reported by Llamas et al.,(2014)¹³³

Sigma	Anti-Sigma	Function	Reference
AlgU	MucA	Alginate production	Mathee <i>et al.</i> , 1997 ¹³⁴
FemI	FemR	Mycobactin-mediated iron uptake	Llamas <i>et al.</i> , 2008 ¹³⁵
FiuI	FiuR	Ferrichrome-mediated iron uptake	Llamas <i>et al.</i> , 2006 ¹³⁶
VreI	VreR	virulence regulator	Llamas <i>et al.</i> , 2009 ¹³⁷
FoxI	FoxR	Ferrioxamine-mediated iron uptake	Llamas <i>et al.</i> , 2006 ¹³⁶
FpvI	FpvR	Pyoverdine-mediated iron uptake	Beare <i>et al.</i> , 2003 ¹³⁸
HasI	HasS	Heme uptake	Ochsner <i>et al.</i> , 2000 ¹³⁹
PA0149	PA0150	Metal uptake	Llamas <i>et al.</i> , 2008 ¹³⁵
PA1300	PA1301	Putative Haem uptake	Llamas <i>et al.</i> , 2008 ¹³⁵
PA1363	PA1364	Putative siderophore-mediated iron uptake	Llamas <i>et al.</i> , 2008 ¹³⁵
PA2050	PA2051	Metal uptake	Llamas <i>et al.</i> , 2008 ¹³⁵
PA2093	PA2094	Siderophore-mediated iron uptake	Llamas <i>et al.</i> , 2008 ¹³⁵
SigX	-	Virulence and biofilm	Gicquel <i>et al.</i> , 2013 ¹⁴⁰
PA4896	PA4895	Siderophore-mediated iron uptake	Llamas <i>et al.</i> , 2008 ¹³⁵
PvdS	FpvR	Pyoverdine synthesis and virulence	Lamont <i>et al.</i> , 2002 ¹⁴¹

ECF sigma factor-associated pathways are also subjected to additional levels of regulation and cross-talk controls. Indeed, *P. aeruginosa* also encodes a remarkable number of transcriptional regulators as found by Stover *et al* (2000) when they sequenced the first *Pseudomonas* genome (*P. aeruginosa* PAO1) and predicted that 7.2% of predicted genes were involved in transcriptional regulation ¹²⁸. Some are well characterized, such as the major iron acquisition regulator Fur (ferric uptake regulator) which modulates expression of other regulators (e.g. ECF sigma factors, two component systems or small regulatory RNAs).

This extensive, integrated and complex transcriptional regulation is well illustrated in Figure 11 where transcriptional regulatory proteins are represented in orange.

Extracted from Balasubramanian *et al* (2012)¹⁴²

Small non-coding RNAs are now acknowledged to be important modulators of various bacterial processes. In particular, bacterial small RNAs (sRNAs) impact gene expression through a variety of mechanisms such as (i) interaction with proteins, (ii) interaction with double-stranded DNA or RNA, (iii) binding to small effector molecule (riboswitch) and (iv) base-pairing with target mRNA¹⁴³.

With a different mode of action, PrrF1/PrrF2 sRNAs mediate their regulatory function through base-pairing with the ribosome binding site of targeted mRNAs, resulting in translational inhibition. Their expression is induced under iron-starvation conditions while otherwise repressed by Fur, and leads to the repression of a large number of genes notably involved in iron storage or aerobic and anaerobic metabolism¹⁴³.

60

been suggested that sRNA conservation across different species is limited, with many being strain-specific¹⁴⁴.

Two examples of adaptation mechanisms characteristic of *P. aeruginosa*

Following the broad overview of regulatory mechanisms reported above, I focus here on two hallmarks of *P. aeruginosa* environmental adaptability.

- Alginate Production

P. aeruginosa is notably responsible for chronic pulmonary infections in cystic fibrosis (CF) patients. The microbiology of CF lungs has received an extensive attention and it is well established that one important mechanism used by *P. aeruginosa* to thrive this complex environment is to overproduce alginate exopolysaccharide¹⁴⁵. Consequently, CF-associated isolates form typical mucoid colonies. Molecular mechanisms regulating the production of alginate have been described and mutations responsible for the conversion towards mucus-producing variants have been identified¹⁴⁶. Overall, the major operon involved in alginate biosynthesis (*alg* operon) comprises 12 genes and is primarily transcribed from a single promoter (P_{algD}). This promoter is dependent on the alternative ECF sigma factor AlgU (also known as AlgT or σ^{22}) which is more widely involved in cell wall-associated stress response. Alginate biosynthesis (and by extension, AlgU-mediated stress response) is controlled by a complex regulatory network (Figure 12). Briefly, AlgU is normally sequestered at the membrane by the anti-sigma factor MucA. Upon stress conditions (e.g. cell wall stress), an intra-membrane proteolytic cascade results in MucA cleavage and AlgU is released and consequently free to bind to the RNAP core and activate the expression of a large variety of genes. AlgU-regulon contains about 300 genes globally involved in adaptation and protection such as osmotic shock proteins, or other regulators such as AlgR that negatively regulates QS or Vfr¹⁴⁷.

Conversion to mucoidy is most commonly due to loss-of-function mutations in *mucA*. For instance, the frequent clinical allele *mucA22* encodes a C-terminal truncated protein which leads to AlgU derepression and thus AlgU-regulated genes are constitutively expressed (including *alg* operon)¹⁴⁸. This alginate barrier confers many advantages to the microbial community such as an increased ability to make biofilms, an increased resistance to phagocytosis and protection against several antibiotics.

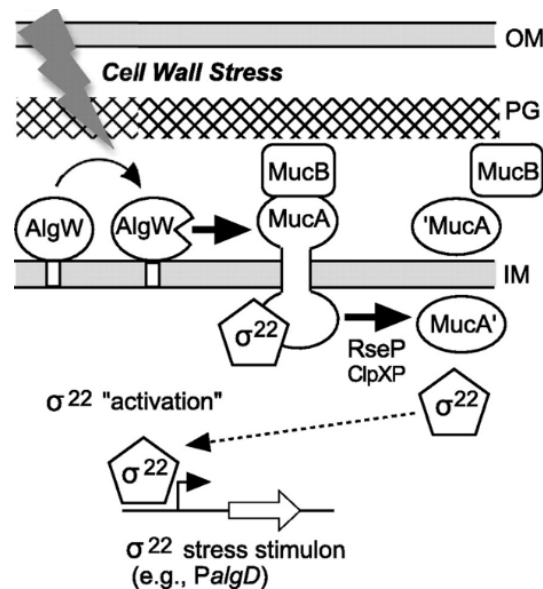


Figure 12: Model of activation of AlgU (σ^{22}) regulon

As reported by Wood and Ohman¹⁴⁷

Model for the “activation” of AlgU activity by regulated intramembrane proteolytic (RIP) degradation of the anti-sigma factor MucA. Under nonstress conditions, AlgU activity is low due to its sequestration at the inner membrane by the MucA-MucB complex. In response to cell wall stress, AlgW cleaves MucA in the C terminus and then RseP and ClpXP cleave the MucA N-terminal region, releasing σ^{22} from posttranslational repression. This allows σ^{22} to complex with core RNAP, thus increasing transcription of target promoters in the σ^{22} stimulon, which includes *PalgD* of the alginate biosynthetic operon. OM, outer membrane; PG, peptidoglycan; IM, inner membrane.

- Iron uptake

Iron is an essential metal for virtually all organisms, as a crucial co-factor of many enzymes involved in several key metabolic processes. However, the poor solubility of iron ions and their reactivity (leading to the generation of harmful reactive oxygen species) require the development of elaborate high-affinity and tightly controlled iron uptake systems. *P. aeruginosa* has evolved efficient strategies for iron acquisition and to control iron homeostasis¹⁴⁹.

Notably, it produces two siderophores, pyoverdine and pyochelin, that bind extracellular iron and transport them back into the cell through TonB-dependent receptors at the surface on the cell. Within the cytoplasm, the concentration of potentially harmful iron ions is controlled by the repressor Fur, which modulates expression of targeted genes (e.g. receptors, binding proteins and transporters for iron chelating molecules) and other regulators such as pyochelin uptake regulator PchR, ECF iron-starvation sigma factors (note that 12 out of 15 ECF sigma factors are involved in iron uptake regulation, see Table 4) or small regulatory RNAs. One major ECF sigma factor in iron-dependent regulation is PvdS, which triggers the production of pyoverdine and its cognate receptor under iron-limited conditions, along with extracellular virulence factors (i.e. protease and exotoxin A)^{133,149}.

In addition to siderophores, *P. aeruginosa* also uptakes iron by capturing hemes contained in molecules produced by other organisms (e.g. xenosiderophores, host iron proteins: hemoglobin, lactoferrin).

Therefore, iron is at the center of a complex network which entails (i) multiple interconnected systems regulating its homeostasis but also (ii) many pathways that are regulated by iron, such as QS or biofilm formation.

In conclusion, *P. aeruginosa* cells are able to pragmatically sense their environment and accordingly modulate their physiology through interconnected regulatory networks, thereby providing them with great versatility.

3. Diversity and genome plasticity

Core genome, accessory genome and pangenome

Like in any other bacteria, genomes of individual *P. aeruginosa* comprise a core genome associated to accessory genetic material. The **core genome** is defined as the set of genes that is present in virtually all *P. aeruginosa* strains. It constitutes about 90% of individual genomes, is highly conserved across strains and encodes a remarkably high genetic and functional diversity (metabolic and virulence factors, as illustrated above)¹⁵⁰.

In contrast, **accessory genome** contains genes that are not present in all strains and appear to be clustered in specific loci, called “regions of genomic plasticity” (RGPs). Accessory genes are usually acquired by horizontal gene transfer and tRNA genes are known to be hot spots for their insertion.

The accessory genome largely contributes to *P. aeruginosa* intra- and inter-clonal diversity as illustrated by the composition of the **pangenome** of 181 *P. aeruginosa* strains² recently analysed by Mosquera-Rendón and co-workers¹⁵¹. Indeed, among the 16 820 non-redundant genes constituting this pangenome, 2503 (15%) define the core genome. Among the remaining 14 317 accessory genes, 5209 (31%) were found to be uniquely present, *i.e.* present in only one of the 181 strains¹⁵¹. Consequently, the accessory genome is a primary contributor to *P. aeruginosa* genome evolution and confers specific phenotypes that possibly grant selective advantages (e.g. new catabolic pathways, virulence factors, antibiotic resistance). Depending on the strain, this accessory genome is more or less prominent. For instance, PAO1 genome display inserts smaller than 14 kb while the LESB58 “Liverpool epidemic strain” (characterized by its high transmissibility among CF patients and its high virulence) contains five genomic islands and five inserted prophages of 14 to 111 kb¹⁵².

Different types of accessory genetic elements and role of HGT

Accessory genetic elements integrated in *P. aeruginosa* genomes roughly fall into four main categories: (i) prophages and phage-like elements, (ii) integrative and conjugative elements, (iii) transposons, insertion sequences and integrons and (iv) replacement islands. As suggested by this classification, the majority of accessory genetic material originates from foreign mobile DNA acquired by the strain at some point in its life history. Prophages were already discussed, so I will focus on the last three elements.

² The so called “pangenome” is the result of the sum of all gene features found in each of the 181 sequenced strains, *i.e.* the core genome common to all strains + accessory genome of each strain.

Integrative and conjugative elements are self-transmissible genetic elements that can transiently exist as plasmid-like elements, transferred by self-mediated conjugation and usually become integrated into the bacterial chromosome (commonly adjacent to tRNA genes) through site-specific recombination. In addition to genes encoding the apparatus necessary to their dissemination (excision, self-transfer, reintegration), they harbour so-called “cargo ORFs” that can confer diverse phenotypes¹⁵⁰.

Transposable elements encode transposase complexes mediating their self-translocation to different sites on a given DNA molecule. In addition to transposase-coding gene, they can carry additional genetic material that may encode supplementary functions (and are then categorized as “transposons”). Some have been found to contain large and complex structures called integrons. An integron is a structure encoding an integrase which is able to capture exogenous genes and integrate them at a recombination site, located downstream of the integron promoter, thereby ensuring expression of the captured gene. Antibiotic resistance cassettes are frequently found within integrons¹⁵⁰.

Replacement islands represent a peculiar category of accessory genetics elements as they are actually present in nearly all *P. aeruginosa* genomes. However, these islands contain horizontally acquired components and are highly divergent from one strain to another. This term was originally coined to describe O-antigen biosynthesis, pyoverdine, pilin modification and flagellin glycosylation gene clusters. For instance, O antigen biosynthesis cluster is highly divergent across strains, consistently with the variation in structure and length of O-antigen among *P. aeruginosa* strains (about 20 different O antigen serotypes). The G+C contents of some of these biosynthesis genes are significantly lower than that of core genome, witnessing an acquisition through HGT¹⁵⁰.

Besides HGT, which mostly contributes to the evolution of accessory genomes, several other processes shape *P. aeruginosa* genomes. In addition to strain-specific mutations, genetic rearrangements and genetic duplication, it has also been observed, for example, that the adaptation of strains to the lungs of CF patients during chronic infection can be associated to large chromosomal deletions, as observed in pyomelanin-producing mutants^{153,154}.

In this chapter, I have summarized how *P. aeruginosa* can adapt to variable environments through the use of a large panel of genetic and metabolic resources coupled to an integrative regulatory network of gene expression. Altogether, these mechanisms make this bacterium a successful human opportunistic pathogen. In addition, misuse and overuse of antibiotics have favoured the emergence and spread of strains resistant to all currently available antibiotics. For these reasons, alternative therapeutic solutions have been investigated for the past 20 years. In particular, our research team has investigated the potential use of phages to treat infections caused by this versatile pathogen.

CHAPTER 5 - MODEL OF STUDY AND RESEARCH GOALS

In this final chapter, I will briefly describe previous results obtained in our research group that led to the formulation of the biological questions addressed during this PhD work.

1. Isolation of new phages with therapeutic potential

Originally, the research project developed in Laurent Debarbieux's lab aimed at assessing the therapeutic potential of phages to treat pulmonary infections due to *P. aeruginosa*.

Consequently, in 2006, five phages (namely PAK_P1 to PAK_P5, hereafter referred as PAK_Px) were isolated from sewage water using the PAK strain of *P. aeruginosa* as a host. They all displayed distinct restriction fragment length patterns while electron microscopy visualization revealed no significant morphological differences and allowed their classification within the *Myoviridae* family. Genome sequencing predicted their strictly lytic nature (no integrase or recombinase are detected among their coding sequences). BLASTP searches with the sequence encoding their identified major capsid protein (through mass spectrometry analysis) against the non-redundant protein database of the NCBI revealed distant similarities with the major capsid protein of FelixO1 *Salmonella* phage. Therefore, they represented a new group of *Pseudomonas* phages.

Their curative potential was then assessed in a mouse model of pulmonary infection. Briefly, mice were infected with a bioluminescent derivative PAK strain, allowing the monitoring of the progression of the infection in real time, and then treated with phages via intranasal instillation (Figure 13). Experiments showed that phages were able to rescue infected mice when administered 2h post infection using a ratio phage-bacteria (also called multiplicity of infection MOI) of 10:1. Phage treatments resulted in a large reduction of bacterial load along with a reduction of the inflammatory response in the lungs. By comparison with other phages infecting *P. aeruginosa*, *i.e.* the myovirus LBL3 and the podoviruses PhiKZ and LUZ19, PAK_Px phages were proven to have the best *in vivo* curative efficacy.

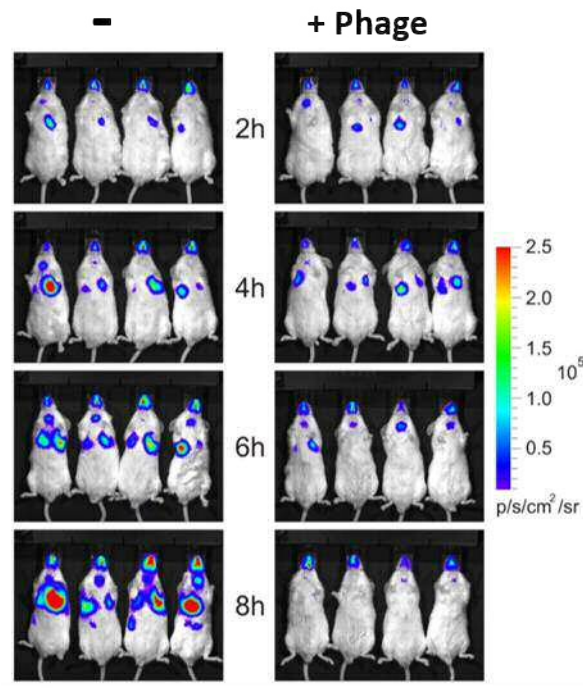


Figure 13: Phage treatment of pulmonary infection

Extracted from Ref 155. Example of time-course images of mice infected with bioluminescent *Pseudomonas aeruginosa* and treated with PBS (left) or treated with the PAK_P1 phage at a MOI of 10:1 (right).

Related publications: Debarbieux *et al.* (2010)¹⁵⁵, Morello *et al.* (2011)¹⁵⁶, Henry *et al.* (2013)¹⁵⁷.

2. Definition of two new phage genera

Although their morphologies are nearly identical, a comparative genomic analysis of PAK_Px phages supported their distribution into two distinct subgroups; one comprising PAK_P1, PAK_P2 and PAK_P4, the other comprising PAK_P3 and PAK_P5. They differ by their mean genome length, their number of predicted ORFs and tRNA, and their GC content, but they also share homologous proteins (about 25%), most of which encode predicted structural components¹⁵⁸. Moreover, the two subgroups display a strong synteny and their genomes are typically composed of strongly conserved regions (0.1% single nucleotide polymorphism (SNP) frequency) alternating with heterogeneous regions (up to 20% SNP frequency, scars of HGT and recombination events)¹⁵⁹. In 2013, by using a full genome sequence search, we identified additional *P. aeruginosa* phages related to either PAK_P1 subgroup or PAK_P3 subgroup along with more distantly related phages (including Felix01). With this set of related genomes, we proposed a reconstruction of their phylogeny based on four conserved proteins and revealed that they likely share a relatively recent common ancestor (Figure 14)¹⁵⁸.

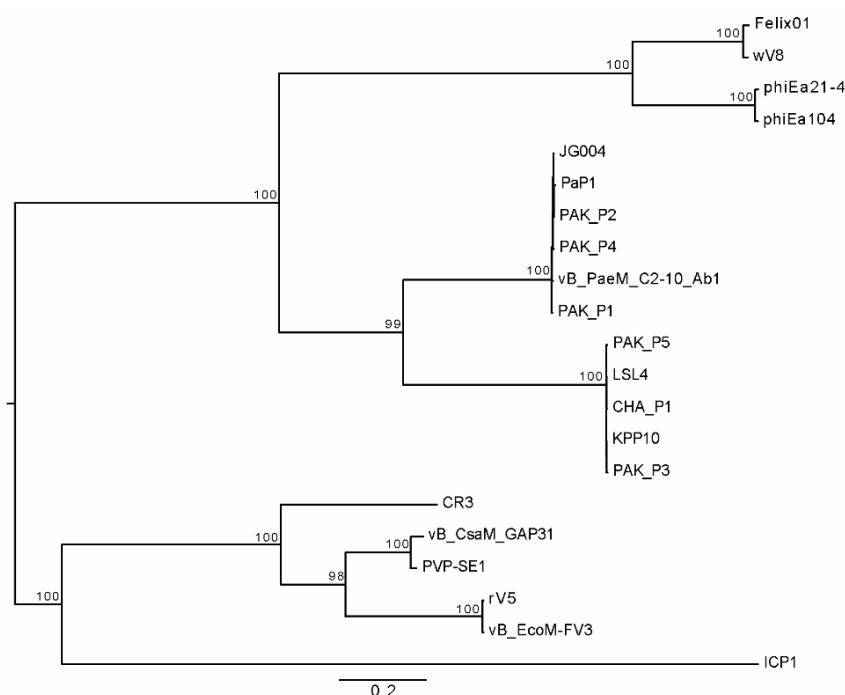


Figure 14: PAK_Px phylogenetic tree

Extracted from Ref 158. The maximum likelihood tree was built from a concatenated alignment of three core proteins (predicted to encode the portal protein, the major capsid protein and the tail sheath protein) common to the 21 bacteriophages. Bootstrap values are indicated and the tree was rooted on the midpoint root.

Overall, genomic contents of the two groups were divergent enough to define two new genera among *Myoviridae*, namely *Kpp10virus* (containing PAK_P3 and PAK_P5) and *Pakpunavirus* (containing PAK_P1, PAK_P2 and PAK_P4)^{158,160}.

At the time of writing (January 2017), we identified, through megablast search, 17 phages belonging to *Pakpunavirus* genus and 11 phages belonging to *Kpp10virus* genus, all of them infecting *P. aeruginosa* and isolated worldwide (e.g. China, Ivory Coast, Germany, Japan, *etc.*). This indicates that phages belonging to these genera are likely to be abundant and ubiquitous.

3. Research goals, scientific questions and manuscript organisation

“The virocell concept should encourage more scientists to go back to the bench and complement shotgun metagenomics and ecological studies with wet studies focusing on interaction between viruses and cells (not only both the entry and exit steps of virions, but also the intracellular stage of the viral life cycle). Only bench work will make sense of the ecological data by progressively revealing the ‘unknown’ of viral information. Besides architectural and replication proteins, viromes are full of genes encoding proteins whose function probably is to regulate the interactions between viruses and their hosts (victims) (...) How many other amazing viral machineries are hidden in the jungle of viral ORFans? Answering this question will require the identification and isolation in the coming years of much more viral or cellular systems from the three domains of life.” - Forterre (2011)¹⁰

This PhD work has been motivated by the very reasons summarized in the above citation extracted from Forterre’s publication defining the *virocell* concept. More specifically, the main objective of this research is to provide a general and mechanistic comprehension of infection cycles of therapeutically valuable phages, and more largely, of new genera of abundant phages. An extensive molecular characterization, on top of representing a further step towards their potential acceptance as therapeutic agents, could lead to the discovery of novel mechanisms of phage-bacteria intracellular interactions.

We therefore investigated the main steps of viral lifecycle with an emphasis on the intracellular stage, especially on the transcriptional dynamic of the *virocell* over the course of infection. Notably, we intended to draw a global scheme of infection cycles by addressing the following questions:

- What is the timing of the infection process and how is phage gene expression regulated?
- What is the effect of phage infection on host physiology?
- I also explored the “jungle of viral ORFans” by trying to elucidate the functions of phage early proteins, in order to have insights on the host reprogramming strategy set up by these phages.

This study is based on two phages, PAK_P3 and PAK_P4, as representatives of the above described genera *Kpp10virus* and *Pakpunavirus*, respectively.

Chapter 1 of the results section provides a classical characterization of the two phages by assessing their genomic content, lifecycle parameters, host ranges and host receptor. The next chapters follow the progression of the infection cycle and are mostly based on data obtained from RNA-Sequencing technology. First, phage transcriptional programs are deciphered in **Chapter 2**.

Chapter 3 focuses on the early stage of infection and aims to understand the function of one phage early protein, namely Gp92, and relies on a deep molecular investigation.

Then, the impact of phage infection on the global host physiology and metabolism is assessed within **Chapter 4**. Finally, **Chapter 5** is dedicated to the characterization of the host transcriptional response to phage invasion.

All along, differences and similarities between the two model phages will be highlighted. For clarity purposes, specific findings will be discussed within the result section.

“Conclusions and perspectives” section is dedicated to an integrative overview of all results obtained in this three-level study and provides a model of transformation of *P. aeruginosa ribocell* into a *virocell*. Finally, conducting this work on two phages from two genera gave us the opportunity to question the conservation of viral strategies during phage evolution.

PART 2

-

RESULTS AND DISCUSSION

CHAPTER 1 - GENERAL DESCRIPTION OF PHAGES PAK_P3 AND PAK_P4

As a prerequisite to investigating the intracellular steps of phage infection cycles, we first intended to precisely characterize PAK_P3 and PAK_P4 properties respectively belonging to the newly defined genera *Kpp10virus* and *Pakpunavirus*¹⁵⁸. Therefore, this chapter aims at presenting a description of their genomic content, growth parameters, host ranges and bacterial receptor. All along, we will highlight their similarities and differences.

1. Comparative genomic: PAK_P3 and PAK_P4 share common life-history traits

Definition of core genes

When comparing PAK_P3 and PAK_P4 genomes, about 13% of their nucleotide sequences can be confidently aligned, with an average identity of 76% ($\pm 9\%$). More precisely, 9% of their sequences are strictly identical (Figure 15). Sixty-nine homologous coding sequences, hereafter referred as core genes, have been defined (*i.e.* approximately 40% of their proteomic content, according to a CoreGenes analysis with default parameters¹⁶¹). The sequence identity for each pair of homologous predicted proteins ranges from 24% to 85% (with a mean of $49 \pm 14\%$). Most of them belong to the structural (49%) and nucleic acid metabolism (32%) modules, while 19% are encoded by early-expressed ORFs (defined in Chapter 2, paragraph 2.1.) which are acknowledged to be the least conserved phage proteins, often responsible for host metabolism subversion³⁴ (Figure 15).

Among coding sequences that do not meet the criteria required to be considered as homologous (*i.e.* blastp score < 75), some appear to have the same predicted functions. For example, protein sequences of putative ribonucleoside-reductase beta subunits (PAK_P3_gp069 and PAK_P4_gp078) share only ~30% of similarity (~17% of identity).

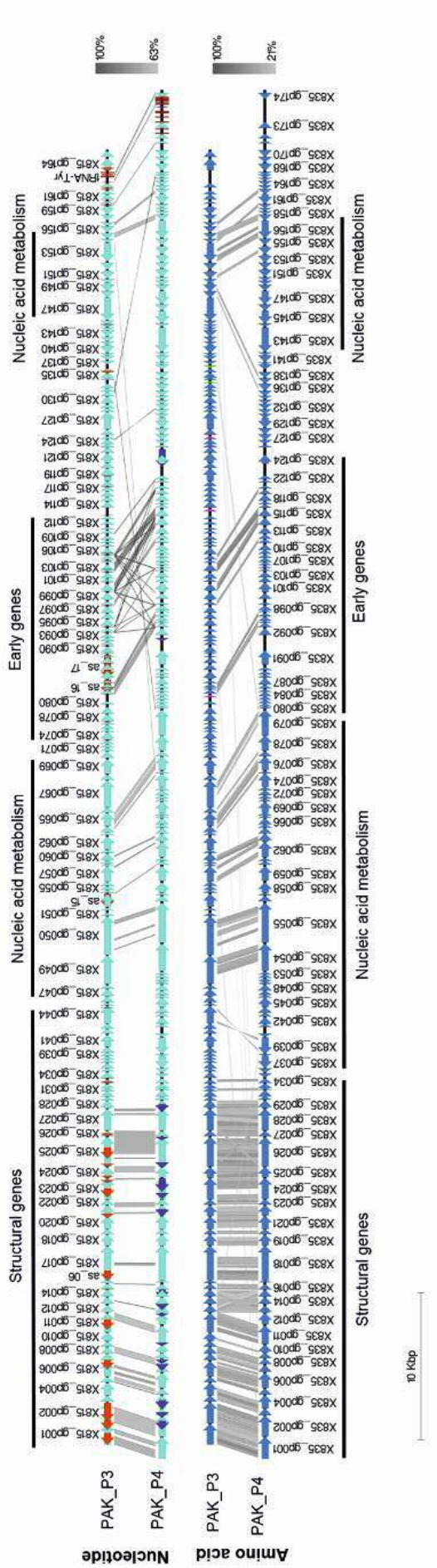


Figure 15: Genomic comparison of phages PAK_P3 and PAK_P4
Upper panel: nucleotide comparison. ORFs are depicted as cyan arrows, antisense RNAs (defined in Chapter 2) are represented as red and dark blue arrows, and tRNAs appear as red squares.
Lower panel: translated nucleotide comparison. CDS are colored in blue.

Predicted functions

Most of the predicted ORFs (~80%) of these two phages could not be associated with a putative function. Amongst those, we can distinguish (i) ORFs having relatives in various other phage genomes, (ii) ORFs similar to predicted gene features of closely related phages (*i.e.* in *Kpp10virus* or *Pakpunavirus* genera) (~50%), and (iii) ORFs with no similarity to any other sequence in public databases, also called ORFans (~5%).

In a large majority, predicted genes with a putative function belong to structural and nucleic acid metabolism-related clusters (and, by extension, to the core gene set). Most of the **structural components** that have been predicted through homology searches (e.g. major capsid protein, tape-measure protein, tail and baseplate proteins) were confirmed with a mass spectrometry analysis of PAK_P3 virions (Figure 16, Table 5). Among the 25 identified proteins, 24 were encoded within the genomic structural cluster while one (Gp72) is located in the vicinity of the early genes cluster.

Table 5: Proteomic analysis of PAK_P3 virions

Proteins identified by mass spectrometry analysis (ESI-MS/MS). ^atheoretical molecular weight; ^bnumber of unique peptides identified; ^ctotal number of spectra.

gp	Predicted function	^a Mw (kDa)	Max sequence coverage (%)	^b Np	^c Ns	Ns/Mw
5	putative decoration protein	13,5	76,6	10	265	19,6
6	major capsid protein	40,3	61,3	20	527	13,1
166	putative head fiber protein	13,9	67,4	6	101	7,3
12	major tail protein	18,8	52,6	6	107	5,7
7	unknown	18,7	34,8	6	91	4,9
11	tail sheat	48,0	35	14	220	4,6
27	tail fiber protein	52,9	46,5	14	229	4,3
2	portal protein	56,3	39,6	21	199	3,5
14	unknown	18,8	36,5	5	42	2,2
23	baseplate related protein	53,4	35	12	85	1,6
25	tail fiber protein	70,2	44	18	88	1,3
4	unknown	31,9	29,8	9	34	1,1
20	tail protein	32,6	44	9	33	1,0
9	head tail adaptator	14,3	41,5	3	11	0,8
19	unknown	14,7	26,4	3	10	0,7
21	baseplate assembly protein	27,2	23,8	3	16	0,6
17	tape measure protein	84,9	24	16	40	0,5
24	unknown	28,2	16,1	3	13	0,5
22	putative tail lysozyme	14,1	20,7	2	6	0,4
18	unknown	28,0	9	2	10	0,4
160	unknown	20,8	18,1	3	7	0,3
13	unknown	18,3	41,3	5	6	0,3
72	unknown	8,1	30,3	2	2	0,2
10	unknown	23,4	20,9	3	5	0,2
8	unknown	17,6	22,6	2	2	0,1

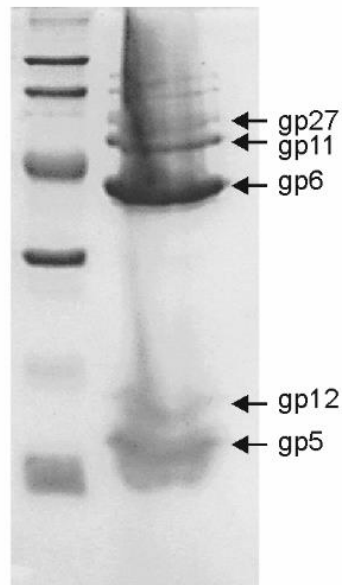


Figure 16: Proteomic analysis of PAK_P3 virions

Denaturing polyacrylamide gel of PAK_P3 virion proteins.

Regarding the **nucleic acid metabolism-related cluster**, bioinformatics predictions revealed that PAK_P3 and PAK_P4 encode their own DNA polymerase. Consequently they are not expected to entirely rely on host DNA polymerase for replication. They are also predicted to encode DNA and RNA ligases, primase and helicase thus supporting further the hypothesis of a host-independent DNA replication. Conversely, they do not seem to encode their own transcriptional apparatus.

Additional potential nucleotide metabolism-related enzymes have been detected by homology searches (e.g. nicotinamide phosphoribosyl transferase, dCMP deaminase, thymidylate synthase, ribonucleotide-reductase and deoxyribonuclease). This indicates that these phages may carry **auxiliary metabolic genes (AMGs)**¹⁶², *i.e.* genes that are not essential, but likely beneficial for replication.

A putative **lysis cassette** has been identified in both genomes. Indeed, PAK_P4_gp29 and its homolog PAK_P3_gp28 are similar to putative endolysins annotated in other phages. Moreover, downstream ORFs are predicted to have transmembrane domains and thus represent holins and/or spanins candidates. Consequently, we hypothesize that PAK_P4 and PAK_P3 use an endolysin/holin-based mechanism of lysis.

2. Lifecycle parameters

To delineate the timing of the main steps of PAK_P3 and PAK_P4 infection cycles, we determined their growth parameters.

PAK_P3 and PAK_P4 have similar adsorption abilities since $\geq 90\%$ of the virions adsorbed on the host strain PAK within 4.6 (± 0.7) min ($k_a = 2.2 \cdot 10^{-9}$ ($\pm 5.1 \cdot 10^{-10}$) mL.min⁻¹) and 4.8 (± 1.7) min ($k_a = 2.4 \cdot 10^{-9}$ ($\pm 1.4 \cdot 10^{-9}$) mL.min⁻¹), respectively (Figure 17).

Standard one-step growth experiments showed that the first functional new virions are rapidly assembled (eclipse periods: 12.3 (± 0.4) min and 13.3 (± 2.0) min, respectively) and almost immediately released in the case of PAK_P3 (latency period: 13 (± 2.1) min) while the process is more delayed in PAK_P4 infected cells (latency period: 18.2 (± 2.9) min). PAK_P3 and PAK_P4-pregnant cells eventually produce an average of 53 (± 21) or 13 (± 5) progeny virions, respectively, within mean infection cycle durations as short as 18 (± 0.6) min for PAK_P3 and 21.4 (± 1.8) min for PAK_P4 (Figure 17).

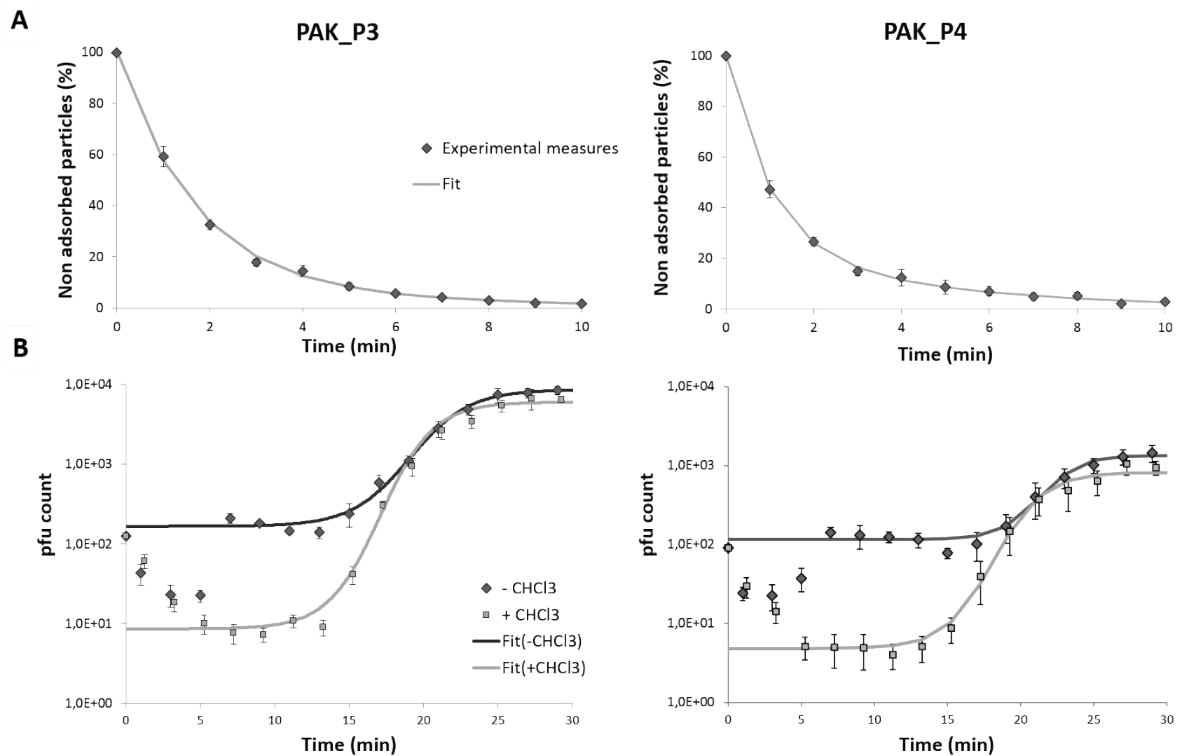


Figure 17: PAK_P3 and PAK_P4 rapidly adsorb to their host and efficiently produce new progenies
(A) Adsorption assays of PAK_P3 and PAK_P4 on *P. aeruginosa* strain PAK. **(B)** One-step growth curve of PAK_P3. Samples treated with (grey squares) or without (black diamonds) CHCl₃. A logistic regression was used to fit the data. Four independent experiments were combined and data are presented as means with standard deviations.

3. Host range

Establishing the range of bacteria one phage can infect is a preliminary step of phage characterization. To evaluate the breadth of PAK_P3 and PAK_P4 host ranges, we assessed their ability to lyse a collection of 142 *P. aeruginosa* clinical strains (obtained from S. Häussler, Helmholtz centre, Germany).

Evaluation of strain collection diversity

The *P. aeruginosa* isolates were collected from diverse infection sites and various countries across Europe. We first assessed the diversity of the strain collection in terms of phage sensitivity (phage typing). We used a set of 15 phages available in our laboratory, isolated on different strains (PAK, PAO1, cystic fibrosis isolates RP73 and CHA), at different times and in distinct environments by several research groups.

Eighty percent of the strains were infected by at least one phage out of the 15 tested. In other words, only 28 out of 142 strains were completely resistant to the phage. We observed a continuum of phage sensitivity phenotypes: ranging from highly sensitive (e.g. isolate Psae0613 is infected by 14 out of 15 phages) to strictly resistant, including strains with intermediate sensitivity (Figure 18). Isolates lysed by two phages are more represented (~20%). We found no correlation between phage sensitivity phenotype and infection site (e.g. respiratory tract or urine tract). Since it is composed of such a diversity of isolates, this collection is well adapted to determine and compare PAK_P3 and PAK_P4 host ranges.

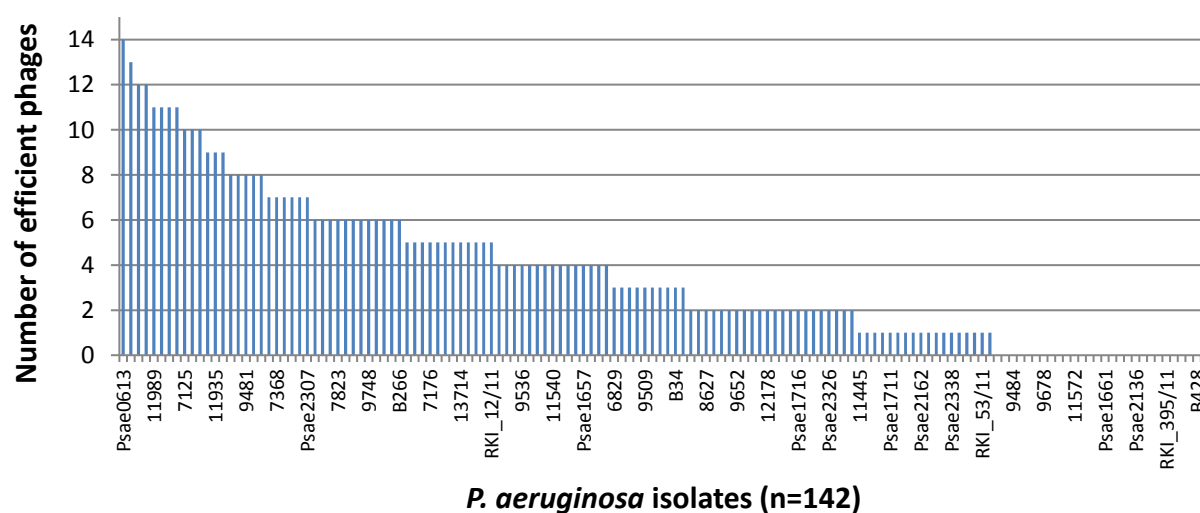


Figure 18: Number of phages leading to a productive infection on a given isolate

The number of phages able to produce plaques on each of the 142 tested isolates have been counted. Isolates have been sorted from the most permissive (left) to the full resistant (right).

PAK_P3 and PAK_P4 have similar host ranges

PAK_P3 and PAK_P4 are able to produce plaques on 26.5 and 32.7% of the isolates, respectively. Among the 53 strains that are sensitive to at least one of the two phages, more than half (56.6%) can be infected by both phages (common host range), 32% are lysed by PAK_P4 only and 13% are specifically targeted by PAK_P3 (Figure 19). No correlation between host ranges and origins of isolates could be drawn.

Therefore, the large overlap of host range, which is first supported by highly homologous structural proteins, may also be supported by similar subversion and replication mechanisms. The proportion of strains specifically infected by each phage suggests that both viruses developed specific characteristics (e.g. PAK_P4 may have evolved toward broadening its host range while PAK_P3 favoured a larger burst size), questioning whether there is a relation between an increased host range and a decreased burst size.

Strain	Origin	City	CF	Bacteriophage	
				PAK_P3	PAK_P4
PAK			no		
CHA	Reference strains		yes		
PAO1			no		
6887	tracheal secrete	Hannover	no		
7032	venous catheter	Hannover	no		
7125	tonsil swab	Hannover	no		
7135	bronchial secrete	Hannover	no		
9157	wound swab abdomen	Hannover	no		
11148	tonsil swab	Hannover	no		
13714	permanent catheter urine	Hannover	no		
14322	bronchial rinsing	Hannover	no		
RKI_12/11	nd	nd	nd		
B214	nd	nd	nd		
B271	nd	nd	nd		
B266	nd	nd	nd		
8044	tracheal secrete	Hannover	yes		
9229	tonsil swab	Hannover	no		
9481	bronchial rinsing	Hannover	no		
11935	bronchoalveolar lavage	Hannover	no		
11989	tonsil swab	Hannover	yes		
12269	sputum	Hannover	yes		
13224	bronchial rinsing	Hannover	no		
Psae0613	nd	nd	nd		
Psae1715	respiratory tract	Freiburg	nd		
Psae2305	respiratory tract	Sassari	nd		
RKI_100/12	nd	nd	nd		
7055	bronchoalveolar lavage	Hannover	no		
12274	bronchoalveolar lavage	Hannover	no		
6827	midstream urine	Hannover	no		
6870	midstream urine	Hannover	no		
6964	wound swab abdomen	Hannover	no		
7084	permanent catheter urine	Hannover	no		
7091	lung transplant recipient	Hannover	yes		
7176	midstream urine	Hannover	no		
7261	permanent catheter urine	Hannover	no		
7368	nasal swab	Hannover	no		
7823	tonsil swab	Hannover	no		
7863	bronchial secrete	Hannover	no		
8349	midstream urine	Hannover	no		
8478	midstream urine	Hannover	no		
8481	bronchial secrete	Hannover	yes		
8613	ear swab	Hannover	no		
9748	tonsil swab	Hannover	no		
10047	tonsil swab	Hannover	no		
11572	midstream urine	Hannover	no		
Psae1640	urine	Muenchen	nd		
Psae1758	respiratory tract	Limburg	nd		
Psae1766	respiratory tract	Bremen	nd		
Psae2307	respiratory tract	Sassari	nd		
7444	bronchoalveolar lavage	Hannover	yes		
11540	midstream urine	Hannover	no		
12207	bronchoalveolar lavage	Hannover	yes		
14103	swab heel	Hannover	no		
Psae1655	respiratory tract	Muenchen	nd		
Psae2319	nd	Palermo	nd		
RKI_360/11	nd	nd	nd		

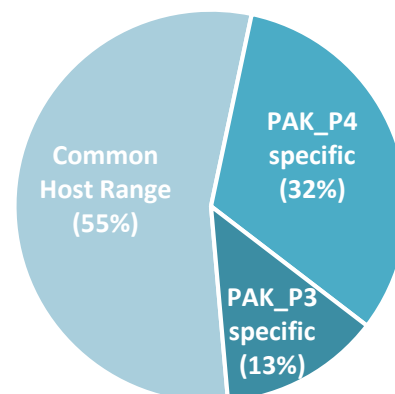


Figure 19: Extract of PAK_P3 and PAK_P4 host range

PAK_P3 and PAK_P4 host ranges were assessed by determining their efficiency of plating (EOP)³ on 142 clinical isolates. Only sensitive strains (53) are represented. Left columns indicate isolate names, their association to cystic fibrosis and infection sites they were isolated from.

Green represents sensitive strains ((EOP) >0.1) and orange represents weakly sensitive strains (0<EOP<0.1) while grey represents resistant strains (no production of plaques). The chart pie provides a graphical representation of the common and specific host ranges of the two phages.

³ EOP is obtained by simultaneously titrating a given phage solution on a bacterial strain of interest X and on a reference strain Y (usually phage primary host). EOP is calculated as follows: $EOP = \text{titer strain X} / \text{titer strain Y}$

4. Phage receptor characterization

To identify PAK_P3 and PAK_P4 bacterial receptor, we screened a transposon (Tn) mutant library of *Pseudomonas* strain PAK (see Methods). Previous work based on a transposon mutant library screen done by another research group established that phage JG004, a close homolog to PAK_P4 (88% of their ORFs are homologous), uses LPS as a receptor¹⁶³. Other groups have later confirmed these findings, with an additional PAK_P4-like phage¹⁶⁴. Consequently, we mainly focused on identifying the receptor recognized by PAK_P3.

Estimation of spontaneous mutation frequencies leading to the emergence of phage resistant variants

Resistant clones were selected by infecting a liquid culture of mixed Tn mutants with PAK_P3 (the library was not ordered) and subsequently recovering surviving clones. Prior to this assay, we first set up appropriate experimental conditions to ensure that clones recovered upon the phage-bacteria incubation period are indeed resistant variants and not just uninfected cells. We subsequently estimated the range of spontaneous mutation rates yielding phage resistance.

Among several conditions tested, we established that infecting a liquid culture of strain PAK using a MOI of 5 for 1 to 3h at 37°C yielded the lowest amount of colonies. Deduced rates of spontaneous resistance were largely variable, ranging from 8.10^{-8} to 2.10^{-6} (mean of $9.3.10^{-7} \pm 6.0.10^{-7}$), consistent with Luria and Delbrück's so called fluctuation test¹⁶⁵. Interestingly, a significant proportion of such mutants displayed a red pigmentation (Figure 20). During the course of these experiments, another group published similar results with *P. aeruginosa* strain PA1/phage PaP1 and demonstrated that this red phenotype is due to a 220kb genomic fragment deletion including genes *hmgA* (leading to pyomelanin hyperproduction, responsible for red phenotype¹⁶⁶) and *galU* (responsible for phage resistance as *galU* mutant synthesizes LPS lacking the O-antigen)¹⁶⁷.

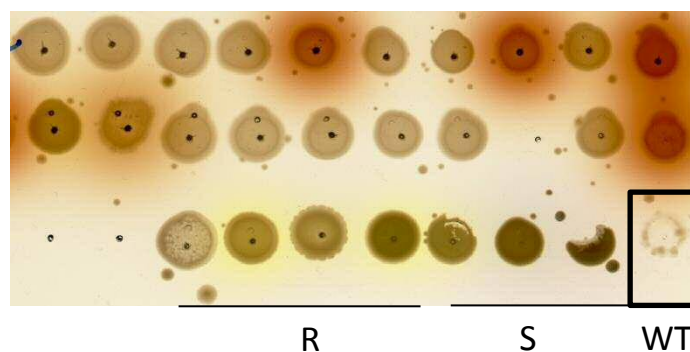


Figure 20: *P. aeruginosa* PAK strain spontaneous mutants resistant to PAK_P3 infection

Twenty spontaneous resistant clones selected in liquid infected culture were re-isolated and spotted on agar plate (top two rows). To confirm their resistant phenotype, a double-spot of PAK_P3 lysate was added. Control strains (resistant “R” and sensitive “S”) and PAK wild-type (WT) strain were tested in parallel (bottom row).

Identification of Tn insertion mutants resistant to phage infection

Using experimental conditions described above, 177 Tn insertion mutants resistant to PAK_P3 were isolated from 3 replicates. Briefly, infected cultures were plated after 1h30 incubation and surviving colonies were then re-isolated on fresh agar plates. A double-spot test was performed to confirm phage resistant phenotype (as described in Figure 20). Semi-random colony PCRs were performed on ‘true’ resistant clones and subsequently sequenced to map the insertion of the transposon (see Methods). This analysis revealed that a large majority (~75%) of the resistant clones displayed a Tn insertion within a single operon belonging to the LPS core oligosaccharide biosynthesis cluster: specifically in genes PAK_4981 (N-acetyl glucosaminyl-phosphatidyl-inositol deacetylase), PAK_4982 (Mig14 family protein) and PAK_4980 (glycosyltransferase) in 42, 24 and 12% of the cases, respectively. These results strongly suggest that PAK_P3 receptor is also the LPS (Table 6).

Table 6: Genes identified with a Tn insertion

ND: Not determined

Tn localisation	% Hits	Function	PAO1 homolog	Pathway
PAK_4981	42	GlcNAc-PI de N acetylase	<i>dnpA</i>	
PAK_4982	24	Mig14 family protein	<i>PA5003</i>	LPS biosynthesis
PAK_4980	12	Ssg protein or glycosyltransferase	<i>ssg</i>	
PAK_1515	5	UDP glucose deshydrogenase	<i>galU</i>	LPS biosynthesis, Nucleotide sugar metabolism
PAK_353	2	fimbrial Usher family protein	<i>cupC3</i>	Transporter activity, Fimbrial Biogenesis
PAK_472	2	bacterial flagellin N-terminal helical region family protein	<i>fliC</i>	Cell motility
PAK_537 - PAK_538 intergenic region	2	PAK_537: Amino acid permease	<i>PA1147</i>	Amino acid transmembrane transport
		PAK_538: NAD ⁺ -diphthamide ADP-ribosyltransferase activity	<i>toxA</i>	Exotoxin A (Inhibition of protein synthesis)
PAK_4463	1	<i>Neisseria</i> PilC family protein	<i>pilY1</i>	Pilus biosynthesis
PAK_4570	1	uracil-xanthine permease family protein	<i>upp - uraA</i>	pyrimidine metabolism (uracil)
mutL	1	DNA mismatch repair protein MutL	<i>mutL</i>	DNA mismatch repair
PAK_3030	1	FKBP-type peptidyl-prolyl cis-trans isomerase family protein	<i>PA3262</i>	Protein folding helper
PAK_2877 - PAK_2878 intergenic region	1	PAK_2877: Semialdehyde dehydrogenase, NAD binding domain protein	<i>PA3116</i>	Amino acid biosynthesis
		PAK_2878: type IV pilus	<i>fimV</i>	Cell motility
PAK_4978	1	O-Antigen ligase family protein	<i>waal</i>	LPS biosynthesis
PAK_2910 - PAK_2911 intergenic region	1	PAK_2910 = NAD-dependent dehydratase/UDP-glucose 4-epimerase	ND	Cell wall biosynthesis
		PAK_2911 = glycosyl transferases group 1 family protein	ND	
PAK_5023	1	shikimate kinase family protein	<i>aroK</i>	Aromatic amino acid biosynthesis, surrounding genes involved in fimbrial biosynthesis
PAK_1513	1	bacterial regulatory s, tetR family protein	<i>mexZ</i>	Transcriptional repressor of efflux system

Adsorption assays were performed on the 3 main Tn mutants affected in LPS biosynthesis and confirmed that phage resistance is due to a lack a surface recognition since PAK_P3 is not able to adsorb to these mutant cells anymore (Figure 21A).

To establish whether these three genes were individually responsible for the resistance phenotype (*i.e.* to rule out potential polar effect of the Tn insertion), scarless deletion mutants were generated¹⁶⁸. The sensitivity to the phage was restored in strain PAK Δ 4981, indicating that Tn insertion resulted in an affected expression of downstream genes (*i.e.* PAK_4980) (Figure 22). Although deletion of PAK_4982 (upstream PAK_4981) was not obtained, it is expected to provide similar results. Therefore, the only gene responsible for the resistant phenotype is the glycosyltransferase encoding gene PAK_4980. We also confirmed that PAK_P4 is not able to infect strain PAK Δ 4980. These results thus indicate that PAK_P3 and PAK_P4 both use LPS as a receptor.

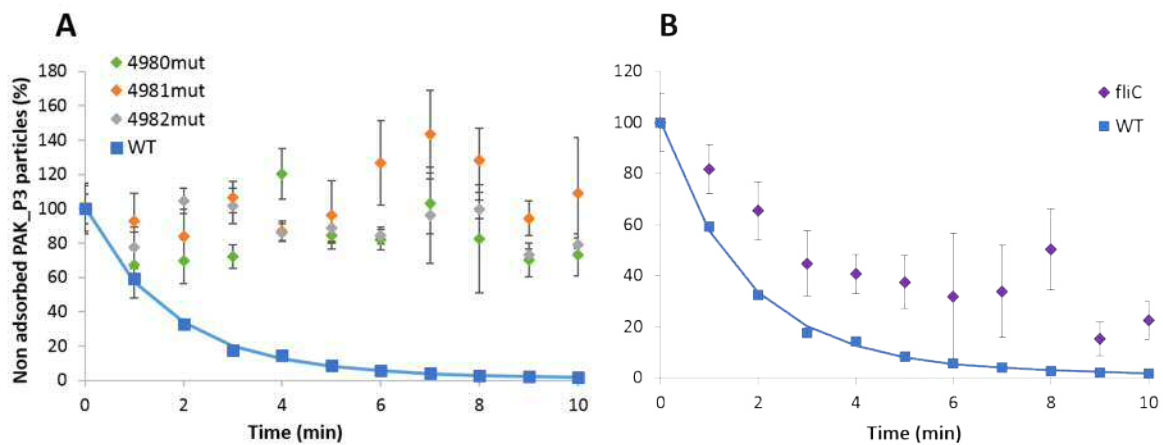


Figure 21: Adsorption assay of PAK_P3 on LPS- and flagellin-mutant strains

Adsorption assays of PAK_P3 on (A) the three strains displaying a Tn insertion within genes PAK_4980 (4980mut), PAK_4981 (4981mut), PAK_4982 (4982mut) and (B) on a deletion mutant of flagellin encoding *fliC* gene. In each case, the adsorption curve of PAK_P3 on the WT strain is displayed. Three independent experiments were combined and data are presented as means with standard deviations.

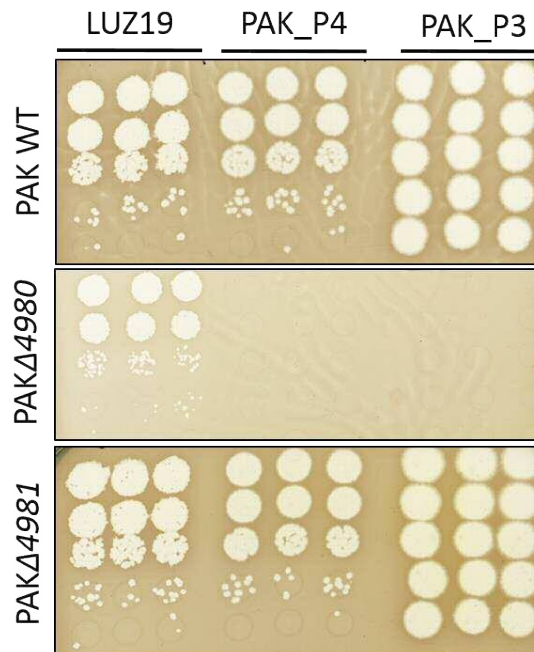


Figure 22: Strain PAKΔ4980 is resistant to phage infection

Ten-fold dilutions of phages PAK_P4, PAK_P3 and LUZ19 (recognizes type IV pilus as a receptor¹⁶⁹) were spotted on bacterial lawn of the indicated strains.

In addition to the three genes discussed above, other genes were identified in this screen. Overall, most of these additional genes are involved in the biosynthesis of cell surface structures (e.g. LPS, pili, flagellum), indicating that their expression is probably necessary in the process of surface recognition by the phage. Adsorption of PAK_P3 on strain PAKΔ*fliC* (in frame deletion of *PAK_472/fliC* gene, strain provided by E. Morello, Tours) was not abolished but less effective, suggesting that the phage might bind to the flagellum during adsorption process (Figure 21B).

Interestingly, other genes involved in metabolic processes such as amino acid biosynthesis, pyrimidine metabolism, DNA mismatch repair and protein folding have been identified. This might indicate that PAK_P3 requires these pathways to be functional to complete its infection cycle. However, given the low frequency of detection of these specific insertions, we cannot discriminate whether the clone is resistant because the gene carrying the Tn insertion is disrupted or because other spontaneous mutation(s) arose in distinct loci.

Within this first chapter we presented a general characterization and comparison of PAK_P3 and PAK_P4: (i) Analyses of their genomic properties allowed us to formulate hypotheses about the

infection mechanisms they use (replication, transcription, lysis), (ii) a precise idea of the timing of their cycles was given by the determination of their growth parameters, (iii) Investigation of their host ranges led us to propose that they may occupy distinct ecological niches, and (iv) bacterial genes necessary for the infection were identified. Assessing these basic properties was crucial to investigate intracellular phage lifecycles, as reported in the following chapters.

CHAPTER 2 - PHAGE TRANSCRIPTIONAL PROGRAMS

To study the dynamics of phage transcriptional programs, we first set up experimental conditions allowing 95% of the cell population to be infected in less than 5 min. The RNA content of the synchronously infected cell population was then extracted at two time points representative of the different steps of a single infection cycle (*i.e.* early and late), and subsequently studied by RNA-Sequencing.

Given the short eclipse period durations characterizing both phages, we focused on 3.5 min and 13 min time points as representative snapshots of the beginning (early) and the end (late) of one infection cycle at the transcription level. One additional sample was collected at 6.5 min to assess the presence of an identifiable middle phase of transcription.

1. Temporal regulation of phage gene expression

During the course of infection, early, middle and late transcripts of both phage genomes were identified. The early transcribed region encompasses genes gp74 through gp112 in PAK_P3 and genes encoding gp080 through gp124 in PAK_P4, all of which encode hypothetical proteins with low or no sequence similarity to any other amino acid sequence in databases (ORFans) (Figure 23). Transcripts produced at middle time point (not represented) correspond to two regions that each contains gene features related to nucleic acid metabolism.

As expected, the structural region appears to be mostly transcribed in late infection. In PAK_P4, the identified 'late region' encompasses genes encoding gp001 to gp034. Similarly in PAK_P3, this region comprises genes encoding gp001 to gp033 and gp049 to gp069. Strikingly, five ORFs (*i.e.* gp34, gp37, gp38, gp45 and gp46), although located within PAK_P3 structural region, are overexpressed early compared to late time point (Figure 23).

Intergenic transcription is observed throughout the genome, highlighting the great compaction of viral genomes where every single gene is expressed, conversely to bacterial genomes. This property is further illustrated by the large amount of antisense transcripts detected, as reported below.

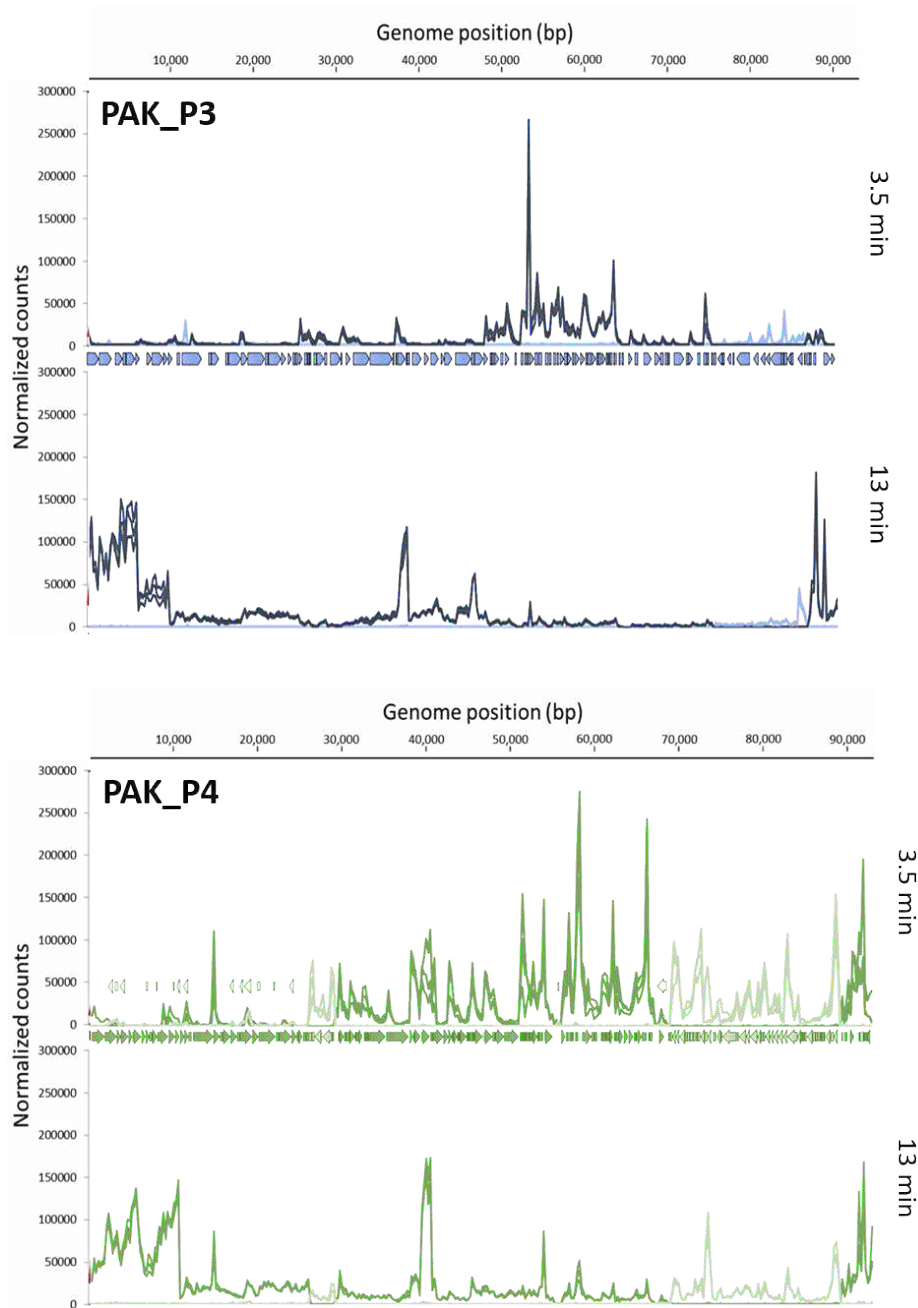


Figure 23: Phage transcription is temporally regulated

Upper panel: PAK_P3 transcription map. Lower panel: PAK_P4 transcription map.

Mapped reads were summarized into stranded count tables of Total Gene Reads that align to every 250bp of the phage genomes using the CLC Genomics Workbench. These read counts were then normalized against each other by the Total Count of reads that align to both phage and host genomes for each sample and then plotted. This allows us to show the relative abundance of phage transcripts over time in the context of the total transcript population. Dark and light coloured (blue or green) graphs represent reads mapping to the forward and reverse strands, respectively, for each replicate in each condition. Phage genomes are represented in between the graphs corresponding to the two time points, with arrows indicating defined coding sequences.

2. Production of non-coding RNAs

Antisense transcription of the structural region during early infection

Analysis of antisense transcripts (asRNAs) produced by PAK_P3 revealed 20 putative asRNAs, 8 of which are small asRNA (mean length 176 ± 30 bp) and 12 are longer than 300 bp. All but one are encoded within genes, suggesting they may act as *cis*-encoded antisense RNAs. These asRNAs are predominantly (15 out of 20) located in the structural genomic region and are significantly more strongly transcribed during early infection compared to late infection, with fold changes ranging between 2 and 17 (Figure 24). PAK_P4 also produces asRNAs that are encoded within structural genes for a large majority (15 out of 17). These asRNAs are also significantly more expressed (2 to 3-fold) 3.5 min after infection compared to late infection stage (13 min).

As it is well known for model phages, gene expression is usually temporally regulated through various mechanisms involving protein-protein or protein-DNA interactions. In particular, late gene expression is the result of a tight regulation driven by phage early proteins. Early expression of antisense RNAs could represent an additional regulation mechanism preventing transcriptional ‘leaks’ from strong promoters controlling expression of late structural genes. Under this hypothesis, the temporal distribution and the location of PAK_P3 and PAK_P4 asRNAs could be consistent with the shut-off of structural gene expression observed 3.5 min post infection. Although *cis*-antisense RNAs appeared to be a common form of regulation in bacterial genomes, with the detection of hundreds of antisense transcripts in multiple genomes these last years, they have not been extensively described in phage genomes. Until recently, the only few reported phage asRNAs (or regulatory RNA-RNA interactions) were exclusively found in lambdoid phages such as the regulatory *oop* RNA known for over 40 years (reviewed in Nejman-Falenczyk *et al.* (2015)¹⁷⁰) or, more recently, the anti-small RNA AgvB¹⁷¹. Here we report evidence of antisense transcription in a virulent phage along with a hypothesis on its function, strengthening the observation made by Wagemans and coworkers (2014) in N4-like *Pseudomonas* podovirus LIT1¹⁷². The use of antisense transcription as a regulatory mechanism by phages might therefore be a more common feature than previously thought.

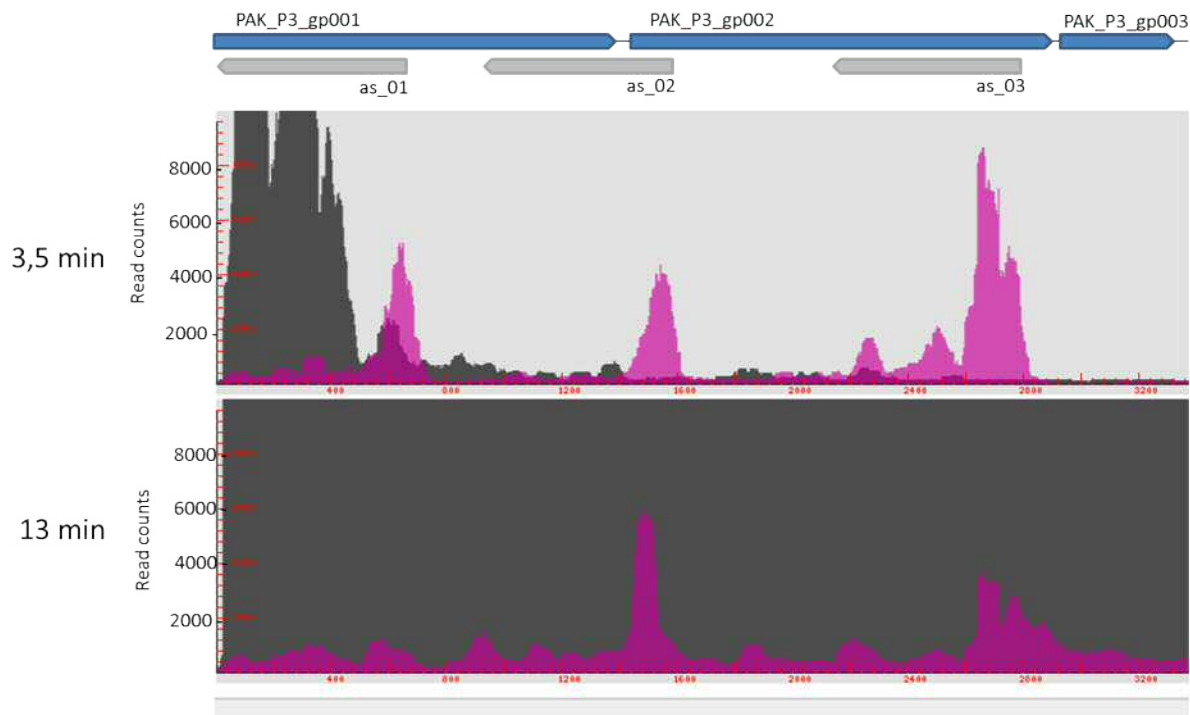


Figure 24: Example of PAK_P3 antisense transcripts

Mapped reads were formatted into graph files for visualization in a strand-specific manner (black and pink represent reads mapping the forward and the reverse strands, respectively) using COV2HTML visualization tool. The annotated *cis*-antisense RNAs genes are indicated as grey arrows and open-reading frames annotated in PAK_P3 are shown as blue arrows. Data obtained 3.5 min and 13 min following PAK_P3 infection are presented in upper and lower part, respectively.

PAK_P3 displays strong expression of temporally regulated small non-coding RNAs

Following the observation of abundant antisense transcripts, we looked for other unusual transcriptional profiles within phage transcriptomes and detected two small (~100bp) transcripts abundantly expressed in PAK_P3 late infected cells. These two transcripts, hereafter referred as sRNA1 and sRNA2, were found in two neighbouring intergenic regions: sRNA1 is encoded within a 200bp-region, between two genes encoding hypothetical proteins, whereas sRNA2 is part of a larger region between two phage-encoded tRNAs (Figure 25). They respectively display a 91- and 12-fold change in their 'late versus early' expression.

We hypothesized that they could represent *trans*-encoded small RNAs, acting by base-pairing a target mRNA. As such, we looked for potential targets in both host and PAK_P3 genomes. Potential targets found on host genome were not differentially expressed 13 min post infection, indicating that sRNA1 and sRNA2 would not act through mRNA degradation but rather have a role in translational silencing, if any.

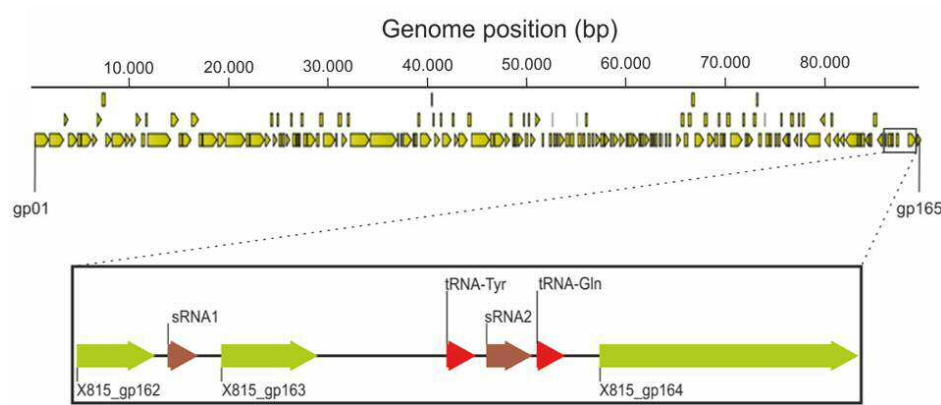


Figure 25: PAK_P3 small non coding RNAs

The genome of PAK_P3 is represented, with yellow arrows indicating predicted coding sequences. An enlargement of the region encoding predicted small RNAs is provided.

Interestingly, a stretch of 11 nucleotides on sRNA2 was found to be repeated eight times in host genome and is systematically located within tRNAs, more particularly within the T ψ C-loop. We propose that it could be involved in translational repression by binding, and thereby blocking, bacterial ribosomes. Alternatively, as this 11bp-stretch is also conserved in closely related phages (*PAK_P1-like* genus), it may represent a starting point leading to the discovery of new phage non-coding RNAs.

On PAK_P3 genome, the only potential targets (11 consecutive nucleotides matching perfectly) are located in the early ORF *gp78* for sRNA1 and in the late gene encoding the putative ribonucleotide-diphosphate reductase *Gp67* for sRNA2.

To date, only few phage-encoded small RNAs have been described in the literature and most of them derive from prophages^{173,174}. The only examples of phage sRNAs encoded by a virulent phage, T4 band C and band D RNAs, were described in the 1970's and their functions have remained unknown ever since¹⁷⁵.

In a primary attempt to characterize the function of these sRNAs, we inserted an IPTG-inducible expression cassette carrying the ~500bp regions encoding these sRNAs in single copy into the host chromosome, but could not detect any effect of these artificial overexpressions on bacterial growth in rich medium, nor on phage lysis kinetics.

3. Core genes are similarly expressed in both infections but display phage-specific adjustments

In previous paragraphs, we globally investigated PAK_P3 and PAK_P4 transcriptional programs and highlighted two main results: (i) viral gene expression is temporally regulated and (ii) antisense transcripts are produced, particularly during early infection. These characteristics are common to both phages.

To determine more precisely how the regulation of phage gene expression is adjusted, we compared the differential expression of the set of 69 core genes. While most conserved gene features still undergo the same pattern of temporal regulation, the strength of that differential expression appears to vary between the two viruses (Figure 26). For instance, PAK_P3 tends to differentially express *gp002-gp006* more drastically than PAK_P4 does, while PAK_P4 favours the differential expression of *gp007-gp010*. Strikingly, one gene is strongly up-regulated during PAK_P4 late infection while it is not temporally regulated in PAK_P3. The corresponding predicted protein (Gp014 in PAK_P4) possesses a wide range of homologs in other unrelated phages and also in bacteria, however its function remains unknown. We also found that genes encoding DNA-metabolism related proteins, as well as early genes, are overall more strongly temporally regulated in PAK_P3 than they are in PAK_P4.

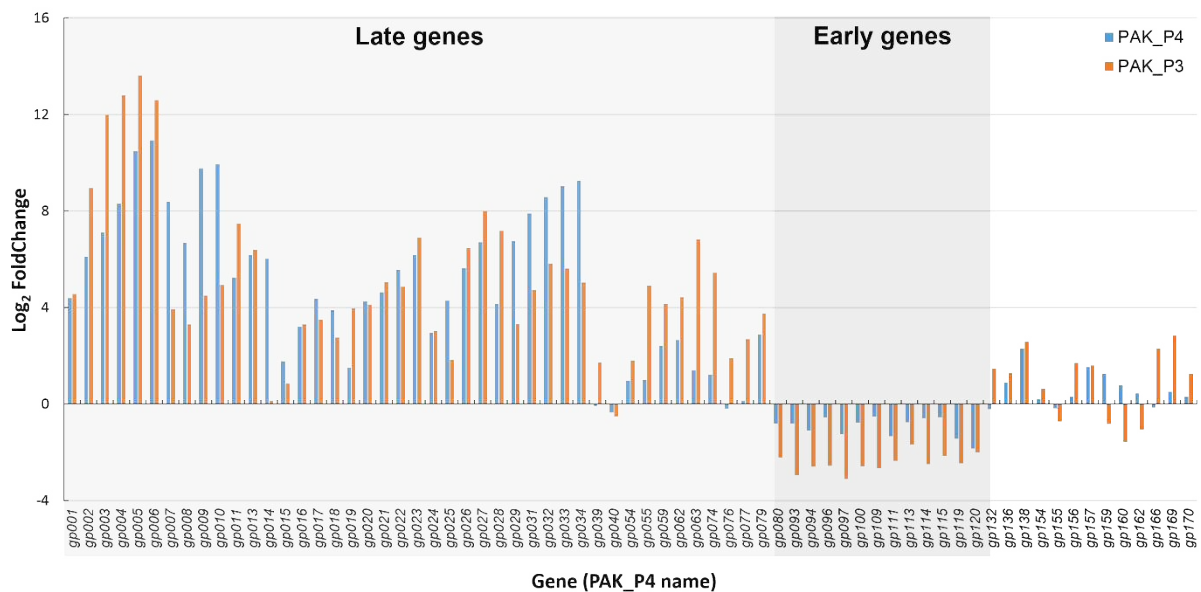


Figure 26: Comparison of PAK_P3 and PAK_P4 homologous genes expression

Homologous genes are listed and their differential expression (late versus early) during PAK_P3 and PAK_P4 infections are indicated as Log₂(Fold Change) values in orange and blue, respectively.

Assuming homologous proteins have similar functions in both phages, these results would indicate that the common protein-coding potential is not used in the same way by PAK_P3 and PAK_P4. By modulating the regulation of one given set of core genes, each phage may orchestrate its infection cycle in a different way compared to the other.

As already mentioned, core genes mostly comprise genes belonging to the structural, lately expressed cluster. Interestingly, some belong to the early cluster which is mostly composed of ORFans. Given that early proteins are usually responsible for hijacking host machineries, the conservation of some of them may indicate that PAK_P3 and PAK_P4 have common mechanisms to co-opt their host. Therefore, in the next chapter, we focus on investigating the functions of such proteins.

CHAPTER 3 - EXAMPLE OF A VIRAL STRATEGY TO CO-OPT THE BACTERIAL SYSTEM DURING EARLY INFECTION

This chapter investigates the functions of one protein encoded by an early-expressed core ORF, namely Gp92, produced by PAK_P3.

1. Identification of a small phage protein inhibiting *P. aeruginosa* growth

As a first step toward the functional characterization of PAK_P3 ORFs, we looked for phage proteins inhibiting bacterial growth. Such deleterious proteins are often responsible for redirecting host metabolism toward viral production and are usually small and precociously produced during the infection cycle⁸¹.

Thus, we selected 10 ORFs from phage PAK_P3 which are predicted to encode proteins (i) smaller than 250 amino acids, (ii) conserved in *Kpp10virus* and *Pakpunavirus* genera and (iii) located outside the lately expressed gene cluster encoding phage structural proteins. These ORFs were then inserted in single copy on *P. aeruginosa* PAK chromosome under the control of an IPTG-dependent promoter and tested for their ability to inhibit bacterial growth.

Interestingly, among all ORFs tested, only one had a significant effect on bacterial growth. Indeed, the strain expressing ORF *gp92* yielded colonies with reduced sizes, suggesting a slower growth rate. This growth inhibitory effect was also tested in other *P. aeruginosa* strains such as the mucoid cystic fibrosis clinical strain CHA and the laboratory strain PAO1 which are respectively sensitive and resistant to phage PAK_P3 itself (Figure 27). The same effect was observed in these strains, as well as in *E. coli* (BW25113 strain), revealing that the mechanism of toxicity of Gp92 is not strain-specific and suggests that this phage protein targets metabolic pathways that are conserved between *P. aeruginosa* and *E. coli*, probably by interacting with strongly conserved bacterial proteins.

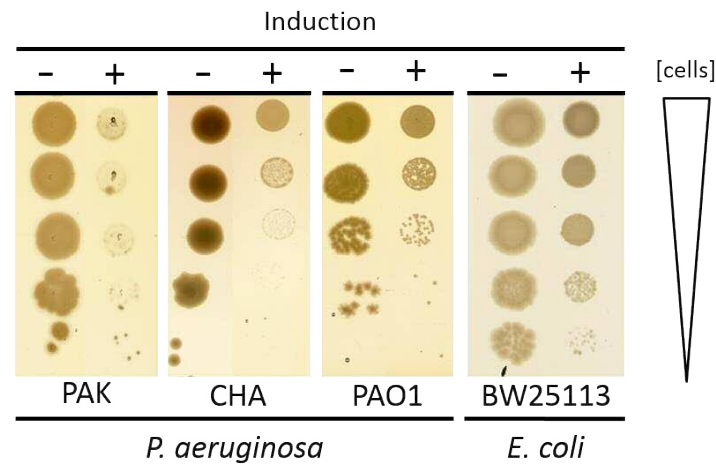


Figure 27: Expression of phage gene *gp92* affects growth of *P. aeruginosa* and *E. coli*

Serial ten-fold dilutions of *P. aeruginosa* strains PAK, CHA, PAO1 and *E. coli* strain BW25113 were spotted on LB medium with (+) or without (-) IPTG (1mM). *Pseudomonas* strains are carrying an expression cassette containing *gp92* gene under the control of an IPTG-dependent promoter integrated in single copy into their chromosome. In *E. coli*, *gp92* is expressed from the same expression cassette but carried on a high copy number plasmid.

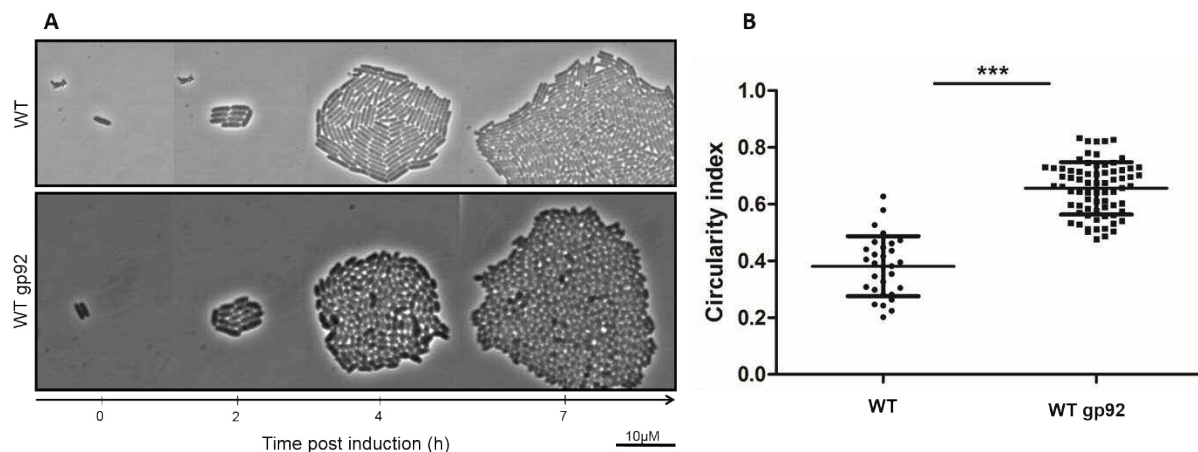


Figure 28: Gp92 affects bacterial cell morphology

(A) Time-lapse optical microscopy. Strain PAK carrying an IPTG-inducible expression cassette containing *gp92* (WTgp92) or a control strain PAK carrying an empty cassette (WT) were dropped on acrylamide pads containing IPTG (t=0). Successive divisions of single cells were monitored by taking pictures every 20min for 8h. The microscope was equipped with an incubation chamber set up at 37°C. **(B)** Measurement of cell *circularity index of 30 WT cells and 70 WTgp92 cells (total) after 4h of IPTG induction.

*Defined as $\text{Circ} = 4\pi \times [\text{Area}] / [\text{Perimeter}]^2$ and ranges from 0 (infinitely elongated polygon) to 1 (perfect circle), assessed with ImageJ software.

2. Gp92 causes cell morphology modification

As bacteria expressing *gp92* form small colonies, we hypothesized that Gp92 production could result in impaired cell motility and/or morphology. Thus, we monitored bacterial growth upon induction of *gp92* expression using time lapse microscopy. Upon two hours of induction, we observed that the classical rod-shaped morphology of *Pseudomonas* cells changed, displaying a cocci-like aspect (Figure 28A). Detailed analysis of cell shape 4h post induction revealed that cells producing Gp92 have a circularity index significantly higher than wild type cells (0.66 ± 0.09 versus 0.38 ± 0.11 , $p\text{-value} < 0.0001$, Mann Whitney test) (Figure 28B). Therefore, Gp92 seems to disrupt bacterial envelope homeostasis.

3. Gp92 transmembrane location is crucial for its activity

Given its effect on cell morphology, we hypothesized that Gp92 may interact with host membrane proteins or be itself localised at the membrane, despite the fact that *in silico* prediction tools could not detect a canonical transmembrane segment or a signal peptide. We then experimentally investigated Gp92 cellular localisation by using a dual *pho-lac* reporter system¹⁷⁶. Activity of a chimeric protein consisting of full length Gp92 fused to PhoA-LacZ reporter was analysed in *E. coli* on an indicator LB medium containing both a blue chromogenic substrate for phosphatase periplasmic activity and a purple substrate for galactosidase cytosolic activity. Cells expressing this construct displayed a Pho⁺ phenotype (blue coloration), revealing that Gp92 has a periplasmic or transmembrane location (Figure 29A-B).

Then we sought to delineate Gp92 domains by fusing only the 26 amino acids N-terminal domain (Gp92_N) or the 56 amino acids C-terminal domain (Gp92_C) to PhoA-LacZ. Cells expressing *gp92_N* construct yielded blue colonies and, consistently, when the N-terminal part of Gp92 is missing, Gp92_C-PhoA-LacZ fusion remained in the cytosol as indicated by the purple phenotype (Lac⁺) of the cells (Figure 29A-B). Altogether, these results indicate that Gp92 is localised at the membrane (or in the periplasm) through its N-terminal part, which constitutes an unusual transmembrane domain (or signal peptide). We also artificially exported Gp92_C in the periplasm by constructing a chimeric protein consisting in PhoA signal peptide fused to Gp92_C (Figure 29A-B).

Next, the toxicity of these constructs was tested and none of the truncated (or exported) versions of Gp92 cause the phenotype observed with the full length protein, demonstrating that its membrane location is crucial to mediate its toxic activity and ruling out that the N-ter domain is a signal peptide (Figure 29C).

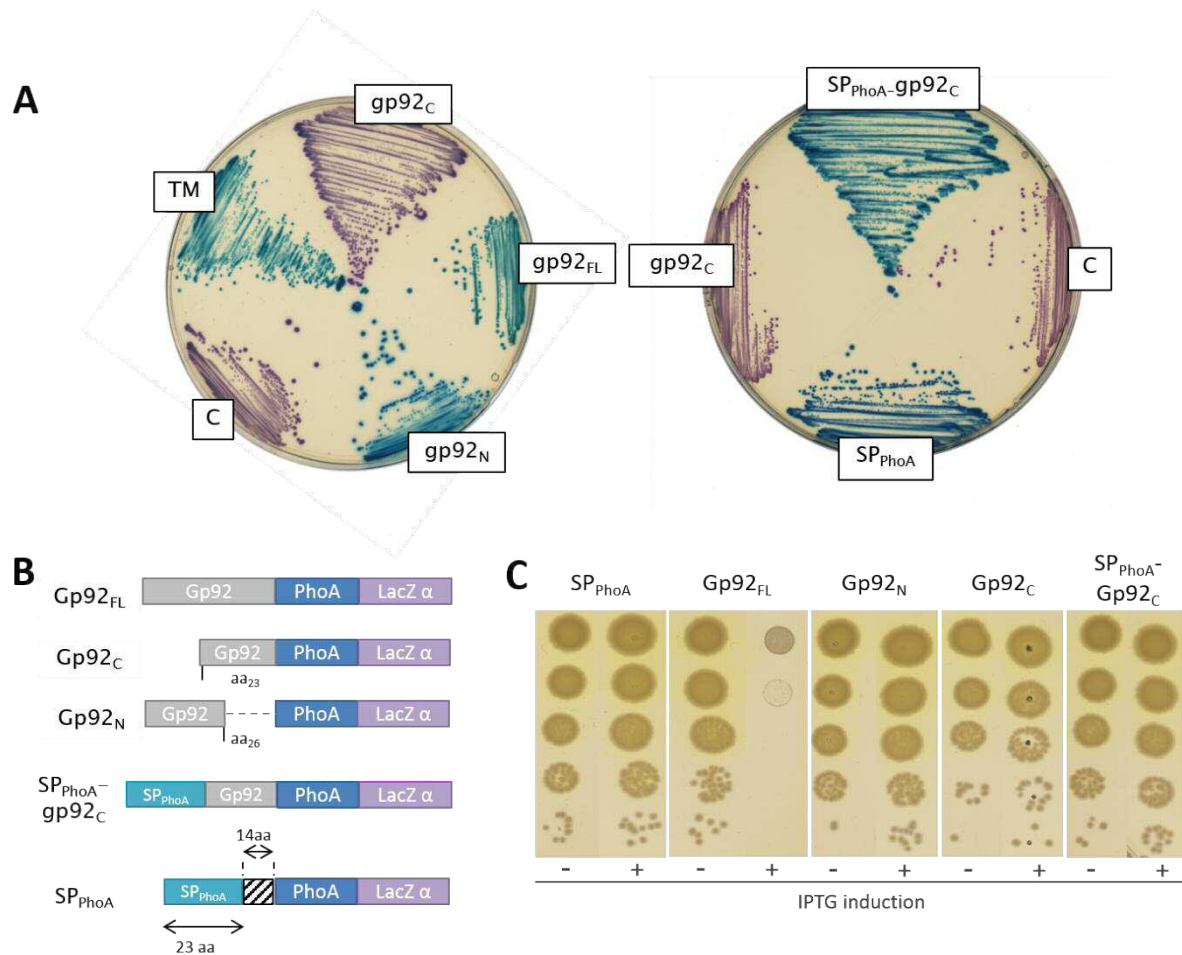


Figure 29: Topology analysis of Gp92

(A) Experimental determination of Gp92 topology. DH5α *E. coli* cells expressing different Gp92 fusion- or control fusion-proteins (C: PhoA₂₂₋₄₇₁-LacZ fusion located in the cytosol, TM: Blr-PhoA₂₂₋₄₇₁-LacZ fusion, transmembrane construct¹⁷⁷) were plated on an indicator medium containing two chromogenic substrates: Magenta-Gal (β-galactosidase activity) and BCIP (Phosphatase activity).

(B) Schematic representation of Gp92 fusion proteins. Full length protein (Gp92_{FL}), C-terminus domain (Gp92_C) or N-terminus domain (Gp92_N) of Gp92 were fused with a chimeric protein consisting of alkaline phosphatase (PhoA) fused with β-galactosidase (LacZ α). Gp92_C lacks 22 amino acids at the N-terminus. Gp92_N consists in the first 26 amino acids of Gp92_{FL}. SP_{PhoA}, the signal peptide of PhoA, was fused to Gp92_C to artificially export it in the periplasm (SP_{PhoA}-Gp92_C).

(C) Toxicity assay was performed on PAK strains carrying IPTG-inducible expression cassettes containing indicated fusions.

4. Gp92 interacts with the stress response-associated transcriptional regulators AlgU and MucA

To identify Gp92 interaction partners, we carried out two independent bacterial adenylate cyclase two-hybrid (BACTH) screens¹⁷⁸ using *E. coli* and *Salmonella typhi* peptides libraries (see Methods), the results of which are summarized in Table 7. Among 13 potential partners identified, the anti- σ^E factor RseA was identified in 35% of the positive clones and was also the only candidate identified in the two independent screens.

Table 7: Gp92-binding proteins identified with bacterial two-hybrid screen

^aIndicates the frequency at which the protein was found (43 clones were analysed). White and grey boxes refer to the screens made against *E. coli* and *S. typhi* libraries, respectively. ND: not determined.

Conserved domain	Protein	Frequency (%) ^a		Homolog in <i>P. aeruginosa</i>
RseA superfamily	RseA	23		MucA
		12	35	
MFS superfamily	EntS	21		PA4622
	ProP	12		PA4343
	YhhS	2		PA1993
	YicJ	2	37	ND
COG2976	YfgM	5		ND
YhhN superfamily	YhhN	5		ND
PgaD superfamily	PgaD	2		ND
Bacterial BAX inhibitor	YbhL	2		ND
DUF1469 superfamily	YqjE	5		ND
DUF423	YgdD	5		ND
DUF2167	STM0777	2		ND
MAPEG family	YecN	2	28	ND

We subsequently confirmed the *in vivo* interaction between Gp92 and RseA, and also tested the interaction with the homologous protein encoded by *P. aeruginosa* (namely MucA), by fusing each interaction partner to one domain of the adenylate cyclase. The simultaneous presence of Gp92 and RseA (or MucA) resulted in maltose assimilation by the cells (assessed by colorimetric assay on MacConkey plates), similar to the positive control, whereas cells only producing one of the two interaction partners remained unable to ferment maltose, as expected (Figure 30). Therefore, we confirmed that Gp92 is interacting *in vivo* with the anti- σ^E factors RseA (*E. coli*) and MucA (*P. aeruginosa*).

As introduced in the first part of this manuscript, the primary role of anti-sigma factors is to sequester sigma factors, preventing them to bind to RNAP core and activate the transcription of the regulons they control. MucA is a well described transmembrane protein binding the sigma factor AlgU, involved

in stress response and notably responsible for alginate overproduction. Using the same assay described above, we found that Gp92 is also able to interact with AlgU.

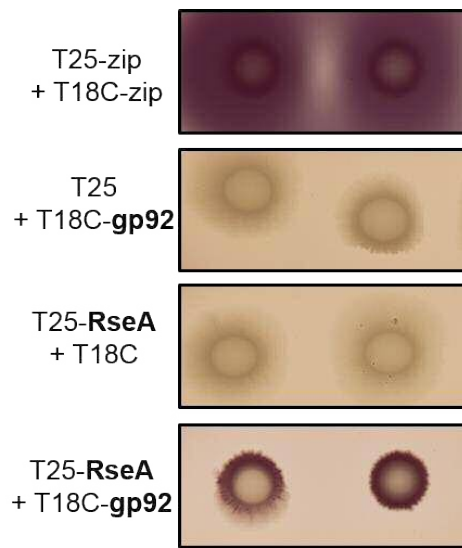


Figure 30: In vivo interaction of Gp92 and anti-sigma factors (RseA/MucA)

Interaction assay using BACTH system. *E.coli* Δcya cells were co-transformed with indicated plasmids and subsequently grown overnight in LB with appropriate antibiotics. As a positive control, we used the two domains of adenylate cyclase (T25 and T18) fused to leucine zipper domains (zip). As negative controls we tested the interaction between the fusion proteins (T18C-gp92 or T25-RseA) and the corresponding complementary domains of the adenylate cyclase only. Five microliters saturated cultures were spotted on MacConkey agar supplemented with maltose, IPTG and antibiotics. Plates were incubated 48h at 30°C. Similar results were obtained by using T25-MucA and T25-AlgU constructs.

5. Gp92 leads to cell death in strains with impaired stress response

Next, we hypothesized that if Gp92-associated phenotype (reduction of colony size) is specifically due to the interaction with MucA, wild-type phenotype should be restored in a strain with an impaired MucA protein. Thus we assessed Gp92-mediated growth inhibition in a strain lacking functional MucA proteins¹³². We used a PAK strain containing the *mucA22* allele, a mutation frequently identified in cystic fibrosis-associated isolates. This strain produces a truncated non-functional MucA protein (*i.e.* not sequestering AlgU, but still located in the membrane), notably resulting in conversion to a mucoid colony phenotype (alginate overproduction). Upon induction of *gp92* expression, *mucA22* strain does not form smaller colonies contrary to wild-type strain (Figure 31), indicating that Gp92 toxic activity is dependent of the interaction with a functional MucA protein.

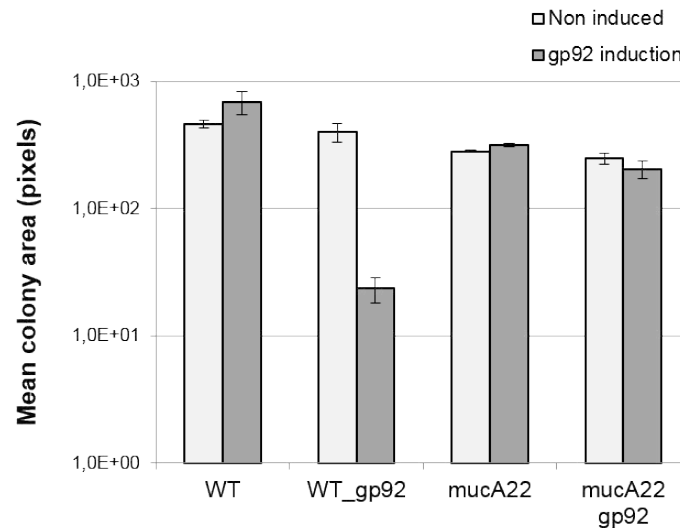


Figure 31: Effect of gp92 expression in *mucA22* strain

Gp92 toxicity assay in *P. aeruginosa* PAK wild type or *mucA22* strains carrying an empty expression cassette (WT and *mucA22*) or a *gp92* expression cassette (WT_gp92 and *mucA22_gp92*) (as described in Figure 27) were plated on LB agar medium without (light gray bars) or with (dark gray bars) 1mM IPTG. After 16h incubation at 37°C, plates were scanned and areas of colonies were evaluated using ImageJ software. Experiments were performed in triplicates and mean colony areas as well as standard deviations are indicated.

It has to be noted that, in absence of MucA, AlgU is available to constitutively activate the transcription of its stress-response regulon, thereby potentially counteracting Gp92-effect.

Since we found that Gp92 interacts with AlgU as well, we also tested the effect of *gp92* expression in a strain lacking AlgU-associated stress response ($\Delta algU$).

In this context, we found that the production of Gp92 results in a significantly reduced growth rate in liquid culture, starting 4h post induction and leading to cell death as we observed a 5-log reduction of bacterial count upon induction (Figure 32A-B). Investigating this phenomenon in more details with time lapse microscopy, we observed that bacterial morphology is dramatically affected: cells became branched and twisted, as soon as 2h post induction, and eventually burst 4h post induction (Figure 32C). Finally, expression of *algU* was restored (by trans-complementation) and resulted in the recovery of the initial Gp92-associated phenotype (*i.e.* small colonies), thereby confirming that Gp92-lethality is revealed by the absence of AlgU (Figure 32D).

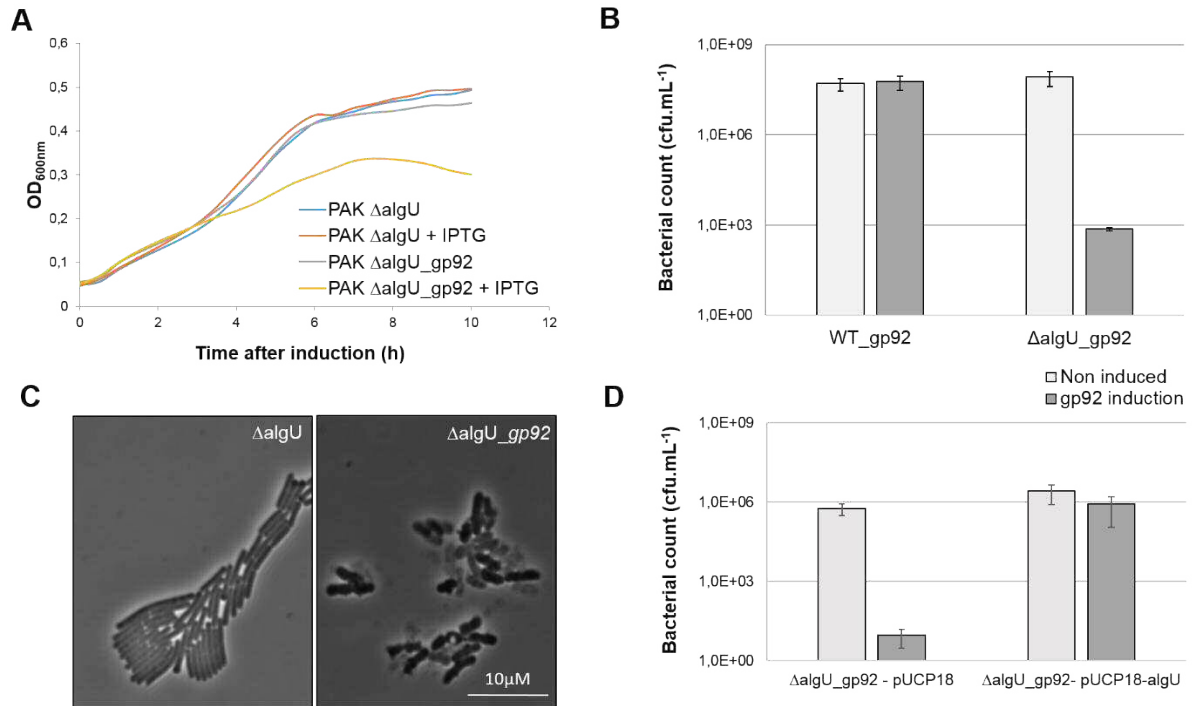


Figure 32: Effect of gp92 expression in Δ algU strain

(A) Growth curves of indicated strains. OD_{600nm} was followed over time upon addition of 1mM IPTG (t=0) in pre-cultures of exponentially growing strains. Experiments were performed in triplicate and a mean curve is represented.

(B, D) Exponentially growing strains carrying *gp92* expression cassette (OD₆₀₀=0.3) were plated on LB agar medium without (light grey bars) or with (dark grey bars) 1mM IPTG. Survival of strains WT, Δ algU or Δ algU carrying complementation plasmid pUCP18-*algU* or control plasmid pUCP18 was estimated by counting colonies growing on both media after 16h incubation at 37°C. Means and standard deviations corresponding to 3 independent experiments are represented.

(C) Optical microscopy images of Δ algU strains with empty (left) or *gp92* (right) expression cassettes, growing on acrylamide pads soaked in LB IPTG (1mM). Pictures were taken 4h post IPTG induction.

The alternative sigma factor AlgU can regulate the expression of a large repertoire of genes. In response to a cell wall stress, it has been demonstrated to influence the transcription of 293 ORFs, including genes involved in envelope biogenesis and remodelling^{147,179}. In particular, AlgU is the major positive regulator of alginate biosynthesis, as it allows the RNAP to bind to the *PalgD* promoter and subsequently transcribe *alg* operon comprising 12 genes¹⁴⁶.

Although we established that the presence of AlgU reduces Gp92-associated toxicity, we could not discriminate whether this attenuation directly results from AlgU-Gp92 interaction or indirectly from the activation of the associated stress response.

Consequently, we next tested whether alginate production (totally absent in strain Δ algU) participates to the limitation of Gp92 toxicity. We observed that *gp92* expression was still deleterious in a Δ algD

mutant strain, even though AlgU and MucA are present and functional. Indeed, we found that *gp92* expression also leads to cell death (3-log decrease of bacterial count, Figure 33A).

In addition, cell morphology was also affected upon induction, although differently from what has been observed with $\Delta algU$ cells. Indeed, $\Delta algD$ cells formed filaments prior to burst (Figure 33B). Consistently, when monitoring bacterial growth in liquid culture, no differences in optical density values could be detected between induced and non-induced cultures, due to this filamentation phenomenon (Figure 33C).

In conclusion, these results suggest that alginate production may be a major pathway (but not the only one), within the AlgU-associated response, responsible for limiting the stress imposed by Gp92.

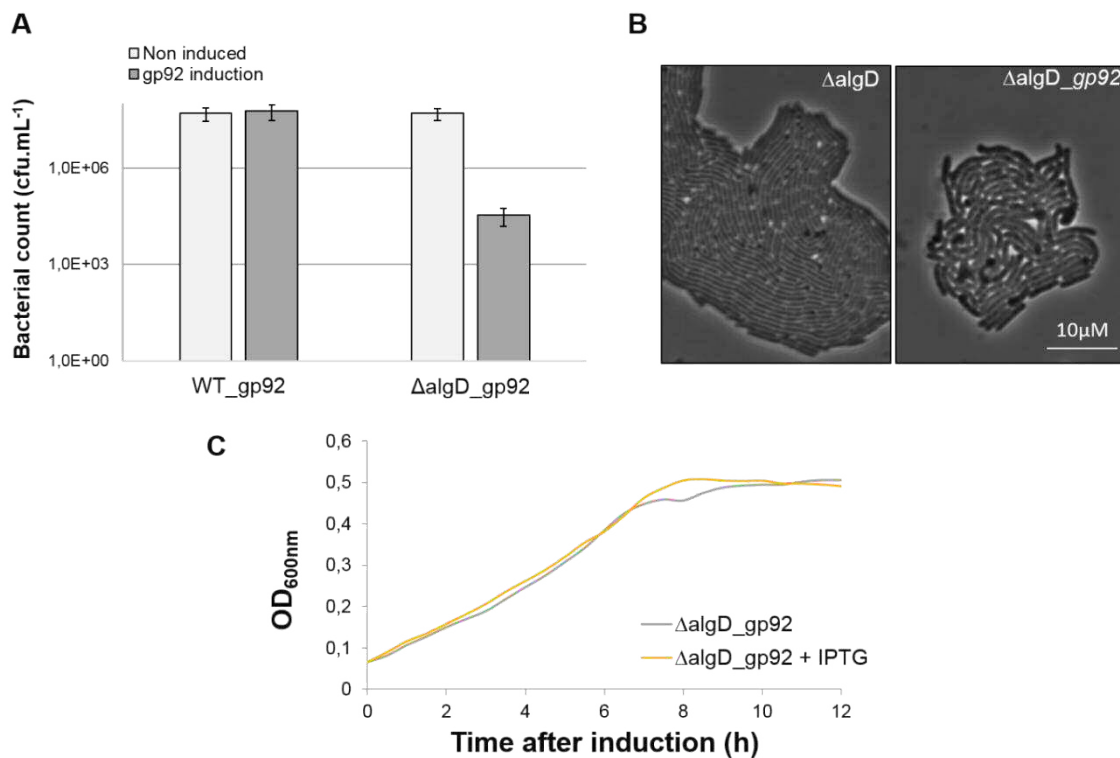


Figure 33: Effect of *gp92* expression in $\Delta algD$ strain

(A) Exponentially growing indicated strains carrying *gp92* expression cassette ($OD_{600}=0.3$) were plated on LB agar medium without (light gray bars) and with (dark gray bars) 1mM IPTG. Survival of WT and $\Delta algD$ strains was estimated by counting colonies growing on both media after 16h incubation at 37°C. Means and standard deviations corresponding to 3 independent experiments are represented.

(B) Optical microscopy images of $\Delta algD$ strains with empty (left) or *gp92* (right) expression cassettes growing on acrylamide pads soaked in LB IPTG (1mM). Pictures were taken 6h post IPTG induction.

(C) Growth curves of indicated strains. OD_{600nm} was followed over time upon addition of 1mM IPTG ($t=0$) in pre-cultures of exponentially growing strains. Experiments were performed in triplicate and a mean curve is represented.

6. Gp92 homologous protein encoded by PAK_P4 is not toxic for *P. aeruginosa* cells

Gp92 shares 44% identity (amino acid level) with its homologous protein encoded by PAK_P4, namely Gp109, also produced early during infection (Figure 34).

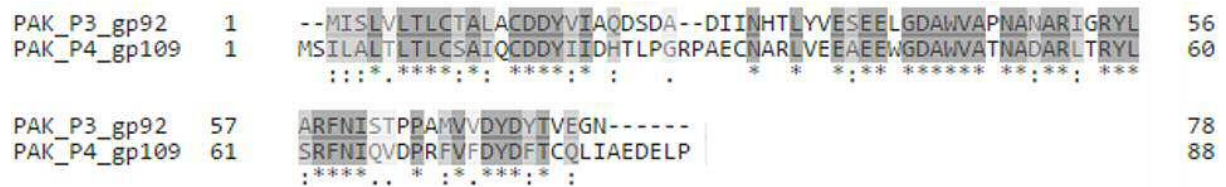


Figure 34 Alignment of Gp92 and Gp109 protein sequences

Alignment was generated by Clustal Omega. Identical and similar residues are shaded in dark and light grey, respectively.

In addition of being conserved in PAK_P4, 'Gp92-like' proteins have been identified in all phages belonging to *Pakpunavirus* and *Kpp10virus*. They are part of the 30 core proteins defined in these two genera¹⁵⁸. Consequently, given this broad conservation, we hypothesized that Gp92 and Gp109 may have the same function and thus expected that induction of *gp109* expression in wild type strain PAK would result in size-reduced colonies. However, we observed that Gp109 does not affect bacterial growth, as reported in Figure 35. Therefore despite strongly conserved domains, Gp109 does not act as Gp92.

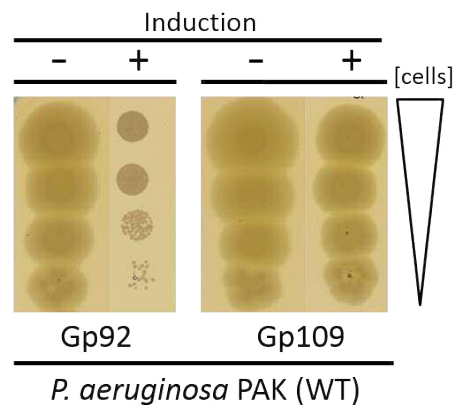


Figure 35: *gp109* expression does not lead to cell toxicity

Serial ten-fold dilutions of *P. aeruginosa* strain PAK, carrying expression cassettes containing *gp92* or *gp109* under the control of an IPTG-dependent promoter integrated in single copy into their chromosome, were spotted on LB medium with (+) or without (-) IPTG (1mM).

7. Proposed model

Altogether, the data presented in this chapter led us to propose the following model of Gp92 function. We hypothesize that Gp92 strengthens AlgU-MucA complex, thereby preventing the release of AlgU and the resulting transcriptional activation of stress response.

Sustaining this hypothesis, our transcriptomics data showed that cells infected by phage PAK_P3 (encoding Gp92) do not display an increased expression of the *algU-mucA* operon (*algU* induces its own expression once activated). In contrast, we observed an increased expression of *algU* and *mucA* upon PAK_P4 (encoding Gp109) infection, which may indicate that phage infected cells attempt to trigger an AlgU-mediated stress response. However, this attempt is probably aborted downstream as typical genes belonging to *algU* regulon (e.g. alginate operon) do not display an increased expression upon PAK_P4 infection (Table 8).

In addition, due to its localisation, we hypothesize that Gp92 generates a membrane stress that is partially counteracted by a low-grade AlgU-stress response (mainly involving alginate production) in wild-type cells. This stress becomes lethal in alginate- or AlgU-deficient strains.

To corroborate this hypothesis, we exposed cells expressing *gp92* to sub-inhibitory concentration of cell wall inhibitor D-cycloserine, previously shown to activate AlgU stress response¹⁸⁰. We chose the highest concentration of D-cycloserine that did not affect the growth of non-induced wild-type strain encoding Gp92. While the optical density of a bacterial culture is not affected by *gp92* expression, addition of D-cycloserine impaired bacterial growth as shown in Figure 36. This suggests that *gp92* expression and D-cycloserine have a synergistic deleterious effect on bacterial growth, thereby strengthening the hypothesis of a membrane stress caused by Gp92.

Table 8: Differential expression of AlgU regulon upon PAK_P4 infection

Expression values of indicated genes at t=0 (Mean 0M) or at t=13min (Mean 13M) following phage infection. Fold change values are indicated as Log₂(FC) and significantly up-regulated (UP) or down-regulated (DOWN) genes are indicated. - : not significant variation, ND: not determined.

AlgU stimulon		PAK_P3					PAK_P4				
algU operon		Mean 0M	Mean 13M	Log2 (FC)	pvalue	Differentially expressed?	Mean 0M	Mean 13M	Log2 (FC)	pvalue	Differentially expressed?
PAK_93 algU	RNA polymerase sigma factor RpoE	798	404	-0,86	0,28	-	851,16	2929,00	1,73	0	UP
PAK_94 mucA	anti sigma-E RseA, N-terminal domain protein	402	181	-1	0,25	-	362,25	1466,80	1,97	0	UP
PAK_95 mucB	mucB/RseB family protein	282	148	-0,8	0,42	-	270,85	681,80	1,31	0	UP
PAK_96 mucC	positive regulator of sigma(E), RseC/MucC family protein	114	33	-1,4	0,03	DOWN	106,03	124,26	0,22	0,64	-
PAK_97 mucD	peptidase Do family protein	1354	416	-1,5	0,05	-	1491,80	1145,89	-0,37	0,44	-
alginate biosynthesis											
PAK_3342 algD	nucleotide sugar dehydrogenase family protein	7	10	0,58	0,6	-	7,93	21,78	1,27	0,14	-
PAK_3343 alg8	hypothetical protein	2	6	1,03	1	-	2,50	7,79	1,34	ND	-
PAK_3344 alg44	pilZ domain protein	1	6	1,79	1	-	1,07	3,73	1,08	ND	-
PAK_3345 algK	sel1 repeat family protein	1	4	1,15	1	-	0,56	2,11	0,75	ND	-
PAK_3346 algE	hypothetical protein	1	4	1,4	1	-	1,07	3,35	1,15	ND	-
PAK_3347 algG	hypothetical protein	2	4	0,56	1	-	1,73	3,39	0,65	ND	-
PAK_3348 algX	hypothetical protein	2	4	0,34	1	-	1,49	4,71	1,12	ND	-
PAK_3349 algL	alginate lyase family protein	2	4	0,56	1	-	2,43	1,63	-0,38	ND	-
PAK_3350	hypothetical protein	1	4	0,54	1	-	0,97	0,00	-0,69	ND	-
PAK_3351 algI	MBOAT family protein	4	4	-0,6	1	-	4,83	5,50	0,04	ND	-
PAK_3352	hypothetical protein	0	4	1,56	1	-	0,00	0,00	ND	ND	-
PAK_3353 algJ	hypothetical protein	1	4	0,64	1	-	1,35	1,76	0,18	ND	-
PAK_3354 algF	alginate O-acetyl transferase AlgF family protein	2	4	0	1	-	2,51	0,00	-1,54	ND	-
PAK_3355 algA	mannose-1-phosphate guanylyltransferase/mannose-6-phosphate isomerase	2	9	1,84	0,04	UP	3,99	7,92	0,84	ND	-
peptidoglycan biosynthesis											
PAK_5061 mdoH	glycosyl transferase 2 family protein	842	481	-0,7	0,48	-	908,12	1511,02	0,72	0,01	-
PAK_4639 mrcB	penicillin-binding protein 1B	1646	314	-2,2	0	DOWN	1864,54	640,09	-1,51	0	DOWN
PAK_3872 mpl	L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase	150	87	-0,6	0,4	-	173,30	179,67	0,06	0,92	-
LPS biosynthesis											
PAK_2917 wzz	chain length determinant family protein	476	79	-2,4	0	DOWN	531,04	196,84	-1,39	0,00	DOWN
PAK_5156 rfbD	dTDP-4-dehydroxymannose reductase	314	106	-1,3	0,09	-	324,54	250,84	-0,36	0,33	-
PAK_5476 wbpH	glycosyl transferases group 1 family protein	165	46	-1,6	0	DOWN	145,13	108,66	-0,42	0,44	-

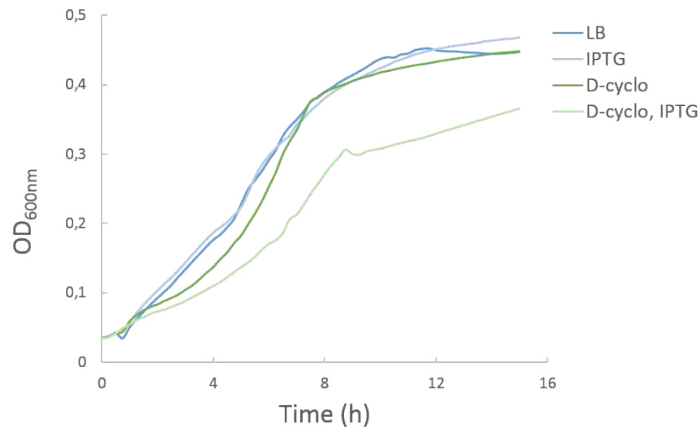


Figure 36: Effect of D-cycloserine treatment on the growth of cells expressing gp92

Strain WT_gp92 was pre-cultured with, or without, 1mM IPTG for 1h before addition of 20 μ g.mL⁻¹ D-cycloserine (D-cyclo) at t=0. OD_{600nm} was then followed over time.

In conclusion, we identified a small early produced phage protein targeting a major host regulatory system involved in stress response, fitting within the acknowledged model of host subversion mediated by phage early proteins. In this particular case, we can hypothesize that AlgU-associated sensors detect phage infection, resulting in a direct or indirect activation of defense mechanisms. By interfering with AlgU-mediated stress response, Gp92 may contribute to the ‘silent’ establishment of phage hijacking strategy.

Interactions between phage early proteins and sigma factors have already been demonstrated but were associated to the inhibition and/or enslavement of host transcription (Table 2). Here we report a different mechanism. Concomitant to this work, Zhao and co-workers found that Gp70.1, encoded by *P. aeruginosa* temperate phage PaP3, causes a similar small-colony phenotype when ectopically produced in the host strain and interacts with the stationary phase-associated and stress response σ S factor¹⁰⁷. However, expression of *orf70.1* did not lead to modification of cell morphogenesis suggesting this protein does not cause membrane stress. No conserved amino acid sequences could be identified between of Gp92 and Gp70.1. Expression of *orf70.1* in *P. aeruginosa* leads to the up-regulation of genes encoding proteins related to extracellular functions (e.g. cell wall, motility, secretion/export proteins). Given our hypothesis, we expect that *gp92* expression would rather down-regulate the expression of such genes that are generally positively regulated by AlgU. These hypothetical opposite functions of Gp70.1 and Gp92 could be explained by the temperate nature of phage PaP3, which implies a different strategy to co-opt the host in comparison to a lytic phage such as PAK_P3.

In conclusion, targeting broad-activity transcriptional regulators of bacterial stress response might be a key step for an optimal phage-programmed transformation of a *ribocell* into a *virocell*.

CHAPTER 4 - IMPACT OF PHAGE INFECTION ON HOST METABOLISM

Following the study of phage transcriptional programs and co-option mechanisms, in this chapter we will focus on the impact of phage infection on host physiology. In particular we investigate variations of cellular content in RNA and metabolites.

1. Dramatic redirection of RNA metabolism by PAK_P3 and PAK_P4

We first looked at variations in the composition of cellular whole transcriptome over a single infection cycle and revealed a progressive and dramatic replacement of host mRNAs with phage transcripts.

While PAK_P3 induces a progressive depletion of host transcripts that eventually represent only 13% of non-ribosomal RNAs in the cell 13 min post infection, PAK_P4 is much more rapid as only 20% of the total transcripts matches the host genome as soon as 3.5 min post infection (Figure 37). However, even in the context of this dramatic depletion of host transcripts, a response to phage infection at the transcription level was observed (developed further in Chapter 5), suggesting a globally accelerated degradation of unstable mRNA species rather than a global transcriptional repression as described for phage T4.

Therefore, both phages shut down host cell transcription however with different kinetics. We hypothesize that they probably have a different control of the machineries dedicated to (i) degrade host mRNAs and/or (ii) catalyse the transcription of viral genes.

The rapid and overwhelming replacement of host transcripts with viral transcripts could indicate that both phages rely on their own transcription apparatus even though they are not predicted to encode a viral RNA polymerase (RNAP) based on sequence homology searches. Therefore, we experimentally assessed the dependency of PAK_P3 and PAK_P4 infections on host RNAP using rifampicin (Rif), which is known to inactivate this enzyme. We found that addition of Rif abolished PAK_P3 and PAK_P4 amplification while it did not affect phage Φ KZ transcription, as previously shown (Figure 38)³⁵. This indicates that both phages appear to efficiently hijack the host transcription machinery to complete their infectious cycle.

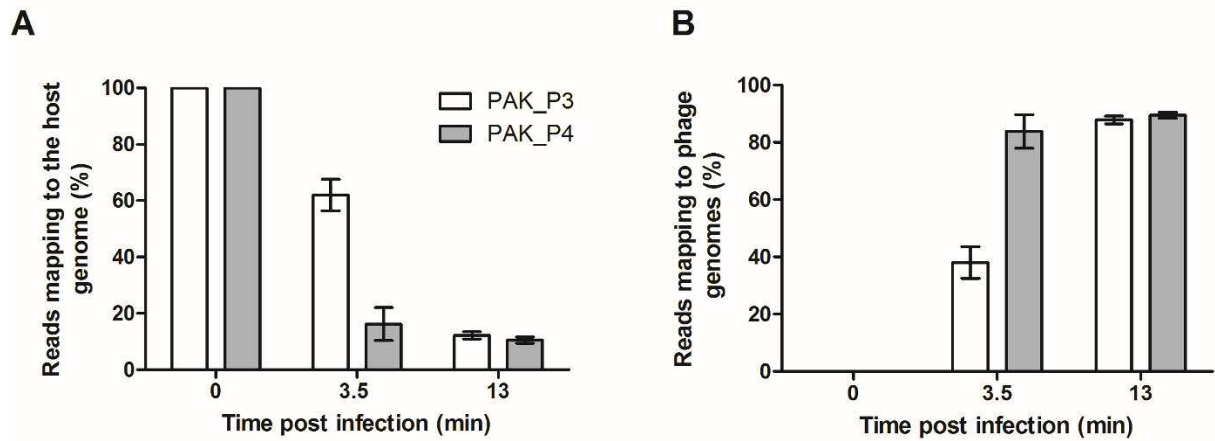


Figure 37: PAK_P4 takes over the host cell transcription faster than PAK_P3

Three independent biological replicates of RNA extracts were harvested from PAK_P4 and PAK_P3 infected cells at 0min, 3.5min, and 13min post infection and subsequently sequenced. Percentages of reads mapping to phage genomes (**A**) and to the host genome (**B**) over the course of PAK_P4 infection (grey bars) or PAK_P3 infection (white bars) are plotted.

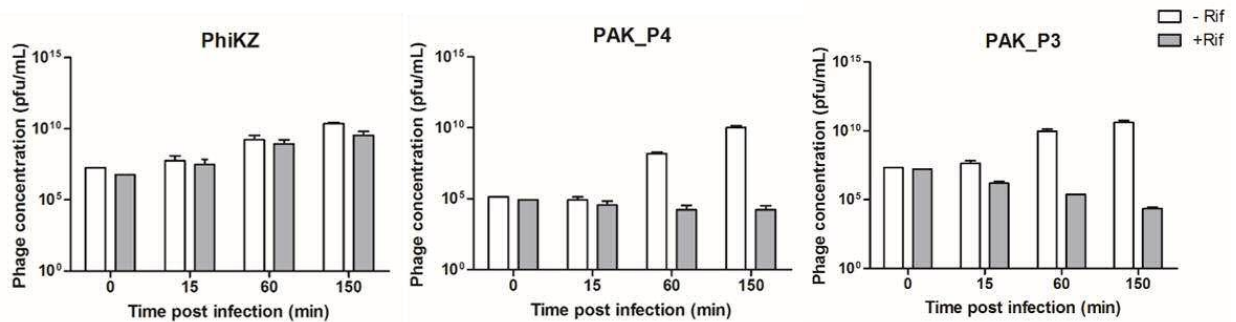


Figure 38: PAK_P3 and PAK_P4 amplification are dependent of host RNAP

Phages (PAK_P3, PAK_P4 or PhiKZ) were added (MOI=1) at t=0 in cultures supplemented with rifampicin (Rif; $100\mu\text{g}.\text{mL}^{-1}$) (grey bars) or not (white bars). Samples were collected after 0, 15, 60 and 150min of phage-bacteria incubation and phage were titred. Means and error bars of 3 independent experiments are represented.

2. PAK_P3 induces changes in host metabolome

We next investigated variations in the composition of cellular metabolome occurring during viral infection. These experiments were performed on PAK_P3 infected cells only.

Global variation of metabolite content

Viruses depend on host cell metabolic resources to complete their intracellular parasitic development³⁴. However, the effects of phage infection on host metabolism are still poorly understood. We thus investigated whether the phage completely shuts off host metabolism, as it may burden efficient phage replication, or if it specifically influences particular pathways.

To assess the impact of PAK_P3 infection on strain PAK metabolism, high-coverage metabolomics analysis was applied to monitor metabolite dynamics during infection¹⁸¹. Comparison of the metabolite levels at different time points post infection to uninfected samples revealed significant metabolic changes upon phage infection. Within the first 5 min of infection, 22% of measured metabolites display altered levels with 13.8% increased and 8.5% decreased ($p\text{-value} \leq 0,05$, $|\text{Log}_2(\text{fold change})| \geq 0,5$). The proportion of metabolites with increased levels gradually rises up to 22% at 25 min post infection, while the proportion of metabolites with decreased levels temporarily drops to 3% to finally increase back to 13% during bacterial lysis (Figure 39).

These variations indicate that PAK_P3 does not simply deplete available host metabolites but relies on an active metabolism in agreement with recent observations identifying phage-specific physiological alterations^{12,182}.

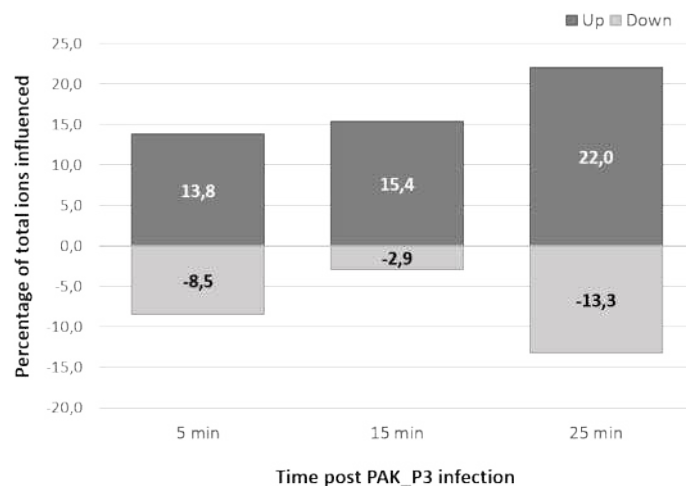


Figure 39: PAK_P3 alters *P. aeruginosa* metabolite content over the course of infection

Percentage of altered *Pseudomonas* metabolite ions during the course of infection ($p\text{-value} \leq 0,05$, $|\text{Log}_2(\text{fold change})| \geq 0,5$), y-axis shows percentage and x-axis shows the time points during infection. In total 377 ions were measured.

Drastic changes in amino acid, nucleotide-sugar and pyrimidine pathways

Next, to investigate whether PAK_P3 targets specific metabolic pathways, a metabolite set enrichment analysis was performed. Overall, metabolites from amino/nucleotide sugar and pyrimidine metabolic pathways were found over-represented among increasing metabolites, while amino acid-related pathways were enriched among decreasing metabolites at later stages of infection (Figure 40).

- **Pyrimidine metabolism**

Intriguingly, about 50% of the detected (deoxy)nucleotides-phosphates have at least two-fold increased levels during infection (Table 9). This can be explained by the global degradation of host mRNA, revealed by our RNA-Seq experiment, which eventually produces an excess of free ribonucleotides that are likely converted into deoxynucleotides by ribonucleotidases.

Interestingly, PAK_P3 encodes a putative ribonucleotide-diphosphate reductase (alpha and beta subunits respectively Gp67 and Gp69) that could catalyze such a reaction. An alternative explanation for the observed increase of (deoxy)nucleotides-phosphates relies on the putative deoxyribonuclease (Gp57) encoded by PAK_P3, which could be responsible for host genome degradation during middle and late infection stages, as observed for phage LUZ19 ¹¹⁶. Supporting these hypotheses, these 3 phage-encoded auxiliary metabolic genes (AMGs) are strongly expressed by PAK_P3 during late infection stage as they are respectively the 7th, 22nd and 19th most expressed genes over 86 late genes. It is noteworthy that a fourth predicted phage-encoded AMG, a CMP deaminase (Gp155), is also involved in nucleotide metabolism and also expressed during late infection although less intensely than the previous ones. Altogether, these predicted AMGs involved in nucleotide metabolism highlight the central need for nucleotides during PAK_P3 infection, in accordance with the short infection cycle span during which about 50 genomes of 88 kb have to be synthesized.

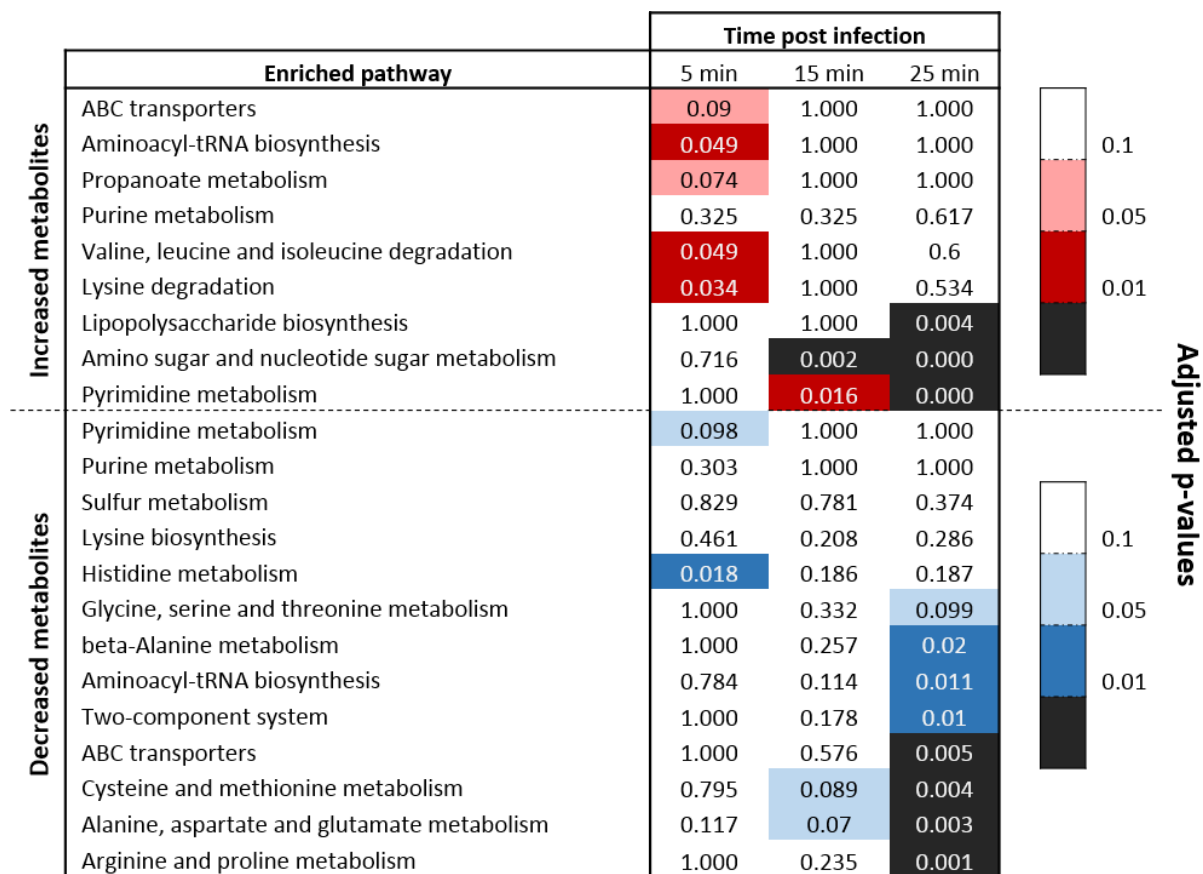


Figure 40: Pathway enrichment analysis during PAK_P3 infection revealed its requirement on pyrimidine metabolism

Values of the enrichment analysis are shown for the significantly enriched pathways (rows) at the different time points (columns). The color scale indicates the cut-off for the p-values, where red and blue are used for, respectively, the pathways found enriched among increased and decreased metabolites.

Table 9: PAK_P3 infection causes a dramatic increase in free deoxynucleotides

This table shows the $\text{Log}_2(\text{fold change})$ for all measured nucleotides over the course of PAK_P3 infection. In grey are those not significantly influenced, the color scale indicates the level of increase (green) or decrease (red) of the metabolite at the specific time point. Metabolite levels displaying more than 2-fold increase are highlighted in bold

Class	Compound	5min		15min		25min	
		$\text{Log}_2(\text{FC})$	p-value	$\text{Log}_2(\text{FC})$	p-value	$\text{Log}_2(\text{FC})$	p-value
Pyrimidine	CMP	0,54	0,00	0,38	0,04	0,73	0,01
	CDP	-0,03	0,81	0,59	0,00	1,28	0,00
	CTP	-0,63	0,00	0,53	0,01	1,61	0,00
	dCMP	0,87	0,00	0,80	0,00	2,04	0,00
	dCDP	0,47	0,00	0,98	0,00	2,75	0,00
	dCTP	-0,07	0,58	1,05	0,00	3,18	0,00
	dTMP	0,60	0,04	1,02	0,02	2,20	0,00
	dTTP	0,01	0,95	0,33	0,01	1,22	0,00
	dUMP	-0,04	0,71	0,08	0,38	0,42	0,01
	UMP	0,31	0,11	0,46	0,01	0,62	0,03
	UDP	-0,37	0,00	0,44	0,04	0,99	0,00
	UTP	-1,06	0,00	0,45	0,10	1,38	0,00
Purine	AMP	0,97	0,00	-0,02	0,97	-0,33	0,26
	ADP	0,48	0,00	0,29	0,01	0,38	0,01
	ATP	-0,38	0,00	0,19	0,32	0,83	0,00
	dAMP	0,87	0,00	0,96	0,00	1,73	0,00
	dADP	0,69	0,00	1,33	0,00	2,59	0,00
	dATP	0,23	0,00	1,01	0,00	2,46	0,00
	GMP	0,62	0,01	0,34	0,18	0,28	0,21
	GDP	0,00	0,99	0,39	0,00	0,81	0,00
	dGMP	0,97	0,00	-0,02	0,97	-0,33	0,26
	dGDP	0,48	0,00	0,29	0,01	0,38	0,01
	dGTP	-0,38	0,00	0,19	0,32	0,83	0,00

- Amino acid and nucleotide/sugar metabolism

Among accumulating metabolites belonging to amino/nucleotide sugar metabolism and to lipopolysaccharide biosynthesis pathway (Figure 40), it is worth noting that the levels of cell wall precursors such as UDP-N-acetyl-D-glucosamine or UDP-N-acetyl-D-galactosaminuronic acid show a significant two- and three-fold increase, respectively, during late infection (Table 10). This increase is not accompanied by altered expression of host genes involved in this pathway (see next paragraph). Therefore, this effect is likely a direct consequence of infection-triggered peptidoglycan degradation and the subsequent release of its precursors. Consistent with this hypothesis, PAK_P3 has a potential AMG (Gp151) similar to a bacterial cell wall hydrolase, which could explain such cell-wall degradation.

Most enriched pathways among decreasing metabolites involve amino acid biosynthesis (Figure 40), more specifically Arg, Pro, Ala, Asn, Glu, Cys and Met metabolism were found significantly enriched (p-value < 0.005). These observed decreases may indicate drainage of amino acid pools in the cell during phage particle formation, due to an imbalance between cellular amino acid biosynthesis and consumption by the phage.

Table 10: PAK_P3 infection influences amino sugar metabolism

This table shows the Log₂(Fold Change) for all indicated metabolites over the course of PAK_P3 infection. In grey are those not significantly influenced, the green color scale indicates the level of increase of the metabolite at the specific time point.

Name	Log ₂ (FC)		
	5 min	15 min	25 min
D-Glucosamine 6-phosphate	0,68	0,85	1,37
alpha-D-Glucosamine 1-phosphate	0,68	0,85	1,37
UDP-N-acetyl-D-galactosamine	0,09	0,45	1,07
UDP-N-acetyl-D-mannosamine	0,09	0,45	1,07
UDP-N-acetyl-D-glucosamine	0,09	0,45	1,07
N-Acetylmuramate	-	0,61	0,78
UDP-N-acetyl-2-amino-2-deoxy-D-glucuronate	-	0,68	1,50
UDP-N-acetyl-D-mannosaminouronate	-	0,68	1,50
UDP-N-acetyl-D-galactosaminuronic acid	-	0,68	1,50
Chitobiose	-	0,16	-
GDP-L-galactose	-	0,77	1,59
GDP-L-gulose	-	0,77	1,59
GDP-6-deoxy-D-talose	0,64	0,96	1,42
GDP-6-deoxy-D-mannose	0,64	0,96	1,42

Variations in host metabolome composition are not otherwise mediated through differential expression of host genes

We initially hypothesized that the observed changes in metabolome composition upon infection would largely be the result of a differential expression of host genes induced by the phage. This would indicate that PAK_P3 mainly interferes with cellular transcription to alter host physiological processes. To address this question, we investigated if the variations at the metabolome level could be directly linked to transcriptional changes. We thus analysed all metabolites belonging to pathways highlighted by the pathway enrichment analysis (see above) that display significant variations ($|\text{Log}_2(\text{fold change})| > 0.5$, $p\text{-value} < 0.05$) as well as differential expression of coding sequences related to the corresponding pathways with a stringent cut-off point ($|\text{Log}_2(\text{fold change})| > 1.3$, $p\text{-value} < 0.05$) (Figure 41).

Only few genes linked to these pathways were significantly differentially expressed upon late infection, indicating that the phage influence on host metabolism is not primarily mediated through differential gene expression. In fact, several pathways with increased metabolite levels have a decreased transcription of the involved genes or vice versa (e.g. lipopolysaccharide biosynthesis). An explanation can be found in the limited turnover of host metabolic enzymes during the short phage infection cycle, rendering the impact of differential expression of the genes encoding essential metabolic enzymes limited. Alternatively, as mentioned above, PAK_P3 encodes several AMGs but also a large number of proteins that share no homology to any known enzyme. Therefore, it is very likely that more AMGs influencing host metabolism remain to be characterized.

Based on these complementary “-omics” approaches, it can be concluded that PAK_P3 does not otherwise redirect host metabolism towards viral reproduction through modification of host gene expression. The general degradation of host RNA observed likely ensures sufficient building blocks for viral genome replication and transcription. The metabolic content of PAK_P3 infected cells shows both increased and decreased metabolite levels. We hypothesize these changes are either the direct consequence of an increased viral consumption of metabolites (e.g. amino acid metabolism) or are likely triggered by phage-encoded AMGs (e.g. pyrimidine metabolism).

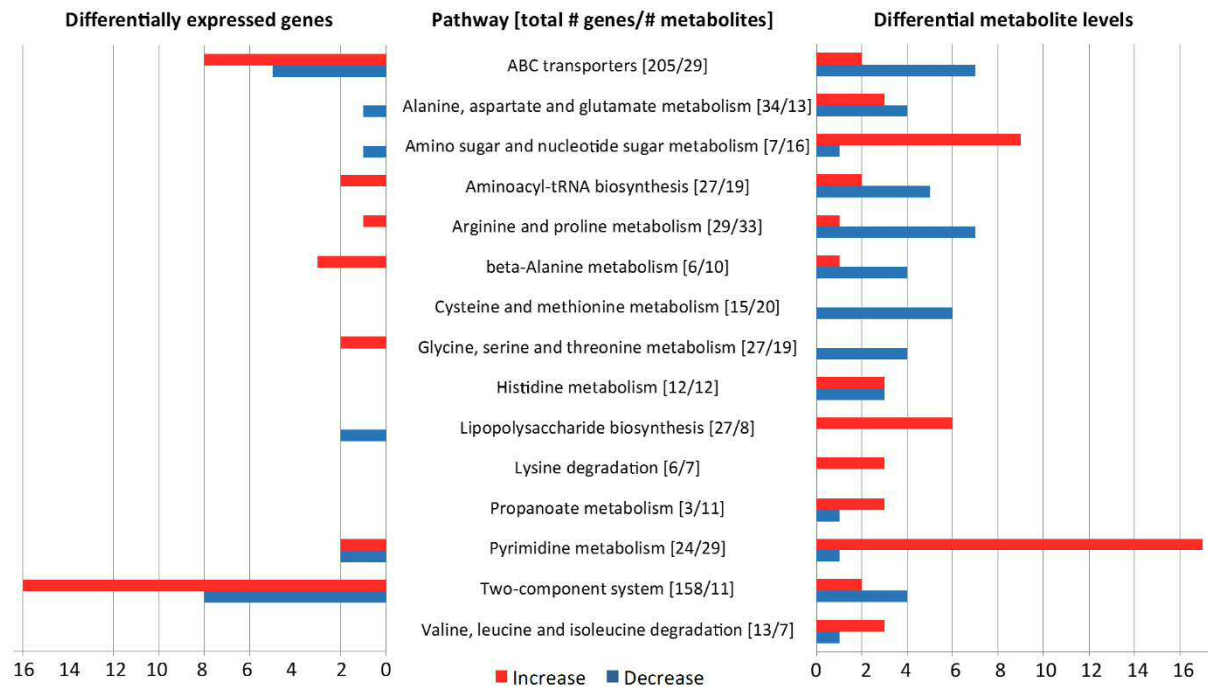


Figure 41: Comparison of significant changes in transcriptomics and metabolomics data from PAK_P3 infected cells reveals no direct correlation

On the left, the number of genes with a significant differential expression ($|\text{Log}_2(\text{fold change})| > 1.3$, $p\text{-value} < 0.05$) is shown. On the right, an overview of the number of metabolites with significantly changed levels ($|\text{Log}_2(\text{fold change})| > 0.5$, $p\text{-value} < 0.05$) is shown. The middle column entails all studied metabolic pathways and indicated between brackets are, respectively, the total number of genes and metabolites involved in this pathway. Red = increase in transcript/metabolite level, Blue = a decrease in transcript/metabolite level.

3. Modulation of the expression of iron uptake-related host genes upon PAK_P4 infection

In contrast to PAK_P3, it appears that PAK_P4 infection influences the expression of host genes involved in specific metabolic pathways.

Indeed, we found that PAK_P4 strongly upregulates the expression of three host operons predicted to be related to siderophore synthesis and transport, which are not upregulated during PAK_P3 infection (Table 11):

- *PAK_4094-4104* (mean $\text{Log}_2(\text{FC}) = 6.1 \pm 1.1$) and *PAK_4106-09* (mean $\text{Log}_2(\text{FC}) = 6.3 \pm 0.4$) are respectively homologous to the *pchEFG* and *pchABCD* operons of *P. aeruginosa* PAO1¹⁸³, which are involved in pyochelin biosynthesis, one of the major siderophores of *P. aeruginosa*
- *PAK_4090-93* (mean $\text{Log}_2(\text{FC}) = 5.4 \pm 0.5$) is homologous to *fpt* system, encoding the corresponding siderophore receptor.

Interestingly, another up-regulated operon, *PAK_1971-78* (mean $\text{Log}_2(\text{FC}) = 3.0 \pm 0.3$) is homologous to the *pvdIJD* operon¹⁸³, encoding non ribosomal peptide synthetases involved in the biosynthesis of the other major *P. aeruginosa* siderophore pyoverdine.

Moreover, *PAK_4645-51* and *PAK_4652*, respectively homologous to the *phuSTUVW* operon and *phuR* receptor gene¹⁸³, are also up-regulated, although to a lesser extent (mean $\text{Log}_2(\text{FC}) = 2.3 \pm 1.0$). These genes encode an ABC transporter involved in heme uptake¹³⁹.

Finally, 14 other genes putatively involved in iron acquisition (mostly tonB-dependent receptors) are up-regulated upon infection (mean $\text{Log}_2(\text{FC}) = 2.0 \pm 0.4$) while *PAK_4711*, a Fur-like protein (ferric uptake regulator), which represses the expression of genes involved in iron uptake, storage and metabolism in the presence of sufficient iron¹⁸⁴, is downregulated (Table 11).

As this up-regulation of iron uptake-related genes may be the downstream consequence of a general activation of stress responses, we investigated the level of expression of alternative sigma factors. Among the 19 alternative sigma factors that are involved in orchestrating stress responses, AlgU is the only sigma factor that is meaningfully transcriptionally up-regulated upon PAK_P4 infection. We subsequently looked at the expression of *algU* regulon that typically includes genes involved in alginate, peptidoglycan and LPS biosynthesis, and found that none of them were up-regulated as previously reported (see Table 8 in the previous chapter). Therefore, the differential expression of iron-related genes does not appear to be part of a more general stress response.

In addition, the iron starvation-associated alternative sigma factor PvdS positively regulates the expression of genes involved in pyoverdine biogenesis and transport under iron-limited conditions, along with genes involved in extracellular virulence factors (*i.e.* protease and exotoxin A¹⁸⁵). While pyoverdine genes are induced upon PAK_P4 infection, neither *pvdS*, nor other genes regulated by PvdS are differentially expressed. Also, other sigma factors involved in iron uptake are not differentially

expressed either. Since no upstream regulators are activated, it is likely that pyoverdinin up-regulation is indeed driven by PAK_P4 rather than part of a host response.

Besides iron-related genes, we observed that genes *rpoB* and *rpoC* encoding the two beta subunits of RNAP (mean $\text{Log}_2(\text{FC})=2.0$), along with 10 genes encoding ribosomal proteins (mean $\text{Log}_2(\text{FC})=1.7$) are up-regulated, while they are not differentially expressed in PAK_P3 infected cells. Overall, it appears that PAK_P4 infection causes a transcriptional activation of several host operons involved in metabolic processes such as *sucABCD-lpdA* (TCA cycle, amino acid metabolism, mean $\text{Log}_2(\text{FC})=2.1$) or *nqrCDF* (energy metabolism, mean $\text{Log}_2(\text{FC})=1.7$), *ccoOP-acnA* (energy metabolism, mean $\text{Log}_2(\text{FC})=1.5$), among other examples.

In summary, we demonstrated in this chapter that RNA metabolism is central in both PAK_P3 and PAK_P4 infections, since both phages trigger a rapid degradation of host mRNA coupled with their replacement with phage transcripts, although with different kinetic. Consistently, metabolomics analysis of PAK_P3 infected cells revealed an activation of pyrimidine metabolism pathway.

In addition, we hypothesize that PAK_P4 is likely to redirect host metabolism processes through manipulation of the expression of corresponding genes, while PAK_P3 would rather rely on its own auxiliary metabolic genes and/or co-opts host enzymes through post-transcriptional manipulations.

Altogether, these results evidence that both viruses actively reprogram the metabolism of their host cell and, consequently, raise the question of the persistency of a host-driven transcriptional response which we are exploring in the next chapter.

Table 11: Differential expression of genes involved in iron uptake upon phage infection

Values of differential expression ($\text{Log}_2(\text{FoldChange})$) upon 13min infection compared to control (non-infected cells) are indicated. Significantly (*i.e.* p value < 0.05 and $|\text{Log}_2(\text{FC})| > 1.3$) up-regulated and down-regulated genes are coloured in green and red, respectively. "0" indicates that the indicated $\text{Log}_2(\text{FC})$ is not significant.

Gene	Annotation	UP DOWN	PAK_P4		PAK_P3		PAO1 homologs
			52 4		4 7		
			$\text{Log}_2(\text{FC})$	Variation	$\text{Log}_2(\text{FC})$	Variation	
PAK_n016	PrrF1 (RyhB analog)		-1,4	DOWN	-2,6	DOWN	PrrF1
PAK_n017	PrrF2 (RyhB analog)		0,6	0	-2,7	DOWN	PrrF2
PAK_4645	conserved hypothetical protein		0,8	0	-0,5	0	phuW
PAK_4646	hypothetical protein		1,9	UP	-0,8	0	phuV
PAK_4647	ABC transporter family protein		2,1	UP	-0,9	0	
PAK_4648	fecCD transport family protein		2,2	UP	-0,9	0	
PAK_4649	fecCD transport family protein		1,5	UP	-1,0	0	phuU
PAK_4650	periplasmic binding family protein		2,6	UP	-0,4	0	phuT
PAK_4651	haemin-degrading HenS,ChuX domain protein		3,6	UP	0,3	0	phuS
PAK_4652	tonB-dependent hemoglobin/transferrin		3,7	UP	-1,5	0	phuR
PAK_4653	rieske [2Fe-2S] domain protein		-0,8	0	0,4	0	PA4711
PAK_4654	putative lipoprotein		-0,7	0	-0,9	0	PA4712
PAK_4090	siderophore transporter, RhtX/FptX family protein		5,8	UP	0,1	0	fptX
PAK_4091	putative membrane protein		5,8	UP	0,2	0	ampO
PAK_4092	putative membrane protein		4,7	UP	0,4	0	
PAK_4093	tonB-dependent siderophore receptor family protein		5,3	UP	-0,1	0	fptA
PAK_4094	hypothetical protein		3,3	UP	1,5	0	
PAK_4095	hypothetical protein		5,2	UP	1,1	0	
PAK_4096	ABC transporter family protein		6,2	UP	1,6	UP	?
PAK_4097	hypothetical protein		5,6	UP	1,3	0	
PAK_4098	thiazolyl imide reductase family protein		6,4	UP	0,9	0	pchG
PAK_4099	methyltransferase domain protein		6,9	UP	1,2	0	
PAK_4100	amino acid adenylation domain protein		7,0	UP	0,7	0	pchF
PAK_4101	condensation domain protein		6,8	UP	0,3	0	
PAK_4102	phosphopantetheine attachment site family protein		7,0	UP	0,7	0	
PAK_4103	amino acid adenylation domain protein		6,4	UP	0,3	0	pchE
PAK_4104	phosphopantetheine attachment site family protein		6,3	UP	2,1	UP	
PAK_4106	AMP-binding enzyme family protein		6,3	UP	0,8	0	pchD
PAK_4107	alpha/beta hydrolase family protein		6,3	UP	1,4	0	pchC
PAK_4108	chorismate mutase related enzymes family protein		5,7	UP	0,4	0	pchB
PAK_4109	isochorismate synthase family protein		6,7	UP	0,2	0	pchA
PAK_1970	tonB-dependent siderophore receptor family protein		3,9	UP	-0,8	0	fpvA
PAK_1971	alpha/beta hydrolase family protein		2,9	UP	0,3	0	
PAK_1972	condensation domain protein		2,6	UP	0,7	0	
PAK_1973	phosphopantetheine attachment site family protein		3,3	UP	3,1	UP	
PAK_1974	amino acid adenylation domain protein		2,8	UP	0,6	0	
PAK_1975	phosphopantetheine attachment site family protein		2,7	0	2,3	UP	pvd DJI
PAK_1976	non-ribosomal peptide synthase TIGR01720 domain protein		3,4	UP	0,8	0	
PAK_1977	non-ribosomal peptide synthase TIGR01720 domain protein		3,4	UP	0,4	0	
PAK_1978	AMP-binding enzyme family protein		2,8	UP	0,2	0	
PAK_112	tonB-dependent Receptor Plug domain protein		2,0	UP	0,8	0	
PAK_2194	tonB-dependent Receptor Plug domain protein		1,4	UP	1,2	0	
PAK_294	tonB-dependent Receptor Plug domain protein		2,2	UP	-0,2	0	
PAK_295	tonB-dependent receptor family protein		2,2	UP	0,0	0	
PAK_3205	tonB-dependent hemoglobin/transferrin/lactoferrin receptor family prot		1,5	UP	-0,8	0	
PAK_3738	tonB-dependent siderophore receptor family protein		1,7	UP	-0,3	0	
PAK_4023	tonB-dependent Receptor Plug domain protein		1,8	UP	-0,5	0	
PAK_4036	tonB-dependent siderophore receptor family protein		2,5	UP	-0,3	0	
PAK_4253 [feol]	ferrous iron transport protein B		1,5	UP	0,5	0	
PAK_4610	tonB-dependent siderophore receptor family protein		2,2	UP	-0,6	0	
PAK_6085	tonB-dependent siderophore receptor family protein		2,0	UP	-0,2	0	
PAK_7	secretin and tonB N terminus short domain protein		2,9	UP	-0,3	0	
PAK_703	tonB-dependent hemoglobin/transferrin/lactoferrin receptor family prot		1,5	UP	-0,7	0	
PAK_770	tonB-dependent siderophore receptor family protein		1,7	UP	0,0	0	
PAK_1317	iron-sulfur cluster biosynthesis family protein		-1,4	DOWN	0,6	0	
PAK_4711	ferric uptake regulator family protein		-2,0	DOWN	-0,9	0	
PAK_6053	tonB family C-terminal domain protein		-2,2	DOWN	-0,4	0	

CHAPTER 5 - IDENTIFICATION OF A COMMON HOST TRANSCRIPTIONAL RESPONSE TO PHAGE INFECTION

Here, we determine how phage infections modulate host gene expression more broadly (not only influence on metabolic genes). Since we established that both PAK_P3 and PAK_P4 trigger a dramatic global decline of host transcription, we examine whether we can still detect specific modifications in gene expression that could suggest the setting up of a host response. In particular, we investigate how similar is the host transcriptional landscape over the course of PAK_P3 and PAK_P4 infections.

1. Prerequisite: reannotation of strain PAK genome using transcriptomic data

Prior to investigate the effect of phage infection on host transcription in details, it was first crucial to exhaustively characterize the host genome of strain PAK. Initially, collaborators (P. Jorth and M. Whiteley) provided us with a draft genome they assembled and annotated (6.28 Mbp, 66.3% GC content and 6,267 predicted ORFs). Using our RNA-Seq data, generated from exponentially growing uninfected PAK cells, we performed a detailed reannotation of this genome.

We manually reannotated 32 ORFs (by detection of an alternative start codon) and defined 63 new putative coding sequences. Among these 63, 39 have been previously annotated in other *P. aeruginosa* genomes while the other 24 are new hypothetical coding sequences that display no homology to sequences in databases and may be considered as strain-specific.

We also identified a total of 75 small ncRNAs encoded in the PAK genome, 26 of which correspond to known functional classes (Figure 42). Among these 26, 12 are similar to uncharacterized ncRNA conserved within the *Pseudomonas* genus and the other 14 have predicted functional assignments, according to Rfam database. The majority of ncRNAs (49 out of 75) could not be assigned to any functional class, and have not been identified in previous RNA-Seq investigations carried out on *P. aeruginosa* strains PAO1 and PA14¹⁸⁶⁻¹⁸⁸, suggesting that they may represent novel ncRNAs regulators.

Eighteen long antisense RNAs (asRNAs) were also identified within genes. As they do not display any consistent ORFs, they are not likely to contain overlapping protein-coding genes and may therefore *cis*-interfere with the expression of gene they are encoded in (Figure 42). Finally, 32 potential riboswitches were identified by looking at intergenic transcription events starting at a significant distance from a downstream gene, usually involved in a metabolic pathway and displaying a characteristic RNA-Seq pattern. Eleven of them were confirmed by Rfam searches (Figure 42).

With more than 50% of new ncRNAs amongst total ncRNAs identified, along with the identification of new putative riboswitches and evidence of antisense transcripts, strain PAK exemplifies the great diversity of bacterial RNA-based regulation¹⁴⁴. Such in-depth annotation, including new strain-specific RNA elements, was mandatory for the subsequent transcriptomic analysis of phage infected cells in order to assess the impact of phage infection on host physiology.

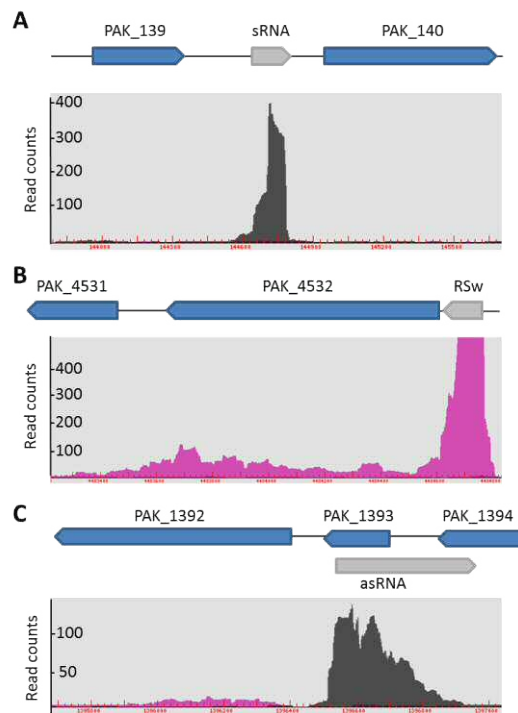


Figure 42: Strain PAK displays high number of non-coding RNAs

A representative example of detection of **(A)** intergenic small RNA (sRNA); **(B)** riboswitch (RSw) and **(C)** *cis*-antisense RNA (asRNA). The mapped reads were formatted into graph files for visualization in a strand-specific manner (black and pink represent reads mapping the forward and the reverse strands, respectively) using COV2HTML. The annotated non-coding RNA genes are indicated as grey arrows and open-reading frames annotated in strain PAK are shown as blue arrows.

2. A common host response is elicited during infection by PAK_P3 and PAK_P4

Although host transcripts are globally replaced by phage transcripts, we could still analyze changes in host mRNA population by normalizing the host transcript counts before infection to the counts after infection, artificially depleting counts before infection and enriching reads after infection. This allows us to look for specific differential expression of host gene features in response to the stress induced by phage infection.

Overall, we found 95 host genes with an increased expression while 975 genes were down-regulated upon PAK_P3 infection. In comparison, the impact of phage infection on gene expression appears to be different in PAK_P4 infected cells, with a ratio of up- versus down-regulated genes more balanced (352 versus 578, respectively). These gene sets contain common elements: 55 upregulated and 445 downregulated host coding sequences have been identified in common in response to PAK_P3 and PAK_P4 infections (Figure 43). While the latter group may mainly be the result of host mRNA degradation, the former could more likely represent a general transcriptional host response to phage infection.

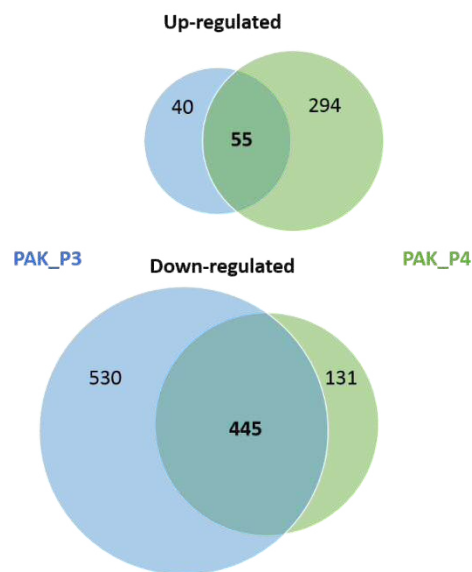


Figure 43: Overview of differentially expressed genes upon PAK_P3 and PAK_P4 infections

Venn diagrams indicate the number of up-regulated (upper panel) and down-regulated (bottom panel) host genes upon PAK_P3 (blue) or PAK_P4 (green) infections. Numbers of common differentially regulated genes are indicated.

Specifically, both infections prompt the transcriptional up-regulation of a predicted P2-like prophage (Figure 44). However, in both cases, viral infection induces global host mRNA depletion faster than the activated prophage-specific transcription can compete with, eventually leaving the infected cell with marginally fewer prophage transcripts than during exponential growth. This indicates that the transcriptional activation of the prophage may be aborted by the non-specific global mRNA degradation.

Moreover, one operon linked to RNA processing (*PAK_4493-4499*) is upregulated during both phage infections, albeit more weakly upon *PAK_P4* infection compared to *PAK_P3* (mean $\text{Log}_2(\text{FC}) = 2.5$ vs mean $\text{Log}_2(\text{FC}) = 6.4$, respectively) (Figure 44). Therefore, the level of upregulation of this operon between both phages inversely correlates with the extent and speed of RNA degradation.

The question of the origin of this phenomenon then arises: is it a phage-driven mechanism, which both viruses have in common, or is it a host defense mechanism?

This operon encodes a RNA 3'-phosphate cyclase *RtcA* (*PAK_4496*) that has been described as being involved in the processing of RNA transcripts such as priming RNA strands for adenylation to protect them from exonucleases, or to tag them as «substrates» for downstream reactions performed by additional enzymes^{189,190}. Therefore, we could hypothesize that this operon may be uniquely co-opted by the phage in order to participate in the global degradation of RNAs by tagging transcripts for degradation by host RNA degradosome or phage-encoded enzymes.

An alternative hypothesis to explain this dramatic up-regulation of *PAK_4493-4499* relies on *RtcB* (*PAK_4494*), a predicted RNA ligase. Together, the *RtcAB* system has been shown to play a role in tRNA repair after stress-induced RNA damage (e.g. viral infection) in *E. coli*¹⁹¹. Since a weaker upregulation of this operon is correlated to a more extensive degradation of host transcripts upon *PAK_P4* infection compared to *PAK_P3*, it could reflect an attempt of the host to respond to extensive RNA damages imposed by both phages. This reparation system would be less active, and thus more weakly protective, in *PAK_P4* infected cells, which may explain the different kinetics of RNA degradation upon both infections.

To solve this question, we generated a deletion mutant of the RNA ligase *rtcB* gene and expected to observe either a decreased or an increased efficiency of phage infection. However, we detected no differences in lysis kinetics or ability to form plaques on this $\Delta\textit{rtcB}$ strain. Therefore we cannot fully discriminate between the two scenarios yet, and we believe that a double mutant $\Delta\textit{rtcAB}$ is likely to provide more meaningful conclusions.

Among other genes up-regulated in common, either their very low expression (less than 30 reads) suggests that their increased expression may not be biologically relevant, or they are predicted to encode hypothetical proteins preventing us to draw meaningful conclusions.

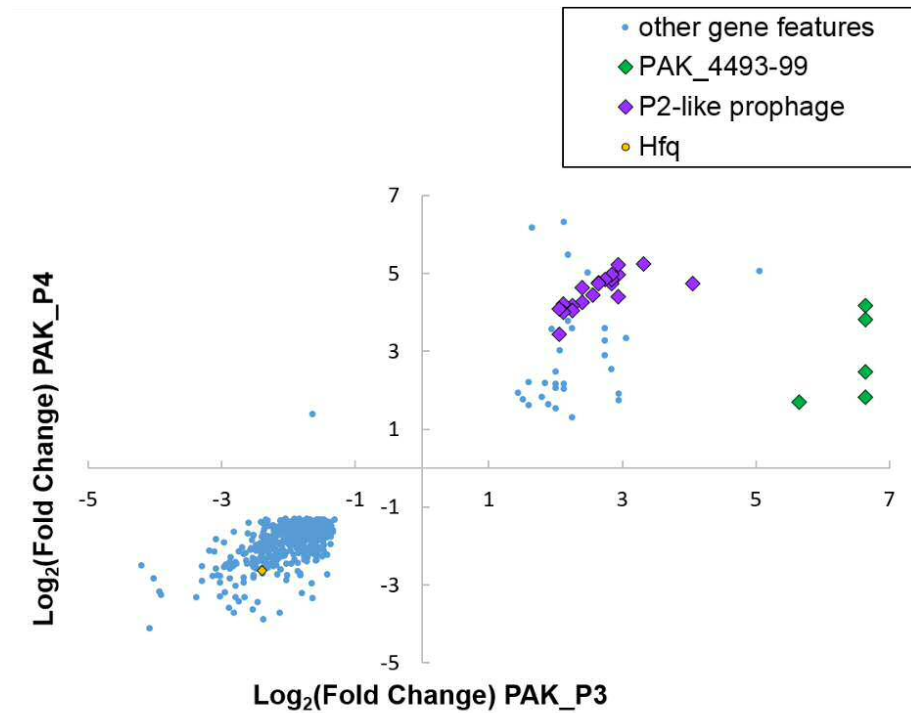


Figure 44: Host strain specifically up-regulates the expression of one operon involved in RNA processing in response to both phage infections.

Host genes significantly ($p < 0.05$) differentially expressed ($|\text{Log}_2(\text{FC})| > 1.3$) between $t=0$ and $t=13$ min upon both PAK_P3 and PAK_P4 infections were listed and their $\text{Log}_2(\text{FC})$ values upon each infection are plotted.

To conclude, despite the global degradation of host RNAs, specific modulation of host gene expression could still be detected, notably with the activation of one particular operon in response to both phages that likely represent a host defence reaction. Apart from this common feature, transcriptional landscapes of PAK_P3- and PAK_P4-infected cells are rather distinct. Indeed, PAK_P4 uniquely up-regulates genes involved in various metabolic pathways (as reported previously), while genes specifically up-regulated in PAK_P3-infected cells have either a very low expression or an uncharacterized function. Overall, PAK_P3 infection does not extensively stimulates expression of specific gene, as it can be seen on Figure 45 (above upper dash line), contrary to PAK_P4.

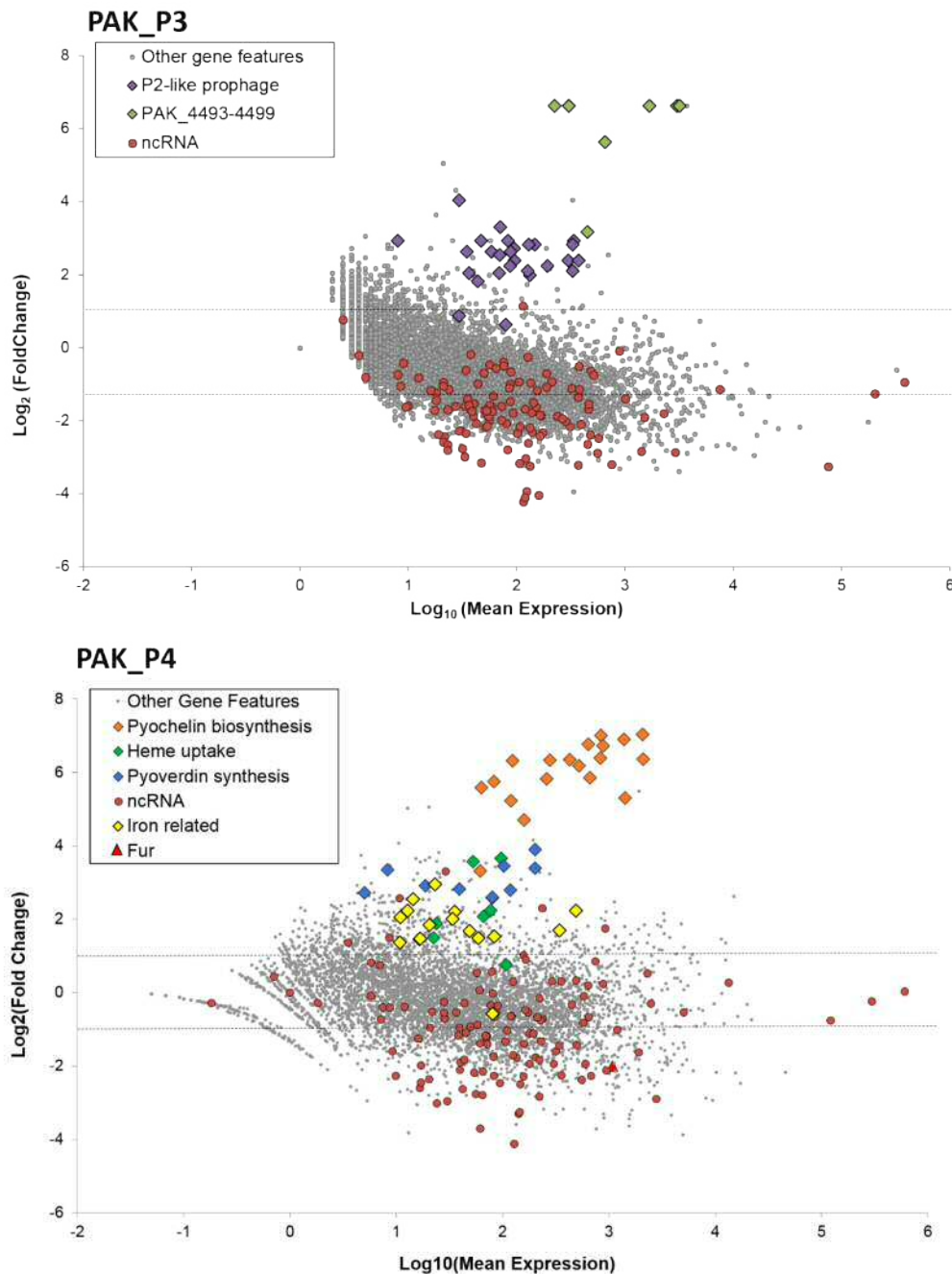


Figure 45: Phage-induced alteration of host gene expression during late infection

Differential expression analyses of host gene features comparing transcript abundance between phage negative controls (t=0 min) and late infections (t= 13min) were performed (PAK_P3: Upper panel, PAK_P4: bottom panel). Dash lines indicate the cut-off used to consider genes as differentially expressed ($|\text{Log}_2(\text{FC})| > 1.3$). This comparison was made after normalizing the read counts that map to each host gene feature between both conditions, ignoring reads that map to the phage, which artificially enriches host transcripts in late infection relative to the total transcripts in the cell. This method allows specific effects on host transcription to be tested for independently of the global replacement of host RNAs by phage RNA

PART 3

-

CONCLUSIONS
AND
PERSPECTIVES

My PhD work investigates the infection cycles of therapeutically valuable phages, namely PAK_P3 and PAK_P4, belonging to the newly described genera *Kpp10virus* and *Pakpunavirus*. Based on an integrative threefold approach combining transcriptomics, metabolomics and molecular biology, this work provides a comprehensive view of the mechanisms underlying the transformation of a *P. aeruginosa* “ribocell” into a “virocell”.

In this final part, I will summarize the main findings reported in this manuscript, through the description of an integrated model of phage infection cycle, and further discuss specific interesting mechanisms discovered during our investigations. I will also take the opportunity to emphasize the evolutionary aspects emerging from this study.

1. OVERVIEW OF PAK_P3 AND PAK_P4 INFECTION CYCLES

A summarizing illustration of PAK_P3 and PAK_P4 infection cycles is given in Figure 46.

The ‘virocellular’ conversion begins when virions encounter their ‘ribocellular’ host and irreversibly attach to it. During this so called adsorption process, the infecting phage engages molecular interactions with surface structures (e.g. flagellum, pilus) and eventually binds to its main receptor (LPS).

Once phage genomic material has penetrated the bacterial host, cellular transcriptional landscape enters into a transition phase during which virocell-associated mRNAs progressively replace ribocell-derived transcripts. First, conserved σ^{70} -like promoter motifs specifically present upstream phage early genes are recognized by the host RNAP¹⁵⁸. Early gene products are then predicted to orchestrate the redirection of bacterial physiological and metabolic processes towards the production of progeny phages. In particular, some of these phage early proteins are dedicated to the limitation of host defensive reactions (e.g. restriction enzymes) while others mediate the takeover of host machineries through direct interactions with (or modifications of) bacterial proteins.

- In particular, interactions between phage early proteins and host housekeeping sigma factors have been demonstrated to shutoff host transcription, thereby making the RNAP core available for phage-directed transcription. By contrast, interactions between phage proteins and alternative sigma factors leading to modulation of associated stress responses have not been reported (at least to my knowledge). Such interplays are yet very likely to occur since it is well established that host stress responses are activated upon phage infection. Indeed, in addition to transcriptional activations of various stress-associated genes that have been recorded in recent transcriptomic studies^{117,192}, direct modulations of the expression of these alternative sigma factors have been

evidenced. For instance, *B. subtilis* σ^W regulon is induced by phage SPP1 infection¹⁹³ while *sigF* expression is down-regulated by mycobacteriophage SWU1¹²². Moreover, a direct interaction of *P. aeruginosa* core RNAP with SigX induced by LUZ19 infection has been demonstrated¹⁶⁹. The discovery of early proteins such as Gp92 and Gp70.1¹⁰⁷ may explain how phages shut off host stress responses.

During the early stage of infection, we detected antisense transcription in the structural genomic region which contains lately expressed genes. Phage-encoded antisense RNAs are increasingly being discovered^{115,117,172,194}, however their functions remain largely unknown. In the particular case of PAK_P3 and PAK_P4, their specific location together with their temporal expression led us to hypothesize that they may represent an additional regulatory level of viral gene expression. Indeed, we propose that these asRNAs inhibit a promoter leakage-associated expression of late structural genes, by base-pairing with corresponding transcripts, thereby preventing their translation and/or inducing their degradation.

- Alternatively, we can hypothesize different functions for phage-encoded asRNAs. They could represent a way to protect phage mRNAs against massive RNA degradation, as it has been suggested for cyanophage P-SSP7¹⁹⁴. Indeed, Stazic and co-workers proposed the following model: Transcription of phage asRNAs covering the entire phage genome allows the extensive formation of dsRNA duplexes. Concurrently, a phage-induced increase in RNase E protein level leads to the degradation of host transcripts while phage mRNAs are protected from RNase E activity by asRNAs^{194,195}.

These findings are of major importance since the central question of how phage mRNA stability is maintained during infection has remained unanswered here.

In this work, we reported an extensive RNA degradation, more or less rapid depending on the phage, occurring over the course of infection. Such phenomenon has been observed in numerous other phage-host models (as introduced in Part1 – Chapter 3) and it has consistently been shown that half-lives of host mRNAs were significantly reduced whereas phage mRNAs were stabilized in T4-infected cells (and also in T2 and T7¹⁹⁶). However, mechanisms by which a phage discriminates between its own transcripts and host-derived transcripts have remained unclear. The work of Stazic and colleagues described above provided one possible mechanism.

- Other mechanisms for regulating phage mRNA stability have been suggested. Jian *et al* (2016)¹⁹⁷ have reported the existence of long 5'- untranslated regions enhancing the stability of phage SW1 transcripts and proposed that such posttranscriptional mechanism may be involved in phage regulatory strategies. Alternatively, phages have also been found to stabilize their mRNAs by

inhibiting host RNA degradation machinery as exemplified by T7-induced inhibitory phosphorylation of RNase E¹⁹⁸ or by the direct inhibition of *P. aeruginosa* RNA degradosome mediated by the PhiKZ-encoded protein Dip¹⁰⁹. In the latter system, Van den Bossche and co-workers proposed a two-step model of RNA degradation in PhiKZ-infected cells: (i) During the early infection, the decrease of bacterial mRNA level may involve the inhibition of host RNAP along with an active decay of host RNA mediated either directly by an unknown PhiKZ-encoded enzyme or indirectly by a PhiKZ-activator of host RNA degradation machinery. This hypothesis is sustained by the discovery of Srd protein of phage T4 which enhances the activity of RNase E⁹⁸. (ii) In a second step, newly synthesized phage RNAs would be protected as a result of the inhibition of RNA degradosome mediated by Dip.

Once phage early proteins have sabotaged cellular processes, the next step in phage lifecycle consists in producing elements required for the assembly of new virions. Co-opted host machineries, together with phage-encoded auxiliary metabolic enzymes would then catalyze the reactions yielding the specific metabolites required for '*virocellular*' life.

Specifically, nucleotide turnover is a central need for viral production. Based on our observations, we proposed that phage-induced host RNAs degradation yields a pool of available ribonucleotides that are likely partly converted into deoxynucleotides by phage-encoded ribonucleotidases. These recycled molecules can then be used to generate the required building blocks for viral replication and late transcription. Similar interference strategies are expected to occur with other metabolic pathways. For instance, the detected decrease in metabolites related to amino acid metabolism during PAK_P3 infection may result from an active diversion of amino acids for the assembly of new viral particles.

However, the following questions can be raised: Does hijacking the host metabolism yield enough building blocks required for such rapid virus production? Is a further acquisition of resources from the environment necessary to be used for *de novo* synthesis?

The intensified transcription of iron-acquisition genes upon PAK_P4 infection could result in increased concentrations of intracellular iron thereby supplying cofactors for an optimal functioning of viral auxiliary metabolic enzymes, eventually favoring key processes such as nucleotide turnover.

- Alternatively, this increased expression of iron uptake genes could potentially be explained by the recent '*Ferrojan Horse hypothesis*'¹⁹⁹, proposing that phage tail fibers are bound to iron ions which allow them to use siderophores as receptors and consequently take advantage of bacterial iron uptake mechanisms. However, no HxH motifs, known to facilitate iron binding, have been found in PAK_P4 putative tail fibers and, more generally, PAK_P4 does not possess more HxH motifs throughout its proteome than PAK_P3, which does not up-regulate iron-related genes (14 motifs versus 10, respectively).

A third hypothesis about the role of iron in phage infection is related to host defense shut off. High concentrations of iron have been observed to repress expression of *pqs* operon (*Pseudomonas* quinolone signal) in PAO1²⁰⁰. Recent studies have suggested that a functional quorum sensing system (PQS in particular) increased the ability of *P. aeruginosa* to grow in the presence of phages and thus may participate to phage resistance^{201,202}. Consistently, *P. aeruginosa* PAO1-infected cells displayed increased levels of PQS upon infection by phages YuA, LUZ24 and 14-1, but not upon infections with phages phiKZ, LUZ19 or PEV2¹⁴. This suggests that latter phages have probably evolved unknown mechanisms to block PQS synthesis. Altogether these findings may support the hypothesis suggesting that PAK_P4 induces iron uptake to inhibit a potential activation of the PQS system that would have otherwise resulted in an increased phage resistance in bacterial population and limit viral propagation.

Our data may also have revealed a novel mechanism of bacterial defense against phage infection. Indeed, we hypothesized that the high up-regulation of *rtcAB* expression represent an attempt of the host to repair phage-induced RNA damages.

- A study of RtcAB physiological role in *E. coli* have reported that *rtcAB* genes are activated by agents impairing the translation apparatus (e.g. tetracycline, gentamicin). Authors concluded that this system is probably involved in the maintenance of ribosome homeostasis by favoring rRNA stability²⁰³, thus strengthening our hypothesis for *rtcAB* function.

In summary, we showed in this work that bacterial cell and phage-infected bacterial cell are easily distinguishable by their radically different metabolic orientation and gene expression. Our results support the view of a cellular metabolism almost exclusively dedicated to the production of phages and, given the extent of bacterial gene expression decay, it is very likely that there is not much of bacterial-centered processes left. In other words, the integration of the results obtained in this work are well supporting the concept of *virocell*.

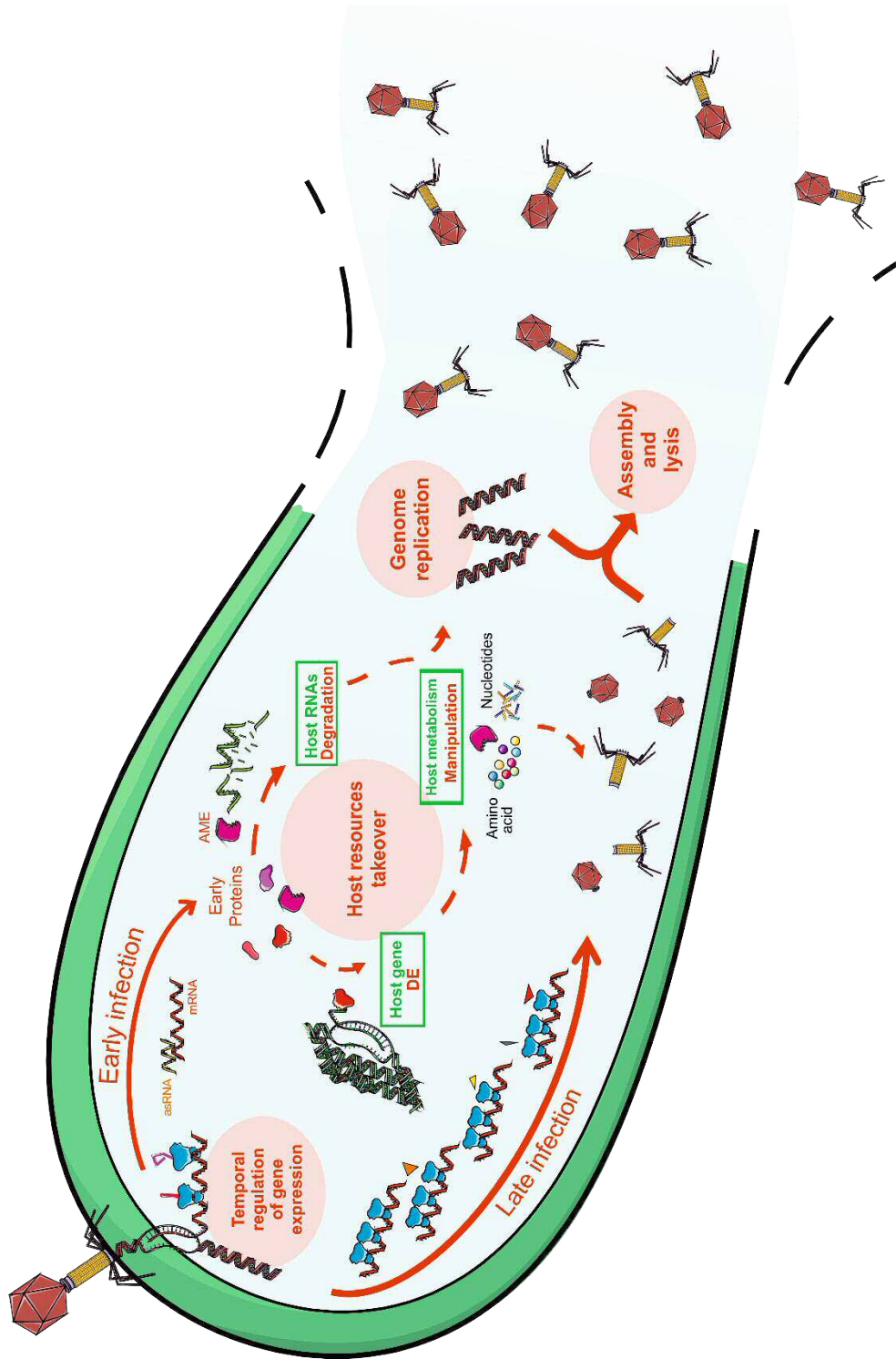


Figure 46: Schematic progression of PAK_P3 and PAK_P4 infection cycles

Red and green colors correspond to phage and host elements respectively. A description of the figure is given in the text (see Discussion). Full arrows depict the main step of phage infection cycle. Dashed arrows represent hypothetical relations between metabolic pathways highlighted by our “-omics” analyses. AME = Auxiliary Metabolic Enzymes, DE= Differential Expression.

2. EVOLUTIONARY PERSPECTIVE EMERGING FROM PAK_P3 AND PAK_P4 COMPARISON

While in cellular organisms, the increasing amount of genomic information is strengthening comparative analyses delineating evolutionary relationships between cellular lineages, the lack of universal genes has hampered such analyses in viruses. In addition, the diversity of phage sequences is such that most genomes contain numerous ORFs displaying no homology to any other sequences in databanks. Therefore, reconstruction of viral evolution cannot be based solely on genomic sequences and requires further information such as elucidation of mechanistic processes ²⁰⁴.

PAK_P3 and PAK_P4 originate from a common ancestor and subsequently diverge enough to be considered as belonging to two distinct genera. By choosing them as models of study, we were able to question whether two phages, with divergent proteomic contents and infecting the same host strain, conserved the infectious strategy inherited from their common viral ancestor during their evolution.

First, we observed that both phages share similar infection parameters although PAK_P4 displays a smaller burst size and a longer latency period than PAK_P3. These observations contrast with the 'optimal lysis timing' model which suggests that longer latency periods result in larger burst sizes²⁰⁵, and require an alternative explanation. Comparison of expressions of putative endolysin-coding genes (*PAK_P4gp29* vs *PAK_P3gp28*), as well as the four downstream ORFs (holins and/or spanins candidates), revealed that these genes are more efficiently transcribed in PAK_P4, which may also appear contradictory with the longer latency period. However, PAK_P4 encodes an asRNA (*PAK_P4as15*), not present in PAK_P3, matching the endolysin gene that could act as a translational repressor, preventing a rapid accumulation of endolysins and delaying lysis. With this supplementary level of lysis control, it would be anticipated that PAK_P4 should be able to produce higher number of progeny than PAK_P3. However, as it is not the case, we hypothesized that both phages may have evolved mechanisms (of host resources co-option or assembly, for example) that are unequally efficient in a given environment. The consequences of these adjustments for the lysis efficiency of each phage may seem of little importance but actually appear to have much broader consequences outside the test tube. This is exemplified by the distinct ability of PAK_P3 to more efficiently treat murine pneumonia compared to PAK_P4 ²⁰⁶.

When exploring further phage molecular lifecycles, we found that (i) they both induce a severe degradation of host mRNAs and (ii) they both redirect host metabolism rather than shutting it off. This could represent the vestiges of an ancestral program that have then been adjusted during evolution giving rise to these two genera (Figure 47).

In addition, we compared how these phages regulate their own gene expression, which appears to be broadly conserved and thus also inherited from their common ancestor. Indeed, we found that both phages display similar temporal regulatory schemes of their syntenic gene sets and that this regulation appears to partially rely on the production of antisense transcripts (Figure 47).

However, beside genes unique to each phage (expected to specifically modulate each infection process), we noted that differences in the transcriptional programs of conserved genes (core genes) may also impact infection cycles, as exemplified with endolysin-coding genes. These observations highlight the extensive panel of solutions that phages have developed to orchestrate bacterial infections (Figure 47).

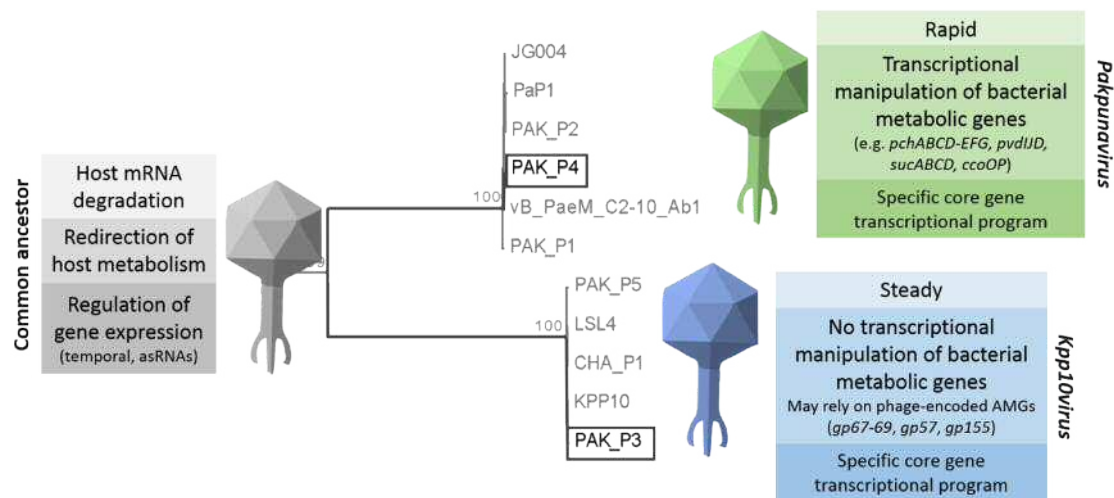


Figure 47: Hypothetical evolution of a phage ancestral program

We hypothesize that ancestral infection mechanisms (induced host RNA degradation, redirection of host metabolism, regulation of phage gene expression) have been differently adjusted within the lineages leading to *Pakpunavirus* (PAK_P4) and *Kpp10virus* (PAK_P3).

In summary, even though they have almost identical morphologies and are able to infect the same host, both phages display substantial variability of their proteomic content, especially within early proteins. Thus, it was unclear whether such related phages would share common mechanisms of action. We showed that the global subversion scheme underlying infection cycle is conserved between phages deriving from a common ancestor although with distinct evolutionary trajectories. These phage-specific adjustments of an ancestral viral program probably explain why they display distinct killing efficiency on a given host strain. The acquisition of unique processes in each phage (*i.e.* mediated by ~100 non-conserved proteins) may also indicate that these viruses adapted to different ecological niches as suggested by their distinct host ranges.

AFTERWORD

This research work, by its global approach, has cleared our general understanding regarding the infection cycles followed by this sub-family of phages and can be used as a basis for numerous future research projects. For example, confirming the function of discovered antisense RNAs may reveal a new mechanism of gene regulation in phages. The role of iron uptake in phage lifecycles, especially its potential link with quorum sensing and phage resistance, would be interesting to investigate as there is growing evidence that phages interfere with (and/or use) communication systems.

Overall, further investigations are now required to fully associate transcriptomics and metabolomics data to viral gene functions, a process which is currently hampered by the lack of versatile genetic tools to construct mutants of virulent phage.

Such effort to deeply characterize particular phage genera is also motivated by the great therapeutic potential of these phages as demonstrated in animal models and recently strengthened by the identification of such phages in the 'Intesti phage' cocktail, a key commercial product of the Eliava Institute in Georgia²⁰⁷.

The knowledge of phage biology provided by next-generation "-omics" approaches not only enlighten viral mechanisms of infection but can also open an array of biotechnological applications based on regulatory elements and proteins found in this new sub-family of phages.

PART 4

-

METHODS

STRAINS AND GROWTH CONDITIONS

P. aeruginosa and *E. coli* strains were usually cultured in LB medium and incubated at 37°C under agitation unless stated otherwise. When necessary, antibiotics were used at the following concentrations: 30µg.mL⁻¹ gentamicin and 300µg.mL⁻¹ carbenicillin for *P. aeruginosa*; 10µg.mL⁻¹ gentamicin, 100µg.mL⁻¹ carbenicillin and 50 µg.mL⁻¹ kanamycin for *E. coli*.

Strains used (or constructed) during this PhD work are listed in Annex 1.

Previously deposited reference genomes of PAK_P3 and PAK_P4 were reannotated during this work. Updated phage genome sequences are available in Genbank and can be accessed via accession numbers NC_022970 (PAK_P3), NC_022986 (PAK_P4).

Strain PAK genome can be consulted via Bioproject accession no. PRJNA232360.

PHAGE BIOLOGY

Host range

The collection of 142 *P. aeruginosa* strains was obtained from Prof. Susanne Häussler (Helmholtz centre for infection research) and has been described in Khaledi *et al.* (2016)²⁰⁸.

Strain collection diversity was evaluated by phage typing using the set of phages indicated below.

Phage	<i>P. aeruginosa</i> isolation strain	Reference
PAK_P1		155
PAK_P2, PAK_P5	PAK	157
LUZ19-PAK, CHA_P1-PAK		Laboratory collection
P3_CHA		156
CHA_P1	CHA	157
LUZ19		209
PhiKZ	PAO1	210
LBL3	Aa245	211
C1A1, C3A1, V1A1	RP73	Laboratory collection

Phage lysates were diluted (ten-fold serial dilutions) in TN buffer (Tris-HCl 10 mM, NaCl 150 mM, pH 7,5) to obtain phage solutions with concentrations ranging from 10^7 to 10^3 pfu.mL⁻¹. Exponentially growing ($0.3 < \text{OD}_{600\text{nm}} < 0.5$) liquid cultures of strains from the collection were used to generate bacterial lawn on agar plates. Five microliters of the successive ten-fold dilutions of each phage lysate were subsequently spotted on these bacteria-covered agar plates (robotically assisted with a liquid-handling robot: Genesis, Tecan). The lysis profile of each tested strain was observed after 16h incubation at 37°C and results were estimated as described in reference 45. Same experiment was performed with PAK_P3 and PAK_P4 to determine their host ranges.

Receptor identification

We obtained a mutant library of *P. aeruginosa* strain PAK generated by random transposon (Tn) mutagenesis using mariner transposon from Dr. Stephen Lory (Harvard Medical School). An exponentially growing ($OD_{600nm} = 0.25$) culture of all mixed mutants (the library was not ordered) was infected with PAK_P3 (MOI = 5) for 1h30 at 37°C and subsequently diluted and plated on LB agar plates supplemented with gentamicin $30\mu g.mL^{-1}$ (mariner Tn carries gentamicin resistance). The same experiment was performed in parallel with a wild type strain PAK. After overnight incubation, colonies growing on agar plates (so called resistant clones) were streaked on fresh LB gentamicin plates. These re-isolated clones were then used to inoculate 150 μL liquid LB gentamicin (in 96-well plates) the next day and grown for 3h at 37°C (with agitation). To ensure that selected purified clones are truly phage resistant (and not false positive), five microliters of these fresh 'resistant' cultures were spotted on agar plates (LB gentamicin) and, once spots were dried, 1 μL (3.10^6 pfu) of phage PAK_P3 or PAK_P4 lysate was dropped on bacterial spots (so called double-spot technique, see ⁴⁵ for additional details). Confirmed resistant clones were then analysed for localisation of Tn insertion by semi-random colony PCR. The first round of PCR is a regular colony PCR using Random PA-1 and Tn-SAM1 primers and the cycles indicated below. Two microliters of the first PCR are then used as template for the second round of PCR (regular PCR with Round2-PA and Tn-SAM2 primers and 3min-extension step). Finally, PCR products are loaded on a 1% agarose gel and amplicons >500bp were gel-purified (NucleoSpin96 PCR Clean Up kit, Macherey Nigel) and sequenced using Tn-SAM Seq primer.

94°C	3'	X10	RandomPA-1	GGCCACGCGTCGACTAGTACNNNNNNNNNNCGATG
94°C	30"		Tn-SAM1	TTGTCGGTGAACGCTCTCCTGAGTAG
50°C*	40"			
72°C	3'		Round2-PA	GGCCACGCGTCGACTAGTAC
94°C	30"	X25	Tn-SAM 2	CCCGCCATAAACTGCCAGGCATC
60°C	40"			
72°C	3'		Tn-SAM Seq	AAGCAGAAGGCCATCCTGAC
4°C	∞			

* Reduce the annealing temperature by 1 °C at each subsequent cycle

Adsorption assay and one-step growth experiment

Bacteriophage growth parameters were assessed as described previously²¹². Briefly, a culture of strain PAK was infected at low MOI (0.1) and incubated 5 min at 37°C with agitation allowing bacteriophage particles to adsorb. Following a 1000-fold dilution, two 100 µL samples were collected every 2 min and either kept on ice until titration, or mixed with CHCl₃. For each time point we thus determined the free bacteriophage count (samples with CHCl₃) as well as the number of free bacteriophages and infective centres (samples without CHCl₃) to calculate eclipse and latency periods respectively.

Experimental data were fitted with a logistical function:

$$(1) \quad f(x) = \frac{a}{1 + e^{-k(x-x_c)}}$$

a : ordinate corresponding to the asymptote when $x \rightarrow +\infty$, represents the maximal pfu count.

x_c : abscissa of the inflection point, represents the mean duration of the infectious cycle.

k : slope of the tangent line to the exponential part of the curve.

Eclipse and latency periods were determined as the value corresponding to $f(x) > 0.05a$

The burst size was determined as:

$$(2) \quad \frac{Phage_titer_{(t=0, -CHCl_3)} - Phage_titer_{(t=0, +CHCl_3)}}{Phage_titer_{(t=30, \pm CHCl_3)}}$$

$Phage_titer_{(t=0, +CHCl_3)}$ and $Phage_titer_{(t=0, -CHCl_3)}$: values of initial phage titers (t=0 min) measured in samples treated or not with chloroform, respectively. The numerator represents the number of intracellular phages.

$Phage_titer_{(t=30, \pm CHCl_3)}$: Mean of phage titers measured in samples treated and not treated with chloroform at t=30 min.

Four independent adsorption assays were performed in the conditions described above with a lower MOI (10^{-3}) and omitting the dilution step. Data could be approximated using an exponential function and adsorption time was defined as the time required to reach a threshold of 10% non-adsorbed bacteriophage particles.

Dependence to host RNA polymerase assay

The experiment was performed according to the methods described by Ceyssens *et al.* (2014). Strain PAK was grown in LB medium until it reached $OD_{600nm} \sim 0.3$, then the culture was split: one half was supplemented with 100 µg.mL⁻¹ rifampicin (Rif) and second half with the same volume of water. Ten minutes later, phages were added at MOI of 1. Samples were taken immediately (t=0), 15, 60 and 150 min post infection and phages were titrated.

“-OMICS” EXPERIMENTS

Strains and growth conditions

For RNA-Seq experiments, cells were infected with phage PAK_P3 or PAK_P4 using a MOI of 25 in order to ensure the synchronicity of the infection (95% of the bacterial population killed after 5 min phage-bacteria incubation). Strain PAK was cultivated in LB supplemented with 10mM MgCl₂ or CaCl₂ to increase PAK_P3 and PAK_P4 adsorption efficacy, respectively.

Whole transcriptome sequencing

RNA-Seq analysis was performed on an exponentially growing culture that was synchronously infected with PAK_P3. Three independent biological replicates were harvested at 0min, 3.5min and 13min to represent, respectively, a phage negative control and early and late transcription.

The preparation of cDNA libraries was performed as described in Blasdel *et al.*²¹³. Briefly, samples were collected at two time points, representing early and late infection, from a synchronously infected culture, with <5% of bacteria remaining uninfected after 3.5 minutes, and halted by rapid cooling in 1/10 volume of ‘stop solution’ (10% phenol, 90% ethanol). Cells were then lysed in TRIzol®, total RNA was purified through a standard organic extraction and ethanol precipitation, and remaining genomic DNA was removed using TURBO DNase®. DNA removal was confirmed with PCR before rRNA was depleted using the Ribo-Zero rRNA Removal Kit (Gram-Negative Bacteria). This rRNA depleted total RNA was then processed into cDNA libraries using Illumina’s TruSeq® Stranded Total RNA Sample Prep Kit® according to manufacturer’s instructions and sequenced using an Illumina NextSeq 500 desktop sequencer on the High 75 cycle.

In PAK_P3 infected cells: more than 11 million 75bp reads mapping to non-ribosomal regions were obtained from each library with the exception of one early sample and one late sample providing 1,221,867 and 941,631 mapped reads respectively due to incomplete rRNA removal.

In PAK_P4 infected cells: over 12 million 75 bp reads mapping to non-ribosomal regions were obtained from each library with the exception of one early sample (2.4 million reads).

After trimming, sequencing reads were aligned separately to both the phage and host genomes with the CLC Genomics workbench v7.5.1. These alignments were then summarized into count tables of Unique Gene Reads that map to phage or host gene features respectively.

RNA-Seq data have been deposited in NCBI-GEO with accession no. GSE76513 for PAK_P3 and no. GSE86022 for PAK_P4

RNA-Seq coverage visualization is available through the COV2HTML software²¹⁴:

Host gene expression upon phage infection (0min versus 13min)	
PAK_P4	https://mmonot.eu/COV2HTML/visualisation.php?str_id=-36
PAK_P3	https://mmonot.eu/COV2HTML/visualisation.php?str_id=-32
Phage gene expression (3.5min versus 13min)	
PAK_P4	https://mmonot.eu/COV2HTML/visualisation.php?str_id=-38
PAK_P3	https://mmonot.eu/COV2HTML/visualisation.php?str_id=-34

Statistical analysis

Each statistical comparison presented was performed using the DESeq2²¹⁵ R/Bioconductor package to normalize samples to each other and then test for differential expression. Notably, we have chosen to normalize the population of reads that map to each genome independently of the other. In the context of a phage infected cell rapidly replacing host transcripts with phage transcripts, this has artificially enriched host reads and depleted phage reads progressively over the course of infection by normalizing away the biologically relevant shift in each organism's proportion of the total reads in the cell. However, this has also allowed us to both show and test for differential expression of both phage and host gene features independently of the more global swing towards phage transcription.

Strain PAK reannotation and prediction of ncRNAs with RNA-Seq data

RNA-Seq data of uninfected strain PAK were visualized using COV2HTML²¹⁴. Reads mapping forward and reverse strands were manually scanned over the whole genome. Both coding regions and intergenic regions displaying an unexpected transcription profile were examined using CLC Genomics Workbench 7.5.1 and Blastp (default parameters) to annotate putative new coding sequences or RNA central (<http://rnacentral.org/sequence-search/>), Rfam search (<http://rfam.xfam.org/>) and RNAfold web server (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>) with default parameters to predict putative small RNAs and riboswitches.

High coverage metabolomics analysis

P. aeruginosa strain PAK cells, grown in minimal medium (30 mM Na₂HPO₄, 14 mM KH₂PO₄, 20 mM (NH₄)₂SO₄, 20 mM glucose, 1 mM MgSO₄, 4 μM FeSO₄), were infected with PAK_P3 at OD₆₀₀ = 0.3 (approx. 1.25.10⁸ cfu). At 0, 5, 10, 15, 20, and 25 minutes post infection, cells were collected by fast filtration²¹⁶. The biomass quantity was adjusted to match the biomass of a 1 mL culture at OD₆₀₀ = 1.0 (approx. 4.10⁸ cfu) by following the OD₆₀₀ and adjustment of the sampling volume. Four biological replicates were sampled and two technical repeats were made of each independent biological sample.

The metabolic content was extracted as described by De Smet *et al*¹⁸². The samples were profiled using only negative mode flow injection-time-of-flight mass spectrometry and detected ions were annotated as previously reported¹⁸¹. Metabolite annotation and statistical analysis was performed using Matlab R2013b (Mathworks, Natick, MA, United States) according to the ion annotation protocol described by Fuhrer *et al.*¹⁸¹. With this method, 6006 ions were detected and 918 of them could be assigned to known *P. aeruginosa* metabolites. After removal of ion adducts, 377 ions were retained that were annotated as 518 metabolites (including mass isomers).

Differential analysis was performed for each time point versus time point zero using a t-test for two samples with unequal variances (Welch test). For metabolic pathway enrichment, lists of significantly changing metabolites for each time point were created based on the thresholds of $|\text{Log}_2(\text{fold change})| \geq 0.5$ and adjusted p-value < 0.1 . In each list, metabolites were sorted by the adjusted p-value, and the pathway enrichment procedure was performed for each subset of size 1 to the size of the significant list using Fischer test as described in²¹⁷ and the smallest p-value for each pathway was reported. For differential analysis and pathway enrichment, p-values were adjusted for multiple hypotheses testing with the Benjamini-Hochberg procedure.

MOLECULAR AND CELLULAR MICROBIOLOGY

General and basic molecular biology

PCR amplifications were carried out by using the Phusion™ high-fidelity DNA polymerase (ThermoFisher Scientific) when the amplicon was meant to be subsequently cloned or by using the DreamTaq™ DNA polymerase (ThermoFisher Scientific) for colony PCRs, according to manufacturer's instructions.

Molecular cloning was typically performed using a standard double digestion – ligation protocol. All constructed plasmids were verified by Sanger sequencing.

PCR primers and plasmids constructed during this PhD work are listed in Annex 1.

Generation of knock-out mutants

In-frame, scarless deletion of PAK genes were performed with two-step allelic exchange as described by Hmelo *et al.* (2015)¹⁶⁸. Briefly, upstream and downstream regions (~500bp length) of targeted gene were PCR-amplified and assembled by “splicing by-overlap extension”-PCR. The generated deletion allele was then cloned into the allelic exchange vector pDONRPEX18Gm, subsequently transformed in a conjugative *E. coli* strain (e.g. S17λpir) and ultimately transferred to *P. aeruginosa* PAK by biparental mating. Deletion mutants were finally selected on no-salt LB agar containing 15% sucrose and verified by Sanger sequencing of the region of interest.

Expression of phage predicted genes in *Pseudomonas aeruginosa*

Insertion of phage ORFs on the chromosome of *P. aeruginosa* was performed as described by Wagemans *et al.*¹⁷². Briefly, phage ORFs were cloned in an *E. coli* – *P. aeruginosa* shuttle expression vector pUC18-mini-Tn7T-Lac and verified by DNA sequencing. Expression vector was co-transformed with plasmid pTNS2 (encoding a transposition system) in PAK strain (or derivatives) allowing the integration of the expression cassette from pUC18-mini-Tn7T-Lac containing the phage ORF under the control of an IPTG-inducible *lac* promoter at the unique transposon site Tn7 on PAK genome.

Toxicity assay

Overnight cultures of *P. aeruginosa* strains carrying phage ORFs and corresponding control strains were diluted 100-fold in fresh LB medium and incubated 2h at 37°C under agitation until OD_{600nm} reached 0.25-0.4. Ten-fold serial dilutions were plated on LB agar supplemented with gentamycin with or without 1mM IPTG. Plates were incubated at 37°C for 24h to 48h and scanned with Epson perfection V700 Pro scanner. Colonies areas were measured using ImageJ software.

Bacterial growth curves

Bacterial growth curves were obtained by following the optical density of the cultures over time in 96-well plates using a spectrophotometer microplate reader (Glomax MultiDetection System, Promega). Cultures of exponentially growing strains of interest were diluted (in the same medium) to adjust the OD_{600nm} at 0.25. Upon appropriate treatment (addition of 1mM IPTG, 20µg.mL⁻¹ D-cycloserine), resulting cultures were aliquoted in tri- or quadruplicates (200µL) in 96-well plates. The microplate reader maintained a temperature of 37°C and read the OD_{600nm} every 15min or 30min with orbital agitation (1min, intensity “high”) prior measurements for 20h.

Time Lapse Microscopy

Growth of bacteria was monitored using time lapse microscopy. Briefly, fresh cultures were incubated until reaching exponential growth and subsequently diluted to have a concentration equivalent to 0.1 OD_{600nm} unit. Two microliters were then spotted on an acrylamide pad (previously soaked in LB+ 1mM IPTG for 4h) and this sample is eventually sealed under a coverslip with solvent-free sealant “Valap” (1:1:1 (w/w/w) vaseline, lanolin, paraffin) and placed under the microscope equipped with an incubation chamber set up at 37°C. Successive divisions of single cells were monitored by taking pictures every 20min for 8h. Images were compiled and analysed using NIS-Elements Viewer 4.2 (Nikon) software.

Bacterial Adenylate Cyclase Two-Hybrid assays

Bacterial two hybrid screening against *E. coli* and *Salmonella typhimurium* libraries and interaction assays were performed as previously described¹⁷⁸. In this system, the prey protein (Gp92) is fused to the C-terminus of the *Bordetella pertussis* adenylate cyclase T18 fragment and the bait consists either in a library of *E. coli* MG1655 (or *S. typhimurium*) polypeptides or in a specific protein fused to the complementary domain T25 of the adenylate cyclase. Plasmids carrying the corresponding constructs are co-expressed in an *E. coli* strain lacking its endogenous *cya* gene encoding the adenylate cyclase. Under these conditions, a functional adenylate cyclase protein is reconstituted only if the two domains are brought close to each other by the interaction of the bait and the prey proteins. This functional complementation results in the production of cAMP and turns on the expression of catabolic operons (typically lactose or maltose operons), which is easily detectable on selective and indicator media. Theoretically, the only bacteria able to grow on this medium harbor polypeptides interacting with Gp92, allowing the reconstitution of the chimeric adenylate cyclase and eventually resulting in cAMP production and maltose regulon expression.

Briefly, plasmid pUT18C-gp92 was introduced into *E. coli* Δ*cya* strains BTH101, DHM1 or DHT1.

The resultant strains were then transformed with the indicated plasmids (pKT25-RseA/MucA/AlgU or pKT25-MG1655/*S. typhimurium* library and plated either on minimal medium M63 agar (containing

40 $\mu\text{g}.\text{mL}^{-1}$ X-Gal) or MacConkey agar, supplemented with 0.2% or 1% maltose, 0.5mM IPTG, 25 or 50 $\mu\text{g}.\text{mL}^{-1}$ kanamycin and 50 or 100 $\mu\text{g}.\text{mL}^{-1}$ carbenicillin, respectively. Plates were incubated at 30°C for 2 to 7 days until appearance of coloured colonies. For the screening, colonies were reisolated and further characterized by sequencing the DNA inserts from the *E. coli* or *S. typhimurium* genomic libraries in the corresponding pKT25 plasmids. Overall, about 60 *cya*⁺ clones were analysed.

Gp92 topology analysis

To study Gp92 membrane topology, the following regions of *gp92* were PCR-amplified using PAK_P3 lysate as a template: full length ORF (Gp92_{FL}), 5'-end from nucleotide 1 to 78 (Gp92N) and 3'-end from nucleotide 69 to 234 (Gp92C). Amplicons were then cloned in a plasmid (pKTop) allowing to generate an in-frame fusion of Gp92 with a dual PhoA-LacZ reporter¹⁷⁶. Constructed plasmids were verified by DNA sequencing and used to transform *E. coli* DH5 α strain, subsequently plated on indicator medium consisting of LB agar supplemented with 5-Bromo-4-chloro-3-indolyl phosphate (BCIP, 80 $\mu\text{g}.\text{mL}^{-1}$), 5-Bromo-6-chloro-3-indolyl β -D-galactopyranoside (Magenta Gal, 100 $\mu\text{g}.\text{mL}^{-1}$), carbenicillin (100 $\mu\text{g}.\text{mL}^{-1}$), kanamycin (50 $\mu\text{g}.\text{mL}^{-1}$), IPTG (1mM) and phosphate buffer (50mM, pH 7.0).

REFERENCES

- 1 Iwanowski, D. Concerning the mosaic disease of the tobacco plant. *St. Petersburg Acad. Imp. Sci. Bul.* **35**, 67-70 (1892).
- 2 Beijerinck, M. J. Concerning a contagium vivum fluidum as cause of the spot disease of tobacco leaves. *Verhandelingen der Koninklijke akademie Wetenschappen te Amsterdam* **65**, 3-21 (1898).
- 3 Stanley, W. M. Isolation of a Crystalline Protein Possessing the Properties of Tobacco-Mosaic Virus. *Science* **81**, 644-645, doi:10.1126/science.81.2113.644 (1935).
- 4 Félix d'Hérelle. Sur un microbe invisible antagoniste des bacilles dysentériques. *Compte rendus de l'académie des sciences de Paris* **165**, 373-375 (1917).
- 5 Salmond, G. P. & Fineran, P. C. A century of the phage: past, present and future. *Nature reviews. Microbiology* **13**, 777-786, doi:10.1038/nrmicro3564 (2015).
- 6 La Scola, B. et al. A giant virus in amoebae. *Science* **299**, 2033, doi:10.1126/science.1081867 (2003).
- 7 Arslan, D., Legendre, M., Seltzer, V., Abergel, C. & Claverie, J. M. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 17486-17491, doi:10.1073/pnas.1110889108 (2011).
- 8 Philippe, N. et al. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* **341**, 281-286, doi:10.1126/science.1239181 (2013).
- 9 Raoult, D. & Forterre, P. Redefining viruses: lessons from Mimivirus. *Nature reviews. Microbiology* **6**, 315-319, doi:10.1038/nrmicro1858 (2008).
- 10 Forterre, P. Manipulation of cellular syntheses and the nature of viruses: The virocell concept. *Cr Chim* **14**, 392-399, doi:10.1016/j.crci.2010.06.007 (2011).
- 11 Sanchez, E. L. & Lagunoff, M. Viral activation of cellular metabolism. *Virology* **479**, 609-618, doi:10.1016/j.virol.2015.02.038 (2015).
- 12 Ankrah, N. Y. et al. Phage infection of an environmentally relevant marine bacterium alters host metabolism and lysate composition. *ISME J* **8**, 1089-1100, doi:10.1038/ismej.2013.216 (2014).
- 13 Thompson, L. R. et al. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proceedings of the National Academy of Sciences of the United States of America* **108**, E757-764, doi:10.1073/pnas.1102164108 (2011).
- 14 De Smet, J. et al. High coverage metabolomics analysis reveals phage-specific alterations to *Pseudomonas aeruginosa* physiology during infection. *ISME J* **10**, 1823-1835, doi:10.1038/ismej.2016.3 (2016).
- 15 Chaikerasitak, V. et al. Assembly of a nucleus-like structure during viral replication in bacteria. *Science* **355**, 194-197, doi:10.1126/science.aal2130 (2017).
- 16 Ackermann, H. W. Phage classification and characterization. *Methods in molecular biology* **501**, 127-140, doi:10.1007/978-1-60327-164-6_13 (2009).
- 17 Krupovic, M., Prangishvili, D., Hendrix, R. W. & Bamford, D. H. Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. *Microbiology and molecular biology reviews : MMBR* **75**, 610-635, doi:10.1128/MMBR.00011-11 (2011).
- 18 Sanger, F. et al. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* **265**, 687-695 (1977).
- 19 Smith, G. P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315-1317 (1985).
- 20 Pietila, M. K., Demina, T. A., Atanasova, N. S., Oksanen, H. M. & Bamford, D. H. Archaeal viruses and bacteriophages: comparisons and contrasts. *Trends in microbiology* **22**, 334-344, doi:10.1016/j.tim.2014.02.007 (2014).
- 21 Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 4576-4579 (1990).
- 22 Rohwer, F. & Edwards, R. The Phage Proteomic Tree: a genome-based taxonomy for phage. *Journal of bacteriology* **184**, 4529-4535 (2002).

- 23 Edwards, R. A. & Rohwer, F. Viral metagenomics. *Nature reviews. Microbiology* **3**, 504-510, doi:10.1038/nrmicro1163 (2005).
- 24 Simmonds, P. *et al.* Consensus statement: Virus taxonomy in the age of metagenomics. *Nat Rev Micro advance online publication*, doi:10.1038/nrmicro.2016.177 (2017).
- 25 Hatfull, G. F. Bacteriophage genomics. *Current opinion in microbiology* **11**, 447-453, doi:10.1016/j.mib.2008.09.004 (2008).
- 26 Muhlenhoff, M., Stummeyer, K., Grove, M., Sauerborn, M. & Gerardy-Schahn, R. Proteolytic processing and oligomerization of bacteriophage-derived endosialidases. *The Journal of biological chemistry* **278**, 12634-12644, doi:10.1074/jbc.M212048200 (2003).
- 27 Glonti, T., Chanishvili, N. & Taylor, P. W. Bacteriophage-derived enzyme that depolymerizes the alginic acid capsule associated with cystic fibrosis isolates of *Pseudomonas aeruginosa*. *Journal of applied microbiology* **108**, 695-702, doi:10.1111/j.1365-2672.2009.04469.x (2010).
- 28 Moak, M. & Molineux, I. J. Peptidoglycan hydrolytic activities associated with bacteriophage virions. *Molecular microbiology* **51**, 1169-1183 (2004).
- 29 Roos, W. H., Ivanovska, I. L., Evilevitch, A. & Wuite, G. J. Viral capsids: mechanical characteristics, genome packaging and delivery mechanisms. *Cellular and molecular life sciences : CMLS* **64**, 1484-1497, doi:10.1007/s00018-007-6451-1 (2007).
- 30 Leiman, P. G. & Shneider, M. M. Contractile tail machines of bacteriophages. *Advances in experimental medicine and biology* **726**, 93-114, doi:10.1007/978-1-4614-0980-9_5 (2012).
- 31 Koch, T., Raudonikiene, A., Wilkens, K. & Ruger, W. Overexpression, purification, and characterization of the ADP-ribosyltransferase (gpAlt) of bacteriophage T4: ADP-ribosylation of *E. coli* RNA polymerase modulates T4 "early" transcription. *Gene expression* **4**, 253-264 (1995).
- 32 Choi, K. H. *et al.* Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. *Journal of molecular biology* **378**, 726-736, doi:10.1016/j.jmb.2008.02.059 (2008).
- 33 Weigel, C. & Seitz, H. Bacteriophage replication modules. *FEMS microbiology reviews* **30**, 321-381, doi:10.1111/j.1574-6976.2006.00015.x (2006).
- 34 Calendar, R. *The bacteriophages*. 2nd edn, (Oxford University Press, 2006).
- 35 Ceyssens, P. J. *et al.* Development of giant bacteriophage varphiKZ is independent of the host transcription apparatus. *Journal of virology* **88**, 10501-10510, doi:10.1128/JVI.01347-14 (2014).
- 36 Ponchon, L., Mangenot, S., Boulanger, P. & Letellier, L. Encapsidation and transfer of phage DNA into host cells: from in vivo to single particles studies. *Biochimica et biophysica acta* **1724**, 255-261, doi:10.1016/j.bbagen.2005.04.016 (2005).
- 37 Aksyuk, A. A. & Rossmann, M. G. Bacteriophage assembly. *Viruses* **3**, 172-203, doi:10.3390/v3030172 (2011).
- 38 Young, R. Phage lysis: three steps, three choices, one outcome. *Journal of microbiology* **52**, 243-258, doi:10.1007/s12275-014-4087-z (2014).
- 39 Wegrzyn, G. & Wegrzyn, A. Genetic switches during bacteriophage lambda development. *Progress in nucleic acid research and molecular biology* **79**, 1-48, doi:10.1016/S0079-6603(04)79001-7 (2005).
- 40 Oppenheim, A. B., Kobiler, O., Stavans, J., Court, D. L. & Adhya, S. Switches in bacteriophage lambda development. *Annual review of genetics* **39**, 409-429, doi:10.1146/annurev.genet.39.073003.113656 (2005).
- 41 Erez, Z. *et al.* Communication between viruses guides lysis-lysogeny decisions. *Nature* **541**, 488-493, doi:10.1038/nature21049 (2017).
- 42 Los, M. & Wegrzyn, G. Pseudolysogeny. *Advances in virus research* **82**, 339-349, doi:10.1016/B978-0-12-394621-8.00019-4 (2012).
- 43 Cenens, W. *et al.* Expression of a novel P22 ORFan gene reveals the phage carrier state in *Salmonella typhimurium*. *PLoS genetics* **9**, e1003269, doi:10.1371/journal.pgen.1003269 (2013).

- 44 Latino, L., Midoux, C., Hauck, Y., Vergnaud, G. & Pourcel, C. Pseudolysogeny and sequential mutations build multiresistance to virulent bacteriophages in *Pseudomonas aeruginosa*. *Microbiology* **162**, 748-763, doi:10.1099/mic.0.000263 (2016).
- 45 Dufour, N. *Phage therapy and ventilator-associated pneumonia due to Escherichia coli : a possible therapeutic approach ? Fundamental aspects and element of feasibility*, Université Paris Diderot Paris 7, (2015).
- 46 Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429-432, doi:10.1038/nature11723 (2013).
- 47 Depardieu, F. *et al.* A Eukaryotic-like Serine/Threonine Kinase Protects Staphylococci against Phages. *Cell host & microbe* **20**, 471-481, doi:10.1016/j.chom.2016.08.010 (2016).
- 48 Gandon, S., Buckling, A., Decaestecker, E. & Day, T. Host-parasite coevolution and patterns of adaptation across time and space. *Journal of evolutionary biology* **21**, 1861-1866, doi:10.1111/j.1420-9101.2008.01598.x (2008).
- 49 Lopez Pascua, L. *et al.* Higher resources decrease fluctuating selection during host-parasite coevolution. *Ecology letters* **17**, 1380-1388, doi:10.1111/ele.12337 (2014).
- 50 Prigent, M., Leroy, M., Confalonieri, F., Dutertre, M. & DuBow, M. S. A diversity of bacteriophage forms and genomes can be isolated from the surface sands of the Sahara Desert. *Extremophiles : life under extreme conditions* **9**, 289-296, doi:10.1007/s00792-005-0444-5 (2005).
- 51 Dutilh, B. E. *et al.* A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nature communications* **5**, 4498, doi:10.1038/ncomms5498 (2014).
- 52 Rohwer, F. Global phage diversity. *Cell* **113**, 141 (2003).
- 53 Sime-Ngando, T. Environmental bacteriophages: viruses of microbes in aquatic ecosystems. *Frontiers in microbiology* **5**, 355, doi:10.3389/fmicb.2014.00355 (2014).
- 54 Clokie, M. R., Millard, A. D., Letarov, A. V. & Heaphy, S. Phages in nature. *Bacteriophage* **1**, 31-45, doi:10.4161/bact.1.1.14942 (2011).
- 55 Suttle, C. A. Marine viruses--major players in the global ecosystem. *Nature reviews. Microbiology* **5**, 801-812, doi:10.1038/nrmicro1750 (2007).
- 56 Perez Sepulveda, B. *et al.* Marine phage genomics: the tip of the iceberg. *FEMS microbiology letters* **363**, doi:10.1093/femsle/fnw158 (2016).
- 57 Roux, S. *et al.* Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **537**, 689-693, doi:10.1038/nature19366 (2016).
- 58 Bobay, L. M., Touchon, M. & Rocha, E. P. Pervasive domestication of defective prophages by bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 12127-12132, doi:10.1073/pnas.1405336111 (2014).
- 59 Fortier, L. C. & Sekulovic, O. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* **4**, 354-365, doi:10.4161/viru.24498 (2013).
- 60 Henry, K. A., Arbabi-Ghahroudi, M. & Scott, J. K. Beyond phage display: non-traditional applications of the filamentous bacteriophage as a vaccine carrier, therapeutic biologic, and bioconjugation scaffold. *Frontiers in microbiology* **6**, 755, doi:10.3389/fmicb.2015.00755 (2015).
- 61 Cyranoski, D. CRISPR gene-editing tested in a person for the first time. *Nature* **539**, 479, doi:10.1038/nature.2016.20988 (2016).
- 62 Schmelcher, M. & Loessner, M. J. Application of bacteriophages for detection of foodborne pathogens. *Bacteriophage* **4**, e28137, doi:10.4161/bact.28137 (2014).
- 63 Zhang, D. *et al.* The Use of a Novel NanoLuc -Based Reporter Phage for the Detection of *Escherichia coli* O157:H7. *Scientific reports* **6**, 33235, doi:10.1038/srep33235 (2016).
- 64 Soni, K. A., Nannapaneni, R. & Hagens, S. Reduction of *Listeria monocytogenes* on the surface of fresh channel catfish fillets by bacteriophage Listex P100. *Foodborne pathogens and disease* **7**, 427-434, doi:10.1089/fpd.2009.0432 (2010).

- 65 Meaden, S. & Koskella, B. Exploring the risks of phage application in the environment. *Frontiers in microbiology* **4**, 358, doi:10.3389/fmicb.2013.00358 (2013).
- 66 Abedon, S. T., Kuhl, S. J., Blasdel, B. G. & Kutter, E. M. Phage treatment of human infections. *Bacteriophage* **1**, 66-85, doi:10.4161/bact.1.2.15845 (2011).
- 67 Ross, A., Ward, S. & Hyman, P. More Is Better: Selecting for Broad Host Range Bacteriophages. *Frontiers in microbiology* **7**, 1352, doi:10.3389/fmicb.2016.01352 (2016).
- 68 Francino, M. P. Antibiotics and the Human Gut Microbiome: Dysbioses and Accumulation of Resistances. *Frontiers in microbiology* **6**, 1543, doi:10.3389/fmicb.2015.01543 (2015).
- 69 Leon, M. & Bastias, R. Virulence reduction in bacteriophage resistant bacteria. *Frontiers in microbiology* **6**, 343, doi:10.3389/fmicb.2015.00343 (2015).
- 70 Lusiak-Szelachowska, M. *et al.* Phage neutralization by sera of patients receiving phage therapy. *Viral immunology* **27**, 295-304, doi:10.1089/vim.2013.0128 (2014).
- 71 Kim, K. P. *et al.* PEGylation of bacteriophages increases blood circulation time and reduces T-helper type 1 immune response. *Microbial biotechnology* **1**, 247-257, doi:10.1111/j.1751-7915.2008.00028.x (2008).
- 72 Lin, T. Y. *et al.* A T3 and T7 recombinant phage acquires efficient adsorption and a broader host range. *PloS one* **7**, e30954, doi:10.1371/journal.pone.0030954 (2012).
- 73 Ando, H., Lemire, S., Pires, D. P. & Lu, T. K. Engineering Modular Viral Scaffolds for Targeted Bacterial Population Editing. *Cell systems* **1**, 187-196, doi:10.1016/j.cels.2015.08.013 (2015).
- 74 Bikard, D. *et al.* Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. *Nature biotechnology* **32**, 1146-1150, doi:10.1038/nbt.3043 (2014).
- 75 Wright, A., Hawkins, C. H., Anggard, E. E. & Harper, D. R. A controlled clinical trial of a therapeutic bacteriophage preparation in chronic otitis due to antibiotic-resistant *Pseudomonas aeruginosa*; a preliminary report of efficacy. *Clinical otolaryngology : official journal of ENT-UK ; official journal of Netherlands Society for Oto-Rhino-Laryngology & Cervico-Facial Surgery* **34**, 349-357, doi:10.1111/j.1749-4486.2009.01973.x (2009).
- 76 Sarker, S. A. & Brussow, H. From bench to bed and back again: phage therapy of childhood *Escherichia coli* diarrhea. *Annals of the New York Academy of Sciences* **1372**, 42-52, doi:10.1111/nyas.13087 (2016).
- 77 Roach, D. R. & Donovan, D. M. Antimicrobial bacteriophage-derived proteins and therapeutic applications. *Bacteriophage* **5**, e1062590, doi:10.1080/21597081.2015.1062590 (2015).
- 78 Gerstmanns, H., Rodriguez-Rubio, L., Lavigne, R. & Briers, Y. From endolysins to Artilysin(R)s: novel enzyme-based approaches to kill drug-resistant bacteria. *Biochemical Society transactions* **44**, 123-128, doi:10.1042/BST20150192 (2016).
- 79 Hyman, P., Abedon, S. T. & C.A.B. International. *Bacteriophages in health and disease*. (CABI, 2012).
- 80 Liu, J. *et al.* Antimicrobial drug discovery through bacteriophage genomics. *Nature biotechnology* **22**, 185-191, doi:10.1038/nbt932 (2004).
- 81 Roucourt, B. & Lavigne, R. The role of interactions between phage and bacterial proteins within the infected cell: a diverse and puzzling interactome. *Environmental microbiology* **11**, 2789-2805, doi:10.1111/j.1462-2920.2009.02029.x (2009).
- 82 Sevostyanova, A. *et al.* Temporal regulation of viral transcription during development of *Thermus thermophilus* bacteriophage phiYS40. *Journal of molecular biology* **366**, 420-435, doi:10.1016/j.jmb.2006.11.050 (2007).
- 83 Drulis-Kawa, Z., Majkowska-Skrobek, G., Maciejewska, B., Delattre, A. S. & Lavigne, R. Learning from bacteriophages - advantages and limitations of phage and phage-encoded protein applications. *Current protein & peptide science* **13**, 699-722 (2012).
- 84 Kobiler, O., Rokney, A. & Oppenheim, A. B. Phage lambda CIII: a protease inhibitor regulating the lysis-lysogeny decision. *PloS one* **2**, e363, doi:10.1371/journal.pone.0000363 (2007).
- 85 Court, R., Cook, N., Saikrishnan, K. & Wigley, D. The crystal structure of lambda-Gam protein suggests a model for RecBCD inhibition. *Journal of molecular biology* **371**, 25-33, doi:10.1016/j.jmb.2007.05.037 (2007).

- 86 Yano, S. T. & Rothman-Denes, L. B. A phage-encoded inhibitor of Escherichia coli DNA replication targets the DNA polymerase clamp loader. *Molecular microbiology* **79**, 1325-1338, doi:10.1111/j.1365-2958.2010.07526.x (2011).
- 87 Pavlova, O. *et al.* Temporal regulation of gene expression of the Escherichia coli bacteriophage phiEco32. *Journal of molecular biology* **416**, 389-399, doi:10.1016/j.jmb.2012.01.002 (2012).
- 88 Michalewicz, J. & Nicholson, A. W. Molecular cloning and expression of the bacteriophage T7 0.7(protein kinase) gene. *Virology* **186**, 452-462 (1992).
- 89 Severinova, E. & Severinov, K. Localization of the Escherichia coli RNA polymerase beta' subunit residue phosphorylated by bacteriophage T7 kinase Gp0.7. *Journal of bacteriology* **188**, 3470-3476, doi:10.1128/JB.188.10.3470-3476.2006 (2006).
- 90 Mayer, J. E. & Schweiger, M. RNase III is positively regulated by T7 protein kinase. *The Journal of biological chemistry* **258**, 5340-5343 (1983).
- 91 Marchand, I., Nicholson, A. W. & Dreyfus, M. Bacteriophage T7 protein kinase phosphorylates RNase E and stabilizes mRNAs synthesized by T7 RNA polymerase. *Molecular microbiology* **42**, 767-776 (2001).
- 92 Nechaev, S. & Severinov, K. Inhibition of Escherichia coli RNA polymerase by bacteriophage T7 gene 2 protein. *Journal of molecular biology* **289**, 815-826, doi:10.1006/jmbi.1999.2782 (1999).
- 93 Kashlev, M., Nudler, E., Goldfarb, A., White, T. & Kutter, E. Bacteriophage T4 Alc protein: a transcription termination factor sensing local modification of DNA. *Cell* **75**, 147-154 (1993).
- 94 Skorko, R., Zillig, W., Rohrer, H., Fujiki, H. & Mailhammer, R. Purification and properties of the NAD⁺: protein ADP-ribosyltransferase responsible for the T4-phage-induced modification of the alpha subunit of DNA-dependent RNA polymerase of Escherichia coli. *European journal of biochemistry* **79**, 55-66 (1977).
- 95 Ouhammouch, M., Adelman, K., Harvey, S. R., Orsini, G. & Brody, E. N. Bacteriophage T4 MotA and AsiA proteins suffice to direct Escherichia coli RNA polymerase to initiate transcription at T4 middle promoters. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 1451-1455 (1995).
- 96 Kolesky, S., Ouhammouch, M., Brody, E. N. & Geiduschek, E. P. Sigma competition: the contest between bacteriophage T4 middle and late transcription. *Journal of molecular biology* **291**, 267-281, doi:10.1006/jmbi.1999.2953 (1999).
- 97 Mosig, G., Colowick, N. E. & Pietz, B. C. Several new bacteriophage T4 genes, mapped by sequencing deletion endpoints between genes 56 (dCTPase) and dda (a DNA-dependent ATPase-helicase) modulate transcription. *Gene* **223**, 143-155 (1998).
- 98 Qi, D., Alawneh, A. M., Yonesaki, T. & Otsuka, Y. Rapid Degradation of Host mRNAs by Stimulation of RNase E Activity by Srd of Bacteriophage T4. *Genetics* **201**, 977-987, doi:10.1534/genetics.115.180364 (2015).
- 99 Mason, S. W. & Greenblatt, J. Assembly of transcription elongation complexes containing the N protein of phage lambda and the Escherichia coli elongation factors NusA, NusB, NusG, and S10. *Genes & development* **5**, 1504-1512 (1991).
- 100 Marr, M. T., Datwyler, S. A., Meares, C. F. & Roberts, J. W. Restructuring of an RNA polymerase holoenzyme elongation complex by lambdoid phage Q proteins. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 8972-8978, doi:10.1073/pnas.161253298 (2001).
- 101 Kaczanowska, M. & Ryden-Aulin, M. Ribosome biogenesis and the translation process in Escherichia coli. *Microbiology and molecular biology reviews : MMBR* **71**, 477-494, doi:10.1128/MMBR.00013-07 (2007).
- 102 Wood, L. F., Tszine, N. Y. & Christie, G. E. Activation of P2 late transcription by P2 Ogr protein requires a discrete contact site on the C terminus of the alpha subunit of Escherichia coli RNA polymerase. *Journal of molecular biology* **274**, 1-7, doi:10.1006/jmbi.1997.1390 (1997).
- 103 Wei, P. & Stewart, C. R. Genes that protect against the host-killing activity of the E3 protein of Bacillus subtilis bacteriophage SPO1. *Journal of bacteriology* **177**, 2933-2937 (1995).
- 104 Losick, R. & Pero, J. Cascades of Sigma factors. *Cell* **25**, 582-584 (1981).

- 105 Yuzenkova, Y., Zenkin, N. & Severinov, K. Mapping of RNA polymerase residues that interact with bacteriophage Xp10 transcription antitermination factor p7. *Journal of molecular biology* **375**, 29-35, doi:10.1016/j.jmb.2007.10.054 (2008).
- 106 Dehbi, M. *et al.* Inhibition of transcription in *Staphylococcus aureus* by a primary sigma factor-binding polypeptide from phage G1. *Journal of bacteriology* **191**, 3763-3771, doi:10.1128/JB.00241-09 (2009).
- 107 Zhao, X. *et al.* Transcriptomic and Metabolomic Analysis Revealed Multifaceted Effects of Phage Protein Gp70.1 on *Pseudomonas aeruginosa*. *Frontiers in microbiology* **7**, 1519, doi:10.3389/fmicb.2016.01519 (2016).
- 108 Klimuk, E. *et al.* Host RNA polymerase inhibitors encoded by varphiKMV-like phages of *Pseudomonas*. *Virology* **436**, 67-74, doi:10.1016/j.virol.2012.10.021 (2013).
- 109 Van den Bossche, A. *et al.* Structural elucidation of a novel mechanism for the bacteriophage-based inhibition of the RNA degradosome. *eLife* **5**, doi:10.7554/eLife.16413 (2016).
- 110 Grill, S., Moll, I., Hasenohrl, D., Gualerzi, C. O. & Blasi, U. Modulation of ribosomal recruitment to 5'-terminal start codons by translation initiation factors IF2 and IF3. *FEBS letters* **495**, 167-171 (2001).
- 111 Conter, A., Bouche, J. P. & Dassain, M. Identification of a new inhibitor of essential division gene *ftsZ* as the *kil* gene of defective prophage *Rac*. *Journal of bacteriology* **178**, 5100-5104 (1996).
- 112 Molshanski-Mor, S. *et al.* Revealing bacterial targets of growth inhibitors encoded by bacteriophage T7. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 18715-18720, doi:10.1073/pnas.1413271112 (2014).
- 113 Kiro, R. *et al.* Gene product 0.4 increases bacteriophage T7 competitiveness by inhibiting host cell division. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 19549-19554, doi:10.1073/pnas.1314096110 (2013).
- 114 Doron, S. *et al.* Transcriptome dynamics of a broad host-range cyanophage and its hosts. *Isme J* **10**, 1437-1455, doi:10.1038/ismej.2015.210 (2016).
- 115 Leskinen, K., Blasdel, B. G., Lavigne, R. & Skurnik, M. RNA-Sequencing Reveals the Progression of Phage-Host Interactions between phiR1-37 and *Yersinia enterocolitica*. *Viruses* **8**, 111, doi:10.3390/v8040111 (2016).
- 116 Lavigne, R. *et al.* A multifaceted study of *Pseudomonas aeruginosa* shutdown by virulent podovirus LUZ19. *mBio* **4**, e00061-00013, doi:10.1128/mBio.00061-13 (2013).
- 117 Howard-Varona, C. *et al.* Regulation of infection efficiency in a globally abundant marine Bacteriodes virus. *Isme J* **11**, 284-295, doi:10.1038/ismej.2016.81 (2017).
- 118 Poranen, M. M. *et al.* Global changes in cellular gene expression during bacteriophage PRD1 infection. *Journal of virology* **80**, 8081-8088, doi:10.1128/JVI.00065-06 (2006).
- 119 Ravantti, J. J., Ruokoranta, T. M., Alapuranen, A. M. & Bamford, D. H. Global transcriptional responses of *Pseudomonas aeruginosa* to phage PRR1 infection. *Journal of virology* **82**, 2324-2329, doi:10.1128/JVI.01930-07 (2008).
- 120 Lindell, D. *et al.* Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449**, 83-86, doi:10.1038/nature06130 (2007).
- 121 Ainsworth, S., Zomer, A., Mahony, J. & van Sinderen, D. Lytic infection of *Lactococcus lactis* by bacteriophages Tuc2009 and c2 triggers alternative transcriptional host responses. *Applied and environmental microbiology* **79**, 4786-4798, doi:10.1128/AEM.01197-13 (2013).
- 122 Fan, X. *et al.* The Global Reciprocal Reprogramming between Mycobacteriophage SWU1 and *Mycobacterium* Reveals the Molecular Strategy of Subversion and Promotion of Phage Infection. *Frontiers in microbiology* **7**, 41, doi:10.3389/fmicb.2016.00041 (2016).
- 123 Mojardin, L. & Salas, M. Global Transcriptional Analysis of Virus-Host Interactions between Phage varphi29 and *Bacillus subtilis*. *Journal of virology* **90**, 9293-9304, doi:10.1128/JVI.01245-16 (2016).
- 124 Van den Bossche, A. *et al.* Systematic identification of hypothetical bacteriophage proteins targeting key protein complexes of *Pseudomonas aeruginosa*. *Journal of proteome research* **13**, 4446-4456, doi:10.1021/pr500796n (2014).

- 125 Holt, G. S. *et al.* Shigatoxin encoding Bacteriophage varphi24B modulates bacterial metabolism to raise antimicrobial tolerance. *Scientific reports* **7**, 40424, doi:10.1038/srep40424 (2017).
- 126 Streeter, K. & Katouli, M. *Pseudomonas aeruginosa*: A review of their Pathogenesis and Prevalence in Clinical Settings and the Environment. *Infection, Epidemiology and Medicine*. DOI: 10.7508/iem.2016.01.008 (2016)
- 127 Gross, H. & Loper, J. E. Genomics of secondary metabolite production by *Pseudomonas* spp. *Natural product reports* **26**, 1408-1446, doi:10.1039/b817075b (2009).
- 128 Stover, C. K. *et al.* Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406**, 959-964, doi:10.1038/35023079 (2000).
- 129 Alm, E., Huang, K. & Arkin, A. The evolution of two-component systems in bacteria reveals different strategies for niche adaptation. *PLoS computational biology* **2**, e143, doi:10.1371/journal.pcbi.0020143 (2006).
- 130 Shi, X. *et al.* Bioinformatics and experimental analysis of proteins of two-component systems in *Myxococcus xanthus*. *Journal of bacteriology* **190**, 613-624, doi:10.1128/JB.01502-07 (2008).
- 131 Jimenez, P. N. *et al.* The multiple signaling systems regulating virulence in *Pseudomonas aeruginosa*. *Microbiology and molecular biology reviews : MMBR* **76**, 46-65, doi:10.1128/MMBR.05007-11 (2012).
- 132 Jones, A. K. *et al.* Activation of the *Pseudomonas aeruginosa* AlgU regulon through mucA mutation inhibits cyclic AMP/Vfr signaling. *Journal of bacteriology* **192**, 5709-5717, doi:10.1128/JB.00526-10 (2010).
- 133 Llamas, M. A., Imperi, F., Visca, P. & Lamont, I. L. Cell-surface signaling in *Pseudomonas*: stress responses, iron transport, and pathogenicity. *FEMS microbiology reviews* **38**, 569-597, doi:10.1111/1574-6976.12078 (2014).
- 134 Mathee, K., McPherson, C. J. & Ohman, D. E. Posttranslational control of the algT (algU)-encoded sigma22 for expression of the alginate regulon in *Pseudomonas aeruginosa* and localization of its antagonist proteins MucA and MucB (AlgN). *Journal of bacteriology* **179**, 3711-3720 (1997).
- 135 Llamas, M. A. *et al.* Characterization of five novel *Pseudomonas aeruginosa* cell-surface signalling systems. *Molecular microbiology* **67**, 458-472, doi:10.1111/j.1365-2958.2007.06061.x (2008).
- 136 Llamas, M. A. *et al.* The heterologous siderophores ferrioxamine B and ferrichrome activate signaling pathways in *Pseudomonas aeruginosa*. *Journal of bacteriology* **188**, 1882-1891, doi:10.1128/JB.188.5.1882-1891.2006 (2006).
- 137 Llamas, M. A. *et al.* A Novel extracytoplasmic function (ECF) sigma factor regulates virulence in *Pseudomonas aeruginosa*. *PLoS pathogens* **5**, e1000572, doi:10.1371/journal.ppat.1000572 (2009).
- 138 Beare, P. A., For, R. J., Martin, L. W. & Lamont, I. L. Siderophore-mediated cell signalling in *Pseudomonas aeruginosa*: divergent pathways regulate virulence factor production and siderophore receptor synthesis. *Molecular microbiology* **47**, 195-207 (2003).
- 139 Ochsner, U. A., Johnson, Z. & Vasil, M. L. Genetics and regulation of two distinct haem-uptake systems, phu and has, in *Pseudomonas aeruginosa*. *Microbiology* **146** (Pt 1), 185-198, doi:10.1099/00221287-146-1-185 (2000).
- 140 Gicquel, G. *et al.* The extra-cytoplasmic function sigma factor sigX modulates biofilm and virulence-related properties in *Pseudomonas aeruginosa*. *PloS one* **8**, e80407, doi:10.1371/journal.pone.0080407 (2013).
- 141 Lamont, I. L., Beare, P. A., Ochsner, U., Vasil, A. I. & Vasil, M. L. Siderophore-mediated signaling regulates virulence factor production in *Pseudomonasaeruginosa*. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 7072-7077, doi:10.1073/pnas.092016999 (2002).
- 142 Balasubramanian, D., Schneper, L., Kumari, H. & Mathee, K. A dynamic and intricate regulatory network determines *Pseudomonas aeruginosa* virulence. *Nucleic acids research* **41**, 1-20, doi:10.1093/nar/gks1039 (2013).

- 143 Sonnleitner, E., Romeo, A. & Blasi, U. Small regulatory RNAs in *Pseudomonas aeruginosa*. *RNA biology* **9**, 364-371, doi:10.4161/rna.19231 (2012).
- 144 Gomez-Lozano, M. *et al.* Diversity of small RNAs expressed in *Pseudomonas* species. *Environmental microbiology reports* **7**, 227-236, doi:10.1111/1758-2229.12233 (2015).
- 145 Leid, J. G. *et al.* The exopolysaccharide alginate protects *Pseudomonas aeruginosa* biofilm bacteria from IFN-gamma-mediated macrophage killing. *Journal of immunology* **175**, 7512-7518 (2005).
- 146 Hay, I. D., Wang, Y., Moradali, M. F., Rehman, Z. U. & Rehm, B. H. Genetics and regulation of bacterial alginate production. *Environmental microbiology* **16**, 2997-3011, doi:10.1111/1462-2920.12389 (2014).
- 147 Wood, L. F. & Ohman, D. E. Identification of genes in the sigma(2)(2) regulon of *Pseudomonas aeruginosa* required for cell envelope homeostasis in either the planktonic or the sessile mode of growth. *mBio* **3**, doi:10.1128/mBio.00094-12 (2012).
- 148 Martin, D. W. *et al.* Mechanism of conversion to mucoidy in *Pseudomonas aeruginosa* infecting cystic fibrosis patients. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 8377-8381 (1993).
- 149 Cornelis, P. Iron uptake and metabolism in pseudomonads. *Applied microbiology and biotechnology* **86**, 1637-1645, doi:10.1007/s00253-010-2550-2 (2010).
- 150 Kung, V. L., Ozer, E. A. & Hauser, A. R. The accessory genome of *Pseudomonas aeruginosa*. *Microbiology and molecular biology reviews : MMBR* **74**, 621-641, doi:10.1128/MMBR.00027-10 (2010).
- 151 Mosquera-Rendon, J. *et al.* Pangenome-wide and molecular evolution analyses of the *Pseudomonas aeruginosa* species. *BMC genomics* **17**, 45, doi:10.1186/s12864-016-2364-4 (2016).
- 152 Klockgether, J., Cramer, N., Wiehlmann, L., Davenport, C. F. & Tummeler, B. *Pseudomonas aeruginosa* Genomic Structure and Diversity. *Frontiers in microbiology* **2**, 150, doi:10.3389/fmicb.2011.00150 (2011).
- 153 Ernst, R. K. *et al.* Genome mosaicism is conserved but not unique in *Pseudomonas aeruginosa* isolates from the airways of young children with cystic fibrosis. *Environmental microbiology* **5**, 1341-1349, doi:10.1111/j.1462-2920.2003.00518.x (2003).
- 154 Hocquet, D. *et al.* Pyomelanin-producing *Pseudomonas aeruginosa* selected during chronic infections have a large chromosomal deletion which confers resistance to pyocins. *Environmental microbiology* **18**, 3482-3493, doi:10.1111/1462-2920.13336 (2016).
- 155 Debarbieux, L. *et al.* Bacteriophages can treat and prevent *Pseudomonas aeruginosa* lung infections. *The Journal of infectious diseases* **201**, 1096-1104, doi:10.1086/651135 (2010).
- 156 Morello, E. *et al.* Pulmonary bacteriophage therapy on *Pseudomonas aeruginosa* cystic fibrosis strains: first steps towards treatment and prevention. *PloS one* **6**, e16963, doi:10.1371/journal.pone.0016963 (2011).
- 157 Henry, M., Lavigne, R. & Debarbieux, L. Predicting in vivo efficacy of therapeutic bacteriophages used to treat pulmonary infections. *Antimicrobial agents and chemotherapy* **57**, 5961-5968, doi:10.1128/AAC.01596-13 (2013).
- 158 Henry, M. *et al.* The search for therapeutic bacteriophages uncovers one new subfamily and two new genera of *Pseudomonas*-infecting Myoviridae. *PloS one* **10**, e0117163, doi:10.1371/journal.pone.0117163 (2015).
- 159 Essoh, C. *et al.* Investigation of a Large Collection of *Pseudomonas aeruginosa* Bacteriophages Collected from a Single Environmental Source in Abidjan, Cote d'Ivoire. *PloS one* **10**, e0130548, doi:10.1371/journal.pone.0130548 (2015).
- 160 Krupovic, M. *et al.* Taxonomy of prokaryotic viruses: update from the ICTV bacterial and archaeal viruses subcommittee. *Archives of virology* **161**, 1095-1099, doi:10.1007/s00705-015-2728-0 (2016).
- 161 Mahadevan, P., King, J. F. & Seto, D. CGUG: in silico proteome and genome parsing tool for the determination of "core" and unique genes in the analysis of genomes up to ca. 1.9 Mb. *BMC research notes* **2**, 168, doi:10.1186/1756-0500-2-168 (2009).

- 162 Breitbart, M., Thompson, L. R., Suttle, C. A. & Sullivan, M. B. Exploring the Vast Diversity of Marine Viruses. *Oceanography* **20**, 135-139 (2007).
- 163 Garbe, J., Bunk, B., Rohde, M. & Schobert, M. Sequencing and characterization of Pseudomonas aeruginosa phage JG004. *BMC microbiology* **11**, 102, doi:10.1186/1471-2180-11-102 (2011).
- 164 Pan, X. *et al.* Genetic Evidence for O-Specific Antigen as Receptor of Pseudomonas aeruginosa Phage K8 and Its Genomic Analysis. *Frontiers in microbiology* **7**, 252, doi:10.3389/fmicb.2016.00252 (2016).
- 165 Luria, S. E. & Delbrück, M. MUTATIONS OF BACTERIA FROM VIRUS SENSITIVITY TO VIRUS RESISTANCE. *Genetics* **28**, 491-511 (1943).
- 166 Rodriguez-Rojas, A. *et al.* Inactivation of the hmgA gene of Pseudomonas aeruginosa leads to pyomelanin hyperproduction, stress resistance and increased persistence in chronic lung infection. *Microbiology* **155**, 1050-1057, doi:10.1099/mic.0.024745-0 (2009).
- 167 Le, S. *et al.* Chromosomal DNA deletion confers phage resistance to Pseudomonas aeruginosa. *Scientific reports* **4**, 4738, doi:10.1038/srep04738 (2014).
- 168 Hmelo, L. R. *et al.* Precision-engineering the Pseudomonas aeruginosa genome with two-step allelic exchange. *Nature protocols* **10**, 1820-1841, doi:10.1038/nprot.2015.115 (2015).
- 169 Ceyssens, P.-J. *Isolation and characterization of lytic bacteriophages infecting Pseudomonas aeruginosa.*, KU Leuven, (2009).
- 170 Nejman-Falenczyk, B. *et al.* Small regulatory RNAs in lambdoid bacteriophages and phage-derived plasmids: Not only antisense. *Plasmid* **78**, 71-78, doi:10.1016/j.plasmid.2014.07.006 (2015).
- 171 Tree, J. J., Granneman, S., McAteer, S. P., Tollervey, D. & Gally, D. L. Identification of bacteriophage-encoded anti-sRNAs in pathogenic Escherichia coli. *Molecular cell* **55**, 199-213, doi:10.1016/j.molcel.2014.05.006 (2014).
- 172 Wagemans, J. *et al.* Functional elucidation of antibacterial phage ORFans targeting Pseudomonas aeruginosa. *Cellular microbiology* **16**, 1822-1835, doi:10.1111/cmi.12330 (2014).
- 173 Dedrick, R. M. *et al.* Functional requirements for bacteriophage growth: gene essentiality and expression in mycobacteriophage Giles. *Molecular microbiology* **88**, 577-589, doi:10.1111/mmi.12210 (2013).
- 174 Nejman-Falenczyk, B. *et al.* A small, microRNA-size, ribonucleic acid regulating gene expression and development of Shiga toxin-converting bacteriophage Phi24Beta. *Scientific reports* **5**, 10080, doi:10.1038/srep10080srep10080 [pii] (2015).
- 175 McClain, W. H., Guthrie, C. & Barrell, B. G. Eight transfer RNAs induced by infection of Escherichia coli with bacteriophage T4. *Proceedings of the National Academy of Sciences of the United States of America* **69**, 3703-3707 (1972).
- 176 Karimova, G., Robichon, C. & Ladant, D. Characterization of YmgF, a 72-residue inner membrane protein that associates with the Escherichia coli cell division machinery. *Journal of bacteriology* **191**, 333-346, doi:10.1128/JB.00331-08 (2009).
- 177 Karimova, G., Davi, M. & Ladant, D. The beta-lactam resistance protein Blr, a small membrane polypeptide, is a component of the Escherichia coli cell division machinery. *Journal of bacteriology* **194**, 5576-5588, doi:10.1128/JB.00774-12 (2012).
- 178 Karimova, G., Pidoux, J., Ullmann, A. & Ladant, D. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 5752-5756 (1998).
- 179 Wood, L. F. & Ohman, D. E. Use of cell wall stress to characterize sigma 22 (AlgT/U) activation by regulated proteolysis and its regulon in Pseudomonas aeruginosa. *Molecular microbiology* **72**, 183-201, doi:10.1111/j.1365-2958.2009.06635.x (2009).
- 180 Wood, L. F., Leech, A. J. & Ohman, D. E. Cell wall-inhibitory antibiotics activate the alginate biosynthesis operon in Pseudomonas aeruginosa: Roles of sigma (AlgT) and the AlgW and Prc proteases. *Molecular microbiology* **62**, 412-426, doi:10.1111/j.1365-2958.2006.05390.x (2006).

- 181 Fuhrer, T., Heer, D., Begemann, B. & Zamboni, N. High-throughput, accurate mass metabolome
profiling of cellular extracts by flow injection-time-of-flight mass spectrometry. *Analytical
chemistry* **83**, 7074-7080, doi:10.1021/ac201267k (2011).
- 182 De Smet, J. *et al.* High coverage metabolomics analysis reveals phage-specific alterations to
Pseudomonas aeruginosa physiology during infection. *The ISME journal*,
doi:10.1038/ismej.2016.3 (2016).
- 183 Winsor, G. L. *et al.* Enhanced annotations and features for comparing thousands of
Pseudomonas genomes in the *Pseudomonas* genome database. *Nucleic acids research* **44**,
D646-653, doi:10.1093/nar/gkv1227 (2016).
- 184 Hantke, K. Iron and metal regulation in bacteria. *Current opinion in microbiology* **4**, 172-177
(2001).
- 185 Imperi, F., Tiburzi, F., Fimia, G. M. & Visca, P. Transcriptional control of the *pvdS* iron starvation
sigma factor gene by the master regulator of sulfur metabolism CysB in *Pseudomonas
aeruginosa*. *Environmental microbiology* **12**, 1630-1642, doi:10.1111/j.1462-
2920.2010.02210.x (2010).
- 186 Ferrara, S. *et al.* Comparative profiling of *Pseudomonas aeruginosa* strains reveals differential
expression of novel unique and conserved small RNAs. *PloS one* **7**, e36553,
doi:10.1371/journal.pone.0036553 PONE-D-12-02445 [pii] (2012).
- 187 Gomez-Lozano, M., Marvig, R. L., Molin, S. & Long, K. S. Genome-wide identification of novel
small RNAs in *Pseudomonas aeruginosa*. *Environmental microbiology* **14**, 2006-2016,
doi:10.1111/j.1462-2920.2012.02759.x (2012).
- 188 Wurtzel, O. *et al.* The single-nucleotide resolution transcriptome of *Pseudomonas aeruginosa*
grown in body temperature. *PLoS pathogens* **8**, e1002945, doi:10.1371/journal.ppat.1002945
PPATHOGENS-D-12-00861 [pii] (2012).
- 189 Chakravarty, A. K. & Shuman, S. RNA 3'-phosphate cyclase (RtcA) catalyzes ligase-like
adenylation of DNA and RNA 5'-monophosphate ends. *The Journal of biological chemistry*
286, 4117-4122, doi:10.1074/jbc.M110.196766 (2011).
- 190 Filipowicz, W., Billy, E., Drabikowski, K. & Genschik, P. Cyclases of the 3'-terminal phosphate in
RNA: a new family of RNA processing enzymes conserved in eucarya, bacteria and archaea.
Acta Biochim Pol **45**, 895-906 (1998).
- 191 Tanaka, N. & Shuman, S. RtcB is the RNA ligase component of an *Escherichia coli* RNA repair
operon. *J Biol Chem* **286**, 7727-7731, doi:10.1074/jbc.C111.219022 (2011).
- 192 Fallico, V., Ross, R. P., Fitzgerald, G. F. & McAuliffe, O. 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. *Journal of virology* **85**, 12032-12042, doi:10.1128/JVI.00275-11 (2011).
- 193 Wiegert, T., Homuth, G., Versteeg, S. & Schumann, W. Alkaline shock induces the *Bacillus
subtilis*σW regulon. *Molecular microbiology* **41**, 59-71, doi:10.1046/j.1365-2958.2001.02489.x
(2001).
- 194 Stazic, D., Pekarski, I., Kopf, M., Lindell, D. & Steglich, C. A Novel Strategy for Exploitation of
Host RNase E Activity by a Marine Cyanophage. *Genetics* **203**, 1149-1159,
doi:10.1534/genetics.115.183475 (2016).
- 195 Stazic, D., Lindell, D. & Steglich, C. Antisense RNA protects mRNA from RNase E degradation by
RNA-RNA duplex formation during phage infection. *Nucleic acids research* **39**, 4890-4899,
doi:10.1093/nar/gkr037 (2011).
- 196 Ueno, H. & Yonesaki, T. Phage-induced change in the stability of mRNAs. *Virology* **329**, 134-
141, doi:10.1016/j.virol.2004.08.001 (2004).
- 197 Jian, H., Xiong, L., Xu, G., Xiao, X. & Wang, F. Long 5' untranslated regions regulate the RNA
stability of the deep-sea filamentous phage SW1. *Scientific reports* **6**, 21908,
doi:10.1038/srep21908 (2016).
- 198 Marchand, I., Nicholson, A. W. & Dreyfus, M. Bacteriophage T7 protein kinase phosphorylates
RNase E and stabilizes mRNAs synthesized by T7 RNA polymerase. *Molecular microbiology* **42**,
767-776, doi:10.1046/j.1365-2958.2001.02668.x (2001).

- 199 Bonnain, C., Breitbart, M. & Buck, K. N. The Ferrojan Horse Hypothesis: Iron-Virus Interactions in the Ocean. *Frontiers in Marine Science* **3**, doi:10.3389/fmars.2016.00082 (2016).
- 200 Yang, L. *et al.* Effects of iron on DNA release and biofilm development by *Pseudomonas aeruginosa*. *Microbiology* **153**, 1318-1328, doi:10.1099/mic.0.2006/004911-0 (2007).
- 201 Moreau, P., Diggle, S. P. & Friman, V.-P. Bacterial cell-to-cell signaling promotes the evolution of resistance to parasitic bacteriophages. *Ecology and Evolution*, n/a-n/a, doi:10.1002/ece3.2818 (2017).
- 202 Hoyland-Kroghsbo, N. M., Maerkedahl, R. B. & Svenningsen, S. L. A quorum-sensing-induced bacteriophage defense mechanism. *mBio* **4**, e00362-00312, doi:10.1128/mBio.00362-12 (2013).
- 203 Engl, C., Schaefer, J., Kotta-Loizou, I. & Buck, M. Cellular and molecular phenotypes depending upon the RNA repair system RtcAB of *Escherichia coli*. *Nucleic acids research* **44**, 9933-9941, doi:10.1093/nar/gkw628 (2016).
- 204 Forterre, P., Krupovic, M. & Prangishvili, D. Cellular domains and viral lineages. *Trends in microbiology* **22**, 554-558, doi:10.1016/j.tim.2014.07.004 (2014).
- 205 Wang, I.-N., Dykhuizen, D. E. & Slobodkin, L. B. The evolution of phage lysis timing. *Evolutionary Ecology* **10**, 545-558, doi:10.1007/bf01237884 (1996).
- 206 Henry, M., Lavigne, R. & Debarbieux, L. Predicting In Vivo Efficacy of Therapeutic Bacteriophages Used To Treat Pulmonary Infections. *Antimicrobial Agents and Chemotherapy* **57**, 5961-5968, doi:10.1128/aac.01596-13 (2013).
- 207 Zschach, H. *et al.* What Can We Learn from a Metagenomic Analysis of a Georgian Bacteriophage Cocktail? *Viruses* **7**, 6570-6589, doi:10.3390/v7122958 (2015).
- 208 Khaledi, A. *et al.* Transcriptome Profiling of Antimicrobial Resistance in *Pseudomonas aeruginosa*. *Antimicrobial agents and chemotherapy* **60**, 4722-4733, doi:10.1128/AAC.00075-16 (2016).
- 209 Lammens, E. *et al.* Representational Difference Analysis (RDA) of bacteriophage genomes. *Journal of microbiological methods* **77**, 207-213, doi:10.1016/j.mimet.2009.02.006 (2009).
- 210 Mesyanzhinov, V. V. *et al.* The genome of bacteriophage phiKZ of *Pseudomonas aeruginosa*. *Journal of molecular biology* **317**, 1-19, doi:10.1006/jmbi.2001.5396 (2002).
- 211 Ceyssens, P. J. *et al.* Comparative analysis of the widespread and conserved PB1-like viruses infecting *Pseudomonas aeruginosa*. *Environmental microbiology* **11**, 2874-2883, doi:10.1111/j.1462-2920.2009.02030.x (2009).
- 212 Hyman, P. & Abedon, S. T. Practical methods for determining phage growth parameters. *Methods in molecular biology* **501**, 175-202, doi:10.1007/978-1-60327-164-6_18 (2009).
- 213 Blasdel, B., Ceyssens, P.-J. & Lavigne, R. in *Bacteriophages: Methods and protocols* Vol. 3 (ed M.R.J. *et al.* eds Clokie) (Humana Press, 2016).
- 214 Monot, M., Orgeur, M., Camiade, E., Brehier, C. & Dupuy, B. COV2HTML: a visualization and analysis tool of bacterial next generation sequencing (NGS) data for postgenomics life scientists. *OMICS* **18**, 184-195, doi:10.1089/omi.2013.0119 (2014).
- 215 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
- 216 Zimmermann, M., Thormann, V., Sauer, U. & Zamboni, N. Nontargeted profiling of coenzyme A thioesters in biological samples by tandem mass spectrometry. *Analytical chemistry* **85**, 8284-8290, doi:10.1021/ac401555n (2013).
- 217 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

ANNEX

-

LISTS

STRAIN LIST

Name	Characteristics	Origin
<i>Pseudomonas aeruginosa</i>		
PAK	Wild type PAK strain	Laboratory collection
PAO1	Wild type PAO1 strain	Laboratory collection
CHA	Wild type CHA strain (CF isolate)	Laboratory collection
PAK Δ fliC	In-frame deletion of <i>fliC</i>	Obtained from E. Morello
PAK 4980mut	PAK strain with Tn insertion in gene PAK_4980	This work
PAK 4981mut	PAK strain with Tn insertion in gene PAK_4981	This work
PAK 4982mut	PAK strain with Tn insertion in gene PAK_4982	This work
PAK Δ 4980	In-frame deletion of <i>PAK_4980</i>	This work
PAK Δ 4981	In-frame deletion of <i>PAK_4981</i>	This work
PAK Δ 4493	In-frame deletion of <i>PAK_4983</i>	This work
PAK Δ 4494	In-frame deletion of <i>PAK_4984</i>	This work
WT_sRNA1	PAK WT strain carrying a pUC18mini-derived transcriptional expression cassette containing PAK_P3-encoded 500bp region containing sRNA1	This work
WT_sRNA2	PAK WT strain carrying a pUC18mini-derived transcriptional expression cassette containing PAK_P3-encoded 500bp region containing sRNA2	This work
CHA_Gp92	CHA strain carrying a pUC18mini-derived expression cassette containing ORF <i>gp92</i>	This work
CHA_empty	CHA strain carrying a pUC18mini-derived empty expression cassette	This work
PAO1_Gp92	PAO1 strain carrying a pUC18mini-derived expression cassette containing ORF <i>gp92</i>	This work
PAO1_empty	PAO1 strain carrying a pUC18mini-derived empty expression cassette	This work
*PAK WT	PAK::lacP1 Δ lacII-lacZ; PAK strain devoid of its endogenous <i>lacI</i> regulator gene	Obtained from M. Wolfgang
WT_Gp92	PAK WT strain carrying a pUC18mini-derived expression cassette containing ORF <i>gp92</i>	This work
WT_empty	PAK WT strain carrying a pUC18mini-derived empty expression cassette	This work
WT_Gp92N	PAK WT strain carrying a pUC18mini-derived expression cassette containing ORF <i>gp92</i> (nucleotides 1-78)	This work
WT_Gp92C	PAK WT strain carrying a pUC18mini-derived expression cassette containing ORF <i>gp92</i> (nucleotides 69-234)	This work
WT_SP _{PhoA}	PAK WT strain carrying a pUC18mini-derived expression cassette containing phoA signal sequence	This work
WT_SP _{PhoA} -Gp92C	PAK WT strain carrying a pUC18mini-derived expression cassette containing phoA signal sequence in frame with ORF <i>gp92</i> (nucleotides 69-234)	This work
PAK Δ algU	PAK Δ algU::lacP1 Δ lacI-lacZ	Obtained from M. Wolfgang
Δ algU_Gp92	PAK Δ algU strain carrying a pUC18mini-derived expression cassette containing ORF <i>gp92</i>	This work
Δ algU_empty	PAK Δ algU strain carrying a pUC18mini-derived empty expression cassette	This work
Δ algU_Gp92-pUCP18	Δ algU_Gp92 strain transformed with pUCP18 plasmid	This work

Δ algU_Gp92-pUCP18-algU	Δ algU_Gp92 strain transformed with pUCP18-algU plasmid	This work
PAK Δ algD	PAK Δ algD::lacP1 Δ lacI-lacZ	Obtained from M. Wolfgang
Δ algD_Gp92	PAK Δ algD strain carrying a pUC18mini-derived expression cassette containing ORF gp92	This work
Δ algD_empty	PAK Δ algD strain carrying a pUC18mini-derived empty expression cassette	This work
PAKmucA22	PAKmucA22::lacP1 Δ lacI-lacZ	Obtained from M. Wolfgang
mucA22_Gp92	PAK mucA22 strain carrying a pUC18mini-derived expression cassette containing ORF gp92	This work
mucA22_empty	PAK mucA22 strain carrying a pUC18mini-derived empty expression cassette	This work
**<i>Escherichia coli</i>		
DH5 α	Cloning strain	Invitrogen
TOP10	Cloning strain	ThermoFisher
XL1-blue	Cloning strain	Obtained from G. Karimova
BW25113	E. coli genetic background of Keio collection	Obtained from C. Beloin
DHM1	F-, cya-854, recA1, endA1, gyrA96 (Nalr), thi1, hsdR17, spoT1, rfbD1, glnV44(AS); BACTH system Δ cya strain	Obtained from G. Karimova
BTH101	F-, cya-99, araD139, galE15, galK16, rpsL1 (Strr), hsdR2, mcrA1, mcrB1; BACTH system Δ cya strain	Obtained from G. Karimova
DHT1	[F- glnV44(AS) recA1 endA1 gyrA96 (Nalr)thi-1 hsdR17 spoT1 rfbD1 cya-854 ilv-691 Tn10]; BACTH system Δ cya strain	Obtained from G. Karimova
S17.1 λ pir	Conjugative strain	Obtained from Z. Baharoglu

*All pUC18mini-derived expression cassette insertions were made in this genetic background, referred here as "WT"

** Derivatives of below *E. coli* strains containing all different plasmids constructed during this work are not listed

PRIMER LIST

Primer name	Description	Sequence
RBS_F		TAAGAAGGAGCCCTTCAC
TnT7TR	Check insertion of the expression cassette in P. aeruginosa chromosome	CACAGCATAACTGGACTGATTC
glms_up		GTGCGACTGCTGGAGCTGAA
glms_down		GCTCTCGCCGATCCTCTACA
pUC18-mini_F	Anneal outside the multiple cloning site of pUC18mini	CGGTTCTGGCAAATTTCTGA
pUC18-mini_R		GGAGGGGTGGAAATGGAGTT
M13 uni (-43)	Universal primers	AGGGTTTTCCAGTCACGACGTT
M13 rev (-49)		GAGCGGATAACAATTTACACAGG
pENTR_R2	Anneal outside pENTR recombination site	GTGGGCGCGCCACCCCTT
pENTR_F2		GCGGCCGCCTTGTTTAAC
AC_16F	gp92, to clone in pKT25	GGGCCCCTGCAGGGATGATCAGTCTCGTTCTCACTC
AC_16R	gp92, to clone in pKT25 or pUT18C	GGGCCC GGATCC TCAGTTACCTCCACAGTGAG
AC_17F	gp92, to clone in pKT25 or pUT18 or pUT18C	GGGCCCCTGCAGAAATGATCAGTCTCGTTCTCACTC
AC_17R	gp92ochre, to clone in pKT25 or pUT18	GGGCCC GGATCC TCGTTACCTCCACAGTGAG
AC_18F	Anneals pKT25, pKNT25, pUT18, pUT18C. Check cloning construct	GTT AGC TCA CTC ATT AGG CAC C
AC_18R	Anneals pKT25, pKNT25. Check cloning construct	CAT CAG CGC CAT TCG CCA TTC
AC_19R	Anneals pUT18, pUT18C. Check cloning construct	CAG CTT GTC TGT AAG CGG ATG C
AC_20F	Deletion PAK_4981 Down_F	TCA CGC CTC CCC ATC CAG CGA CTG CTG CTT GCG TGC GCT CAT
AC_20R	Deletion PAK_4981 Down_R	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTA GTT CGC TGC CAT CTA TCT TG
AC_21F	Deletion PAK_4981 Up_F	GGG GAC AAG TTT GTA CAA AAA AGC AGG CTC A TGC CGG TGT ACT TGC AGG GAA TG
AC_21R	Deletion PAK_4981 Up_R	TCG CTG GAT GGG GAG GCG TGA
AC_22F	Deletion PAK_4981 Seq_F	GTA TGG CCC GAA CTG ATC ACC
AC_22R	Deletion PAK_4981 Seq_R	CTA TAA CCA GCG GCG CGA AGT G
AC_23F	Deletion PAK_4980 Down_F	TACGTCGCCAGGTCTTCTCCAG GCG CTG CTC TTT CTG GAC CAG
AC_23R	Deletion PAK_4980 Down_R	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTA CAA TCA CCT GCA CGA CAA CGA TC
AC_24F	Deletion PAK_4980 Up_F	GGG GAC AAG TTT GTA CAA AAA AGC AGG CTC A CTG AGG TCG AAG CCG TAC TTC G
AC_24R	Deletion PAK_4980 Up_R	CTGG AGA AGA CCT GGC GAC GTA
AC_25F	Deletion PAK_4980 Seq_F	GAT CGC CTC GTT GAG CGT TCT C
AC_25R	Deletion PAK_4980 Seq_R	CTG CCT CAT CCG CAT TTC GAT C
AC_26F	Deletion PAK_4982 Down_F	TCAGACCTGGAAGACCGGAGA GGC GCG CAA CCT GTT CAG CAT
AC_26R	Deletion PAK_4982 Down_R	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTA CCT GGA AGA GCA GCT CAA GGA C
AC_27F	Deletion PAK_4982 Up_F	GGG GAC AAG TTT GTA CAA AAA AGC AGG CTC A GTT CAG GTA GCG CAA GCC GTT C
AC_27R	Deletion PAK_4982 Up_R	TCT CCG GTC TTC CAG GTC TGA

AC_28F	Deletion PAK_4982 Seq_F	GAT GAC CAT CAC CCG CCG ATC
AC_28R	Deletion PAK_4982 Seq_R	AAC AGC CTG CTG GCG ATC CTC
AC_29F	gp92 to clone in pUC18mini between PstI and XhoI	GGGCCC CTGCAG ATGATCAGTCTCGTTCTCACTC
AC_29R		GGGCCC CTCGAG TCAGTTACCCTCCACAGTGTAG
AC_30F	gp92 ochre to clone in pUC18mini between BamHI and PstI	GGGCCC GGATCC ATGATCAGTCTCGTTCTCACTC
AC_30R		GGGCCC CTGCAG GTTACCCTCCACAGTGTAG
AC_32F'	mucA to clone in pKT25_F	GGGCCC TCTAGA Gatgagtcgtgaagccctgcag
AC_32R'	mucA to clone in pKT25_R	ATCTTT CCCGGG tcagcgggtttccaggctg
T25-Seq	sequencing primer pKT25 (for library screening)	GGCGCGCAGTTCGGTGACCAGCGGC
AC_37R	gp92 3'-end with SpeI site	GGGCCC ACTAGT TCAGTTACCCTCCACAGTGTAG
AC_38F	gp92 C-term ochre to clone in pKTop (with AC_17R)	GCAT CTGCAGA AAGCGATGCTGACATCATCAATC
AC_43R		CAT GGATCC TCGTCAGCATCGTGTCTTGCGC
AC_50F	PhoA_SS to clone in pKTop	GGCC AAGCTT GGTGAAACAAAGCACTATTGCAC
AC_50R		AT CTGCAG GCTGTCCGGGCTTTTGTCAC
AC_51F	Deletion PAK_4493 Down_F	TCGTGAAGAAGAACGAACGCGCCTGGACCAGGTA
AC_51R	Deletion PAK_4493 Down_R	TTGAACGGAC
AC_52F	Deletion PAK_4493 Up_F	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTA
AC_52R	Deletion PAK_4493 Up_R	GTA CCG CCT GCT CGA TGG CGC TG
AC_53F	Deletion PAK_4493 Seq_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTCACTCGC
AC_53R	Deletion PAK_4493 Seq_R	TATATCCATTTTCGACCGC
AC_55F	For cloning PAK_P3_sRNA1 in pUC18mini (no RBS)	CGCGTTCGTTCTTCTTCACGA
AC_55R		CAGGCATCTCTCCATCGAATTC
AC_56F	Cloning PhoA-92C in pUC18mini_RBS	GTTCGAGAGGCTAGTGGAAGC
AC_56R		CAT GGATCC GGGAAGACCTGGTGAATCACAG
AC_57F	Cloning gp92N in pUC18mini_RBS	AT CTGCAG GCCGAAAAGCTTCCGGATGAC
AC_57R		CAT GGATCC ATGGTGAAACAAAGCACTATTG
AC_61F	Cloning gp92C in pUC18mini_RBS	GGCC CTCGAG TTAGTTACCCTCCACAGTGTAGTC
AC_61R		CAT GGATCC ATGATCAGTCTCGTTCTCACTCTC
AC_62F	Cloning PhoA in pUC18mini_RBS	AT CTGCAG TTAGTCAGCATCGTGTCTTG
AC_62R		CAT GGATCC ATGAGCGATGCTGACATCATC
AC_58F	Sequencing algU	AT CTGCAG TTAGTTACCCTCCACAGTGTAGTC
AC_58R		CAT GGATCC ATGGTGAAACAAAGCACTATTG
AC_60F	Sequencing algD	AT CTGCAG TTATGTCCGGGCTTTTG
AC_60R		GAAGCCCAGTCTATCTTG
AC_63F	Deletion PAK_4494 Down_F	CATAGCGATACCTCTCTTG
AC_63R		GAATCAGCATCTTTGGTTTG
AC_64F	Deletion PAK_4494 Up_F	GCTTGTTGTTATGGCGCAC
AC_64R		AAGCCGATCAAGCTCTGGACCCTGCGTCAGGTGGTG
AC_65F	Deletion PAK_4494 Seq_F	TGCGTG
		GGGGACCACTTTGTACAAGAAAGCTGGGTAGTACG
		GAGTGAAAGGCGCTGC
		GGGGACAAGTTTGTACAAAAAAGCAGGCTCACTCG
		ACGAACTGCTGGAGAAC
		GGTCCAGAGCTTGATCGGCTT
		GAGTATCTGTACCGCGAGC

AC_65R	Deletion PAK_4494 Seq_R	GATGGCGTTCTTCCCTGTTCT
AC_67F	For cloning PAK_P3_sRNA2 in pUC18mini (no RBS)	CAT <u>GGATCCC</u> CAGGGAGCGGAGCCAAAAGAG
AC_67R		AT <u>CTGCAGC</u> CCTCCATTACGGGTTGAGAAAG
AC_73F	Deletion PAK_4496-4497 Down_F	GAACTCGACGGCAGCGAAGGTCATCGCATCGTCATC ACATCG
AC_73R	Deletion PAK_4496-4497 Down_R	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTA GCGTTACAACCGGTCCTGTGAG
AC_74F	Deletion PAK_4496-4497 Up_F	GGG GAC AAG TTT GTA CAA AAA AGC AGG CTC A GAGTGGTATCAGCGCGTCGAG
AC_74R	Deletion PAK_4496-4497 Up_R	CTTCGCTGCCGTCGAGTTCT
AC_75F	Deletion PAK_4496-4497 Seq_F	CGAGAGGCTAGTGGAAAGCCAC
AC_75R	Deletion PAK_4496-4497 Seq_R	GACTCCTCGACCGCCGGGATG
AC_77F	algU to clone in pKT25	GGGCCC <u>TCTAGA</u> Gatgctaaccaggaac
AC_77R		ATCTTT <u>CCCGGGT</u> CAGGCTTCTCGCAACAAAG

PLASMID LIST

Plasmid name	Description	Source
BACTH assays		
pKT25	Backbone vector for bacterial two-hybrid assay	From G. Karimova
pKNT25	Backbone vector for bacterial two-hybrid assay	From G. Karimova
pUT18	Backbone vector for bacterial two-hybrid assay	From G. Karimova
pUT18C	Backbone vector for bacterial two-hybrid assay	From G. Karimova
pKT25-zip	Positive control for bacterial two-hybrid assay	From G. Karimova
pUT18C-zip	Positive control for bacterial two-hybrid assay	From G. Karimova
pKNT25-gp92	pKNT25 cloned with gp92	This study
pUT18-gp92	pUT18 cloned with gp92	This study
pUT18C-gp92	pUT18C cloned with gp92	This study
pKT25-RseA	pKT25 cloned with rseA	From G. Karimova
pKT25-MucA	pKT25 cloned with mucA	This study
pKT25-AlgU	pKT25 cloned with algU	This study
Localisation assays		
pKTop	Backbone vector for membrane topology analysis	From G. Karimova
pKTop-Blr	pKTop cloned with blr encoding a transmembrane protein (control)	From G. Karimova
pKTop-gp92	pKTop cloned with gp92	This study
pKTop-gp92C	pKTop cloned with gp92C (nt 69-234)	This study
pKTop-gp92N	pKTop cloned with gp92N (nt 1-78)	This study
pKTop-SP _{PhoA} -gp92C	pKTop cloned with phoA signal sequence in frame with gp92C (nt 69-234)	This study
pKTop-SP _{PhoA}	pKTop cloned with phoA signal sequence	This study
Toxicity assays		
pENTR™/SD/D/TOPO™	Cloning vector: Gateway™ system entry vector	Invitrogen
pENTR-gp92	pENTR™ cloned with gp92	This study
pUC18mini-GW	pUC18mini backbone vector - compatible with Gateway™ cloning system	From J. Wagemans
pUC18mini-empty	pUC18mini-GW with an empty expression cassette	From J. Wagemans
pUC18mini-gp92	pUC18mini cloned with gp92	This study
pUC18mini-gp92N	pUC18mini cloned with gp92N	This study
pUC18mini-gp92C	pUC18mini cloned with gp92C	This study
pUC18mini-SP _{PhoA} -gp92C	pUC18mini cloned with phoA signal sequence in frame with gp92C (nt 69-234)	This study
pUC18mini-SP _{PhoA}	pUC18mini cloned with phoA signal sequence	This study
Deletion mutants		
pDONRpEX18Gm	Allelic exchange vector replicative in <i>Pseudomonas</i> (to build deletion/insertions/SNPs) (Gateway™ compatible)	From H. Almlad
pDONRpEX18-4980	Allelic exchange vector pDONRpEX18Gm carrying PAK_4980 deletion allele	This study
pDONRpEX18-4981	Allelic exchange vector pDONRpEX18Gm carrying PAK_4981 deletion allele	This study

pDONRpEX18-4982	Allelic exchange vector pDONRpEX18Gm carrying PAK_4982 deletion allele	This study
pDONRpEX18-4493	Allelic exchange vector pDONRpEX18Gm carrying PAK_4493 deletion allele	This study
pDONRpEX18-4494	Allelic exchange vector pDONRpEX18Gm carrying PAK_4493 deletion allele	This study
Expression assays		
pUC18mini-PAK_P3-sRNA1	pUC18mini (without RBS) with 500bp region containing PAK_P3 sRNA1	This study
pUC18mini-PAK_P3-sRNA2	pUC18mini (without RBS) with 500bp region containing PAK_P3 sRNA2	This study
Complementation test		
pUCP18	<i>E.coli</i> - <i>P. aeruginosa</i> shuttle vector	From R.Voulhoux
pUCP18-algU	pUCP18 shuttle vector cloned with algU gene (complementation)	This study

ARTICLE 1

—

**THE SEARCH FOR THERAPEUTIC BACTERIOPHAGES
UNCOVERS ONE NEW SUBFAMILY AND TWO NEW GENERA OF
PSEUDOMONAS-INFECTING *MYOVIRIDAE***

Published in PLoS One

RESEARCH ARTICLE

The Search for Therapeutic Bacteriophages Uncovers One New Subfamily and Two New Genera of *Pseudomonas*-Infecting *Myoviridae*

Marine Henry¹, Louis-Marie Bobay^{2,3,4}, Anne Chevallereau^{1,5}, Emilie Sausseureau^{1,4}, Pieter-Jan Ceyssens^{6,7}, Laurent Debarbieux^{1*}

1 Institut Pasteur, Molecular Biology of the Gene in Extremophiles Unit, Department of Microbiology, Paris, France, **2** Institut Pasteur, Microbial Evolutionary Genomics Unit, Department of Genomes and Genetics, Paris, France, **3** CNRS, UMR3525, Paris, France, **4** Université Pierre et Marie Curie, Cellule Pasteur UPMC, Paris, France, **5** Université Paris Diderot, Sorbonne Paris Cité, Cellule Pasteur, Paris, France, **6** Laboratory of Gene Technology, Division of Gene Technology, Katholieke Universiteit Leuven, Heverlee, B-3001, Belgium, **7** Unit of Bacterial Diseases, Scientific Institute of Public Health (WIV-ISP), Brussels, Belgium

* laurent.debarbieux@pasteur.fr



OPEN ACCESS

Citation: Henry M, Bobay L-M, Chevallereau A, Sausseureau E, Ceyssens P-J, Debarbieux L (2015) The Search for Therapeutic Bacteriophages Uncovers One New Subfamily and Two New Genera of *Pseudomonas*-Infecting *Myoviridae*. PLoS ONE 10 (1): e0117163. doi:10.1371/journal.pone.0117163

Academic Editor: Krystyna Dąbrowska, Institute of Immunology and Experimental Therapy, Polish Academy of Sciences, POLAND

Received: April 8, 2014

Accepted: December 19, 2014

Published: January 28, 2015

Copyright: © 2015 Henry et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by a grant from Vaincre la Mucoviscidose (IC1011) to MH and a grant from the Ministère de l'enseignement supérieur et de la recherche to LMB. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

In a previous study, six virulent bacteriophages PAK_P1, PAK_P2, PAK_P3, PAK_P4, PAK_P5 and CHA_P1 were evaluated for their *in vivo* efficacy in treating *Pseudomonas aeruginosa* infections using a mouse model of lung infection. Here, we show that their genomes are closely related to five other *Pseudomonas* phages and allow a subdivision into two clades, PAK_P1-like and KPP10-like viruses, based on differences in genome size, %GC and genomic contents, as well as number of tRNAs. These two clades are well delineated, with a mean of 86% and 92% of proteins considered homologous within individual clades, and 25% proteins considered homologous between the two clades. By ESI-MS/MS analysis we determined that their virions are composed of at least 25 different proteins and electron microscopy revealed a morphology identical to the hallmark *Salmonella* phage Felix O1. A search for additional bacteriophage homologs, using profiles of protein families defined from the analysis of the 11 genomes, identified 10 additional candidates infecting hosts from different species. By carrying out a phylogenetic analysis using these 21 genomes we were able to define a new subfamily of viruses, the *Felixounavirinae* within the *Myoviridae* family. The new *Felixounavirinae* subfamily includes three genera: *Felixounalikevirus*, *PAK_P1likevirus* and *KPP10likevirus*. Sequencing genomes of bacteriophages with therapeutic potential increases the quantity of genomic data on closely related bacteriophages, leading to establishment of new taxonomic clades and the development of strategies for analyzing viral genomes as presented in this article.

Introduction

In its first report on antibiotic resistance (published on March 31, 2014) the World Health Organization pointed out that, everyone on the planet is now at risk of infection by untreatable multidrug resistant (MDR) bacterial infections (http://www.who.int/iris/bitstream/10665/112642/1/9789241564748_eng.pdf). Proposed solutions to this worldwide threat to public health include better hygiene, access to clean water, infection control in health-care facilities, vaccination and control of antibiotic prescriptions. In addition to these available solutions, discovery of new antibacterial drugs is strongly encouraged. Among current and future solutions, phage therapy—the use of bacteriophages (viruses infecting bacteria) to treat bacterial infections—occupies a singular place. This therapeutic treatment started to be used to treat human bacterial infections before the discovery of antibiotics but was later discontinued in most countries except in Eastern Europe where it was and is still regularly used, in particular in Georgia, Poland and Russia [1,2]. Facing the need for new antibacterial weapons, interest in phage therapy has reignited in the past few years, with an increasing number of publications reporting on the curative efficacy of bacteriophages in various animal models of infection. However, these reports are not always accompanied by molecular studies of the therapeutic bacteriophages.

Nowadays, the reduced cost of sequencing is an incentive to provide access to the raw molecular data of therapeutic bacteriophages [3]. However, analysis of these data is still a major challenge due to the poor conservation of sequences between bacteriophage genomes. In addition, each bacteriophage contains several “orphans” to which a function is difficult to assign. In most cases, a rapid analysis of bacteriophage genomes can determine whether a bacteriophage is temperate or virulent. This is valuable information because temperate bacteriophages would not be recommended for therapeutic use due to their capabilities to exchange genetic material with bacterial strains [4]. Nevertheless, beyond such analysis, one can distinguish between two situations: either a genome reveals its close proximity to published genomes, or no close homolog can be found making it rather difficult to assign a classification with confidence and to elaborate a strategy for molecular characterization. We hypothesize that comparative genomics of closely related bacteriophages may both help their accurate classification and highlight molecular characteristics which could be used to guide further analysis.

From our previous work, we isolated in 2006 and 2009 bacteriophages infecting *Pseudomonas aeruginosa*, a Gram-negative opportunistic pathogen widespread in the environment [5–7]. *P. aeruginosa* is acknowledged as the leading cause of chronic infections in cystic fibrosis patients. It is frequently isolated in cases of ventilation-associated pneumonia, chronic obstructive pulmonary disease, and also on the skin of burns patients and other sites such as urinary tract and ears [8–10]. The therapeutic potential of these bacteriophages was then evaluated using a mouse model of acute lung infection and some were also included in a preclinical study performed on cystic fibrosis sputa samples [5,6,11,12]. The genomes of two of these bacteriophages, namely PAK_P1 and PAK_P3, were sequenced in 2009 and published in 2010 and 2011 respectively, revealing no close relationship to any other published bacteriophage genomes [5,6]. The mass spectrometry of major capsid proteins of these bacteriophages led us to identify a distant homology (less than 30% identity) to the major capsid protein of Felix O1 bacteriophage [5,6]. The relationship to Felix O1 bacteriophage has never been characterized further despite the publication of four other closely related *P. aeruginosa* bacteriophages, namely KPP10, JG004, PaP1 and vB_PaeM_C2-10_Ab1 [13–16]. We report here on the genome sequences of four additional bacteriophages isolated by our group (namely PAK_P2, PAK_P4, PAK_P5 and CHA_P1) and describe their genome organization through the comparison with other closely related bacteriophages infecting *P. aeruginosa*. Our findings, based on analysis of protein family profiles led us to develop a coherent bacteriophage taxonomy comprising two new genera, and

new subfamily of viruses (tentatively named *Felixounavirinae*) within the *Myoviridae* family. Identification of some original and intriguing molecular characteristics was also successful.

Results

Isolation of six new bacteriophages infecting *P. aeruginosa*

In 2006 we isolated five bacteriophages infecting the PAK strain of *P. aeruginosa* (PAK_P1, PAK_P2, PAK_P3, PAK_P4 and PAK_P5) and, in 2009, a bacteriophage (CHA_P1) infecting the CHA strain, a cystic fibrosis isolate [17]. Electron microscopy showed that these six bacteriophages had a similar morphology (see Fig. 1A for images of previously unpublished bacteriophages), with an icosahedral head and criss-cross pattern on the tail, characteristics resembling FelixO1 bacteriophage infecting *Salmonella* [18]. Despite most of them having been isolated on the same host, a restriction fragment length polymorphism analysis revealed that their genomic content was not identical (not shown). The genomes of these six bacteriophages were then sequenced and a Megablast analysis revealed their close similarity to five other bacteriophages infecting also *P. aeruginosa*, JG004 [13], PaP1 [15] vB_PaeM_C2-10_Ab1 [14], KPP10 [16] and LSL4 [19]. An additional bacteriophage, P3_CHA [6], was excluded from this study because the difference between its DNA sequence and that of PAK_P3 was negligible (2 nucleotides). Detailed annotation of the genomes of the six newly sequenced bacteriophages is provided as supplementary information (S1 Text). Briefly, we found that 10 to 15% of the predicted ORFs could be linked to a putative function while 5 to 10% displayed no similarity to any other sequence in current databases. About 50% of each bacteriophage genome displays sequence similarity only to genes encoding hypothetical proteins (unknown functions) present only in the 11 bacteriophages listed above. The remaining ORFs were also annotated as encoding hypothetical proteins but presented similarities to ORFs from mostly bacteriophage or prophage genomes. Position of genome termini of the six bacteriophage genome sequenced was identified using sequence coverage (see S1 Table and S1 Text) [20,21].

Genomic and proteomic analysis of PAK_P1-like and KPP10-like bacteriophages

Several characteristics (genome length, GC content and number of tRNA) of the 11 bacteriophages mentioned above suggest they could be classified into two distinct clades (Table 1). We decided to name these proposed two clades according to the publication date of the genome of the first bacteriophage discovered in each clade: “PAK_P1-like” (including PAK_P1 published in 2010, PAK_P2, PAK_P4, JG004, PaP1 and vB_PaeM_C2-10_Ab1) and “KPP10-like” (including KPP10 published in 2011, PAK_P3, PAK_P5, CHA_P1 and LSL4).

PAK_P1-like bacteriophages have a mean genome size of 92.8 kb (SD = 598 bp), a mean GC% of 49.3% (SD = 0.09%) and carry 11 to 13 tRNAs. By contrast, KPP10-like bacteriophages have a mean genome size of 88.2 kb (SD = 345 bp), a mean GC content of 54.8% (SD = 0.08%) and only three predicted tRNAs: a tRNA-Asn, a tRNA-Tyr and a tRNA-Gln. In both clades GC content is significantly lower than that of the *P. aeruginosa* host (67% GC for the *P. aeruginosa* core genome [22–24]), consistent with previous observations that bacteriophages tend to have a higher proportion of A+T residues than their bacterial host [25]. Most bacteriophages contain one or two tRNAs, but a few (including bacteriophage FelixO1) have been shown to contain more than 20 [26]. In both clades, the tRNAs were found to be located in the close vicinity of the packaging ORFs, upstream from the large terminase subunit, spanning regions of 2.52 kb (PAK_P1-like) and 400 bp (KPP10-like) (Fig. 2). The codon usage of the representative bacteriophages PAK_P1 and PAK_P3 was compared with that of their isolation host (the PAK strain; S2 Table). We

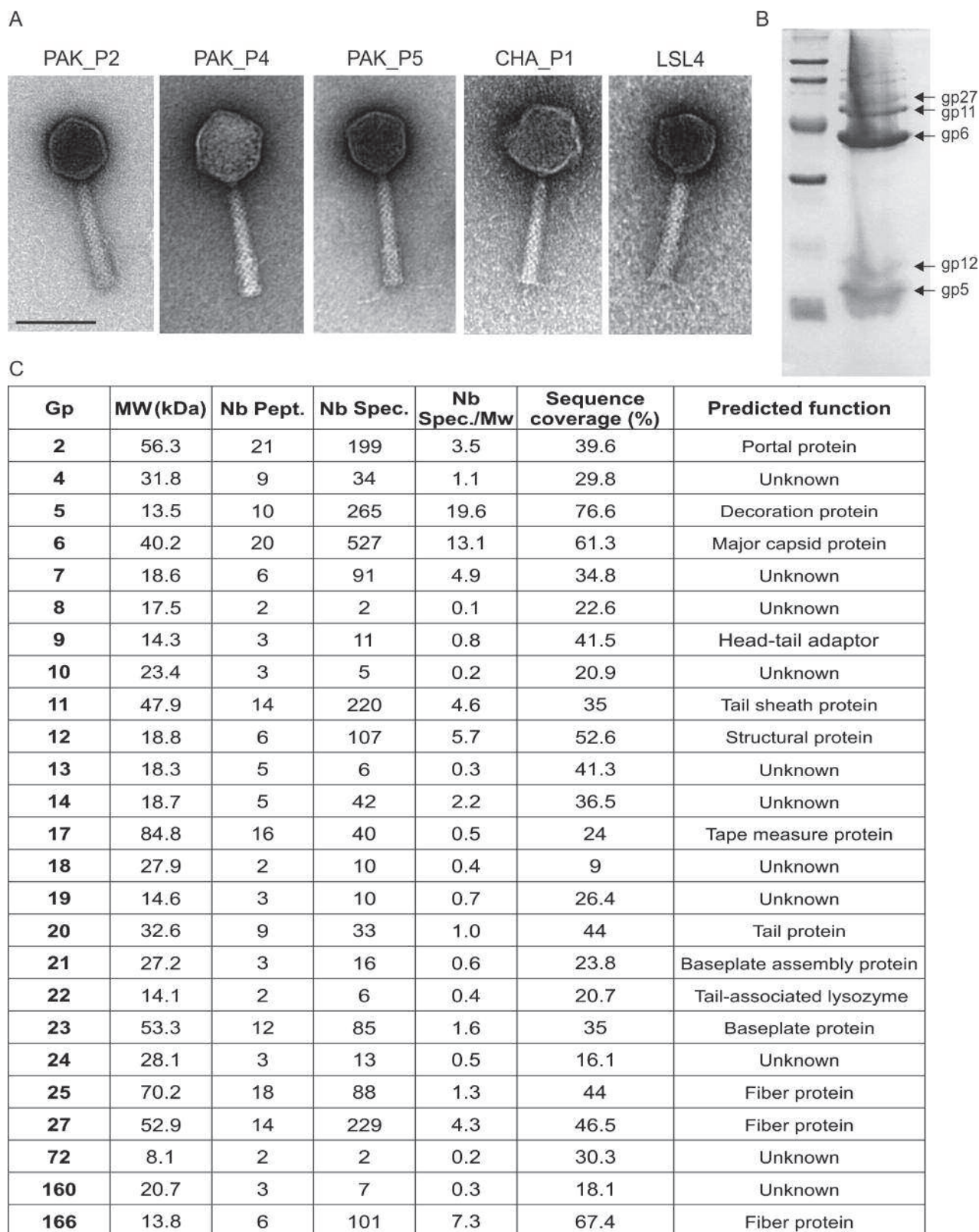


Figure 1. Electron microscopy of bacteriophages and proteomic analysis of PAK_P3 virion. A. Electron micrographs of the indicated bacteriophages (scale bar: 100 nm). B. Denaturing polyacrylamide gel of PAK_P3 virion proteins. C. Proteins identified by mass spectrometry analysis (ESI-MS/MS). MW,

theoretical molecular weight; Nb Pept., number of unique peptides identified; Nb Spec., total number of spectra. Nb Spec/Mw, relative abundance; Sequence coverage, percentage of the protein sequence covered by peptides

doi:10.1371/journal.pone.0117163.g001

found that, in PAK_P1, 13 tRNAs correspond to codons used more frequently than in the host, with tRNA-Leu and tRNA-Arg being the most frequent (respectively 16 and 29 times more). By contrast, in PAK_P3, a higher frequency was found only for the tRNA-Gln.

Published results of a proteomic analysis of PaP1 bacteriophage (member of the PAK_P1-like clade) reported the identification of 12 proteins while a preliminary report on KPP10 (member of the KPP10-like clade) reported only 7 proteins. In order to obtain additional identification of structural proteins from a bacteriophage of the latter clade we analyzed the virion proteins from PAK_P3 (Fig. 1B). A total of 25 proteins were identified using ESI MS/MS analysis (Fig. 1C). Twenty one of them are encoded within the structural region of the genome, two (gp160 and gp166) were relatively close to it and one, gp72, was not. As the abundance of Gp72 was low, this protein may display some affinity for structural proteins, rather than directly taking part in the virion assembly. Gp6, the major capsid protein, was the most abundantly identified protein. Notably, the head decoration protein Gp5 had the highest relative abundance (#spectra/mol. weight) suggesting a prominent role in PAK_P3 capsid morphology.

Identification of putative regulatory elements

Alignments of the nucleotide sequences of bacteriophage genomes of PAK_P1-like clade revealed the presence of a variable region which is 11 kb long (approximately 12% the length of the genome) (Fig. 3A). A blast analysis and the subsequent alignment of regions from all the PAK_P1-like bacteriophages revealed the presence of a 41 nucleotide-long repeated sequence (5 repeats in PAK_P1, PaP1, vB_PaeM_C2-10_Ab1, JG004 and 6 repeats in PAK_P2, PAK_P4). Alignment of the entire set of repeats from the PAK_P1-like clade revealed that 32 of the 41 nucleotides were strictly conserved (Fig. 3B). As this intergenic motif contains a conserved σ^{70} promoter sequence (TTGACA-N₁₇-TAgAAT), it most likely serves to guide the bacterial RNA polymerase to the early phage genes at the onset of phage infection. Similar promoter repeats were identified in the putative early genome regions of the KPP10-like clade, but at a much lower frequency, with each member of this clade containing only two repeats (S1A

Table 1. General characteristics of the genomes of the 11 bacteriophages belonging to either the PAK_P1-like or the KPP10-like clades.

Clade	Name	Length	ORFs predicted	GC content	t-RNA	Accession	Reference
PAK_P1-like	PAK_P1	93398	181	49.50%	13	KC862297	[5]
	PAK_P2	92495	176	49.30%	11	KC862298	[11]
	PAK_P4	93147	174	49.30%	13 + 1 pseudo	KC862300	[11]
	JG004	93017	161	49.30%	12	NC_019450.1	[13]
	C2-10_Ab1	92777	158	49.28%	12	NC_019918.1	[14]
	PaP1	91715	157	49.40%	13	NC_019913.1	[15]
KPP10-like	KPP10	88322	146	54.80%	3	NC_015272.1	[16]
	PAK_P3	88097	165	54.80%	3	KC862299	[6]
	PAK_P5	88789	164	54.70%	3	KC862301	[11]
	CHA_P1	88255	166	54.60%	3	KC862295	[11]
	LSL4 ^a	87739	165	54.80%	3	Not published	[19]

a: the LSL4 bacteriophage was isolated in Sri Lanka, with the Lio12 strain, and its genome sequence was kindly provided by R. Lavigne.

doi:10.1371/journal.pone.0117163.t001

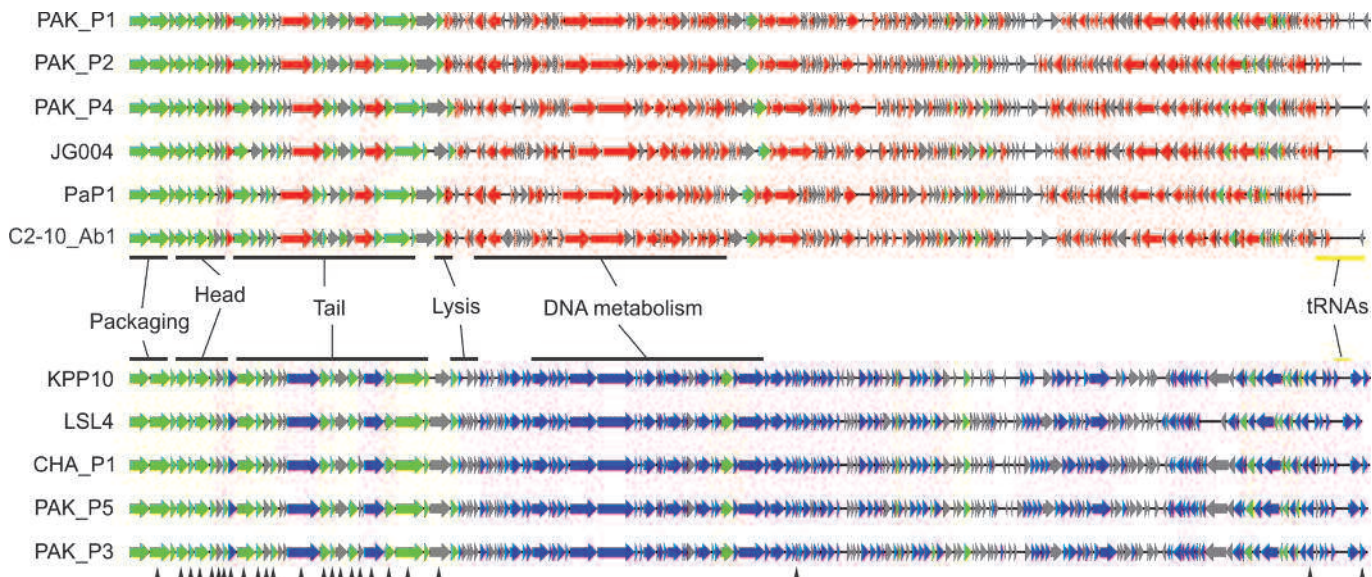


Figure 2. Maps of the 11 genomes involved in this study. In green, the core ORFs homologous in the 11 bacteriophages; red and blue ORFs are common to the PAK_P1-like and KPP10-like bacteriophages, respectively. Arrows (▲) designate the sequences corresponding to proteins identified by mass spectrometry of the PAK_P3 virion.

doi:10.1371/journal.pone.0117163.g002

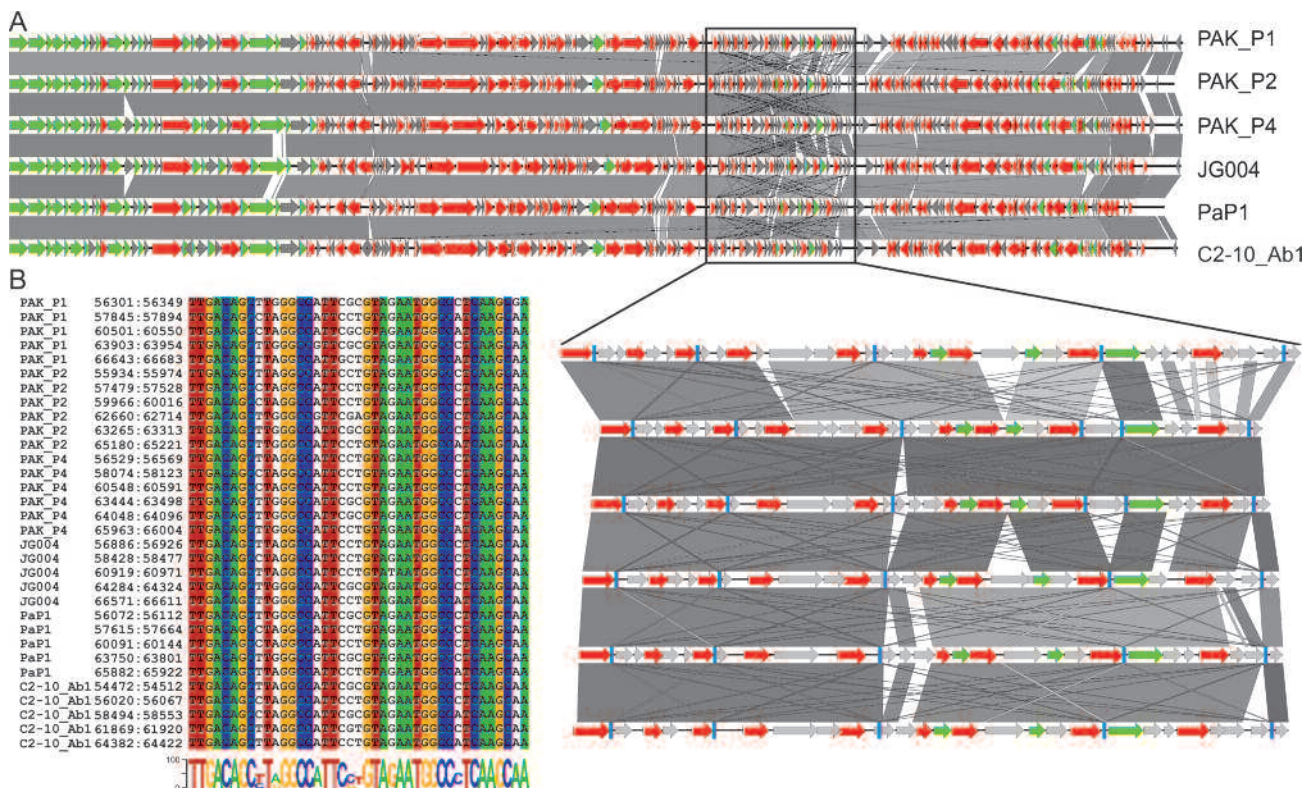


Figure 3. Identification of repeated promoter-like regions of the PAK_P1-like clade. A. Representation of the homology (blastn) between the six PAK_P1-like bacteriophages, with an enlargement of the region in which 41-nt repeats were identified (represented as light blue boxes). B. Alignment of the total of 32 repeats identified in the six genomes, with colors indicating strictly conserved bases, and a WebLogo representation of the consensus (the height of the letters represents their frequency at each position) below the alignment.

doi:10.1371/journal.pone.0117163.g003

Fig.). More unusual is the perfect nucleotide conservation surrounding the-35 and-10 motifs. In a blastn search against the nonredundant database, the consensus sequence of these repeats yielded hits only with bacteriophage genomes from these two clades, with an e-value < 0.0001 (the consensus sequence for repeats of the KPP10-like clade is more conserved; S1B Fig.). Therefore, these repeats are an additional characteristic to these two clades which could be linked to the common bacterial species they infect.

Designation of core genes and identification of specific markers

We determined the proportion of gene products conserved in the PAK_P1-like and KPP10-like clades, by carrying out a comparative genomics analysis with a 40% similarity threshold and a size constraint (Materials and Methods). The two clades were well delimited, with a mean of 86% and 92% of proteins considered homologous within individual clades, and 25% proteins considered homologous between the two clades (S3 Table).

We used these parameters to define protein families, for the subsequent creation of protein profiles for screening against all publicly available bacteriophage sequences, to perform our annotation (Materials and Methods). In total, 404 protein families (S4 Table) were identified, several of which included ORFs specific to only the PAK-P1-like or KPP10-like clade of bacteriophages (represented in red and blue, respectively, in Fig. 2). Many of these ORFs appeared to be conserved between the two clades but were nevertheless too divergent to include them in the set of 26 “core families” with a homolog in each of the 11 bacteriophages. The ORFs corresponding to these 26 families were designated as the core genes of the 11 bacteriophages (S5 Table). Unsurprisingly, most of these core genes belong to the structural module of these genomes (Fig. 2). Despite the search using protein profiles very few predicted ORFs could be assigned with confidence to a function. We therefore carried out additional analyses (transmembrane domains and structural similarities), including iterative searches with the alignment of the 11 homologs from each family (S1 Text). Nevertheless, the majority of ORFs still could not be assigned to a function highlighting the novelty of the bacteriophages.

We then searched the 11 bacteriophage genomes for markers that could be used to detect in a specific manner bacteriophages infecting *P. aeruginosa* belonging to the two clades. Four core families were then selected (families 5, 21, 22 and 25 in S5 Table). The corresponding ORFs from the 11 genomes showed no significant matches (e-value > 0.001) with other elements in blast searches against the nonredundant database. Alignments of ORFs from family 5 displayed regions of strong identity, which we used to design specific degenerated primers that were then tested experimentally (5'-CATCAGCGYCTKAGCAACTGGCT-3' and 5'-CTGGTSWAC YGCGAAGATGTTCT-3'). The detection of as few as 100 pfu of PAK_P1 was achieved in solution containing 1x10⁷ pfu of PhiKZ, an unrelated *Myoviridae* phage infecting *P. aeruginosa* (no PCR product was obtained when using a solution containing only PhiKZ). This set of primers should therefore be sufficient to detect the presence of a bacteriophage from these two clades (we also obtained a PCR product for PAK_P5). Sequencing of the PCR products allows further assignment of the bacteriophage to one of the two clades.

Evolutionary relationships of bacteriophage genomes related to the *Felixounalikevirus* genus

We then attempted to characterize relationships between the bacteriophages of both the PAK_P1-like and KPP10-like clades and more distantly related bacteriophages. Using HMMER we built sequence profiles for the families of homologous proteins defined in the 11 bacteriophages and searched for homologous proteins among the bacteriophage genomes of GenBank. We identified 10 bacteriophages containing a number of genes (>20) with significant matches

Table 2. General characteristics of the 10 bacteriophages most closely related to *Pseudomonas*-infecting bacteriophages of PAK_P1-like and KPP10-like clades.

Clade	Name	Host	Size (bp)	Homologous proteins ^a (%)	Genome ID
FelixO1-like	phiEa104	Erwinia	84565	28 (24%)	NC_015292
	phiEa21-4	Erwinia	84576	26 (22%)	NC_011811
	WV8	Escherichia	88487	27 (19%)	NC_012749
	FelixO1	Salmonella	86155	28 (21%)	NC_005282
rV5-like	PVP-SE1	Salmonella	145964	38 (16%)	NC_016071
	vB_CsaM_GAP31	Cronobacter	147940	37 (14%)	NC_019400
	vB_EcoM-FV3	Escherichia	136947	28 (13%)	NC_019517
	CR3	Cronobacter	149273	30 (11%)	NC_017974
	rV5	Escherichia	137947	27 (12%)	NC_011041
	ICP1	Vibrio	125956	21 (9%)	NC_015157

^a indicates the number of homologous proteins identified with HMMER (e-value<0.001) by comparison to the total set of protein families of the 11 PAK_P1-like and KPP10-like bacteriophages.

(%) indicates the number of homologous proteins identified divided by the total number of proteins encoded by each genome.

doi:10.1371/journal.pone.0117163.t002

(e-value<0.001). These 10 additional bacteriophages belong to *Myoviridae* family of viruses and have genomes larger than 84 kb (Table 2).

In the set of 21 genomes we identified three conserved genes predicted to encode structural proteins (portal, major capsid and tail sheath). The corresponding protein sequences were then aligned, concatenated and a phylogenetic tree was built using the maximum likelihood method (see Materials and Methods). Not surprisingly, the two PAK_P1-like and KPP10-like clades of bacteriophages were found to be most closely related to each other (Fig. 4). The closest clade to those is constituted by the four bacteriophages belonging to the *Felixounalikevirus* genus (FelixO1, wV8, phiEa21-4 and phiEa104). These four bacteriophages have several features in common with the PAK_P1-like and KPP10-like bacteriophages, including a large number of tRNAs (>20), similar genome sizes (~90 kb) and almost identical morphologies. These findings strongly suggest that these three clades are related. An analysis with the CoreGenes program, which is used to define bacteriophage taxonomic groups [18,27], indicated that FelixO1 had only 22 and 21 homologous proteins in common with PAK_P1 and PAK_P3 respectively. These values, corresponding to 14 and 11% of the respective proteomes of these bacteriophages are well below the 40% shared proteins used to define a genus, which confirms that these clades belong to different genera [18]. The second closest clade to the PAK_P1-like and KPP10-like clades contained five bacteriophages, infecting various *Enterobacteria*, which belong to a proposed rV5-like viruses genus (CR3, vB_CsaM_GAP31, PVP-SE1, rV5 and vB_EcoM-FV3) [28,29]. Finally, the most distantly related bacteriophage was the *Vibrio* bacteriophage ICP1. This molecular phylogeny analysis therefore revealed that the *Cronobacter* bacteriophages vB_CsaM_GAP31 and CR3, the *Salmonella* bacteriophage PVP-SE1 and the *Escherichia coli* bacteriophages FV3 and rV5 are closely related to each other, forming an rV5-like genus divergent from *Felixounalikevirus*, PAK_P1-like and KPP10-like bacteriophages and from the *Vibrio* bacteriophage ICP1 (Fig. 4) [28]. It has recently been suggested that this rV5-like genus could be split into three separate genera, rV5-like viruses (rV5 and FV3), PVP-like viruses (PVP-SE1, GAP31 and SSE-121, an as yet unpublished genome from A. Letarov) and Phi92-like viruses [30], which is consistent with our molecular phylogeny analyses.

Horizontal exchange may however affect the organization of bacteriophage genomes and blur phylogenetic reconstructions. As a consequence, genes located in different functional

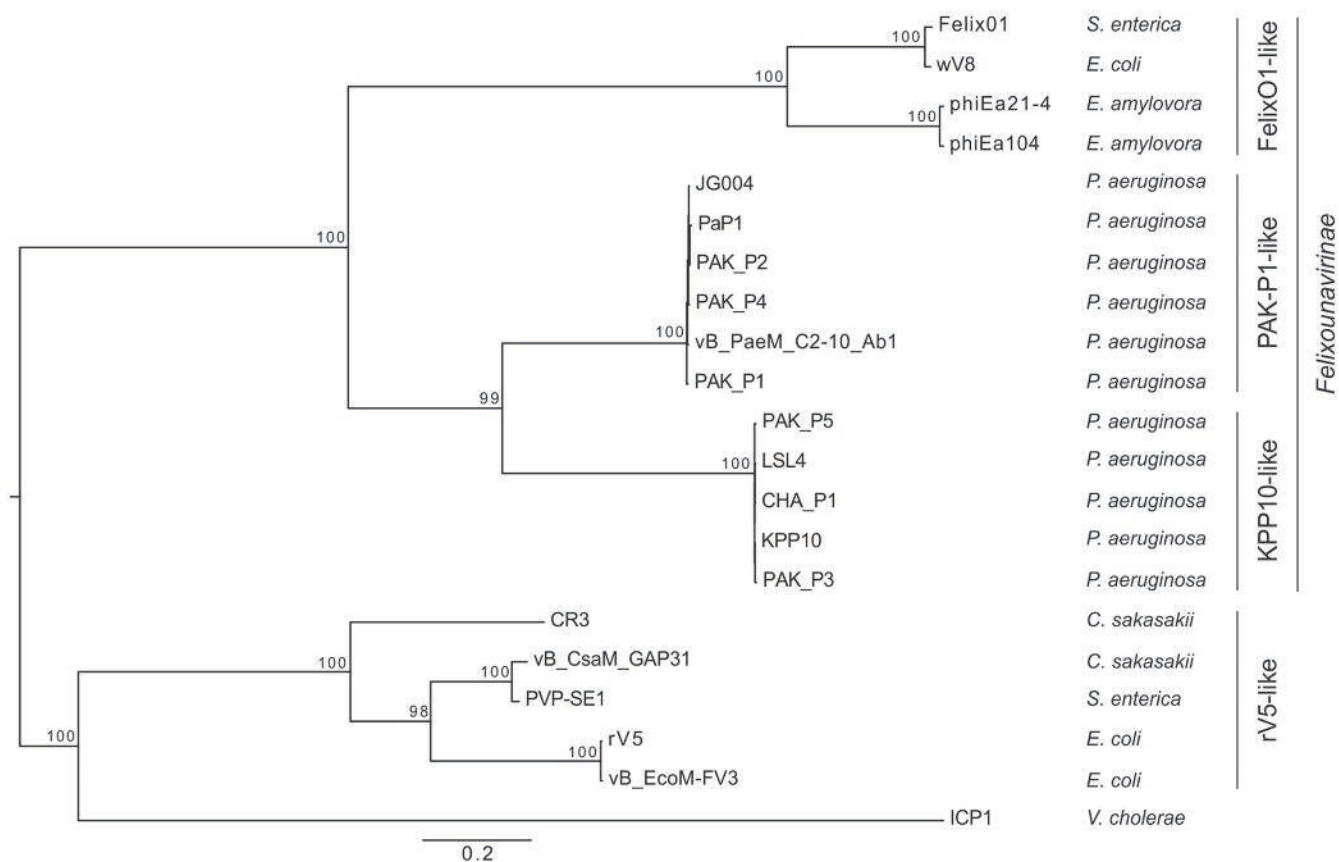


Figure 4. Phylogenetic tree of PAK_P1-like and KPP10-like bacteriophages and their closest relatives. The maximum likelihood tree was built from a concatenated alignment of three core proteins (predicted to encode the portal protein, the major capsid protein and the tail sheath protein) common to the 21 bacteriophages. Bootstrap values are indicated and the tree was rooted on the midpoint root.

doi:10.1371/journal.pone.0117163.g004

modules may have different evolutionary histories [31]. We therefore used a conserved non-structural protein, the primase, to reconstruct the phylogeny of 19 of these bacteriophages (no homolog of the primase was identified in PVP-SE1 and vB_CsaM_GAP31). The phylogenetic tree obtained with primase sequences confirmed the phylogeny based on structural proteins (S2 Fig.). These results suggest that PAK_P1-like and KPP10-like bacteriophages are related to bacteriophages infecting different hosts. These results do not support relationships of these two clades with other *Pseudomonas*-infecting bacteriophages as proposed by Lu *et al.* [15].

Discussion

Worldwide reports on MDR infections and the lack of new antibacterial drugs led to the reevaluation of phage therapy. While *in vivo* data from various animal models are encouraging, molecular studies on candidate therapeutic bacteriophages are still scarce. Following a recent report on *in vitro* and *in vivo* evaluations of six bacteriophages infecting *P. aeruginosa* (PAK_P1, PAK_P2, PAK_P3, PAK_P4 and PAK_P5 and CHA_P1), we performed the *in silico* analysis of these bacteriophage genomes. They were found to be closely related to another five genomes present in the database which infect *P. aeruginosa* (JG004, PaP1, vB_PaeM_C2-10_Ab1, KPP10 and LSL4). Comparative analysis of the general characteristics of these 11 genomes (GC content, genome length, number of tRNAs) suggested that they constitute two clades named PAK_P1-like and KPP10-like viruses. Consistent with these characteristics, the bacteriophages

of these two clades share less than 40% of proteins, despite displaying an almost identical genome organization [18]. It is worth noting that, during the revision of this manuscript, the genome of PhiPsa374, a bacteriophage infecting *Pseudomonas syringae*, was published as being closely related to PAK_P1 and JG004 bacteriophages [32], which suggests it would belong to the PAK_P1 like clade.

An additional characteristic of these 11 bacteriophages was identified with the repetition of promoter-like sequences located in a short region which could correspond to a putative early transcribed region. The consensus sequences of these repeats are very specific suggesting that these bacteriophages may use a particular way to regulate gene expression. As no homologs to these sequences were found in PhiPsa374 bacteriophage infecting *P. syringae*, it is tempting to speculate that they may be specific to *P. aeruginosa*.

The 11 bacteriophages were isolated in various countries (France, Japan, Germany, China and Ivory Coast), at different time periods, from different hosts, but their genomes did not display mosaic structure frequently observed in genomes of bacteriophages [31,33]. This may be due to the lack of recombinases, which are involved in genome mosaicism [34–37]. Finally, in blast searches against the nonredundant database (last check in June 2014) with primer sequences designed to detect specifically bacteriophages belonging to PAK_P1-like or KPP10-like clades, no matches outside this group of 11 bacteriophages were obtained. This suggests that either these primers are too stringent (they indeed excluded PhiPsa374) or that bacteriophages belonging to these clades are not abundant and not yet represented in virome data.

Genomic analysis of the 11 bacteriophages led to the definition of protein profiles, which are more sensitive than sequence-sequence comparisons for the detection of distantly related homologs [38]. This is particularly important when viruses infect different species, as the tendency towards sequence adaptation to hosts leads to considerable divergence. Indeed, we identified 10 distantly related bacteriophages infecting various hosts (Felix O1, wV8, phiEa21-4, phiEa104, CR3, vB_CsaM_GAP31, PVP-SE1, rV5, vB_EcoM-FV3 and ICP1). We were then able to place the two new clades in the virus classification, using three conserved structural proteins to reconstruct phylogeny, revealing new relationships between the entire set of 21 bacteriophages. This reconstruction of phylogeny, in addition to the proteome comparisons of PAK_P1 and PAK_P3 with Felix O1, provided clear support for the creation of two new genera: *PAK_P1likevirus* (including PAK_P1, PAK_P2, PAK_P4, JG004, PaP1 and vB_PaeM_C2-10_Ab1) and *KPP10likevirus* (including KPP10, PAK_P3, PAK_P5 and CHA_P1). We suggest that these two genera could, together with the *FelixO1likevirus* genus (including Felix O1, wV8, phiEa21-4 and phiEa104), be grouped into a new subfamily of the *Myoviridae* named *Felixounavirinae*. Our results also showed that classification methods based on gene content [39,40], give a reliable information despite the limitations imposed by the rapid evolution of divergent sequences in bacteriophage genomes [41,42].

Our *in silico* analysis not only revealed new relationships between bacteriophages but also paves the way for a better molecular characterization of these viruses which display a clear therapeutic potential. Our results clearly highlighted the most conserved genetic elements which could represent the first targets for in depth molecular characterization. For example, out of the nine core genes not associated with putative functions, three (family number 21, 22 and 23) are located outside the structural region and therefore are most likely to carry essential functions for these viruses. Additional molecular analysis based on our results should provide insights on whether both genera rely on similar molecular processes to hijack their host, on the molecular basis of the differences in efficacy observed *in vitro* and *in vivo* towards a same host, or on the specific genes needed for infecting other hosts than *P. aeruginosa* (both closely related like *P. syringae*, or more distant like *Salmonella*, *E. coli* and *Erwinia*). In addition, it would be interesting to check whether the conclusions drawn from other comparative

genomics of *P. aeruginosa* bacteriophages belonging to a different genera, such as the one conducted on PhiKMV bacteriophages on host range and antibodies neutralization, apply to the two new genera [43].

To conclude, it seems likely that years will be needed to achieve a complete molecular characterization of new bacteriophages, since such work has not even yet been completed for model bacteriophages such as T4 or T7. From a clinical perspective, with an increasing number of patients running out of antibiotics-based solutions, the question of the extent to which novel bacteriophages, positively evaluated for their therapeutic potential, should be further characterized before being used in medicine is raised [4,44–46].

Materials and Methods

Sequencing and annotation of bacteriophage genomes

Accession numbers of genomes used in this study are reported in Table 1.

The genomes of the PAK_P1, PAK_P2, PAK_P3, PAK_P4, PAK_P5 and CHA_P1 bacteriophages were sequenced by the 454 technique. The various contigs were assembled with Sequencher software (v4.8, Gene Codes Corporation, Ann Arbor, MI, USA). The genome sequences obtained were then submitted to the Phage RAST program [47] and manually curated in the Artemis Genome Browser [48], with NCBI blast tools (blastp, blastn, tblastx and psi-blast). We identified tRNAs with the tRNA-Scan SE online tool, using the default search mode and the same settings as for a bacterial source [49].

Bioinformatic analysis

For clarity, we present all genome alignments with an arbitrary start at the first base of the ORF predicted to encode the large terminase subunit, but we retained the gene identifiers for genomes already published. MegaBlast analysis was used to identify related bacteriophage genomes. Transmembrane domains were predicted with TMHMM Prediction Server (www.cbs.dtu.dk/services/TMHMM). Structural similarity searches were performed with HHPred [50]. Genome maps were generated with Easyfig [51]. Sequence coverage was assessed with Tablet software [52].

Comparative genomics analysis

Homologous proteins were defined as proteins displaying >40% similarity and a difference of <50% in protein length. A similarity score was calculated with the BLOSUM60 matrix and the Needleman-Wunsch end gap-free alignment algorithm (in house software). Families of homologous proteins were then built by transitivity: a protein belongs to the family if homologous to a protein already present in this family. Sequences were aligned with MUSCLE v3.6 [53] and protein profiles were built with HMMER [54] for each protein family. We then used HMMER to compare each protein family profile with a set of complete genome sequences for 831 non-redundant bacteriophages, downloaded from GenBank (<http://ftp.ncbi.nih.gov/genomes/Viruses/> last accessed April 2013). Bacteriophages presenting at least 20 positive matches (e-value<0.001) were retained for phylogenetic analysis.

Phylogenetic analysis

We assessed the relationship between the 21 bacteriophages, defining homologous proteins as described above, but with a lower threshold (>35% similarity and <50% of difference in protein length). Three homologous proteins were found to be common to all 21 bacteriophage genomes and were used to infer their relationships. We first aligned the sequences for each group

of homologous proteins independently, with MUSCLE v3.6 [53]. Non-informative positions were trimmed with BMGE, using the BLOSUM30 matrix [55]. The three alignments were then concatenated into a single alignment and a maximum likelihood tree was built with PhyML v3.0 with a LG + $\Gamma(4)$ model [56]. The topology of the tree was determined with 100 bootstrap replicates, under the same model. The tree was rooted on the midpoint root.

Electron microscopy and proteomic analysis

Cesium chloride-purified bacteriophage preparations were used for electron microscopy studies in a JEOL 1200 EXII electron microscope, after staining with uranyl acetate. Proteomic analysis of the PAK_P3 virion was performed as previously described [57,58]. Briefly, heat-denatured virions proteins were separated onto a SDS-PAGE gel which was stained with Coomassie. Bands were excised and stain removed before reduction by DTT (101mM), followed by alkylation with iodoacetamide (55mM). Following trypsin digestion peptides were analyzed by electrospray ionisation tandem mass spectrometry (ESI MS/MS).

Supporting Information

S1 Text. This text contains details about annotation and genome termini of PAK_P1, PAK_P2, PAK_P3, PAK_P4, PAK_P5 and CHA_P1 bacteriophages.
(DOCX)

S1 Table. Sequences and coordinates of the boundaries of regions covered by high numbers of reads of the six bacteriophage genomes reported in this study.
(PDF)

S2 Table. Comparison of the codon usage of bacteriophages PAK_P1 and PAK_P3 with that of their host (strain PAK)
(PDF)

S3 Table. Percentage of ORFs homologous between bacteriophages of the PAK_P1-like and KPP10-like clades
(PDF)

S4 Table. List of the 404 protein families deduced from the genomes of PAK_P1-like and KPP10-like bacteriophages
(PDF)

S5 Table. Families of core ORFs and families with putative identified functions of PAK_P1-like and KPP10-like bacteriophages. Question marks indicate that only one analysis provided support for the indicated putative function, while absence of question marks indicate that at least two analysis were concordant.
(PDF)

S1 Fig. Identification of repeated promoter-like regions of the KPP10-like bacteriophages
A. Representation of the homology (blastn) between the five KPP10-like bacteriophages in which 41-nt repeats were detected (represented as light blue boxes). B. Alignment of the 10 repeats identified in the five genomes, with colors indicating strictly conserved bases, with a WebLogo representation of the consensus (the height of the letters represents their frequency at each position) shown beneath the alignment.
(PDF)

S2 Fig. Phylogenetic tree for the primase of PAK_P1-like and KPP10-like bacteriophages and their closest relatives. The maximum likelihood tree was built from a protein alignment

of the primase sequences, a non-structural protein, common to 19 bacteriophages (the primase was not identified in the PVP-SE1 and GAP31 genomes). Bootstrap support is indicated on the tree. The tree was rooted on the midpoint root.

(PDF)

Acknowledgments

We thank Alexis Criscuolo for his help with 454 sequencing files. We thank Rob Lavigne as well as reviewers for critical suggestions.

Author Contributions

Conceived and designed the experiments: MH LMB LD. Performed the experiments: MH LMB AC PJC. Analyzed the data: MH LMB PJC LD. Contributed reagents/materials/analysis tools: ES. Wrote the paper: MH LMB PJC LD.

References

1. Burrowes B, Harper DR, Anderson J, McConville M, Enright MC (2011) Bacteriophage therapy: potential uses in the control of antibiotic-resistant pathogens. *Expert Rev Anti Infect Ther* 9: 775–785. doi: [10.1586/eri.11.90](https://doi.org/10.1586/eri.11.90) PMID: [21905786](https://pubmed.ncbi.nlm.nih.gov/21905786/)
2. Saussereau E, Debarbieux L (2012) Bacteriophages in the experimental treatment of *Pseudomonas aeruginosa* infections in mice. *Adv Virus Res* 83: 123–141. doi: [10.1016/B978-0-12-394438-2.00004-9](https://doi.org/10.1016/B978-0-12-394438-2.00004-9) PMID: [22748810](https://pubmed.ncbi.nlm.nih.gov/22748810/)
3. McCallin S, Alam Sarker S, Barretto C, Sultana S, Berger B, et al. (2013) Safety analysis of a Russian phage cocktail: from metagenomic analysis to oral application in healthy human subjects. *Virology* 443: 187–196. doi: [10.1016/j.virol.2013.05.022](https://doi.org/10.1016/j.virol.2013.05.022) PMID: [23755967](https://pubmed.ncbi.nlm.nih.gov/23755967/)
4. Brussow H (2012) What is needed for phage therapy to become a reality in Western medicine? *Virology* 434: 138–142. doi: [10.1016/j.virol.2012.09.015](https://doi.org/10.1016/j.virol.2012.09.015) PMID: [23059181](https://pubmed.ncbi.nlm.nih.gov/23059181/)
5. Debarbieux L, Leduc D, Maura D, Morello E, Criscuolo A, et al. (2010) Bacteriophages can treat and prevent *Pseudomonas aeruginosa* lung infections. *J Infect Dis* 201: 1096–1104. doi: [10.1086/651135](https://doi.org/10.1086/651135) PMID: [20196657](https://pubmed.ncbi.nlm.nih.gov/20196657/)
6. Morello E, Saussereau E, Maura D, Huerre M, Touqui L, et al. (2011) Pulmonary bacteriophage therapy on *Pseudomonas aeruginosa* cystic fibrosis strains: first steps towards treatment and prevention. *PLoS One* 6: e16963. doi: [10.1371/journal.pone.0016963](https://doi.org/10.1371/journal.pone.0016963) PMID: [21347240](https://pubmed.ncbi.nlm.nih.gov/21347240/)
7. Pier GB, Ramphal R (2009) *Pseudomonas aeruginosa*. In: Mandell GL, editor. *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*. 7th ed. Edinburgh: Churchill Livingstone. pp. 2835–2860. doi: [10.1056/NEJMoa1304839.Two](https://doi.org/10.1056/NEJMoa1304839.Two) PMID: [24450891](https://pubmed.ncbi.nlm.nih.gov/24450891/)
8. Zilberberg MD, Shorr AF (2013) Prevalence of multidrug-resistant *Pseudomonas aeruginosa* and carbapenem-resistant Enterobacteriaceae among specimens from hospitalized patients with pneumonia and bloodstream infections in the United States from 2000 to 2009. *J Hosp Med* 8: 559–563. doi: [10.1002/jhm.2080](https://doi.org/10.1002/jhm.2080) PMID: [24022878](https://pubmed.ncbi.nlm.nih.gov/24022878/)
9. Ciofu O, Hansen CR, Hoiby N (2013) Respiratory bacterial infections in cystic fibrosis. *Curr Opin Pulm Med* 19: 251–258. doi: [10.1097/MCP.0b013e32835f1afc](https://doi.org/10.1097/MCP.0b013e32835f1afc) PMID: [23449384](https://pubmed.ncbi.nlm.nih.gov/23449384/)
10. Williams BJ, Dehnbostel J, Blackwell TS (2010) *Pseudomonas aeruginosa*: host defence in lung diseases. *Respirology* 15: 1037–1056. doi: [10.1111/j.1440-1843.2010.01819.x](https://doi.org/10.1111/j.1440-1843.2010.01819.x) PMID: [20723140](https://pubmed.ncbi.nlm.nih.gov/20723140/)
11. Henry M, Lavigne R, Debarbieux L (2013) Predicting in vivo efficacy of therapeutic bacteriophages used to treat pulmonary infections. *Antimicrob Agents Chemother* 57: 5961–5968. doi: [10.1128/AAC.01596-13](https://doi.org/10.1128/AAC.01596-13) PMID: [24041900](https://pubmed.ncbi.nlm.nih.gov/24041900/)
12. Saussereau E, Vachier I, Chiron R, Godbert B, Sermet I, et al. (2014) Effectiveness of bacteriophages in the sputum of cystic fibrosis patients. *Clin Microbiol Infect*.
13. Garbe J, Bunk B, Rohde M, Schobert M (2011) Sequencing and characterization of *Pseudomonas aeruginosa* phage JG004. *BMC Microbiol* 11: 102. doi: [10.1186/1471-2180-11-102](https://doi.org/10.1186/1471-2180-11-102) PMID: [21569567](https://pubmed.ncbi.nlm.nih.gov/21569567/)
14. Essou C, Blouin Y, Loukou G, Cablanmian A, Lathro S, et al. (2013) The susceptibility of *Pseudomonas aeruginosa* strains from cystic fibrosis patients to bacteriophages. *PLoS One* 8: e60575. doi: [10.1371/journal.pone.0060575](https://doi.org/10.1371/journal.pone.0060575) PMID: [23637754](https://pubmed.ncbi.nlm.nih.gov/23637754/)

15. Lu S, Le S, Tan Y, Zhu J, Li M, et al. (2013) Genomic and Proteomic Analyses of the Terminally Redundant Genome of the *Pseudomonas aeruginosa* Phage PaP1: Establishment of Genus PaP1-Like Phages. *PLoS One* 8: e62933. doi: [10.1371/journal.pone.0062933](https://doi.org/10.1371/journal.pone.0062933) PMID: [23675441](https://pubmed.ncbi.nlm.nih.gov/23675441/)
16. Uchiyama J, Rashel M, Takemura I, Kato S, Ujihara T, et al. (2012) Genetic characterization of *Pseudomonas aeruginosa* bacteriophage KPP10. *Arch Virol* 157: 733–738. doi: [10.1007/s00705-011-1210-x](https://doi.org/10.1007/s00705-011-1210-x) PMID: [22218962](https://pubmed.ncbi.nlm.nih.gov/22218962/)
17. Toussaint B, Delic-Attree I, Vignais PM (1993) *Pseudomonas aeruginosa* contains an IHF-like protein that binds to the *algD* promoter. *Biochem Biophys Res Commun* 196: 416–421. PMID: [8216322](https://pubmed.ncbi.nlm.nih.gov/8216322/)
18. Lavigne R, Darius P, Summer EJ, Seto D, Mahadevan P, et al. (2009) Classification of Myoviridae bacteriophages using protein sequence similarity. *BMC Microbiol* 9: 224. doi: [10.1186/1471-2180-9-224](https://doi.org/10.1186/1471-2180-9-224) PMID: [19857251](https://pubmed.ncbi.nlm.nih.gov/19857251/)
19. Ceyssens PJ, Noben JP, Ackermann HW, Verhaegen J, De Vos D, et al. (2009) Survey of *Pseudomonas aeruginosa* and its phages: de novo peptide sequencing as a novel tool to assess the diversity of worldwide collected viruses. *Environmental Microbiology* 11: 1303–1313. doi: [10.1111/j.1462-2920.2008.01862.x](https://doi.org/10.1111/j.1462-2920.2008.01862.x) PMID: [19207572](https://pubmed.ncbi.nlm.nih.gov/19207572/)
20. Gill JJ, Berry JD, Russell WK, Lessor L, Escobar-Garcia DA, et al. (2012) The *Caulobacter crescentus* phage phiCbK: genomics of a canonical phage. *BMC Genomics* 13: 542. doi: [10.1186/1471-2164-13-542](https://doi.org/10.1186/1471-2164-13-542) PMID: [23050599](https://pubmed.ncbi.nlm.nih.gov/23050599/)
21. Jiang X, Jiang H, Li C, Wang S, Mi Z, et al. (2011) Sequence characteristics of T4-like bacteriophage IME08 genome termini revealed by high throughput sequencing. *Virol J* 8: 194. doi: [10.1186/1743-422X-8-194](https://doi.org/10.1186/1743-422X-8-194) PMID: [21524290](https://pubmed.ncbi.nlm.nih.gov/21524290/)
22. He J, Baldini RL, Deziel E, Saucier M, Zhang Q, et al. (2004) The broad host range pathogen *Pseudomonas aeruginosa* strain PA14 carries two pathogenicity islands harboring plant and animal virulence genes. *Proc Natl Acad Sci U S A* 101: 2530–2535. PMID: [14983043](https://pubmed.ncbi.nlm.nih.gov/14983043/)
23. Wolfgang MC, Kulasekara BR, Liang X, Boyd D, Wu K, et al. (2003) Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* 100: 8484–8489. PMID: [12815109](https://pubmed.ncbi.nlm.nih.gov/12815109/)
24. Tümmler B (2006) Clonal Variations in *Pseudomonas aeruginosa*. In: Ramos J-L, Levesque RC, editors. *Pseudomonas: Volume 4: Molecular Biology of Emerging Issues*. Springer.
25. Rocha EPC, Danchin A (2002) Base composition bias might result from competition for metabolic resources. *Trends in Genetics* 18: 291–294. PMID: [12044357](https://pubmed.ncbi.nlm.nih.gov/12044357/)
26. Bailly-Bechet M, Vergassola M, Rocha E (2007) Causes for the intriguing presence of tRNAs in phages. *Genome Res* 17: 1486–1495. PMID: [17785533](https://pubmed.ncbi.nlm.nih.gov/17785533/)
27. Lavigne R, Seto D, Mahadevan P, Ackermann HW, Kropinski AM (2008) Unifying classical and molecular taxonomic classification: analysis of the Podoviridae using BLASTP-based tools. *Res Microbiol* 159: 406–414. doi: [10.1016/j.resmic.2008.03.005](https://doi.org/10.1016/j.resmic.2008.03.005) PMID: [18555669](https://pubmed.ncbi.nlm.nih.gov/18555669/)
28. Santos SB, Kropinski AM, Ceyssens PJ, Ackermann HW, Villegas A, et al. (2011) Genomic and proteomic characterization of the broad-host-range *Salmonella* phage PVP-SE1: creation of a new phage genus. *Journal of Virology* 85: 11265–11273. doi: [10.1128/JVI.01769-10](https://doi.org/10.1128/JVI.01769-10) PMID: [21865376](https://pubmed.ncbi.nlm.nih.gov/21865376/)
29. Truncaite L, Simoliunas E, Zajanckauskaite A, Kaliniene L, Mankeviciute R, et al. (2012) Bacteriophage vB_EcoM_FV3: a new member of "rV5-like viruses". *Arch Virol* 157: 2431–2435. doi: [10.1007/s00705-012-1449-x](https://doi.org/10.1007/s00705-012-1449-x) PMID: [22907825](https://pubmed.ncbi.nlm.nih.gov/22907825/)
30. Kropinski AM, Waddell T, Meng J, Franklin K, Ackermann HW, et al. (2013) The host-range, genomics and proteomics of *Escherichia coli* O157:H7 bacteriophage rV5. *Virol J* 10: 76. doi: [10.1186/1743-422X-10-76](https://doi.org/10.1186/1743-422X-10-76) PMID: [23497209](https://pubmed.ncbi.nlm.nih.gov/23497209/)
31. Juhala RJ, Ford ME, Duda RL, Youtton A, Hatfull GF, et al. (2000) Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdaoid bacteriophages. *J Mol Biol* 299: 27–51. PMID: [10860721](https://pubmed.ncbi.nlm.nih.gov/10860721/)
32. Frampton RA, Taylor C, Holguin Moreno AV, Visnovsky SB, Petty NK, et al. (2014) Identification of bacteriophages for biocontrol of the kiwifruit canker phytopathogen *Pseudomonas syringae* pv. *actinidiae*. *Appl Environ Microbiol* 80: 2216–2228. doi: [10.1128/AEM.00062-14](https://doi.org/10.1128/AEM.00062-14) PMID: [24487530](https://pubmed.ncbi.nlm.nih.gov/24487530/)
33. Zuber S, Ngom-Bru C, Barretto C, Bruttin A, Brussow H, et al. (2007) Genome analysis of phage JS98 defines a fourth major subgroup of T4-like phages in *Escherichia coli*. *J Bacteriol* 189: 8206–8214. PMID: [17693496](https://pubmed.ncbi.nlm.nih.gov/17693496/)
34. Martinsohn JT, Radman M, Petit MA (2008) The lambda red proteins promote efficient recombination between diverged sequences: implications for bacteriophage genome mosaicism. *PLoS Genet* 4: e1000065. doi: [10.1371/journal.pgen.1000065](https://doi.org/10.1371/journal.pgen.1000065) PMID: [18451987](https://pubmed.ncbi.nlm.nih.gov/18451987/)

35. Lopes A, Amarir-Bouhram J, Faure G, Petit MA, Guerois R (2010) Detection of novel recombinases in bacteriophage genomes unveils Rad52, Rad51 and Gp2.5 remote homologs. *Nucleic Acids Res* 38: 3952–3962. doi: [10.1093/nar/gkq096](https://doi.org/10.1093/nar/gkq096) PMID: [20194117](https://pubmed.ncbi.nlm.nih.gov/20194117/)
36. Bobay LM, Touchon M, Rocha EP (2013) Manipulating or superseding host recombination functions: a dilemma that shapes phage evolvability. *PLoS Genet* 9: e1003825. doi: [10.1371/journal.pgen.1003825](https://doi.org/10.1371/journal.pgen.1003825) PMID: [24086157](https://pubmed.ncbi.nlm.nih.gov/24086157/)
37. De Paepe M, Hutinet G, Son O, Amarir-Bouhram J, Schbath S, et al. (2014) Temperate phages acquire DNA from defective prophages by relaxed homologous recombination: the role of Rad52-like recombinases. *PLoS Genet* 10: e1004181. doi: [10.1371/journal.pgen.1004181](https://doi.org/10.1371/journal.pgen.1004181) PMID: [24603854](https://pubmed.ncbi.nlm.nih.gov/24603854/)
38. Eddy SR (2011) Accelerated Profile HMM Searches. *PLoS Comput Biol* 7: e1002195. doi: [10.1371/journal.pcbi.1002195](https://doi.org/10.1371/journal.pcbi.1002195) PMID: [22039361](https://pubmed.ncbi.nlm.nih.gov/22039361/)
39. Rohwer F, Edwards R (2002) The Phage Proteomic Tree: a genome-based taxonomy for phage. *J Bacteriol* 184: 4529–4535. PMID: [12142423](https://pubmed.ncbi.nlm.nih.gov/12142423/)
40. Mahadevan P, King JF, Seto D (2009) CGUG: in silico proteome and genome parsing tool for the determination of "core" and unique genes in the analysis of genomes up to ca. 1.9 Mb. *BMC Res Notes* 2: 168. doi: [10.1186/1756-0500-2-168](https://doi.org/10.1186/1756-0500-2-168) PMID: [19706165](https://pubmed.ncbi.nlm.nih.gov/19706165/)
41. Krupovic M, Bamford DH (2011) Double-stranded DNA viruses: 20 families and only five different architectural principles for virion assembly. *Curr Opin Virol* 1: 118–124. doi: [10.1016/j.coviro.2011.06.001](https://doi.org/10.1016/j.coviro.2011.06.001) PMID: [22440622](https://pubmed.ncbi.nlm.nih.gov/22440622/)
42. Comeau AM, Tremblay D, Moineau S, Rattei T, Kushkina AI, et al. (2012) Phage morphology recapitulates phylogeny: the comparative genomics of a new group of myoviruses. *PLoS One* 7: e40102. doi: [10.1371/journal.pone.0040102](https://doi.org/10.1371/journal.pone.0040102) PMID: [22792219](https://pubmed.ncbi.nlm.nih.gov/22792219/)
43. Ceyssens PJ, Glonti T, Kropinski NM, Lavigne R, Chanishvili N, et al. (2011) Phenotypic and genotypic variations within a single bacteriophage species. *Virol J* 8: 134. doi: [10.1186/1743-422X-8-134](https://doi.org/10.1186/1743-422X-8-134) PMID: [21429206](https://pubmed.ncbi.nlm.nih.gov/21429206/)
44. Verbeken G, Pirnay JP, Lavigne R, Jennes S, De Vos D, et al. (2014) Call for a dedicated European legal framework for bacteriophage therapy. *Arch Immunol Ther Exp (Warsz)* 62: 117–129. doi: [10.1007/s00005-014-0269-y](https://doi.org/10.1007/s00005-014-0269-y) PMID: [24500660](https://pubmed.ncbi.nlm.nih.gov/24500660/)
45. Huys I, Pirnay JP, Lavigne R, Jennes S, De Vos D, et al. (2013) Paving a regulatory pathway for phage therapy. Europe should muster the resources to financially, technically and legally support the introduction of phage therapy. *EMBO Rep* 14: 951–954. doi: [10.1038/embor.2013.163](https://doi.org/10.1038/embor.2013.163) PMID: [24136414](https://pubmed.ncbi.nlm.nih.gov/24136414/)
46. Verbeken G, Huys I, Pirnay JP, Jennes S, Chanishvili N, et al. (2014) Taking bacteriophage therapy seriously: a moral argument. *Biomed Res Int* 2014: 621316. doi: [10.1155/2014/621316](https://doi.org/10.1155/2014/621316) PMID: [24868534](https://pubmed.ncbi.nlm.nih.gov/24868534/)
47. Aziz RK, Devoid S, Disz T, Edwards RA, Henry CS, et al. (2012) SEED servers: high-performance access to the SEED genomes, annotations, and metabolic models. *PLoS One* 7: e48053. doi: [10.1371/journal.pone.0048053](https://doi.org/10.1371/journal.pone.0048053) PMID: [23110173](https://pubmed.ncbi.nlm.nih.gov/23110173/)
48. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, et al. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* 16: 944–945. PMID: [11120685](https://pubmed.ncbi.nlm.nih.gov/11120685/)
49. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25: 955–964. PMID: [9023104](https://pubmed.ncbi.nlm.nih.gov/9023104/)
50. Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33: W244–248. PMID: [15980461](https://pubmed.ncbi.nlm.nih.gov/15980461/)
51. Sullivan MJ, Petty NK, Beatson SA (2011) Easyfig: a genome comparison visualizer. *Bioinformatics* 27: 1009–1010. doi: [10.1093/bioinformatics/btr039](https://doi.org/10.1093/bioinformatics/btr039) PMID: [21278367](https://pubmed.ncbi.nlm.nih.gov/21278367/)
52. Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L, et al. (2013) Using Tablet for visual exploration of second-generation sequencing data. *Briefings in Bioinformatics* 14: 193–202. doi: [10.1093/bib/bbs012](https://doi.org/10.1093/bib/bbs012) PMID: [22445902](https://pubmed.ncbi.nlm.nih.gov/22445902/)
53. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792–1797. PMID: [15034147](https://pubmed.ncbi.nlm.nih.gov/15034147/)
54. Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39: W29–37. doi: [10.1093/nar/gkr367](https://doi.org/10.1093/nar/gkr367) PMID: [21593126](https://pubmed.ncbi.nlm.nih.gov/21593126/)
55. Criscuolo A, Gribaldo S (2010) BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* 10: 210. doi: [10.1186/1471-2148-10-210](https://doi.org/10.1186/1471-2148-10-210) PMID: [20626897](https://pubmed.ncbi.nlm.nih.gov/20626897/)
56. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59: 307–321. doi: [10.1093/sysbio/syq010](https://doi.org/10.1093/sysbio/syq010) PMID: [20525638](https://pubmed.ncbi.nlm.nih.gov/20525638/)

57. Lavigne R, Noben JP, Hertveldt K, Ceyssens PJ, Briers Y, et al. (2006) The structural proteome of *Pseudomonas aeruginosa* bacteriophage phiKMV. *Microbiology* 152: 529–534. PMID: [16436440](#)
58. Moak M, Molineux IJ (2004) Peptidoglycan hydrolytic activities associated with bacteriophage virions. *Mol Microbiol* 51: 1169–1183. PMID: [14763988](#)

ARTICLE 2





**NEXT-GENERATION “-OMICS” APPROACHES REVEAL A
MASSIVE ALTERATION OF HOST RNA METABOLISM DURING
BACTERIOPHAGE INFECTION OF PSEUDOMONAS
AERUGINOSA**


Published in PLoS Genetics

RESEARCH ARTICLE

Next-Generation “-omics” Approaches Reveal a Massive Alteration of Host RNA Metabolism during Bacteriophage Infection of *Pseudomonas aeruginosa*

Anne Chevallereau^{1,2}, Bob G. Blasdel³, Jeroen De Smet³, Marc Monot⁴, Michael Zimmermann⁵, Maria Kogadeeva⁵, Uwe Sauer⁵, Peter Jorth⁶, Marvin Whiteley⁶, Laurent Debarbieux^{1*}, Rob Lavigne^{3*}

1 Institut Pasteur, Molecular Biology of the Gene in Extremophiles Unit, Department of Microbiology, Paris, France, **2** Université Paris Diderot, Sorbonne Paris Cité, Cellule Pasteur, Paris, France, **3** Laboratory of Gene Technology, Department of Biosystems, KU Leuven, Leuven, Belgium, **4** Institut Pasteur, Laboratoire Pathogénèse des bactéries anaérobies, Département de Microbiologie, Paris, France, **5** Institute of Molecular Systems Biology, Eidgenössische Technische Hochschule (ETH) Zürich, Zürich, Switzerland, **6** Department of Molecular Biosciences, Institute for Cellular and Molecular Biology, Center for Infectious Disease, University of Texas, Austin, Texas, United States of America

 These authors contributed equally to this work.

* laurent.debarbieux@pasteur.fr (LD); rob.lavigne@kuleuven.be (RL)



OPEN ACCESS

Citation: Chevallereau A, Blasdel BG, 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(7): e1006134. doi:10.1371/journal.pgen.1006134

Editor: Mireille Ansaldi, Aix-Marseille Université, Centre National de la Recherche Scientifique, FRANCE

Received: March 16, 2016

Accepted: May 31, 2016

Published: July 5, 2016

Copyright: © 2016 Chevallereau et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All RNA-Seq files are available from the NCBI GEO database (accession number GSE76513) and all other relevant data can be found in the Supporting Information files of this paper.

Funding: This research was supported by the Geconcerteerde Onderzoeks Actie grant ‘Phage Biosystems’ from the KU Leuven (<http://www.kuleuven.be/onderzoek/kernprojecten/goa.htm>). BGB has a PhD scholarship within the framework of an

Abstract

As interest in the therapeutic and biotechnological potentials of bacteriophages has grown, so has value in understanding their basic biology. However, detailed knowledge of infection cycles has been limited to a small number of model bacteriophages, mostly infecting *Escherichia coli*. We present here the first analysis coupling data obtained from global next-generation approaches, RNA-Sequencing and metabolomics, to characterize interactions between the virulent bacteriophage PAK_P3 and its host *Pseudomonas aeruginosa*. We detected a dramatic global depletion of bacterial transcripts coupled with their replacement by viral RNAs over the course of infection, eventually leading to drastic changes in pyrimidine metabolism. This process relies on host machinery hijacking as suggested by the strong up-regulation of one bacterial operon involved in RNA processing. Moreover, we found that RNA-based regulation plays a central role in PAK_P3 lifecycle as antisense transcripts are produced mainly during the early stage of infection and viral small non coding RNAs are massively expressed at the end of infection. This work highlights the prominent role of RNA metabolism in the infection strategy of a bacteriophage belonging to a new characterized sub-family of viruses with promising therapeutic potential.

Author Summary

The increase of the proportion of multidrug resistant bacterial strains is alarming and alternative ways to treat infections are necessary such as the use of the natural enemies of

Onderzoeks Toelage grant of the KULeuven. AC was supported by PhD fellowships from the Ministère de l'Enseignement Supérieur et de la Recherche; ED N° 516 B3MI Paris Diderot University (<http://www.enseignementsup-recherche.gouv.fr/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

bacteria, also known as phage therapy. However, explorations of the molecular mechanisms underlying the viral cycle of bacteriophages have been so far restricted to a small number of viruses infecting model bacteria such as *Escherichia coli*. By combining next-generation transcriptomics and metabolomics approaches, we have now demonstrated that the virulent bacteriophage PAK_P3, infecting the opportunistic pathogen *Pseudomonas aeruginosa*, directly interferes with specific host metabolic pathways to complete its infection cycle. In particular, it triggers a dramatic degradation of host RNAs and stimulates bacterial pyrimidine metabolism to promote a nucleotide turnover. Overall, we found that upon PAK_P3 infection, host metabolism is redirected to generate the required building blocks for efficient viral replication. We also showed that PAK_P3 gene expression relies on RNA-based regulation strategies using small non coding RNAs and antisense RNAs. Our findings highlight the molecular strategies employed by this virulent phage, which is a representative of a new subfamily of viruses shown to display promising therapeutic values.

Introduction

The threat of antibiotic resistance has renewed attention to phage therapy leading to isolation of many bacteriophages (phages) targeting human pathogens such as *Pseudomonas aeruginosa* and, consequently, an increasing number of phage genome sequences are available [1]. Comparative genomics has allowed the implementation of a genome-based taxonomy for tailed phages which reflects their great diversity. However, the lack of knowledge of molecular mechanisms underlying their infectious cycles is slowing down their global acceptance as valid therapeutics. Indeed, outside basic characterizations (e.g. phage growth parameters, identification of bacterial receptors and phage structural proteins) many questions about their infection strategy remain conspicuously unanswered for most phages, mainly because genome annotations cannot provide hints on the functions of many viral genes.

For *Pseudomonas* phages, the introduction of whole-transcriptome studies with RNA-Sequencing (RNA-Seq) has recently led to improved genome annotations, discovery of regulatory elements and elucidation of temporal transcriptional schemes, while at the same time looking at the impact on transcription regulation of host genes upon phage infection. For example, giant phage ϕ KZ is now understood to infect and lyse its host cell as well as produce phage progeny in the absence of functional bacterial transcriptional machinery [2]. The impact of phage infection on the host can also be observed at the metabolome level. Recently, a high coverage metabolomics analysis comparing several viruses that cover most genera of *Pseudomonas* phages infecting strain PAO1, revealed specific phage and infection-stage alterations of the host physiology. These changes often appear mediated by phage-encoded auxiliary metabolic genes (AMGs) and by host gene features that are specifically modulated by the phage [3].

One *Pseudomonas* phage clade that has not yet been studied is comprised of the two newly proposed genera (*PAK_P1-like* and *KPP10-like*) belonging to a new subfamily of viruses, *Felixounavirinae*. Interestingly, these phages display the best therapeutic potential in an experimental murine lung infection model as compared to other *P. aeruginosa* phages belonging to distinct clades [4]. Aside from structural genes, most of their predicted ORFs could not be associated with a putative function and consequently, no meaningful conclusions about their strategy for hijacking host metabolism could be drawn [5].

In this work, we used synergistic next generation approaches to provide the first parallel transcriptomics and metabolomics analyses on phage PAK_P3, a representative of the

KPP10-like genus. We intended to draw a detailed global scheme of PAK_P3 infectious cycle by addressing the following questions: Does PAK_P3 control expression of specific bacterial genes? Does it interfere with bacterial metabolism? How does it regulate its gene expression?

Our major finding is the predominant role of RNA metabolism in PAK_P3 infectious strategy. Beside the dramatic global depletion of host transcripts induced by phage infection, PAK_P3 causes a strong up-regulation of a single specific host operon. Consistently, an increase of pyrimidine metabolism upon infection was revealed by metabolomics analysis showing that, like T-even phages, PAK_P3 actively manages nucleotides scavenged from their hosts [6]. In addition, besides revealing the temporal expression of PAK_P3 genes, we highlighted an unexpected prominent role of RNA-based regulation of phage gene expression. Indeed, PAK_P3 produces early antisense transcripts encompassing structural genes as well as phage-encoded small non coding RNAs.

Results

Reannotation of strain PAK genome using transcriptomic data revealed numerous RNA-based regulatory elements

To study bacterial transcriptional response to PAK_P3 infection, it was first crucial to exhaustively characterize the genome of its host, *P. aeruginosa* strain PAK. Initially, a draft genome was produced and assembled (6.28 Mbp, 66.3% GC content and 6,267 predicted ORFs). Next, a detailed genome reannotation was performed based on RNA-Seq data generated from exponentially growing and uninfected PAK cells (S1 Table).

Using COV2HTML [7] to visualize transcripts, we manually reannotated 32 open reading frames (ORFs) (by detection of an alternative start codon) and defined 63 new putative coding sequences. Among them, 39 have been previously annotated in other *P. aeruginosa* genomes while the other 24 are new hypothetical coding sequences that display no homology to sequences in databases and may be considered as strain-specific (S1 Table).

Recent genome-wide studies based on RNA-Seq led to the discovery of a substantial number of non-coding RNAs (ncRNAs), which are now acknowledged as important modulators of various bacterial processes (for a review of ncRNAs in *Pseudomonas aeruginosa*, see [8]). We identified a total of 75 small ncRNAs encoded in the PAK genome, 26 of which correspond to known functional classes (S1 Fig, S1 Table). Among these 26, 12 are similar to uncharacterized ncRNA conserved within the *Pseudomonas* genus and the other 14 have predicted functional assignments, according to Rfam. The majority of ncRNAs (49 out of 75) could not be assigned to any functional class (see Methods), and have not been identified in previous RNA-Seq investigations carried out on *P. aeruginosa* strains PAO1 and PA14 [9–11], suggesting that they may represent novel ncRNAs regulators.

Eighteen long antisense RNAs (asRNAs) were also identified within genes. As they do not display any consistent ORFs, they are not likely to contain overlapping protein-coding genes and may therefore *cis*-interfere with the expression of gene they are encoded in (S1 Fig, S1 Table). Finally, 32 potential riboswitches were identified by looking at intergenic transcription events starting at a significant distance from a downstream gene, usually involved in a metabolic pathway and displaying a characteristic RNA-Seq pattern. Eleven of them were confirmed by Rfam search (S1 Fig, S1 Table).

With more than 50% of new ncRNAs amongst total ncRNAs identified, along with the identification of new putative riboswitches and evidence of antisense transcripts, strain PAK exemplifies the great diversity of bacterial RNA-based regulation [12]. Such in-depth annotation, including new strain-specific RNA elements, was mandatory for the subsequent transcriptomic

analysis of phage infected cells in order to assess the impact of phage infection on host physiology.

The fast-replicating phage PAK_P3 progressively takes over host cell transcription and tends to dominate over competing mobile genetics elements

To study the dynamics of the transcriptional and metabolic consequences of phage infection, we first selected the most relevant time points, representative of the different steps of the course of infection by determining the growth parameters of PAK_P3. Adsorption assays revealed that $\geq 90\%$ of PAK_P3 virions adsorbed on strain PAK within 4.6 ± 0.7 min ($k_a = 2.2 \cdot 10^{-9} \pm 5.1 \cdot 10^{-10}$ mL.min⁻¹) (Fig 1A). A standard one step growth experiment showed that the first functional new virions are rapidly assembled (eclipse period: 12.3 ± 0.4 min) and almost immediately released (latency period: 13 ± 2.1 min), producing an average of 53 ± 21 progeny phages per infected cell (Fig 1B). With a mean infection cycle duration as short as 18 ± 0.6 min, PAK_P3, with a genome length ≥ 80 kb, is faster than the myoviruses ϕ KZ (60–65 min, 280 kb) [3] and T4 (25–30 min, 168 kb) [13], therefore being among the most rapid *Myoviridae*.

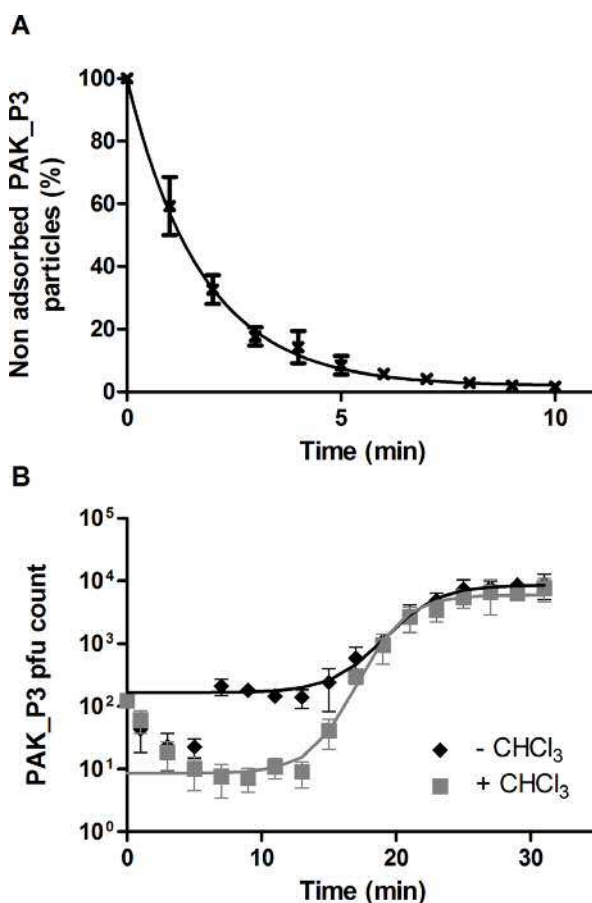


Fig 1. PAK_P3 rapidly adsorbs to its host and efficiently produces new progenies. (A) Adsorption assays of PAK_P3 on *P. aeruginosa* strain PAK. (B) One-step growth curve of PAK_P3. Samples treated with (grey squares) or without (black diamonds) CHCl₃. A logistic regression was used to fit the data. Four independent experiments were combined and data are presented as means with standard deviations.

doi:10.1371/journal.pgen.1006134.g001

Given the short eclipse period duration, we focused on 3.5 min and 13 min time points as representative snapshots of the beginning (early) and the end (late) of one infection cycle at the transcription level. Investigation of the regulation of both viral and host gene expression over a single phage infection cycle by RNA-Seq revealed a progressive and dramatic replacement of host mRNA with phage transcripts. This process eventually results in host transcripts representing fewer than 13% of non-ribosomal RNAs in the cell (Fig 2). However, even in the context of this dramatic depletion of host transcripts, a response to phage infection at the transcription level was observed, suggesting a globally accelerated degradation of unstable mRNA species rather than a global transcriptional repression as described for phage T4 [14].

In addition to providing a transcriptional environment fully co-opted by the phage for optimal infection (*i.e.* making host RNA polymerase available for viral RNAs for instance), this observed host RNA depletion can be expected to suppress host defenses that require host transcripts to function [15] as well as prophage induction attempts. Indeed, PAK_P3 infection appears to activate the transcription of a P2-like prophage (Fig 3, S2 Table) as corresponding transcripts display a 6.8-fold increase in PAK_P3 infected cells at late time point compared to non-infected cells. However, host transcripts overall were depleted by 7.2 fold at the late time point, which would leave the infected cell with marginally fewer prophage transcripts than

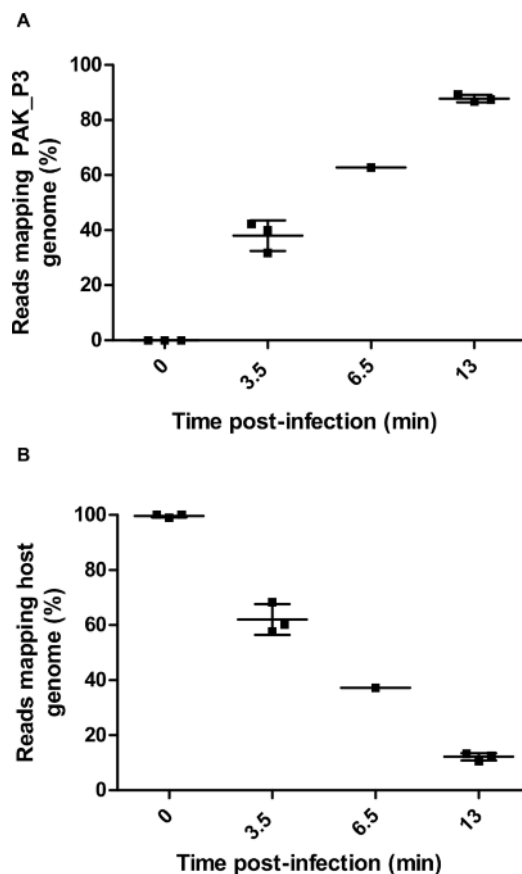


Fig 2. PAK_P3 takes over the host cell transcription over the course of infection. Three independent biological replicates of RNA extracts were harvested from PAK_P3 infected cells at 0min, 3.5min and 13min post infection, as well as a single sample collected at 6.5min and all were subsequently sequenced. Plotting the percentage of reads mapping to the PAK_P3 genome (A) and to the host genome (B) over the course of infection shows that PAK_P3 progressively dominates the transcriptional environment of the cell with phage transcripts.

doi:10.1371/journal.pgen.1006134.g002

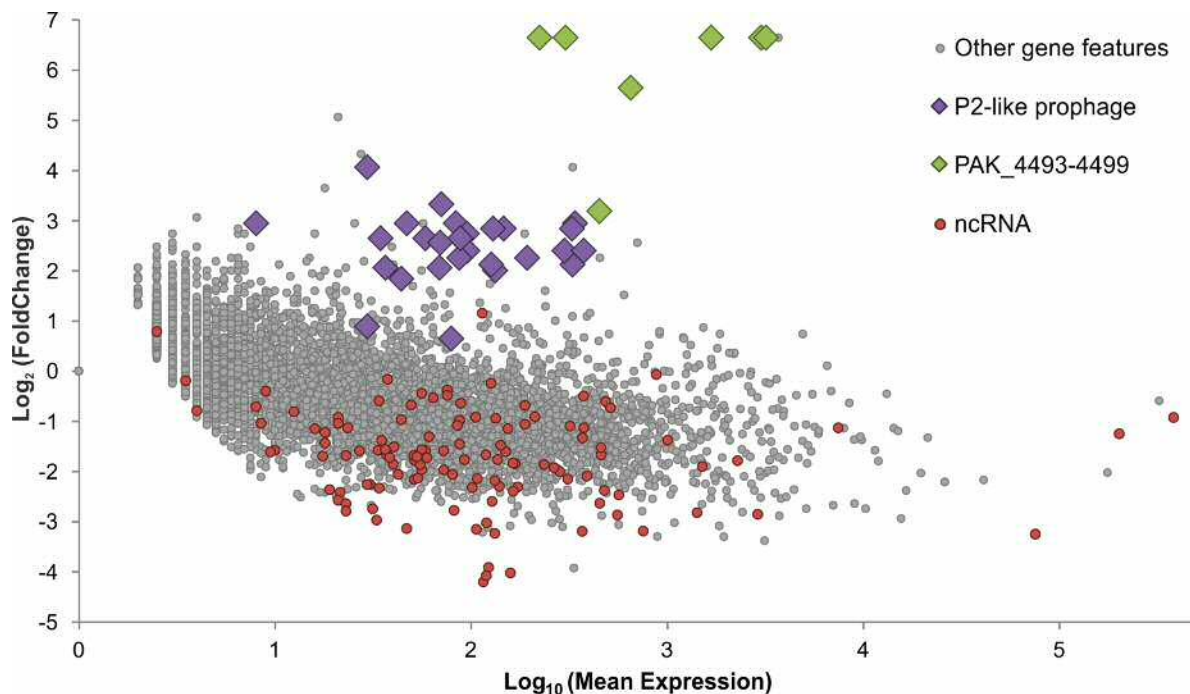


Fig 3. PAK_P3 alters expression of many host gene features by late infection. Differential expression analysis of host gene features comparing transcript abundance between phage negative controls ($t = 0$ min) and late infection ($t = 13$ min) was performed. This comparison was made after normalizing the read counts that map to each host gene feature between both conditions, ignoring reads that map to the phage, which artificially enriches reads in the late condition. This method thus compares the negative control directly to a host transcript population that has been depleted by replacement with phage transcripts during infection, normalizing away the global depletion so that more specific shifts can be reported and independently tested for.

doi:10.1371/journal.pgen.1006134.g003

during exponential growth, indicating that the transcriptional activation of the prophage is suppressed, although not completely blocked.

Phage infection triggers dramatic and specific up-regulation of one bacterial operon linked to RNA processing

Although host transcripts are globally replaced by phage transcripts, we could still analyze the changes in host mRNA population by normalizing the host transcript counts before infection to the counts after infection, artificially depleting counts before infection and enriching reads after infection. This allows us to look for specific differential expression of host gene features in response to the stress of phage infection as well as specific changes in host gene expression imposed by the phage in order to hijack cellular metabolism.

We discovered that one operon, comprising six genes (PAK_4493–4499), has a nearly 80-fold increase in abundance relative to other host genes, which is large enough to strongly enrich its transcript abundance relative to the total RNA in the cell even in the context of global RNA degradation (Fig 3, S2 Table).

RNA-Seq analysis thus provided precise depictions of phage influence on the bacterial transcriptome and host transcriptional response to infection. It also allowed us to decipher the transcriptional strategy adopted by the phage to control its own gene expression (see below). To have a broader view of the consequences of a phage infection on host cell physiology, we performed a complementary metabolomics analysis.

Amino acid, nucleotide/sugar and pyrimidine pathways display drastic changes in phage infected cells

Viruses depend on host cell metabolic resources to complete their intracellular parasitic development [16]. However, the effects of phage infection on host metabolism are still poorly understood. We thus investigated whether the phage completely shuts off host metabolism, as it may burden efficient phage replication, or if it influences specific pathways.

To assess the impact of PAK_P3 infection on strain PAK metabolism, high-coverage metabolomics analysis was applied to monitor metabolite dynamics during infection [17]. Comparison of the metabolite levels at different time points post infection to uninfected samples revealed significant metabolic changes upon phage infection. Within the first 5 min of infection, 22% of measured metabolites display altered levels with 13.8% increased and 8.5% decreased ($p\text{-value} \leq 0,05$, $|\text{Log}_2(\text{fold change})| \geq 0,5$). The proportion of metabolites with increased levels gradually rises up to 22% at 25 min post infection, while the proportion of metabolites with decreased levels temporarily drops to 3% to finally increase back to 13% during bacterial lysis (Fig 4). These variations indicate that PAK_P3 does not simply deplete available host metabolites but relies on an active metabolism in agreement with recent observations identifying phage-specific physiological alterations [3,18].

Next, to investigate whether PAK_P3 targets specific metabolic pathways, a metabolite set enrichment analysis was performed. Overall, metabolites from amino/nucleotide sugar and pyrimidine metabolic pathways were found over-represented among increasing metabolites, while amino acid-related pathways were enriched among decreasing metabolites at later stages of infection (Fig 5). Intriguingly, about 50% of the detected (deoxy)nucleotides-phosphates have at least two-fold increased levels during infection (S3 Table).

Among accumulating metabolites belonging to amino/nucleotide sugar metabolism and to lipopolysaccharide biosynthesis pathway (Fig 5), it is worth noting that the levels of cell wall precursors such as UDP-N-acetyl-D-glucosamine or UDP-N-acetyl-D-galactosaminuronic acid show a significant two- and three-fold increase, respectively, during late infection (S4 Table). This increase is not accompanied by altered expression of host genes involved in this pathway (see below).

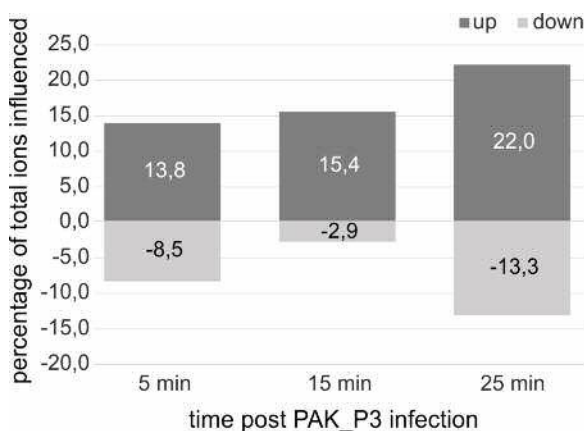


Fig 4. PAK_P3 alters *P. aeruginosa* metabolite content over the course of infection. Percentage of altered *Pseudomonas* metabolite ions during the course of infection ($p\text{-value} \leq 0,05$, $|\text{Log}_2(\text{fold change})| \geq 0,5$), y-axis shows percentage and x-axis shows the time points during infection. In total 377 ions were measured.

doi:10.1371/journal.pgen.1006134.g004

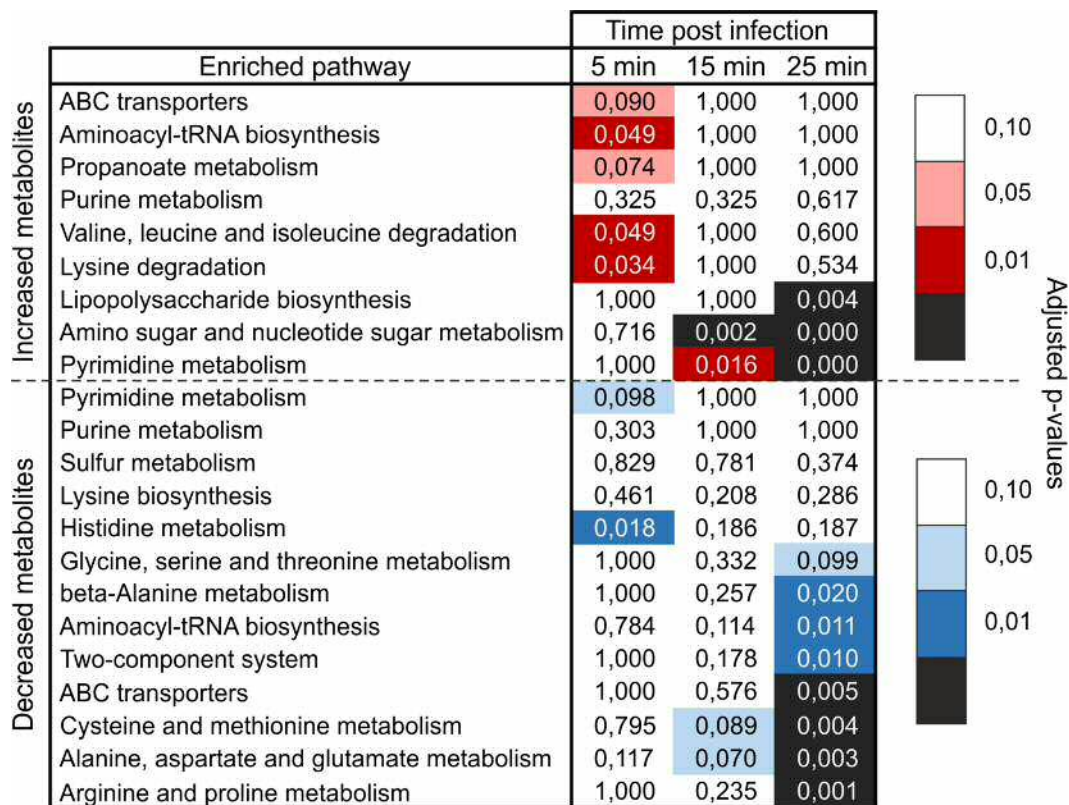


Fig 5. Pathway enrichment analysis during PAK_P3 infection revealed its requirement on pyrimidine metabolism. Values of the enrichment analysis are shown for the significantly enriched pathways (rows) at the different time points (columns). The color scale indicates the cut-off for the p-values, where red and blue are used for, respectively, the pathways found enriched among increased and decreased metabolites.

doi:10.1371/journal.pgen.1006134.g005

Most enriched pathways among decreasing metabolites involve amino acid biosynthesis (Fig 5), more specifically Arg, Pro, Ala, Asn, Glu, Cys and Met metabolism were found significantly enriched ($p\text{-value} < 0.005$). These observed decreases may indicate drainage of amino acid pools in the cell during phage particle formation, due to an imbalance between cellular amino acid biosynthesis and consumption by the phage.

Phage-induced changes of host metabolome composition are not otherwise mediated through differential expression of host genes

We initially hypothesized that the observed changes in metabolome composition upon infection would largely be the result of a differential expression of host genes induced by the phage. This would indicate that PAK_P3 mainly interferes with cellular transcription to alter host physiological processes. To address this question, we investigated if the variations at the metabolome level could be directly linked to transcriptional changes. We thus analyzed all metabolites belonging to pathways highlighted by the pathway enrichment analysis (see above) that display significant variations ($|\text{Log}_2(\text{fold change})| > 0.5$, $p\text{-value} < 0.05$) as well as differential expression of coding sequences related to the corresponding pathways with a stringent cut-off point ($|\text{Log}_2(\text{fold change})| > 1.3$, $p\text{-value} < 0.05$) (Fig 6).

Only few genes linked to these pathways were significantly differentially expressed upon late infection, indicating that the phage influence on host metabolism is not primarily mediated through differential gene expression. In fact, several pathways with increased metabolite levels

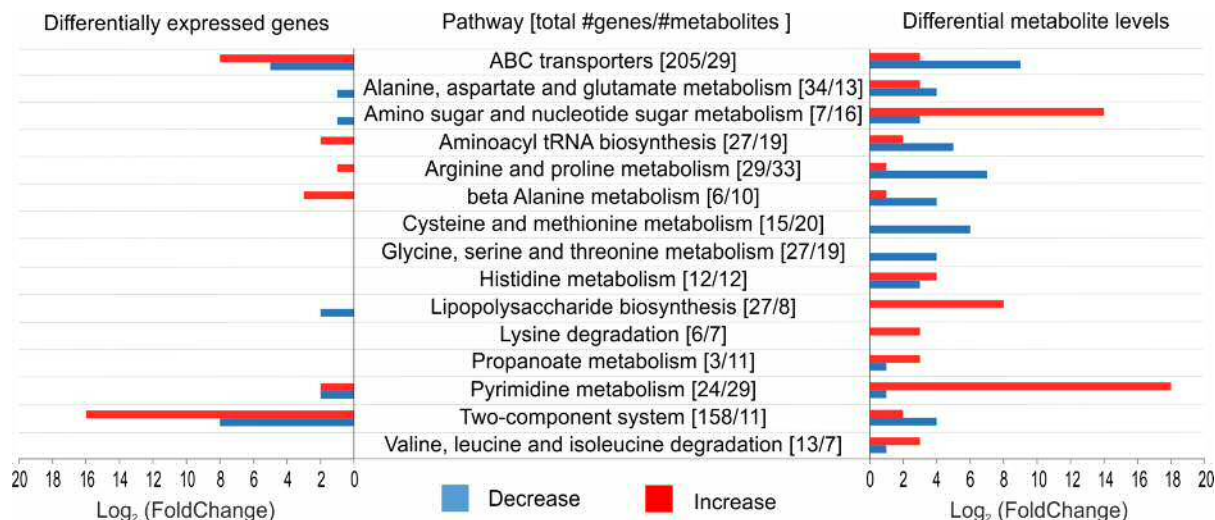


Fig 6. Comparison of significant changes in transcriptomics and metabolomics data from PAK_P3 infected cells reveals no direct correlation. On the left, the number of genes with a significant differential expression ($|\text{Log}_2(\text{fold change})| > 1.3$, $p\text{-value} < 0.05$) is shown. On the right, an overview of the number of metabolites with significantly changed levels ($|\text{Log}_2(\text{fold change})| > 0.5$, $p\text{-value} < 0.05$) is shown. The middle column entails all studied metabolic pathways and indicated between brackets are, respectively, the total number of genes and metabolites involved in this pathway. Red = increase in transcript/metabolite level, Blue = a decrease in transcript/metabolite level.

doi:10.1371/journal.pgen.1006134.g006

have a decreased transcription of the involved genes or vice versa (e.g. lipopolysaccharide biosynthesis).

Based on these complementary “-omics” approaches, it can be concluded that PAK_P3 does not otherwise redirect host physiology towards viral reproduction through modification of host gene expression. The general degradation of host RNA observed likely ensures sufficient building blocks for viral genome replication. The metabolic content of PAK_P3 infected cells shows both increased and decreased metabolite levels. We hypothesize these changes are either the direct consequence of an increased viral consumption of metabolites (e.g. amino acid metabolism) or are likely triggered by phage-encoded AMGs (e.g. pyrimidine metabolism).

Gene expression of PAK_P3 is temporally regulated

Besides redirecting host cell physiology, the phage must also control its own gene expression. Here we intended to investigate the transcriptional strategy of PAK_P3 and also discovered unexpected regulatory mechanisms the phage uses to complete its infection cycle.

During the course of infection, early, middle and late transcripts of PAK_P3 genome were identified (Fig 7, S5 Table). The early transcribed region encompasses genes gp74 through gp112, all of which encode hypothetical proteins with low or no sequence similarity to gene products from other bacteriophages (so-called ‘ORFans’). Transcripts produced at middle time point focus on two regions that each contains gene features related to nucleic acid metabolism. As expected, the structural region appears to be mostly transcribed in late infection. Strikingly, five ORFs (i.e. gp34, gp37, gp38, gp45 and gp46), although located in the structural region, are overexpressed early compared to late time point. Finally, all predicted genes are transcribed, except for gp113, which corresponds to the predicted genome terminus. Intergenic transcription is observed throughout the genome, highlighting the great compaction of viral genomes where every single gene is expressed, in contrast to bacterial genomes. This property is further illustrated by the large amount of antisense transcripts detected, as reported below.

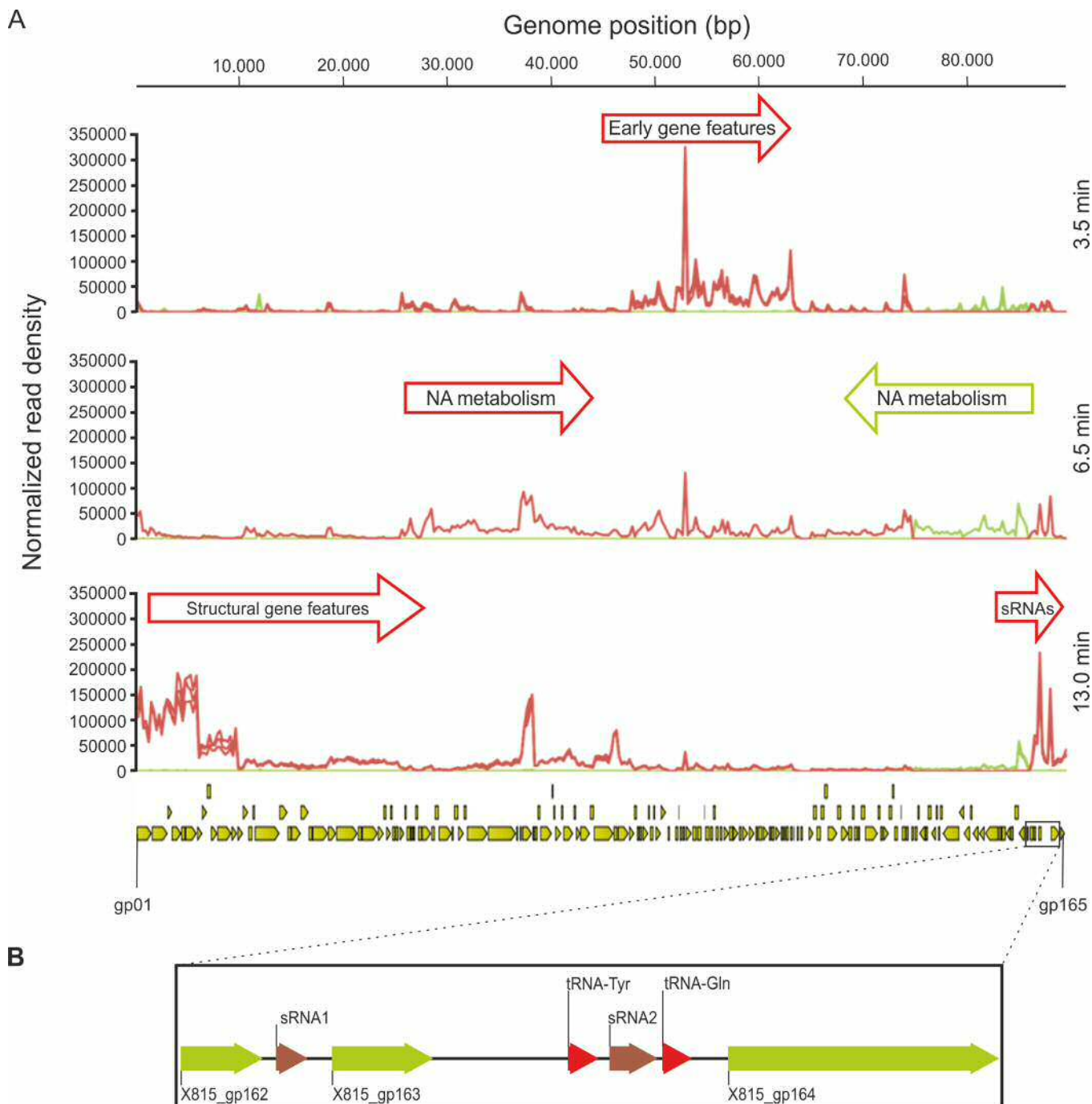


Fig 7. PAK_P3 transcription is temporally regulated. (A) Mapped reads were summarized into stranded count tables of Total Gene Reads that align to every 250bp of the PAK_P3 genome using the CLC Genomics Workbench. These read counts were then normalized against each other by the Total Count of reads that align to both phage and host genomes for each sample and then plotted. This allows us to show the relative abundance of phage transcripts over time in the context of the total transcript population. Red and green graphs represent reads mapping to the forward and reverse strands, respectively, for each replicate in each condition. The PAK_P3 genome is represented at the bottom of the panel with yellow arrows indicating defined coding sequences. (B) Enlargement of the PAK_P3 genome region encoding small RNAs. NA = nucleic acid

doi:10.1371/journal.pgen.1006134.g007

PAK_P3 expresses antisense RNA elements targeting its structural region during the early stage of infection

Analysis of antisense transcripts of PAK_P3 revealed 20 putative asRNAs, 8 of which are small asRNA (mean length 176 ± 30 bp) and 12 are longer than 300 bp ([S6 Table](#)). All but one are encoded within genes, suggesting they may act as *cis*-encoded antisense RNAs. These asRNA are predominantly (15 out of 20) located in the structural region of PAK_P3 genome and are significantly more strongly transcribed during early infection compared to late infection with fold changes ranging between 2 and 17 ([S2](#) and [S3](#) Figs, [S5 Table](#)). These data support the hypothesis that such antisense transcription is used to shut down expression of late structural genes during the early stage of infection.

PAK_P3 displays strong expression of temporally regulated small, non-coding RNAs

Following the observation of abundant antisense transcripts, we looked for other unusual transcriptional profiles within PAK_P3 transcriptome and detected two abundant small (~100bp) transcripts during late infection. These two transcripts, hereafter referred as sRNA1 and sRNA2, were found in two neighboring intergenic regions: sRNA1 is encoded within a 200bp-intergenic region between two genes encoding hypothetical proteins, whereas sRNA2 is part of a larger intergenic region between two phage-encoded tRNAs ([Fig 7](#)). They are temporally regulated since they display a 91- and 12-fold change, respectively, in their 'late versus early' expression. Strikingly, these two small RNAs belong to the most strongly transcribed regions of the phage genome during late infection as they respectively represent the 18th and 24th most expressed gene features over 86 late genes ([S5 Table](#)).

We hypothesized that they could be *trans*-encoded small RNAs, acting by base-pairing on a target mRNA. As such, we looked for potential target regions in both phage and host genomes. The potential targets found on the host genome were not differentially expressed 13 min post infection, indicating that these two phage small RNAs would not act through mRNA degradation but rather have a role in translational silencing, if any.

Interestingly, a stretch of 11 nucleotides on sRNA2 was found to be repeated eight times on the host genome and systematically located within tRNAs, more particularly within the T ψ C-loop. We propose that it could be involved in translational repression by binding, and eventually blocking, bacterial ribosomes. As this 11bp-stretch is also conserved in closely related phages (*PAK_P1-like* genus), it may represent a starting point leading to the discovery of new phage non-coding RNAs.

On the phage genome, the only potential targets (11 consecutive nucleotides matching perfectly) are located in the early ORF product gp78 for sRNA1 and in the late gene encoding the putative ribonucleotide-diphosphate reductase gp67 for sRNA2.

To date, only few phage-encoded small RNAs have been described in the literature and most of them derive from prophages [[19,20](#)]. The only examples of phage sRNAs encoded by a virulent phage, T4 band C and band D RNAs, were described in the 1970's and their functions have remained unknown ever since [[21](#)].

Discussion

Next generation transcriptomics, metabolomics, and classical microbiological techniques have here been integrated for the first time to describe virus/host interactions between the candidate therapeutic bacteriophage PAK_P3 and its host, *P. aeruginosa* strain PAK. By capturing early, middle and late infection time points, we delineated genomic regions of temporally distinct

phage expression. This particularly highlights early gene features, which are typically involved in the shutdown of host metabolism. Like the approximately 50 so called ‘monkey-wrench’ proteins found in phage T4, small early proteins likely have functions reliant on protein-protein interactions to disrupt host systems and could potentially be exploited to aid in small molecule antibiotic design [22–24].

It is well established for model bacteriophages, including T7 and T4, that the temporal regulation of middle and late gene expression is typically the result of a tight regulation driven by phage early proteins through various mechanisms such as redirection of host RNA polymerase to phage middle and late promoters (like phage T4 proteins AsiA-MotA or phage-encoded sigma factor gp28 in SPO1) [16,25]. The early expression of antisense RNAs could represent an additional regulation mechanism preventing transcriptional leaks from strong promoters controlling expression of late structural genes. Consistent with this hypothesis, the temporal distribution and the location of the numerous PAK_P3 asRNAs correlate with the shut-off of structural gene expression observed 3.5 min post infection. Although *cis*-antisense RNAs appear to be a common form of regulation in bacterial genomes, they have not been extensively described in phage genomes. Beside the regulatory *oop* RNA reported over 40 years ago (reviewed in [26]), no other asRNAs were reported until recently [27] and exclusively in lambdoid phages. Moreover, such asRNAs have never been reported for virulent phages until 2014 [28]. Therefore, the high number asRNAs reported for PAK_P3 implies that antisense transcription may be a regulatory mechanism used by phages more frequently than previously thought.

From a phage-host interaction point of view, we found that existing host transcripts are rapidly overwhelmed with viral transcription. This may reflect a globally accelerated degradation of RNA in the cell in a way similar to phage T4. Indeed, it has been previously reported that T4 globally alters the stability of existing mRNAs, in addition to repressing the transcription of cytosine containing DNA [29,30]. This hypothesis is further supported by the drastic overexpression of one host operon (PAK_4493–4499) encoding RNA processing-related proteins. In particular, this operon encodes a RNA 3'-phosphate cyclase RtcA (PAK_4496) that has been described as being involved in the processing of RNA transcripts such as priming RNA strands for adenylation to protect them from exonucleases or to mark them for further processing so they serve as substrates for downstream reactions performed by additional enzymes [31,32]. Therefore, we hypothesize that this operon may be uniquely upregulated by the phage in order to participate in the global degradation of RNAs during infection, which we observe in both the RNA-Seq and metabolomics data, by tagging transcripts for degradation by phage encoded enzymes. An alternative hypothesis to explain this dramatic up-regulation of PAK_4493–4499 relies on RtcB (PAK_4494), a predicted RNA ligase. Together, the RtcAB system has been shown to play a role in tRNA repair after stress-induced RNA damage (e.g. viral infection) in *E. coli* [33]. Also, it has been shown that phage T4 RNA 3'-phosphate cyclase (encoded by *pseT*) and RNA ligase (*rli*) are involved in overcoming resistance [34] by restrictive strains of *E. coli* producing phage induced tRNA anticodon nuclease (encoded by the *prf* locus) which causes abortive infection by preventing effective translation of phage genes [35]. Therefore, it is possible that PAK_P3 upregulates this operon to activate a host repair function to interfere with a yet uncharacterized host restriction system or a *prf*-like locus. However, deleting the RNA ligase *rtcB* gene (PAK_4494), appears to have no toxic effect on the host as well as no effect on the efficiency of plating (S1 Text).

The observation of a phage induced host RNA degradation is further supported by our metabolomics data. Indeed, the increased pyrimidine metabolism confirms that nucleotide turnover is a central viral need to achieve a successful infection cycle. Overall, we showed that upon PAK_P3 infection, the host metabolism is not shutdown but redirected to generate the

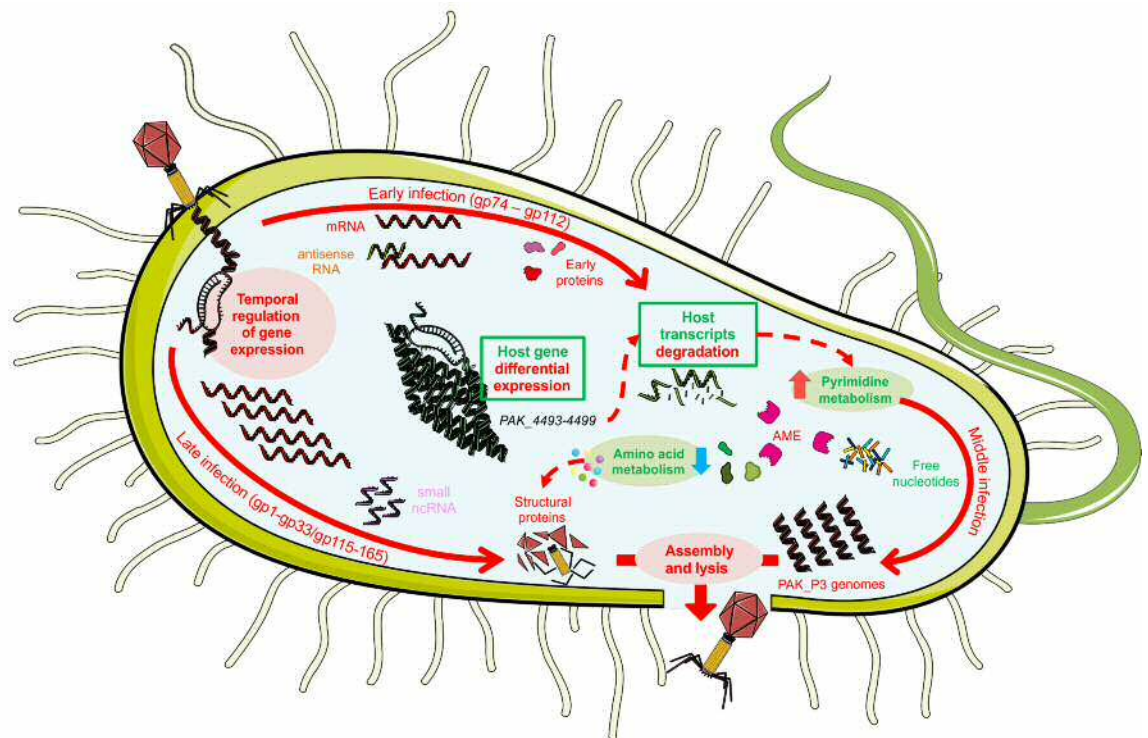


Fig 8. Molecular details of PAK_P3 infection cycle of *P. aeruginosa* strain PAK. Red and green colors correspond to phage and host elements respectively. A description of the figure is given in the text (see [Discussion](#)). Full arrows depict the three temporal stages of PAK_P3 infection cycle. Dashed arrows represent relations between metabolic pathways highlighted by our "-omics" analyses. AME = Auxiliary Metabolic Enzymes.

doi:10.1371/journal.pgen.1006134.g008

required building blocks for viral replication and this redirection is not the result of a phage induced differential host gene expression, aside from the RNA processing operon. An explanation for this metabolic turnover relies on phage-encoded AMG and phage early proteins. We propose that phage early proteins would interfere with host metabolic processes through interactions with bacterial proteins. Once the host machinery is disrupted, phage metabolic enzymes would take over and catalyze the reactions yielding the specific metabolites required for viral replication ([Fig 8](#)). For instance, we hypothesize that the observed global degradation of host mRNA eventually produces an excess of free ribonucleotides that are likely converted into deoxynucleotides by ribonucleotidases. Interestingly, PAK_P3 encodes a putative ribonucleotide-diphosphate reductase (alpha and beta subunits, respectively gp67 and gp69) that could catalyze such a reaction. An alternative explanation for the observed increase of (deoxy) nucleotides-phosphates relies on the putative deoxyribonuclease (gp57) encoded by PAK_P3, which could be responsible for host genome degradation during middle and late infection stages, as observed for phage LUZ19 [[36](#)]. Supporting these hypotheses, these three phage-encoded AMG are strongly expressed by PAK_P3 during late infection stage as they are respectively the 7th, 22nd and 19th most expressed genes over 86 late genes ([S5 Table](#)). It is noteworthy that a fourth predicted phage-encoded AMG, a CMP deaminase (gp155), is also involved in nucleotide metabolism and also expressed during late infection although less intensely than the other AMG mentioned. Altogether, these predicted AMG involved in nucleotide metabolism highlight the central need for nucleotides during PAK_P3 infection, in accordance with the short infection cycle span during which about 50 genomes of 88 kb have to be synthesized. Indeed, the advantage found in precisely manipulating nucleotide

depolymerization and pathways to shut down host mechanisms, provide material for phage DNA synthesis, and prevent osmotic stress appears to be significant across phage clades. For example, *Pseudomonas* phage Lu11 contains ORFs predicted to be involved in nucleotide metabolism [37], *E. coli* phage T5 degrades host DNA before exporting it outside of the cell [38], and T4 even encodes for its own, nearly complete, parallel DNA precursor biosynthesis pathway [6].

Another example of phage-driven interference with host metabolic pathway is given by LPS biosynthesis pathway. The observed accumulation of cell wall precursors, which is not correlated with an altered expression of corresponding host genes, may be a direct consequence of peptidoglycan degradation and the subsequent release of its precursors triggered by the infection. Consistent with this hypothesis, PAK_P3 has a potential AMG (gp151) similar to a bacterial cell wall hydrolase, which could explain such cell-wall degradation.

Further investigations are now required to fully associate transcriptomics and metabolomics data to viral gene functions, a process which is currently hampered by the lack of versatile genetic tools to construct mutants of virulent bacteriophage. Such effort to deeply characterize one particular phage genus (i.e. *KPP10-like*) is also motivated by the great therapeutic potential of these phages as demonstrated in animal models and recently strengthened by the identification of such a phage in the 'Intesti phage' cocktail, a key commercial product of the Eliava Institute in Georgia [4,39–41]. Overall, the knowledge of phage biology provided by next-generation "-omics" approaches not only enlighten viral mechanisms of infection but can also open an array of biotechnological applications based on regulatory elements and proteins found in this new sub-family of phages.

Materials and Methods

Strains and growth conditions

P. aeruginosa strain PAK [42] was cultured in LB medium supplemented with 10mM MgCl₂ at 37°C unless stated otherwise. For RNA-Seq experiments, cells were infected with bacteriophage PAK_P3 using a multiplicity of infection of 25 in order to ensure the synchronicity of the infection (95% of the bacterial population killed after 5 min phage-bacteria incubation).

Adsorption assay and one-step growth experiment

Bacteriophage growth parameters were assessed as described previously [43]. Briefly, a culture of strain PAK was infected at low MOI (0.1) and incubated 5 min at 37°C with agitation allowing bacteriophage particles to adsorb. Following a 1000-fold dilution, two 100 µL samples were collected every 2 min and either kept on ice until titration, or mixed with CHCl₃. For each time point we thus determined the free bacteriophage count (samples with CHCl₃) as well as the number of free bacteriophages and infective centres (samples without CHCl₃) to calculate eclipse and latency periods respectively.

Experimental data were fitted with a logistical function:

$$f(x) = \frac{a}{1 + e^{-k(x-x_c)}} \quad (1)$$

a : ordinate corresponding to the asymptote when $x \rightarrow +\infty$, represents the maximal pfu count.

x_c : abscissa of the inflection point, represents the mean duration of the infectious cycle.

k : slope of the tangent line to the exponential part of the curve.

Eclipse and latency periods were determined as the x value corresponding to $f(x) > 0.05a$

The burst size was determined as:

$$\frac{\text{Phage_titer}_{(t=0, -\text{CHCl}_3)} - \text{Phage_titer}_{(t=0, +\text{CHCl}_3)}}{\text{Phage_titer}_{(t=30, \pm\text{CHCl}_3)}} \quad (2)$$

$\text{Phage_titer}_{(t=0, +\text{CHCl}_3)}$ and $\text{Phage_titer}_{(t=0, -\text{CHCl}_3)}$: values of initial phage titers ($t = 0$ min) measured in samples treated or not with chloroform, respectively. The numerator represents the number of intracellular phages.

$\text{Phage_titer}_{(t=30, \pm\text{CHCl}_3)}$: Mean of phage titers measured in samples treated and not treated with chloroform at $t = 30$ min.

Four independent adsorption assays were performed in the conditions described above with a lower MOI (10^{-3}) and omitting the dilution step. Data could be approximated using an exponential function and adsorption time was defined as the time required to reach a threshold of 10% non-adsorbed bacteriophage particles.

P. aeruginosa strain PAK genome sequencing

Genomic DNA was isolated from *P. aeruginosa* strain PAK and pyrosequencing was performed on a Roche 454 FLX system with Titanium chemistry at the University of Texas Genomic Sequencing and Analysis Facility. The draft assembly of ~6.3 Mbp consists of 9 scaffolds, 490 large contigs, and 616 total contigs and was annotated at the University of Maryland Institute for Genomic Sciences using the IGS Prokaryotic Annotation Pipeline[44]. Scaffolds deposited in GenBank can be accessed via Bioproject accession no. PRJNA232360. More details are available in [S1 Text](#).

Whole transcriptome sequencing

RNA-Seq analysis was performed on an exponentially growing culture that was synchronously infected with PAK_P3. Three independent biological replicates were harvested at 0min, 3.5min and 13min to represent, respectively, a phage negative control and early and late transcription while one additional sample was collected at 6.5 minutes to assess the presence of an identifiable middle phase of transcription.

The preparation of cDNA libraries was performed as described in Blasdel *et al.* (*in press*) [45]. Briefly, samples were collected at three time points, representing early, middle, and late infection, from a synchronously infected culture, with <5% of bacteria remaining uninfected after 3.5 minutes, and halted by rapid cooling in 1/10 volume of 'stop solution' (10% phenol, 90% ethanol). Cells were then lysed in TRIzol, total RNA was purified through a standard organic extraction and ethanol precipitation, and remaining genomic DNA was removed using TURBO DNase. DNA removal was confirmed with PCR before rRNA was depleted using the Ribo-Zero rRNA Removal Kit (Gram-Negative Bacteria). This rRNA depleted total RNA was then processed into cDNA libraries using Illumina's TruSeq Stranded Total RNA Sample Prep Kit according to manufacturer's instructions and sequenced using an Illumina NextSeq 500 desktop sequencer on the High 75 cycle. More than 11 million 75bp reads mapping to non-ribosomal regions were obtained from each library with the exception of one early sample and one late sample providing 1,221,867 and 941,631 mapped reads respectively due to incomplete rRNA removal. After trimming, sequencing reads were aligned separately to both the phage and host genomes with the CLC Genomics workbench v7.5.1. These alignments were then summarized into count tables of Unique Gene Reads that map to phage or host gene features respectively.

RNA-Seq data have been deposited in NCBI-GEO with accession no. GSE76513.

RNA-Seq coverage visualization is available through the COV2HTML software at [https://mmonot.eu/COV2HTML/visualisation.php?str_id=-32] for a comparison of the host (0 min / 13 min) and [https://mmonot.eu/COV2HTML/visualisation.php?str_id=-34] for a comparison of the phage (3.5 min/ 13 min)

Strain PAK reannotation and prediction of ncRNAs with RNA-Seq data

RNA-Seq data of uninfected strain PAK were visualized using COV2HTML [7]. Reads mapping forward and reverse strands were manually scanned over the whole genome. Both coding regions and intergenic regions displaying an unexpected transcription profile were examined using CLC Genomics Workbench 7.5.1 and Blastp (default parameters) to annotate putative new coding sequences or RNA central (<http://rnacentral.org/sequence-search/>), Rfam search (<http://rfam.xfam.org/>) and RNAfold web server (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>) with default parameters to predict putative small RNAs and riboswitches.

Statistical analysis

Each statistical comparison presented was performed using the DESeq2 [46] R/Bioconductor package to normalize samples to each other and then test for differential expression. Notably, we have chosen to normalize the population of reads that map to each genome independently of the other. In the context of a phage infected cell rapidly replacing host transcripts with phage transcripts, this has artificially enriched host reads and depleted phage reads progressively over the course of infection by normalizing away the biologically relevant shift in each organism's proportion of the total reads in the cell. However, this has also allowed us to both show and test for differential expression of both phage and host gene features independently of the more global swing towards phage transcription.

High coverage metabolomics analysis

P. aeruginosa strain PAK cells, grown in minimal medium (30 mM Na₂HPO₄, 14 mM KH₂PO₄, 20 mM (NH₄)₂SO₄, 20 mM glucose, 1 mM MgSO₄, 4 μM FeSO₄), were infected with PAK_P3 at OD₆₀₀ = 0.3 (approx. 1.25.10⁸ CFU). At 0, 5, 10, 15, 20, and 25 minutes post infection, cells were collected by fast filtration [47]. The biomass quantity was adjusted to match the biomass of a 1 mL culture at OD₆₀₀ = 1.0 (approx. 4.10⁸ CFU) by following the OD₆₀₀ and adjustment of the sampling volume. Four biological replicates were sampled and two technical repeats were made of each independent biological sample. The metabolic content was extracted as described by De Smet *et al* [3]. The samples were profiled using only negative mode flow injection-time-of-flight mass spectrometry and detected ions were annotated as previously reported [17]. Metabolite annotation and statistical analysis was performed using Matlab R2013b (Mathworks, Natick, MA, United States) according to the ion annotation protocol described by Fuhrer *et al.* [17]. With this method, 6006 ions were detected and 918 of them could be assigned to known *P. aeruginosa* metabolites. After removal of ion adducts, 377 ions were retained that were annotated as 518 metabolites (including mass isomers).

Differential analysis was performed for each time point versus time point zero using a t-test for two samples with unequal variances (Welch test). For metabolic pathway enrichment, lists of significantly changing metabolites for each time point were created based on the thresholds of $|\text{Log}_2(\text{fold change})| \geq 0.5$ and adjusted p-value < 0.1. In each list, metabolites were sorted by the adjusted p-value, and the pathway enrichment procedure was performed for each subset of size 1 to the size of the significant list using Fischer test as described in [48] and the smallest p-value for each pathway was reported. For differential analysis and pathway enrichment, p-values were adjusted for multiple hypotheses testing with the Benjamini-Hochberg procedure.

Supporting Information

S1 Text. Supplemental methods.
(DOCX)

S1 Table. *P. aeruginosa* strain PAK reannotation based on transcriptomic data of the uninfected sample. Newly annotated features (small RNAs, riboswitches, CDS and antisense RNAs) are listed. Their location on PAK genome, their sequence and comments on their annotation are indicated.
(XLSX)

S2 Table. *P. aeruginosa* strain PAK differential gene expression statistical analysis. This table lists all PAK annotated features as well as their corresponding mean counts normalized with the DESeq2 normalization method while excluding phage counts, the calculated fold change and the results of a DESeq2 test for differential expression comparing the non-infected (0 min) and infected (13 min) conditions.
(XLSX)

S3 Table. PAK_P3 infection causes a dramatic increase in free deoxynucleotides. This table shows the Log₂(fold change) for all measured nucleotides over the course of PAK_P3 infection. In grey are those not significantly influenced, the color scale indicates the level of increase (green) or decrease (red) of the metabolite at the specific time point. Metabolite levels displaying more than 2-fold increase are highlighted in bold.
(XLSX)

S4 Table. PAK_P3 infection influences amino sugar metabolism. This table shows the Log₂(fold change) for all indicated metabolites over the course of PAK_P3 infection. In grey are those not significantly influenced, the green color scale indicates the level of increase of the metabolite at the specific time point.
(XLSX)

S5 Table. PAK_P3 differential gene expression statistical analysis. This table lists all PAK_P3 annotated features and their corresponding total gene reads as well as those phage counts for early (3.5 min) and late (13 min) infection normalized against each other, excluding host reads. Values for each independent biological replicate (R1, R2, R3) are indicated and the results of our statistical analysis are provided.
(XLSX)

S6 Table. PAK_P3 antisense transcripts and small RNAs. Newly annotated RNA features (small RNAs and antisense RNAs) are listed. Their location on PAK_P3 genome, their sequence and comments on their annotation are indicated.
(XLSX)

S1 Fig. Strain PAK displays high number of non-coding RNAs. A representative example of detection of (A) intergenic small RNA (sRNA); (B) riboswitch (Rsw) and (C) *cis*-antisense RNA (asRNA). The mapped reads were formatted into graph files for visualization in a strand-specific manner (black and pink represent reads mapping the forward and the reverse strands, respectively) using COV2HTML. The annotated non-coding RNA genes are indicated as grey arrows and open-reading frames annotated in strain PAK are shown as blue arrows.
(TIF)

S2 Fig. Temporal regulation of bacteriophage PAK_P3 gene expression is revealed by RNA-Seq analysis. Differential expression analysis comparing the expression of phage gene

features between early ($t = 3.5$ min) and late infection ($t = 13$ min) while excluding host reads from the normalization. Blue = Early gene features, Orange = Late gene features, Red = Gene features not significantly differentially expressed. The previously annotated gp113 is also shown with negligible expression throughout infection and has been deleted from the annotation.

(TIF)

S3 Fig. Bacteriophage PAK_P3 expresses many antisense transcripts. Mapped reads were formatted into graph files for visualization in a strand-specific manner (black and pink represent reads mapping the forward and the reverse strands, respectively) using COV2HTML. The annotated *cis*-antisense RNAs genes are indicated as grey arrows and open-reading frames annotated in PAK_P3 are shown as blue arrows. Data obtained 3.5 min and 13 min following PAK_P3 infection are presented in upper and lower part, respectively.

(TIF)

Acknowledgments

We are grateful to Zhou Xu (CNRS UMR8226, Paris, France) and Mart Krupovic (Molecular Biology of the Gene in Extremophiles, Institut Pasteur, Paris, France) for critical comments and helpful suggestions on the manuscript. We thank Henrik Almlad (Department of Immunology and Microbiology, Faculty of Health and Medical Sciences, Costerton Biofilm Center, University of Copenhagen, Copenhagen, Denmark) for providing pDONRPEX18Gm plasmid.

Author Contributions

Conceived and designed the experiments: AC BGB JDS MZ MK US PJ MW LD RL. Performed the experiments: AC BGB JDS MZ MK US PJ MW. Analyzed the data: AC BGB JDS MM MZ MK US PJ MW LD RL. Contributed reagents/materials/analysis tools: AC BGB JDS MM MZ MK US PJ MW LD RL. Wrote the paper: AC BGB JDS MM MZ MK US PJ MW LD RL.

References

1. Reardon S (2014) Phage therapy gets revitalized. *Nature* 510: 15–16. doi: [10.1038/510015a](https://doi.org/10.1038/510015a) PMID: [24899282](https://pubmed.ncbi.nlm.nih.gov/24899282/)
2. Ceyssens PJ, Minakhin L, Van den Bossche A, Yakunina M, Klimuk E, et al. (2014) Development of giant bacteriophage varphiKZ is independent of the host transcription apparatus. *J Virol* 88: 10501–10510. doi: [10.1128/JVI.01347-14](https://doi.org/10.1128/JVI.01347-14) PMID: [24965474](https://pubmed.ncbi.nlm.nih.gov/24965474/)
3. De Smet J, Zimmermann M, Kogadeeva M, Ceyssens PJ, Vermaelen W, et al. (2016) High coverage metabolomics analysis reveals phage-specific alterations to *Pseudomonas aeruginosa* physiology during infection. *ISME J*.
4. Henry M, Lavigne R, Debarbieux L (2013) Predicting In Vivo Efficacy of Therapeutic Bacteriophages Used To Treat Pulmonary Infections. *Antimicrobial Agents and Chemotherapy* 57: 5961–5968. doi: [10.1128/AAC.01596-13](https://doi.org/10.1128/AAC.01596-13) PMID: [24041900](https://pubmed.ncbi.nlm.nih.gov/24041900/)
5. Henry M, Bobay LM, Chevallereau A, Sausseureau E, Ceyssens PJ, et al. (2015) The search for therapeutic bacteriophages uncovers one new subfamily and two new genera of *Pseudomonas*-infecting Myoviridae. *PLoS One* 10: e0117163. doi: [10.1371/journal.pone.0117163](https://doi.org/10.1371/journal.pone.0117163) PMID: [25629728](https://pubmed.ncbi.nlm.nih.gov/25629728/)
6. Matthews CK, Allen JR (1983) DNA precursor biosynthesis. In: Matthews CK, Kutter EM, Mosig G, Berg PB, editors. *Bacteriophage T4*. Washington, D.C.: American Society for Microbiology. pp. 59–70.
7. Monot M, Orgeur M, Camiade E, Brehier C, Dupuy B (2014) COV2HTML: a visualization and analysis tool of bacterial next generation sequencing (NGS) data for postgenomics life scientists. *OMICS* 18: 184–195. doi: [10.1089/omi.2013.0119](https://doi.org/10.1089/omi.2013.0119) PMID: [24512253](https://pubmed.ncbi.nlm.nih.gov/24512253/)
8. Sonnleitner E, Romeo A, Blasi U (2012) Small regulatory RNAs in *Pseudomonas aeruginosa*. *RNA Biol* 9: 364–371. doi: [10.4161/rna.19231](https://doi.org/10.4161/rna.19231) PMID: [22336763](https://pubmed.ncbi.nlm.nih.gov/22336763/)

9. Ferrara S, Brugnoli M, De Bonis A, Righetti F, Delvillani F, et al. (2012) Comparative profiling of *Pseudomonas aeruginosa* strains reveals differential expression of novel unique and conserved small RNAs. *PLoS One* 7: e36553. doi: [10.1371/journal.pone.0036553](https://doi.org/10.1371/journal.pone.0036553) PMID: [22590564](https://pubmed.ncbi.nlm.nih.gov/22590564/)
10. Gomez-Lozano M, Marvig RL, Molin S, Long KS (2012) Genome-wide identification of novel small RNAs in *Pseudomonas aeruginosa*. *Environ Microbiol* 14: 2006–2016. doi: [10.1111/j.1462-2920.2012.02759.x](https://doi.org/10.1111/j.1462-2920.2012.02759.x) PMID: [22533370](https://pubmed.ncbi.nlm.nih.gov/22533370/)
11. Wurtzel O, Yoder-Himes DR, Han K, Dandekar AA, Edelheit S, et al. (2012) The single-nucleotide resolution transcriptome of *Pseudomonas aeruginosa* grown in body temperature. *PLoS Pathog* 8: e1002945. doi: [10.1371/journal.ppat.1002945](https://doi.org/10.1371/journal.ppat.1002945) PMID: [23028334](https://pubmed.ncbi.nlm.nih.gov/23028334/)
12. Gomez-Lozano M, Marvig RL, Molina-Santiago C, Tribelli PM, Ramos JL, et al. (2015) Diversity of small RNAs expressed in *Pseudomonas* species. *Environ Microbiol Rep* 7: 227–236. doi: [10.1111/1758-2229.12233](https://doi.org/10.1111/1758-2229.12233) PMID: [25394275](https://pubmed.ncbi.nlm.nih.gov/25394275/)
13. Vafabakhsh R, Kondabagil K, Earnest T, Lee KS, Zhang Z, et al. (2014) Single-molecule packaging initiation in real time by a viral DNA packaging machine from bacteriophage T4. *Proc Natl Acad Sci U S A* 111: 15096–15101. doi: [10.1073/pnas.1407235111](https://doi.org/10.1073/pnas.1407235111) PMID: [25288726](https://pubmed.ncbi.nlm.nih.gov/25288726/)
14. Kashlev M, Nudler E, Goldfarb A, White T, Kutter E (1993) Bacteriophage T4 Alc protein: a transcription termination factor sensing local modification of DNA. *Cell* 75: 147–154. PMID: [8402894](https://pubmed.ncbi.nlm.nih.gov/8402894/)
15. Abedon ST (2012) Bacterial 'immunity' against bacteriophages. *Bacteriophage* 2: 50–54. PMID: [22666656](https://pubmed.ncbi.nlm.nih.gov/22666656/)
16. Calendar R (2006) The bacteriophages. Oxford; New York: Oxford University Press. xiii, 746 p. p.
17. Fuhrer T, Heer D, Begemann B, Zamboni N (2011) High-throughput, accurate mass metabolome profiling of cellular extracts by flow injection-time-of-flight mass spectrometry. *Anal Chem* 83: 7074–7080. doi: [10.1021/ac201267k](https://doi.org/10.1021/ac201267k) PMID: [21830798](https://pubmed.ncbi.nlm.nih.gov/21830798/)
18. Ankrah NY, May AL, Middleton JL, Jones DR, Hadden MK, et al. (2014) Phage infection of an environmentally relevant marine bacterium alters host metabolism and lysate composition. *ISME J* 8: 1089–1100. doi: [10.1038/ismej.2013.216](https://doi.org/10.1038/ismej.2013.216) PMID: [24304672](https://pubmed.ncbi.nlm.nih.gov/24304672/)
19. Dedrick RM, Marinelli LJ, Newton GL, Pogliano K, Pogliano J, et al. (2013) Functional requirements for bacteriophage growth: gene essentiality and expression in mycobacteriophage Giles. *Mol Microbiol* 88: 577–589. doi: [10.1111/mmi.12210](https://doi.org/10.1111/mmi.12210) PMID: [23560716](https://pubmed.ncbi.nlm.nih.gov/23560716/)
20. Nejman-Falenczyk B, Bloch S, Licznarska K, Dydecka A, Felczykowska A, et al. (2015) A small, micro-RNA-size, ribonucleic acid regulating gene expression and development of Shiga toxin-converting bacteriophage Phi24Beta. *Sci Rep* 5: 10080. doi: [10.1038/srep10080](https://doi.org/10.1038/srep10080) PMID: [25962117](https://pubmed.ncbi.nlm.nih.gov/25962117/)
21. McClain WH, Guthrie C, Barrell BG (1972) Eight transfer RNAs induced by infection of *Escherichia coli* with bacteriophage T4. *Proc Natl Acad Sci U S A* 69: 3703–3707. PMID: [4566457](https://pubmed.ncbi.nlm.nih.gov/4566457/)
22. Citorik RJ, Mimee M, Lu TK (2014) Bacteriophage-based synthetic biology for the study of infectious diseases. *Curr Opin Microbiol* 19: 59–69. doi: [10.1016/j.mib.2014.05.022](https://doi.org/10.1016/j.mib.2014.05.022) PMID: [24997401](https://pubmed.ncbi.nlm.nih.gov/24997401/)
23. Drulis-Kawa Z, Majkowska-Skrobek G, Maciejewska B, Delattre AS, Lavigne R (2012) Learning from bacteriophages—advantages and limitations of phage and phage-encoded protein applications. *Curr Protein Pept Sci* 13: 699–722. PMID: [23305359](https://pubmed.ncbi.nlm.nih.gov/23305359/)
24. Hermoso JA, Garcia JL, Garcia P (2007) Taking aim on bacterial pathogens: from phage therapy to enzybiotics. *Curr Opin Microbiol* 10: 461–472. PMID: [17904412](https://pubmed.ncbi.nlm.nih.gov/17904412/)
25. Rothman-Denes BRLaLB (2015) Structural and Biochemical Investigation of Bacteriophage N4-Encoded RNA Polymerases. *Biomolecules* 5: 647–667. doi: [10.3390/biom5020647](https://doi.org/10.3390/biom5020647) PMID: [25924224](https://pubmed.ncbi.nlm.nih.gov/25924224/)
26. Nejman-Falenczyk B, Bloch S, Licznarska K, Felczykowska A, Dydecka A, et al. (2015) Small regulatory RNAs in lambdoid bacteriophages and phage-derived plasmids: Not only antisense. *Plasmid* 78: 71–78. doi: [10.1016/j.plasmid.2014.07.006](https://doi.org/10.1016/j.plasmid.2014.07.006) PMID: [25111672](https://pubmed.ncbi.nlm.nih.gov/25111672/)
27. Tree JJ, Granneman S, McAteer SP, Tollervy D, Gally DL (2014) Identification of bacteriophage-encoded anti-sRNAs in pathogenic *Escherichia coli*. *Mol Cell* 55: 199–213. doi: [10.1016/j.molcel.2014.05.006](https://doi.org/10.1016/j.molcel.2014.05.006) PMID: [24910100](https://pubmed.ncbi.nlm.nih.gov/24910100/)
28. Wagemans J, Blasdel BG, Van den Bossche A, Uytterhoeven B, De Smet J, et al. (2014) Functional elucidation of antibacterial phage ORFans targeting *Pseudomonas aeruginosa*. *Cell Microbiol* 16: 1822–1835. doi: [10.1111/cmi.12330](https://doi.org/10.1111/cmi.12330) PMID: [25059764](https://pubmed.ncbi.nlm.nih.gov/25059764/)
29. Ueno H, Yonesaki T (2004) Phage-induced change in the stability of mRNAs. *Virology* 329: 134–141. PMID: [15476881](https://pubmed.ncbi.nlm.nih.gov/15476881/)
30. Uzan M (2009) RNA processing and decay in bacteriophage T4. *Prog Mol Biol Transl Sci* 85: 43–89. doi: [10.1016/S0079-6603\(08\)00802-7](https://doi.org/10.1016/S0079-6603(08)00802-7) PMID: [19215770](https://pubmed.ncbi.nlm.nih.gov/19215770/)

31. Chakravarty AK, Shuman S (2011) RNA 3'-phosphate cyclase (RtcA) catalyzes ligase-like adenylation of DNA and RNA 5'-monophosphate ends. *J Biol Chem* 286: 4117–4122. doi: [10.1074/jbc.M110.196766](https://doi.org/10.1074/jbc.M110.196766) PMID: [21098490](https://pubmed.ncbi.nlm.nih.gov/21098490/)
32. Filipowicz W, Billy E, Drabikowski K, Genschik P (1998) Cyclases of the 3'-terminal phosphate in RNA: a new family of RNA processing enzymes conserved in eucarya, bacteria and archaea. *Acta Biochim Pol* 45: 895–906. PMID: [10397337](https://pubmed.ncbi.nlm.nih.gov/10397337/)
33. Tanaka N, Shuman S (2011) RtcB is the RNA ligase component of an *Escherichia coli* RNA repair operon. *J Biol Chem* 286: 7727–7731. doi: [10.1074/jbc.C111.219022](https://doi.org/10.1074/jbc.C111.219022) PMID: [21224389](https://pubmed.ncbi.nlm.nih.gov/21224389/)
34. Schmidt FJ (1985) RNA splicing in prokaryotes: bacteriophage T4 leads the way. *Cell* 41: 339–340. PMID: [2580641](https://pubmed.ncbi.nlm.nih.gov/2580641/)
35. Levitz R, Chapman D, Amitsur M, Green R, Snyder L, et al. (1990) The optional *E. coli* prr locus encodes a latent form of phage T4-induced anticodon nuclease. *EMBO J* 9: 1383–1389. PMID: [1691706](https://pubmed.ncbi.nlm.nih.gov/1691706/)
36. Lavigne R, Lecoutere E, Wagemans J, Cenens W, Aertsen A, et al. (2013) A multifaceted study of *Pseudomonas aeruginosa* shutdown by virulent podovirus LUZ19. *MBio* 4: e00061–00013. doi: [10.1128/mBio.00061-13](https://doi.org/10.1128/mBio.00061-13) PMID: [23512961](https://pubmed.ncbi.nlm.nih.gov/23512961/)
37. Adriaenssens EM, Mattheus W, Cornelissen A, Shaburova O, Krylov VN, et al. (2012) Complete genome sequence of the giant *Pseudomonas* phage Lu11. *J Virol* 86: 6369–6370. doi: [10.1128/JVI.00641-12](https://doi.org/10.1128/JVI.00641-12) PMID: [22570243](https://pubmed.ncbi.nlm.nih.gov/22570243/)
38. McCorquodale DJ, Warner RH (1988) Bacteriophage T5 and related phages. In: Calendar R, editor. *The bacteriophages*. Oxford; New York: Oxford University Press. pp. 439–476.
39. Zschach H, Joensen KG, Lindhard B, Lund O, Goderdzishvili M, et al. (2015) What Can We Learn from a Metagenomic Analysis of a Georgian Bacteriophage Cocktail? *Viruses* 7: 6570–6589. doi: [10.3390/v7122958](https://doi.org/10.3390/v7122958) PMID: [26703713](https://pubmed.ncbi.nlm.nih.gov/26703713/)
40. Debarbieux L, Leduc D, Maura D, Morello E, Criscuolo A, et al. (2010) Bacteriophages can treat and prevent *Pseudomonas aeruginosa* lung infections. *J Infect Dis* 201: 1096–1104. doi: [10.1086/651135](https://doi.org/10.1086/651135) PMID: [20196657](https://pubmed.ncbi.nlm.nih.gov/20196657/)
41. Morello E, Sausseureau E, Maura D, Huerre M, Touqui L, et al. (2011) Pulmonary bacteriophage therapy on *Pseudomonas aeruginosa* cystic fibrosis strains: first steps towards treatment and prevention. *PLoS One* 6: e16963. doi: [10.1371/journal.pone.0016963](https://doi.org/10.1371/journal.pone.0016963) PMID: [21347240](https://pubmed.ncbi.nlm.nih.gov/21347240/)
42. Takeya K, Amako K (1966) A rod-shaped *Pseudomonas* phage. *Virology* 28: 163–165. PMID: [4955194](https://pubmed.ncbi.nlm.nih.gov/4955194/)
43. Hyman P, Abedon ST (2009) Practical methods for determining phage growth parameters. *Methods Mol Biol* 501: 175–202. doi: [10.1007/978-1-60327-164-6_18](https://doi.org/10.1007/978-1-60327-164-6_18) PMID: [19066822](https://pubmed.ncbi.nlm.nih.gov/19066822/)
44. Galens K, Orvis J, Daugherty S, Creasy HH, Angiuoli S, et al. (2011) The IGS Standard Operating Procedure for Automated Prokaryotic Annotation. *Stand Genomic Sci* 4: 244–251. doi: [10.4056/sigs.1223234](https://doi.org/10.4056/sigs.1223234) PMID: [21677861](https://pubmed.ncbi.nlm.nih.gov/21677861/)
45. Blasdel B, Ceyssens P-J, Lavigne R (2016) Preparing cDNA libraries from lytic phage infected cells for whole transcriptome analysis by RNA-Seq. In: eds Clokie MRJea, editor. *Bacteriophages: Methods and protocols*: Humana Press.
46. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550. PMID: [25516281](https://pubmed.ncbi.nlm.nih.gov/25516281/)
47. Zimmermann M, Thormann V, Sauer U, Zamboni N (2013) Nontargeted profiling of coenzyme A thioesters in biological samples by tandem mass spectrometry. *Anal Chem* 85: 8284–8290. doi: [10.1021/ac401555n](https://doi.org/10.1021/ac401555n) PMID: [23895734](https://pubmed.ncbi.nlm.nih.gov/23895734/)
48. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102: 15545–15550. PMID: [16199517](https://pubmed.ncbi.nlm.nih.gov/16199517/)

ARTICLE 3

—

COMPARATIVE TRANSCRIPTOMICS ANALYSES REVEAL THE CONSERVATION OF AN ANCESTRAL INFECTIOUS STRATEGY IN TWO BACTERIOPHAGE GENERA

Published in The ISME Journal

ORIGINAL ARTICLE

Comparative transcriptomics analyses reveal the conservation of an ancestral infectious strategy in two bacteriophage genera

Bob G Blasdel^{1,5}, Anne Chevallereau^{2,3,5}, Marc Monot⁴, Rob Lavigne¹ and Laurent Debarbieux²

¹Laboratory of Gene Technology, Department of Biosystems, Leuven, Belgium; ²Institut Pasteur, Molecular Biology of the Gene in Extremophiles Unit, Department of Microbiology, Paris, France; ³Université Paris Diderot, Sorbonne Paris Cité, Cellule Pasteur, Paris, France and ⁴Institut Pasteur, Laboratoire Pathogénèse des bactéries anaérobies, Département de Microbiologie, Paris, France

Although the evolution of tailed bacteriophages has increasingly been better understood through comparisons of their DNA sequences, the functional consequences of this evolution on phage infectious strategies have remained unresolved. In this study, we comprehensively compared the transcriptional strategies of two related myoviruses, PAK_P3 and PAK_P4, infecting the same *Pseudomonas aeruginosa* host strain. Outside of the conservation of their structural clusters, their highly syntenic genomes display only limited DNA similarity. Despite this apparent divergence, we found that both viruses follow a similar infection scheme, relying on a temporal regulation of their gene expression, likely involving the use of antisense transcripts, as well as a rapid degradation of 90% of the host non-ribosomal mRNA, as previously reported for PAK_P3. However, the kinetics of the mRNA degradation is remarkably faster during PAK_P4 infection. Moreover, we found that each virus has evolved specific adaptations, as exemplified by the distinct patterns of their core genes expression as well as the specific manipulation of the expression of iron-related host genes by PAK_P4. This study enhances our understanding of the evolutionary process of virulent phages, which relies on adjusting globally conserved ancestral infection mechanisms.

The ISME Journal advance online publication, 12 May 2017; doi:10.1038/ismej.2017.63

Introduction

Viruses that infect bacteria, also known as bacteriophages (phages), have crucial roles in a variety of environments including the renewal of bacterial populations, recycling biochemical resources and horizontal gene transfer, as reflected by their abundance and diversity. Multiple comparative genomic analyses have suggested that related phage genomes can be clustered based on their sequence similarity (Kwan *et al.*, 2005; Kropinski *et al.*, 2007; Marinelli *et al.*, 2012; Grose and Casjens, 2014). However, several studies have shown that bacteria and their phages rapidly coevolve, promoting genetic diversity in both populations (Marston

et al., 2012; Williams, 2013; Koskella and Brockhurst, 2014). This genetic diversity is further stimulated by horizontal gene transfer (HGT) events, which in some cases can make it difficult to reconstruct evolutionary history of phages. Consequently, many studies have aimed at understanding how different phage clusters are related to each other or how phage populations are structured (based on comparative (meta)genomics) (Hatfull *et al.*, 2006; Kwan *et al.*, 2006; Grose and Casjens, 2014; Pope *et al.*, 2015; Roux *et al.*, 2015).

Apart from host-range and defense systems evolution, no mechanistic studies have yet experimentally addressed the evolution of phage infection strategies themselves, leaving unresolved questions on the consequences of divergent evolution of phage genomic sequences on the functions of those sequences. In particular, it has not yet been directly investigated how conserved the infection strategy is during the evolution of virulent phages, which can only be weakly predicted using bioinformatics tools. Following the isolation and genomic characterization of several phages infecting the same *Pseudomonas aeruginosa* host strain, an opportunistic pathogen widely present in the environment, we previously found that, although their morphologies were nearly

Correspondence: R Lavigne, Laboratory of Gene Technology, Department of Biosystems, KU Leuven Kasteelpark Arenberg 21, Box 2462, Heverlee 3001, Belgium.

E-mail: rob.lavigne@kuleuven.be

or L Debarbieux, Unité Biologie Moléculaire du Gène chez les Extrémophiles, Département de Microbiologie. Institut Pasteur, 25 rue du Dr Roux, 75015 Paris, France.

E-mail: laurent.debarbieux@pasteur.fr

⁵These authors contributed equally to this work.

Received 25 October 2016; revised 22 February 2017; accepted 22 March 2017

identical, their genomic contents were divergent enough to support their classification within two new genera among *Myoviridae*, namely *Kpp10virus* and *Pakpunavirus* (Henry *et al.*, 2015; Krupovic *et al.*, 2016). Moreover, an additional study analyzing phages belonging to these genera pointed out that their genomes are typically composed of strongly conserved regions (0.1% single-nucleotide polymorphism (SNP) frequency) alternating with heterogeneous regions (up to 20% SNP frequency, scars of HGT and recombination events; Essoh *et al.*, 2015). Overall, our previous comparative genomic analysis revealed that phages belonging to these distinct genera differ by their mean genome length, their number of predicted open reading frames (ORFs) and tRNA, and their GC content, but they also share homologous proteins (about 25%), most of which encoding structural components (Henry *et al.*, 2015). Despite this significant genomic divergence, they display a strong synteny and the reconstruction of their phylogeny, based on four conserved proteins, revealed that they likely share a relatively recent common ancestor (Henry *et al.*, 2015).

By selecting PAK_P3 and PAK_P4 as representatives of *Kpp10virus* and *Pakpunavirus* genera respectively, we aimed to investigate whether the transcriptional strategy is conserved during evolution of phages by performing a transcriptomic-based comparative analysis (RNA-sequencing). Our investigation was driven by the following questions. Do phages that have a common origin and subsequently diverge, still share common mechanisms to infect their host? How similar are the host transcriptional responses to infections by two different phages?

We found that the global infection process is conserved (temporal regulation of gene expression, production of antisense transcripts, rapid degradation of host mRNA) with few phage specific tweaks (specific manipulation of host gene expression, custom patterns of phage core gene expression). Altogether, these results suggest that, despite genomic divergence, phages are bound to their ancestral mode of action.

Materials and methods

Strains and growth conditions

P. aeruginosa strain PAK (Takeya and Amako, 1966) was grown in LB medium supplemented with 10 mM CaCl₂ at 37 °C unless stated otherwise. For RNA-seq experiments, cells were infected with bacteriophage PAK_P4 (Henry *et al.*, 2013) using a multiplicity of infection (MOI) of 25 to ensure the synchronicity of the infection (95% of the bacterial population killed after 5 min phage-introduction). Bacteriophage PAK_P3 was isolated and described by Morello *et al.* (2011). Updated phage genome sequences are available in Genbank and can be accessed via accession numbers NC_022970 (PAK_P3), NC_022986 (PAK_P4). Updated annotations of strain PAK genome are available via GenBank accession number CP020659.

Adsorption assay and one-step growth experiment
PAK_P4 adsorption assays (MOI 10⁻³) and one-step growth experiments (MOI 10⁻¹) were performed as described in Chevallereau *et al.* (2016). Briefly, a liquid culture of strain PAK was infected at indicated MOI and incubated at 37 °C during 5 min with agitation to allow phage particles to adsorb. The culture was diluted (1/1000) and two 100 µl samples were collected every 2 min and either kept on ice until titration, or mixed with CHCl₃. For each sample, free phage titer (samples with CHCl₃) as well as the number of free phages and infective centers (samples without CHCl₃) were recorded to calculate eclipse and latency periods, respectively.

Dependence to host RNA polymerase assay

The experiment was performed according to the methods described by Ceyssens *et al.* (2014). The strain PAK was grown in LB medium until it reached OD_{600nm} ~0.3, then the culture was split with one half supplemented with 100 µg ml⁻¹ rifampicin (Rif) and second half with the same volume of water. Ten minutes later, phages were added at MOI of 1. Samples were taken immediately (*t*=0), 15, 60 and 150 min post infection and phages were titrated.

Whole transcriptome sequencing and analysis

RNA-sequencing analysis was performed as described in Chevallereau *et al.* (2016). Over 12 million 75 bp reads mapping to non-ribosomal regions were obtained from each library with the exception of one early sample (2.4 million reads). After trimming, sequencing reads were aligned separately to both the phage and host genomes with the CLC Genomics workbench v7.5.1 (QIAGEN Bioinformatics, Aarhus, Denmark) These alignments were then summarized into count tables of Total Gene Reads that map to phage or host gene features respectively. 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.

RNA-seq data have been deposited in NCBI-GEO with accession no. GSE86022.

RNA-seq coverage visualization is available through the COV2HTML software at https://mmonot.eu/COV2HTML/visualisation.php?str_id=-36 for a comparison of the host (0 min/13 min) and https://mmonot.eu/COV2HTML/visualisation.php?str_id=-38 for a comparison of the phage (3.5 min/13 min).

Results

PAK_P3 and PAK_P4 share common life-history traits

To compare the molecular mechanisms of infection used by virulent phages that recently diverged, we chose two phages infecting the same *P. aeruginosa* host, strain PAK. These two myoviruses, PAK_P3 and

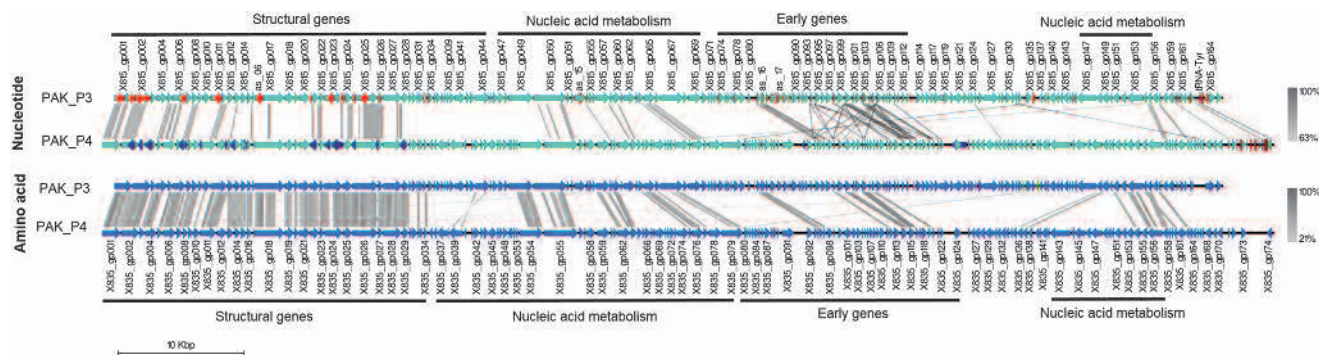


Figure 1 Genomic comparison of phages PAK_P3 and PAK_P4. Upper panel: nucleotide comparison. ORFs are depicted as cyan arrows, antisense RNAs are represented as red and dark blue arrows, and small RNAs (ncRNA, tRNA) appear as red squares. Lower panel: translated nucleotide comparison. Coding sequences are colored in blue.

PAK_P4, respectively belong to the newly defined genera *Kpp10virus* and *Pakpunavirus* (Henry *et al.*, 2015) and display distinct host ranges (Saussereau *et al.*, 2014). PAK_P4 shares an average 76% ($\pm 9\%$) identity over 13% of PAK_P3 genome, in other words, only 9% of the nucleotide sequence is identical between the two phages (Supplementary Table 1; Figure 1). Sixty-nine homologous coding sequences, hereafter referred as core genes, have been defined (that is, $\sim 40\%$ of the proteomic content of the two phages, according to CoreGenes analysis with default parameters (Mahadevan *et al.*, 2009)). The sequence identity for each pair of homologous predicted proteins ranges from 24 to 85% (with a mean of $49 \pm 14\%$). Most of these homologous proteins belong to the structural (49%) and nucleic acid metabolism (32%) modules, while 19% of them are encoded by early-expressed ORFs (defined in the third paragraph) which are generally considered to be the least conserved phage proteins, often responsible for the subversion of host metabolism (Calendar, 2006) (Supplementary Table 1; Figure 1). It has to be noted that among the coding sequences that do not meet the criteria required to be considered as homologous (that is, blastp score < 75), some appear to have the same predicted functions. For example, the putative ribonucleoside-reductase beta subunits (PAK_P3_gp069 and PAK_P4_gp078) share only $\sim 30\%$ of similarity ($\sim 17\%$ of identity) in their protein sequences. Nevertheless, most of the predicted ORFs could not be associated with a putative function.

Altogether, these observations raise the question of whether the infectious mechanisms and strategies used by PAK_P3 and PAK_P4 are conserved, given their divergent evolution. To answer this, we first compared their growth parameters. PAK_P4 adsorbs onto the strain PAK within 4.8 ± 1.7 min ($k_a = 2.4 \times 10^{-9} \pm 1.4 \times 10^{-9}$ ml min $^{-1}$) and rapidly assembles new functional virions (eclipse period: 13.3 ± 2.0 min), although they are not immediately released (latency period: 18.2 ± 2.9 min). PAK_P4 infected cells eventually produce an average of 13 ± 5 progeny phages within a mean infection cycle duration as short as 21.4 ± 1.8 min (Supplementary Figure 1). In comparison with PAK_P3, PAK_P4

displays a smaller burst size and a longer latency period (Chevallereau *et al.*, 2016). These results hint that, if both phages have largely similar infection strategies (as shown below), they may have evolved unique processes for enacting them that may be variably effective. To investigate this further, we studied the transcriptomic dynamics of both phages.

PAK_P4 and PAK_P3 rapidly overwhelm the transcriptional environment of the host by efficiently hijacking bacterial RNA polymerase

In a previous study, we investigated the dynamics of host strain PAK transcriptome throughout PAK_P3 infection focusing on early (3.5 min) and late (13 min) stages of infection (Chevallereau *et al.*, 2016). As PAK_P4 has a comparable eclipse period as PAK_P3, we performed a transcriptomic analysis using the same time points. While PAK_P3 induces a progressive depletion of host transcripts that eventually represent only 13% of non-ribosomal RNAs in the cell 13 min post infection, PAK_P4 is much more rapid as only 20% of the total transcripts matches the host genome as soon as 3.5 min post infection (Figure 2). Both phages shut down host cell transcription but following different kinetics. We hypothesize that they may have a different control of the machineries dedicated to (i) degrade host mRNAs and/or (ii) catalyze the transcription of viral genes.

The rapid and overwhelming replacement of host transcripts with viral transcripts could indicate that both phages rely on their own transcription apparatus even though they are not predicted to encode a viral RNA polymerase (RNAP) based on sequence homology searches. Therefore, we experimentally assessed the dependency of PAK_P3 and PAK_P4 infections on host RNAP using rifampicin (Rif), which is known to inactivate this enzyme. We found that addition of Rif abolished PAK_P3 and PAK_P4 amplification while it did not affect phage PhiKZ transcription, as previously shown (Supplementary Figure 2; Ceyssens *et al.*, 2014). This indicates that both phages appear to

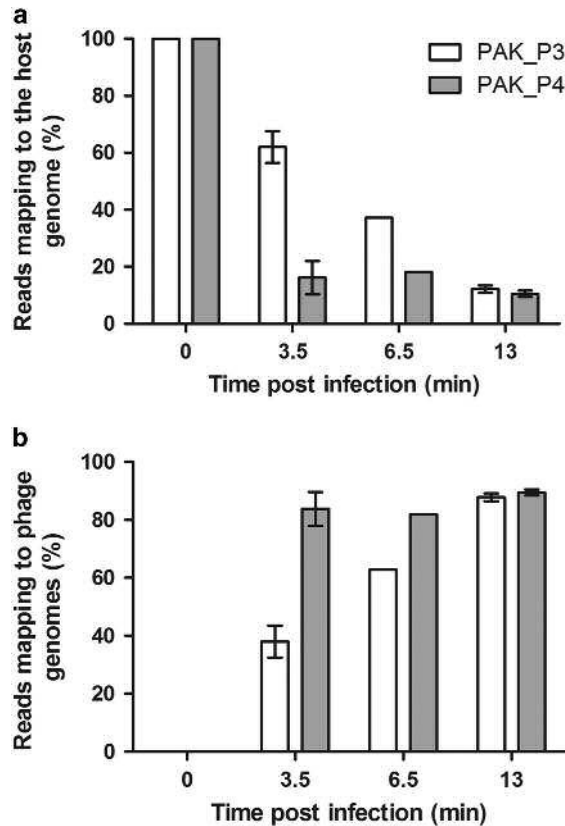


Figure 2 PAK_P4 takes over the host cell transcription, faster than PAK_P3, over the course of infection. Three independent biological replicates of RNA extracts were harvested from PAK_P4 and PAK_P3 infected cells at 0 min, 3.5 min and 13 min post infection, as well as a single sample collected at 6.5 min and all were subsequently sequenced. Percentages of reads mapping to phage genomes (a) and to the host genome (b) over the course of PAK_P4 infection (gray bars) or PAK_P3 infection (white bars) are plotted. The 6.5 min time point was not included in further analyses as it was performed as a single replicate.

efficiently hijack the host transcription machinery to complete their infectious cycle.

The transcriptional strategy of PAK_P4 is broadly similar to PAK_P3

PAK_P4 temporally regulates its transcription over the course of infection, as previously found for PAK_P3 (Chevallereau *et al.*, 2016). Most of the gene features encoding unique hypothetical proteins (that is, with low or no homology to any other amino acid sequence in databases) are expressed during the early stage of infection (3.5 min). This ‘early region’ encompasses gene features encoding gp080 through gp124, while the ‘late region’ encompasses genes encoding gp001 to gp034. These gene features, which are massively expressed 13 min post infection, primarily encode the predicted phage structural proteins (Supplementary Table 2; Figure 3). PAK_P4 also produces antisense RNAs (asRNAs) that appear to be acting as *cis*-encoded asRNAs since most of them (15 out of 17) are encoded within structural genes (Supplementary Table 2). These asRNAs are

significantly more expressed (2 to 3-fold change) 3.5 min after infection compared with late infection stage (13 min). We hypothesize that the transcription of these non-coding RNAs act as repressors to attenuate the expression of the structural genes during the early stage of infection. Since both phage strategies display similar characteristics, our data support an overall conservation of regulatory mechanisms to control viral gene expression, but does not exclude specific differences (see below).

Core genes are similarly expressed in both infections but display extensive phage specific adjustment

To determine more precisely how the regulation of phage gene expression is adjusted, we compared the differential expression of the set of 69 core genes. While most conserved gene features still undergo the same temporal regulation, the strength of this differential expression appears to have changed between the two viruses (Figure 4; Supplementary Table 3). For instance, PAK_P3 tends to differentially express gp002-gp006 more drastically than PAK_P4, while PAK_P4 favors the differential expression of gp007-gp010. Strikingly, one gene is strongly upregulated during PAK_P4 late infection while it is not temporally regulated in PAK_P3. The corresponding hypothetical protein (gp014 in PAK_P4) possesses a wide range of homologs in other unrelated phages and in some bacteria. We also found that genes encoding DNA-metabolism related proteins as well as early genes are overall more strongly temporally regulated in PAK_P3 than they are in PAK_P4.

This comparative analysis of gene expression highlights that evolution has readily selected for small modifications of the mechanisms regulating a variety of specific systems, potentially having a major impact on the way the virus interacts with its host. To address this, we compare how PAK_P3 and PAK_P4 manipulate host transcription and investigate whether the host similarly responds to both phages.

PAK_P4 specifically evolved mechanisms to modulate the expression of genes involved in iron uptake

Phage PAK_P4 strongly upregulates the expression of three host operons predicted to be related to siderophore synthesis and transport, which are not upregulated during PAK_P3 infection (Figure 5; Supplementary Table 4). Indeed, *PAK_4094-4104* (mean $\text{Log}_2(\text{FC}) = 6.1 \pm 1.1$) and *PAK_4106-09* (mean $\text{Log}_2(\text{FC}) = 6.3 \pm 0.4$) are respectively homologous to the *pchEFG* and *pchABCD* operons of *P. aeruginosa* PAO1 (Winsor *et al.*, 2016), which are involved in pyochelin biosynthesis, one of the major siderophores of *P. aeruginosa* while *PAK_4090-93* (mean $\text{Log}_2(\text{FC}) = 5.4 \pm 0.5$) is homologous to *fpt* system, encoding the corresponding siderophore receptor. Interestingly, another upregulated operon, *PAK_1971-78* (mean $\text{Log}_2(\text{FC}) = 3.0 \pm 0.3$) is homologous

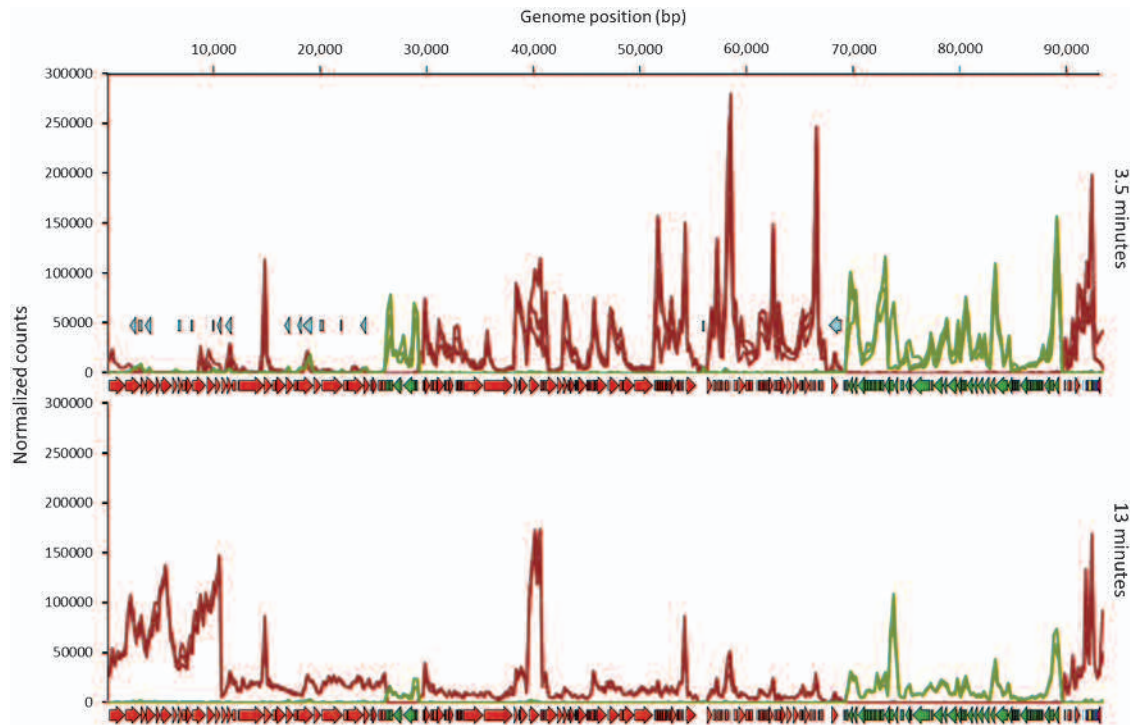


Figure 3 PAK_P4 Transcription map. Transcript densities for early and late infections, normalized by the Total Count of reads that align to the phage genome for each sample. Read density and coding sequence annotations on the Watson and Crick strands are highlighted in red and green respectively while antisense RNA, tRNA and other non-coding RNA annotations are depicted in light blue, dark blue and yellow.

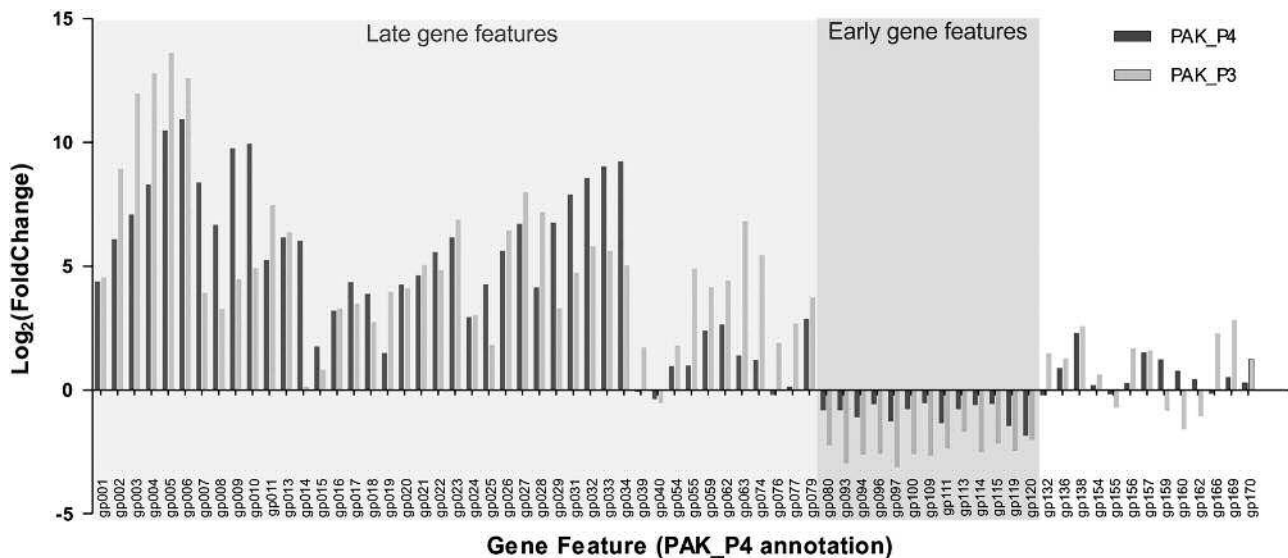


Figure 4 Comparison of PAK_P3 and PAK_P4 homologous genes expression. Phage homologous genes are listed and their differential expression (late versus early) during PAK_P3 and PAK_P4 infections are indicated as $\text{Log}_2(\text{fold change})$ values in light and dark gray, respectively.

to the *pvdIJD* operon (Winsor *et al.*, 2016), encoding non-ribosomal peptide synthetases involved in the biosynthesis of the other major *P. aeruginosa* siderophore pyoverdine. Moreover, *PAK_4645-51* and *PAK_4652*, respectively homologous to the *phuS-TUVW* operon and *phuR* receptor gene (Winsor *et al.*, 2016), are also upregulated, although to a lesser extent (mean $\text{Log}_2(\text{FC}) = 2.3 \pm 1.0$). These

genes encode an ABC transporter involved in heme uptake (Ochsner *et al.*, 2000). Finally, 14 other genes putatively involved in iron acquisition (mostly TonB-dependent receptors) are upregulated upon infection (mean $\text{Log}_2(\text{FC}) = 2.0 \pm 0.4$) while *PAK_4711*, a Fur-like protein (ferric uptake regulator), which represses the expression of genes involved in iron uptake, storage and metabolism in the presence

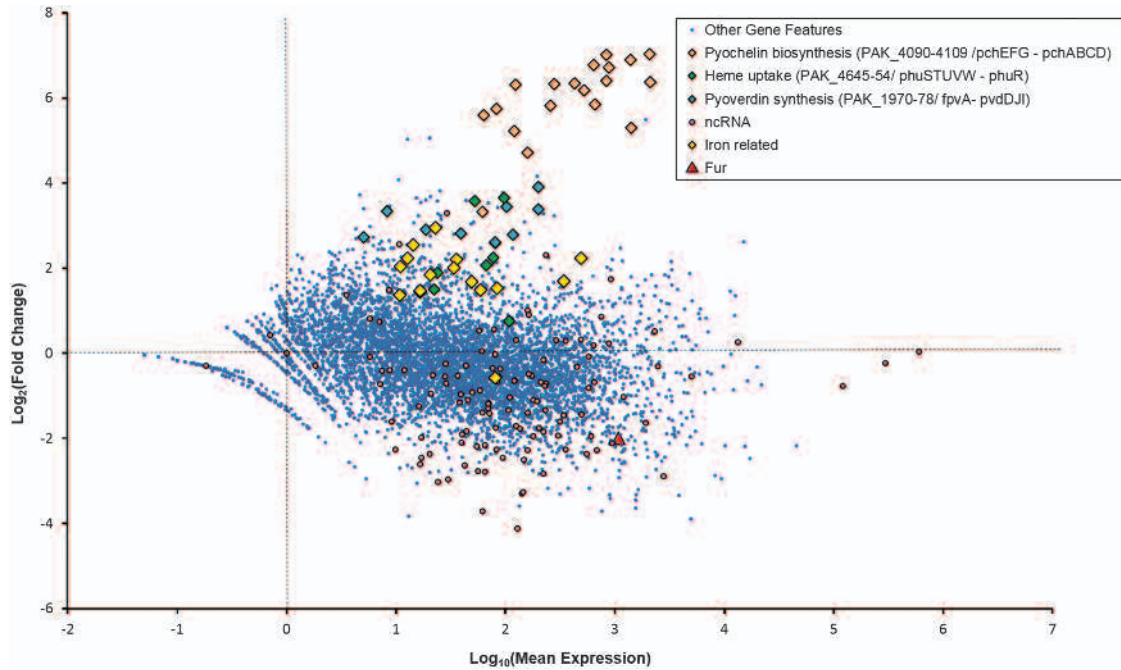


Figure 5 PAK_P4 specifically modulates transcript abundance of many host gene features. An MA plot depicting a differential expression analysis of host gene features demonstrates how the host transcriptional content changes from pre-infection exponential growth ($t=0$) to late infection ($t=13$ min). Notably, this analysis was performed after normalizing the host read counts to host read counts, ignoring the rapidly accumulating phage transcripts in a way that artificially enriches host transcripts in late infection relative to the total transcripts in the cell. This method allows specific effects on host transcription to be tested for independently of the global replacement of host RNAs by phage RNA.

of sufficient iron (Hantke, 2001), is downregulated (Figure 5; Supplementary Table 4).

As this upregulation of iron uptake-related genes may be the downstream consequence of a general activation of stress responses, we investigated the level of expression of alternative sigma factors which are acknowledged to orchestrate such responses. The only sigma factor that is meaningfully transcriptionally upregulated is *algU*, which may reflect an increase of AlgU-associated bacterial stress response. We subsequently looked at the expression of *algU* regulon that typically includes genes involved in alginate, peptidoglycan and LPS biosynthesis, and found that none of them were upregulated upon PAK_P4 infection (Supplementary Table 4). Overall, the differential expression of iron-related genes does not appear to be part of a more general stress response.

Altogether, these results indicate that, unlike PAK_P3, which only triggers the specific and meaningful upregulation of a single operon (Chevallereau *et al.*, 2016), PAK_P4 has specifically evolved processes leading to the complex manipulation of iron homeostasis in the host cell. We next investigate whether we could detect common host transcripts variations to both phage infections which would reflect a conserved host response against this group of viruses.

A common host response is elicited during infection by PAK_P3 and PAK_P4

We identified 55 host coding sequences upregulated in response to both PAK_P3 and PAK_P4 infection while

445 were downregulated (Supplementary Table 5). While the latter may mainly be the result of host mRNA degradation, the former could more likely represent a transcriptional general host response to phage infection, suggesting that some mechanisms for manipulating host transcription may have been conserved. Specifically, infections by both PAK_P3 and PAK_P4 prompt the marked transcriptional upregulation of a predicted P2-like prophage (Figure 6; Supplementary Table 5). However, like PAK_P3, PAK_P4 transcripts globally replace all host mRNAs faster than the increased prophage specific transcription can compete with (Chevallereau *et al.*, 2016). Moreover, one operon linked to RNA processing (*PAK_4493-4499*), which was previously found to be specifically upregulated during PAK_P3 infection, is also upregulated during PAK_P4 infection, albeit more weakly (mean $\text{Log}_2(\text{FC})=6.4$ vs mean $\text{Log}_2(\text{FC})=2.5$, respectively; Figure 6; Supplementary Table 5). The level of upregulation of this operon between both phages also inversely correlates with the extent and speed of RNA degradation. Additionally, we found that both infections also specifically deplete Hfq transcripts, similar to PhiK2 infected PAO1 cells (Ceyssens *et al.*, 2014; Figure 6 and Supplementary Table 5). Future experiments will address the biological consequences of these transcriptional modifications.

Discussion

While in cellular organisms, the increasing amount of genomic information is strengthening comparative

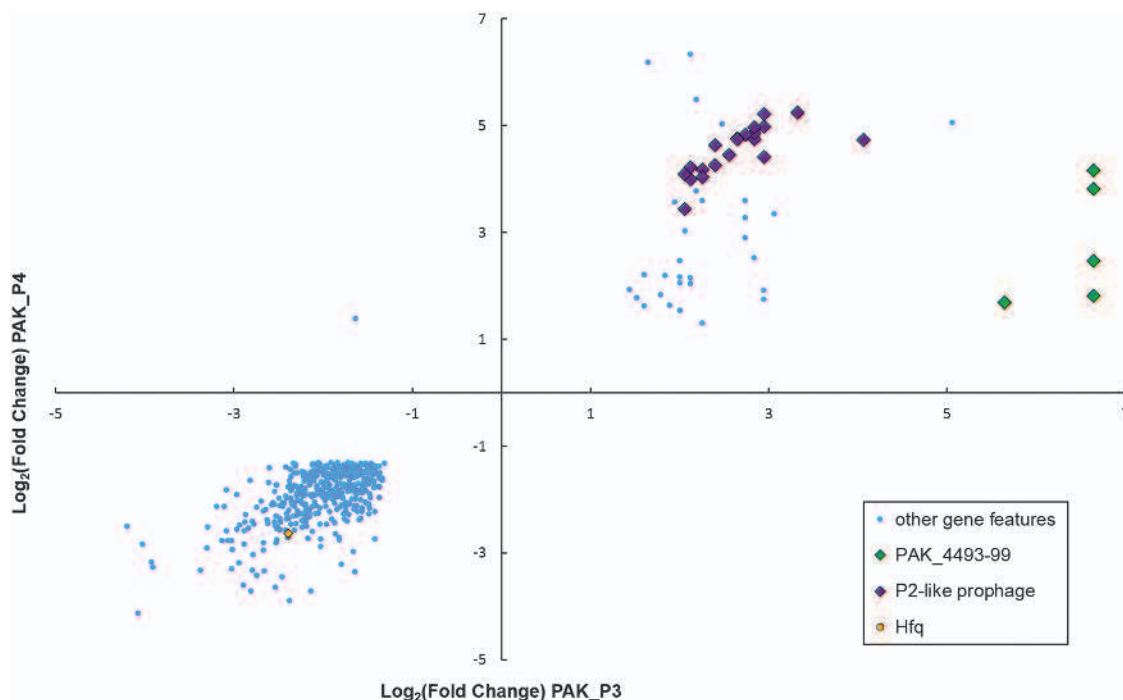


Figure 6 Host strain specifically upregulates the expression of one operon involved in RNA processing in response to both phage infections. Host genes significantly ($P < 0.05$) differentially expressed ($|\text{Log}_2(\text{FC})| > 1.3$) between $t = 0$ and $t = 13$ min upon both PAK_P3 and PAK_P4 infections were listed and their $\text{Log}_2(\text{FC})$ values upon each infection are plotted.

analyses delineating evolutionary relationships between cellular lineages, the lack of universal genes has hampered such analyses in viruses. In addition, the diversity of phage sequences is such that most genomes contain numerous ORFs displaying no homology to any other sequences in databanks. Therefore, reconstruction of viral evolution cannot be based solely on genomic sequences and requires further information such as elucidation of mechanistic processes (Forterre *et al.*, 2014). In this context, we have assessed whether two phages (PAK_P3 and PAK_P4), with divergent proteomic contents and infecting the same host strain, conserved the infectious strategy of their common ancestor during their evolution.

First, we observed that both phages share similar infection parameters although PAK_P4 displays a smaller burst size and a longer latent period than PAK_P3. This would not be predicted by the optimal lysis timing model which suggests that longer latency periods result in larger burst sizes (Wang *et al.*, 1996), and requires an alternative explanation. Comparison of expressions of putative endolysin-coding genes (*PAK_P4gp29* vs *PAK_P3gp28*), as well as the four downstream ORFs (holins and/or spanins candidates), revealed that these genes are more efficiently transcribed in PAK_P4, which also appears contradictory with the longer latency period. However, PAK_P4 encodes an asRNA (*PAK_P4as15*), not present in PAK_P3, matching the endolysin gene that could act as a translational repressor, preventing a rapid accumulation of endolysins and delaying lysis. With a supplementary level of lysis control,

it was anticipated that PAK_P4 should be able to produce higher number of progeny than PAK_P3. However, as it is not the case, we hypothesized that both phages may rely on distinct infection mechanisms, consistent with their largely divergent protein content, that could explain the observed differences in latency period and burst size. The consequences of these adjustments in the efficiency of each phage may seem of little importance but actually appear to have much broader consequences outside the test tube. This is exemplified by the distinct ability of PAK_P3 to more efficiently treat murine pneumonia compared with PAK_P4 (Henry *et al.*, 2013).

Therefore, we further explored phage molecular lifecycles and found that they both massively degrade host RNA with little manipulation of host gene expression, while relying on the host RNAP which is hijacked to efficiently transcribe phage gene features throughout infection. However: (i) the kinetics of RNA degradation are different and (ii) each phage co-opts the host transcription in a unique way. Remarkably, among host gene features differentially expressed in common between the two phage infections, the *PAK_4493-4499* operon encoding an RtcAB system was previously proposed to be upregulated by PAK_P3 to use the RNA 3'-phosphate cyclase RtcA to prime host transcripts for adenylation as part of a degradation pathway (Chevallereau *et al.*, 2016). However, upon PAK_P4 infection, the upregulation of this operon is much weaker while the degradation of host transcripts is more extensive. Thus, the upregulation of this operon could instead be a reflection of the host attempting to respond to

the extensive RNA damage imposed by both phages, which is more weakly active and thus more weakly protective in PAK_P4 infected cells, which may explain the different kinetics of RNA degradation upon both infections.

As a result, we anticipated that co-option of nucleotide metabolism must be central in PAK_P4 infection process, as it has been shown for other phages, including PAK_P3 (Chevallereau *et al.*, 2016; De Smet *et al.*, 2016). The strong and extensive increased transcription of iron acquisition genes upon PAK_P4 infection supports this hypothesis as increased concentrations of intracellular iron would supply cofactors for viral auxiliary metabolic enzymes, eventually favoring nucleotide turnover.

Alternatively, this increased expression of iron uptake genes could potentially be explained by the recent 'Ferrojan Horse hypothesis', proposing that phage tail fibers are bound to iron ions which allow them to use siderophores as receptors and consequently take advantage of bacterial iron uptake mechanisms (Bonnain *et al.*, 2016). However, no HxH motifs, known to facilitate iron binding, have been found in PAK_P4 putative tail fibers and, more generally, PAK_P4 does not possess more HxH motifs throughout its proteome than PAK_P3, which does not up-regulate iron-related genes (14 motifs versus 10, respectively).

In addition to studying direct phage-host interaction, we compared how these phages regulate their own gene expression, which appears to be broadly conserved and thus inherited from their common ancestor. Notably, recent work by Doron *et al.* has elegantly determined that the Syn9 phage produces a nearly identical transcriptional program when infecting three evolutionarily divergent hosts. In contrast, host response to phage infection was found to be host-specific (Doron *et al.*, 2016). Similar results were obtained by Howard-Varona *et al.* using the ϕ 38:1 phage on two *Cellulophaga baltica* hosts (Howard-Varona *et al.*, 2016) indicating that neither of these two phages adjust their transcriptional program to different hosts. We have now added another piece to the puzzle by conversely studying the infection processes of two divergent phages on one common host.

We found that both phages display similar temporal regulatory schemes of their syntenic gene sets. In addition, this regulation appears to partially rely on the production of antisense transcripts, indicating that this mechanism probably already existed in the common ancestor. However, beside the genes unique to each phage (expected to specifically modulate each phage infection), we noted that differences in the transcriptional program of conserved genes (core genes) also impact infection cycles, as exemplified by the endolysin. These observations highlight the extensive panel of solutions that phages have developed to orchestrate bacterial infections. In addition, such transcriptomics studies revealing infection strategies could help implementing phage

taxonomy. Indeed, including genome transcription patterns add a novel layer of properties that could be considered to cluster and differentiate phage species and genera.

In summary, even though they have almost identical morphologies and are able to infect the same host, both phages display substantial variability of their proteomic content, especially within early proteins. Thus, it was unclear whether such related phages would share common mechanisms of action. We showed that the global subversion scheme underlying infection cycle is conserved between phages deriving from a common ancestor although with distinct evolutionary trajectories. We hypothesize that this common mechanism is most likely mediated by the set of the 13 homologous phage early proteins. More intuitive is the observation that both phages have evolved unique processes to infect one particular host, probably explaining why they display distinct killing efficiency. These processes may also indicate that both phages adapted to different ecological niches (for example, low iron environment for PAK_P4).

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

This research was supported by the Geconcerteerde Onderzoeks Actie grant 'Phage Biosystems' from the KU Leuven. (<http://www.kuleuven.be/onderzoek/kernprojecten/goa.htm>). BB has a PhD scholarship within the framework of an Onderzoeks Toelage grant of the KU Leuven. AC was supported by a PhD fellowship from the Ministère de l'Enseignement Supérieur et de la Recherche (ED N°516 B3MI/ ED N°157 BioSPC Paris Diderot Université). RL and LD are members of the « PhageBiotics research community », supported by the FWO Vlaanderen. We deeply thank Damien Mornico (Hub Bioinformatique et Biostatistique, Institut Pasteur – C3BI, USR 3756 IP CNRS) for his invaluable help to successfully update the annotations of strain PAK.

References

- Bonnain C, Breitbart M, Buck KN. (2016). The Ferrojan horse hypothesis: iron-virus interactions in the ocean. *Front Mar Sci* **3**: 82.
- Calendar R. (2006). *The Bacteriophages*. 2nd edn. Oxford University Press: Oxford; New York.
- Ceyssens PJ, Minakhin L, Van den Bossche A, Yakunina M, Klimuk E, Blasdel B *et al.* (2014). Development of giant bacteriophage ϕ KZ is independent of the host transcription apparatus. *J Virol* **88**: 10501–10510.
- Chevallereau A, Blasdel BG, 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.
- De Smet J, Zimmermann M, Kogadeeva M, Ceyssens PJ, Vermaelen W, Blasdel B *et al.* (2016). High coverage metabolomics analysis reveals phage-specific alterations to *Pseudomonas aeruginosa* physiology during infection. *ISME J* **10**: 1823–1835.
- Doron S, Fedida A, Hernandez-Prieto MA, Sabehi G, Karunker I, Stazic D *et al.* (2016). Transcriptome dynamics of a broad host-range cyanophage and its hosts. *ISME J* **10**: 1437–1455.
- Essouh C, Latino L, Midoux C, Blouin Y, Loukou G, Nguetta SP *et al.* (2015). Investigation of a large collection of *Pseudomonas aeruginosa* bacteriophages collected from a single environmental source in Abidjan, Cote d'Ivoire. *PLoS One* **10**: e0130548.
- Forterre P, Krupovic M, Prangishvili D. (2014). Cellular domains and viral lineages. *Trends Microbiol* **22**: 554–558.
- Grose JH, Casjens SR. (2014). Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family Enterobacteriaceae. *Virology* **468–470**: 421–443.
- Hantke K. (2001). Iron and metal regulation in bacteria. *Curr Opin Microbiol* **4**: 172–177.
- Hatfull GF, Pedulla ML, Jacobs-Sera D, Cichon PM, Foley A, Ford ME *et al.* (2006). Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform. *PLoS Genet* **2**: e92.
- Henry M, Lavigne R, Debarbieux L. (2013). Predicting in vivo efficacy of therapeutic bacteriophages used to treat pulmonary infections. *Antimicrob Agents Chemother* **57**: 5961–5968.
- Henry M, Bobay LM, Chevallereau A, Saussereau E, Ceyssens PJ, Debarbieux L. (2015). The search for therapeutic bacteriophages uncovers one new subfamily and two new genera of *Pseudomonas*-infecting Myoviridae. *PLoS One* **10**: e0117163.
- Howard-Varona C, Roux S, Dore H, Solonenko NE, Holmfeldt K, Markillie LM *et al.* (2016). Regulation of infection efficiency in a globally abundant marine Bacteriodes virus. *ISME J* **11**: 284–295.
- Koskella B, Brockhurst MA. (2014). Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol Rev* **38**: 916–931.
- Kropinski AM, Sulakvelidze A, Konczy P, Poppe C. (2007). Salmonella phages and prophages—genomics and practical aspects. *Methods Mol Biol* **394**: 133–175.
- Krupovic M, Dutilh BE, Adriaenssens EM, Wittmann J, Vogensen FK, Sullivan MB *et al.* (2016). Taxonomy of prokaryotic viruses: update from the ICTV bacterial and archaeal viruses subcommittee. *Arch Virol* **161**: 1095–1099.
- Kwan T, Liu J, DuBow M, Gros P, Pelletier J. (2005). The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. *Proc Natl Acad Sci USA* **102**: 5174–5179.
- Kwan T, Liu J, Dubow M, Gros P, Pelletier J. (2006). Comparative genomic analysis of 18 *Pseudomonas aeruginosa* bacteriophages. *J Bacteriol* **188**: 1184–1187.
- Love MI, Huber W, Anders S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.
- Mahadevan P, King JF, Seto D. (2009). CGUG: in silico proteome and genome parsing tool for the determination of 'core' and unique genes in the analysis of genomes up to ca. 1.9 Mb. *BMC Res Notes* **2**: 168.
- Marinelli LJ, Fitz-Gibbon S, Hayes C, Bowman C, Inkeles M, Loncaric A *et al.* (2012). Propionibacterium acnes bacteriophages display limited genetic diversity and broad killing activity against bacterial skin isolates. *MBio* **3**: e00279–12.
- Marston MF, Pierciey FJ Jr, Shepard A, Gearin G, Qi J, Yandava C *et al.* (2012). Rapid diversification of coevolving marine Synechococcus and a virus. *Proc Natl Acad Sci USA* **109**: 4544–4549.
- Morello E, Saussereau E, Maura D, Huerre M, Touqui L, Debarbieux L. (2011). Pulmonary bacteriophage therapy on *Pseudomonas aeruginosa* cystic fibrosis strains: first steps towards treatment and prevention. *PLoS One* **6**: e16963.
- Ochsner UA, Johnson Z, Vasil ML. (2000). Genetics and regulation of two distinct haem-uptake systems, phu and has, in *Pseudomonas aeruginosa*. *Microbiology* **146**: 185–198.
- Pope WH, Bowman CA, Russell DA, Jacobs-Sera D, Asai DJ, Cresawn SG *et al.* (2015). Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. *Elife* **4**: e06416.
- Roux S, Hallam SJ, Woyke T, Sullivan MB. (2015). Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. *Elife* **4**: e08490.
- Saussereau E, Vachier I, Chiron R, Godbert B, Sermet I, Dufour N *et al.* (2014). Effectiveness of bacteriophages in the sputum of cystic fibrosis patients. *Clin Microbiol Infect* **20**: O983–O990.
- Takeya K, Amako K. (1966). A rod-shaped *Pseudomonas* phage. *Virology* **28**: 163–165.
- Wang I-N, Dykhuizen DE, Slobodkin LB. (1996). The evolution of phage lysis timing. *Evolut Ecol* **10**: 545–558.
- Williams HT. (2013). Phage-induced diversification improves host evolvability. *BMC Evol Biol* **13**: 17.
- Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA, Brinkman FS. (2016). Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic Acids Res* **44**: D646–D653.



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>

© The Author(s) 2017

Supplementary Information accompanies this paper on The ISME Journal website (<http://www.nature.com/ismej>)

POPULARISATION OF SCIENCE

—

**LES BACTERIOPHAGES: COMMENT CES VIRUS ALLIES
FONCTIONNENT-ILS?**

Published in Biofutur

Les bactériophages : comment ces virus alliés fonctionnent-ils ?

Les bactériophages sont des virus qui infectent spécifiquement les bactéries. Le dicton « *les ennemis de nos ennemis sont nos amis* » pourrait dès lors s'appliquer à l'utilisation thérapeutique des bactériophages pour combattre les bactéries résistantes aux antibiotiques. En ligne de mire, l'éradication de maladies bactériennes à l'origine d'épidémies ou orphelines de traitement.

© PHOTOTAKE/KUNKEL/BSIP

Les bactériophages sont omniprésents dans les écosystèmes microbiens, des plus vastes – les océans – aux plus réduits – le tube digestif d'un insecte microscopique, par exemple. Nous-mêmes en hébergeons des quantités impressionnantes, aussi bien à la surface de notre peau que dans notre appareil digestif. Les bactériophages sont ainsi les entités les plus abondantes et les plus diversifiées sur la Terre. La quantité estimée de 10^{30} bactériophages est un chiffre qui donne le vertige lorsqu'on le compare à celui de toutes les cellules constituant les quelque 7 milliards d'humains, soit « seulement » 10^{22} cellules ! En termes de diversité, plus difficile à appréhender, on estime qu'il y aurait au moins 100 millions de génomes de bactériophages dans la nature (1). Cette diversité est à mettre en parallèle avec la très grande spécificité des bactériophages, qui font d'eux des virus capables d'infecter seulement un nombre restreint de bactéries au sein d'une même espèce, à quelques exceptions près. À titre de comparaison, le spectre d'hôte

d'un bactériophage est beaucoup plus étroit que celui d'un antibiotique à spectre étroit comme la pénicilline. Cette étroitesse de spectre bactérien peut, dans une approche thérapeutique, être compensée par l'association de plusieurs bactériophages en cocktail, chacun possédant des spectres complémentaires et non redondants.

CLASSIFICATION

Suite aux premières observations microscopiques effectuées dans les années 1940, la classification des bactériophages s'est d'abord basée sur la morphologie des virus puis sur le type d'acide nucléique qu'ils renferment (ADN ou ARN, simple ou double brin) et la présence ou non d'une enveloppe (figure p. 32). Il est à noter que 96 % des bactériophages appartiennent à l'ordre des *Caudovirales*, des virus à ADN double brin non enveloppés. D'une taille variable – de 25 à 800 nanomètres sur l'axe tête-queue –, leur organisation générale comporte : une tête (capside) à symétrie icosa-

édrique, à l'intérieur de laquelle est protégé le génome – taille variant de 17 à 498 kilobases – ; une queue de longueur variable – réduite chez les *Podoviridae*, flexible chez les *Siphoviridae* et contractile chez les *Myoviridae* –, issue de la polymérisation de protéines. Elle est organisée de façon tubulaire ce qui permet le transit de l'ADN viral lors de son éjection vers le cytoplasme de l'hôte infecté ; enfin, des fibres de queue, fixées à une plate-forme basale, permettent aux bactériophages d'interagir avec des récepteurs spécifiques présents à la surface de l'hôte qui reconnaissent précisément le type de bactérie dont ils sont les prédateurs.

Les principaux bactériophages modèles, dont l'étude a permis l'éclosion et l'essor de la biologie moléculaire, appartiennent à l'ordre des *Caudovirales* (figure p. 33). Les *Podoviridae*, dont le bactériophage T7 est le représentant, sont les moins représentés au sein des *Caudovirales* (10 %). Les *Myoviridae* représentent environ 25 % des phages et l'on

Les auteurs

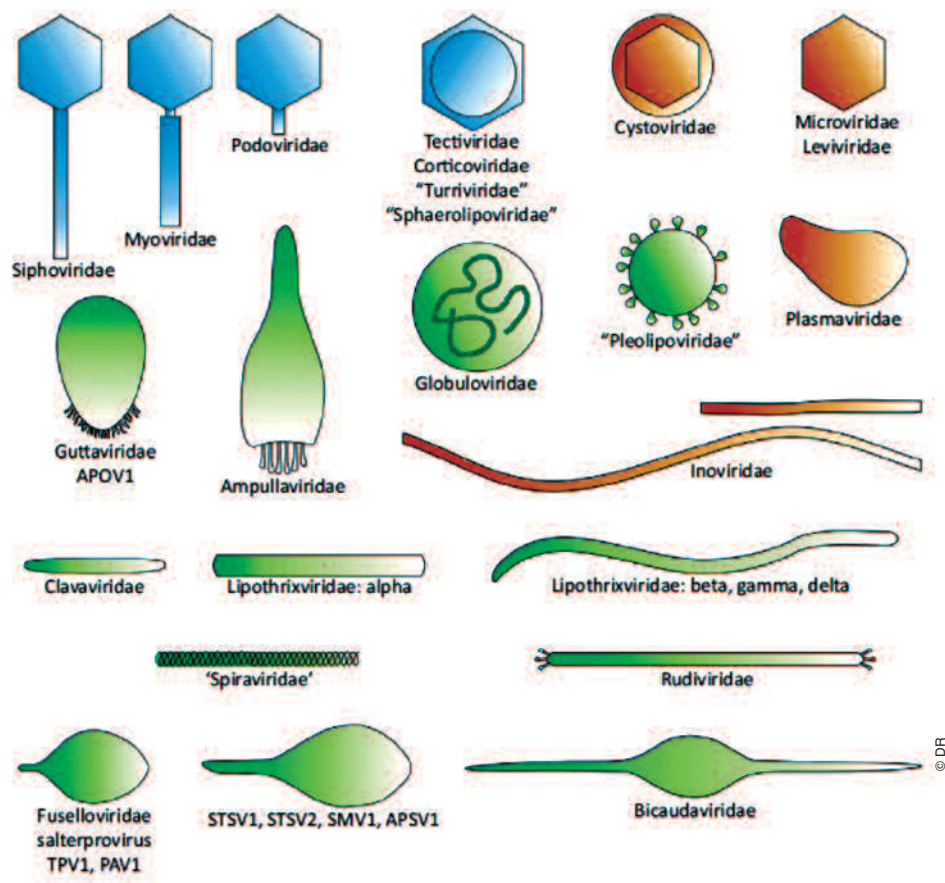
Nicolas Dufour^{1,2},
Anne Chevallereau^{1,3}
et Laurent Debarbieux¹

¹ Institut Pasteur,
Unité de Biologie moléculaire du
Gène chez les extrémophiles,
Département de microbiologie,
Paris

² AP-HP,
Hôpital Louis Mourier,
Service de Réanimation
médico-chirurgicale,
Colombes

³ Université Paris Diderot,
Sorbonne Paris Cité,
Cellule Pasteur,
Paris

Morphotypes des virus infectant les bactéries seulement (en bleu), les bactéries et les archées (marron) ou les archées seules (vert). Ces représentations n'ont pas été réalisées à l'échelle. D'après (2).



On retrouve le bactériophage T4 qui est sans aucun doute le bactériophage le plus étudié. Quant aux *Siphoviridae*, ils constituent la famille la plus représentée (65 %) avec le bactériophage lambda (λ) comme fer de lance.

LES CYCLES INFECTIEUX

Les bactériophages utilisent différents types de cycles infectieux, dépendants des informations codées dans leurs génomes et de l'état métabolique des cellules infectées. On distingue quatre types de cycles. Deux d'entre eux occupent le devant de la scène : les cycles lytique et lysogénique. Plus rarement – donc moins étudiés –, les bactériophages utilisent un cycle pseudolyso-génique ou un cycle infectieux chronique (figure p. 34) (4).

La première étape, commune à l'ensemble de ces cycles, comprend l'adsorption du bactériophage à la surface de la cellule hôte via un ou plusieurs récepteurs spécifiques (extrémité polysaccharidique du lipopolysaccharide, protéines membranaires, capsules...), suivie de l'éjection de l'acide nucléique viral (ADN ou ARN) dans le cytoplasme.

LE CYCLE VIRULENT, GOLD STANDARD POUR LA PHAGOTHÉRAPIE

Dès l'introduction du matériel génétique viral, les gènes qualifiés de « précoces » sont les premiers transcrits et traduits, le plus souvent par la machinerie de l'hôte. Ils codent des protéines dont les fonctions principales consistent à détourner le métabolisme cellulaire au profit de la production coordonnée de nouveaux virions tout en déjouant les systèmes de défense bactériens. Par exemple, les protéines précoces du bactériophage T7 comprennent, entre autres, des protéines inhibant les enzymes de restriction et la machinerie transcriptionnelle bactériennes, des nucléases destinées à fragmenter le chromosome bactérien ou encore une ARN- et une ADN-polymérase accordant au virus plus d'autonomie vis-à-vis de la machinerie cellulaire. Une fois ce piratage finement orchestré dans le but de transformer la bactérie hôte en véritable usine productrice de virus, le génome du bactériophage est répliqué en de multiples copies. Vient ensuite une phase de synthèse de grandes quantités de protéines structurales (protéines « tardives ») qui s'auto-

assemblent et encapsident les nouvelles copies du génome, aboutissant finalement à la formation de virions matures. La dernière étape consiste à provoquer, à un moment bien précis, la lyse active de la bactérie, de façon à libérer le maximum de virions assemblés.

Ce schéma général du cycle lytique s'accompagne de très nombreuses subtilités puisque chaque famille de bactériophages adopte sa propre stratégie de piratage. Des études globales des transcriptomes, des métabolomes et des protéomes commencent tout juste à se développer et permettront d'avoir une meilleure vision des mécanismes mis en place par les bactériophages pour prendre le contrôle de leurs hôtes. Ce cycle virulent est sans nul doute le plus intéressant en termes d'application thérapeutique.

LA LYSOGÉNIE, FORCE D'ÉVOLUTION BACTÉRIENNE

Par opposition aux bactériophages virulents, les bactériophages adoptant un cycle lysogénique sont dits « tempérés ». Cette distinction repose sur une capacité que seuls ces bactériophages

possèdent : l'intégration de leur génome au sein du chromosome bactérien. Sous l'influence de différents facteurs, notamment l'état métabolique de la bactérie, un bactériophage tempéré peut s'engager soit dans un cycle lytique, soit dans un cycle lysogénique. Une décision qui s'effectue juste après l'éjection du matériel génétique dans le cytoplasme et qui nécessite un jeu d'enzymes spécifiques que ne possèdent pas les bactériophages virulents. Une fois intégré dans le chromosome bactérien, au niveau de sites spécifiques, le génome viral est appelé « prophage » et la bactérie est dite « lysogène ». Immobilisé dans le chromosome bactérien, le génome viral est alors dupliqué au cours de chaque division cellulaire et est ainsi transmis à la descendance. La transcription, à bas bruit, de gènes viraux codant des protéines régulatrices permet au virus de percevoir les modifications physiologiques de l'hôte et, le cas échéant, déclenche l'excision du prophage (phénomène d'induction). Le virus engage alors un cycle infectieux lytique aboutissant à la destruction de la cellule hôte.

Une bactérie lysogène ne peut pas être nouvellement infectée par le bactériophage qu'elle contient (immunité) mais peut l'être par d'autres bactériophages tempérés et ainsi devenir une bactérie polylysogène. Au cours de son évolution, la bactérie peut garder un prophage de manière définitive lorsque celui-ci perd les protéines nécessaires à son excision, par mutations par exemple. Répétée à l'envi, cette situation aboutit à affecter durablement les

génomés bactériens à un point tel qu'il est estimé que les éléments prophagiques peuvent constituer jusqu'à 10 % du chromosome de certaines souches bactériennes.

Les prophages expriment aussi des « morons » – de l'anglais *more*, plus –, des gènes apportant une fonction supplémentaire à la bactérie mais qui n'ont pas de rôle dans l'infection virale. De cette façon, les prophages confèrent souvent un avantage aux souches qui les hébergent, contrepartie nécessaire pour assurer leur maintien dans les populations bactériennes. Nombreux sont les exemples en infectiologie où la pathogénicité des souches bactériennes est dépendante d'un facteur de virulence codé par un prophage (shigatoxine, toxine cholérique ou diphtérique, leucocidine de Panton-Valentine...).

LES CYCLES PSEUDOLYSOGÉNIQUE ET CHRONIQUE

La pseudolysogénie peut être interprétée comme un état de non-choix entre un cycle lytique et un cycle lysogénique (5). Après son injection, le matériel génétique viral reste quiescent sous forme extrachromosomique dans la cellule hôte, à la manière d'un plasmide. Sa transmission à la descendance est en général asymétrique car il n'est pas répliqué : seule une des cellules filles en hérite. Souvent observée lors de conditions de croissance défavorables pour l'hôte, un manque de nutriments ou un stress chimique par exemple, la pseudolysogénie peut s'étendre sur plusieurs géné-

rations ou cesser, conduisant soit à l'immobilisation par intégration dans le chromosome bactérien, soit à la dispersion via un cycle lytique.

Employé par certains bactériophages, dont le célèbre bactériophage filamenteux M13, très utilisé en biotechnologie, dans la technique de *phage display** notamment, le cycle d'infection chronique est, quant à lui, un cycle lytique sans lyse. Les particules virales produites à faible débit sont, en effet, excrétées par bourgeonnement ou extrusion et les cellules infectées continuent de se diviser, transmettant le virus à la descendance.

DES VÉHICULES GÉNÉTIQUES MOTEURS DE L'ÉVOLUTION

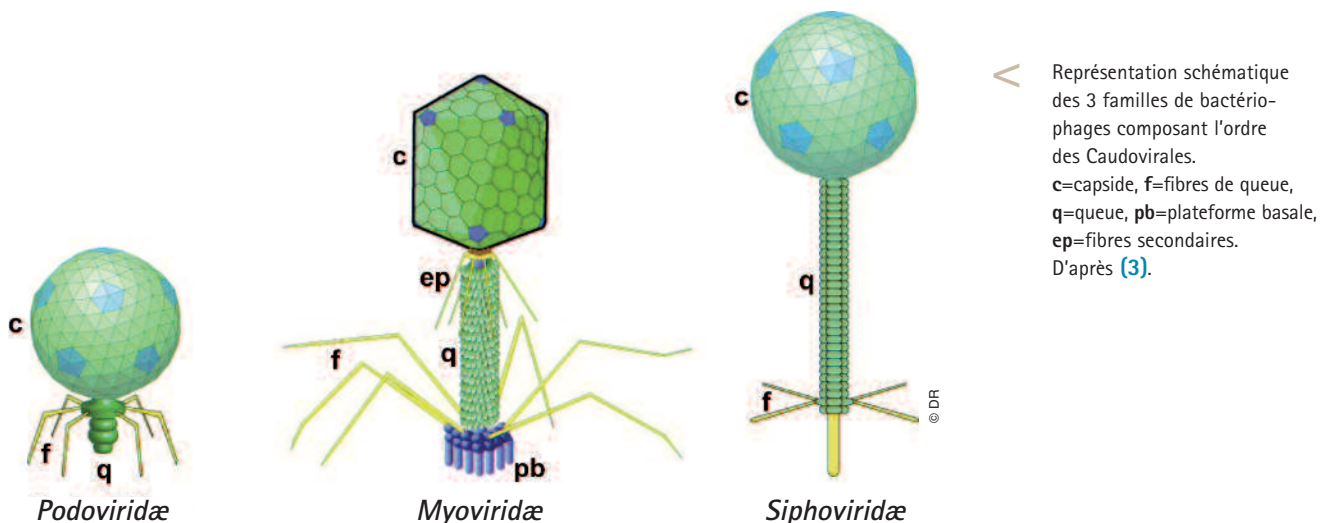
Lors de l'étape d'encapsulation, les bactériophages peuvent commettre des erreurs et emprisonner du matériel génétique bactérien en plus ou à la place du matériel génétique viral. Ils servent alors de véhicule à ce matériel et assurent son transfert vers d'autres bactéries dans un processus appelé transduction. Celle-ci peut être généralisée ou spécialisée.

La transduction d'éléments génétiques codant, par exemple, des fonctions en relation avec la virulence ou la résistance aux antibiotiques procure un avantage

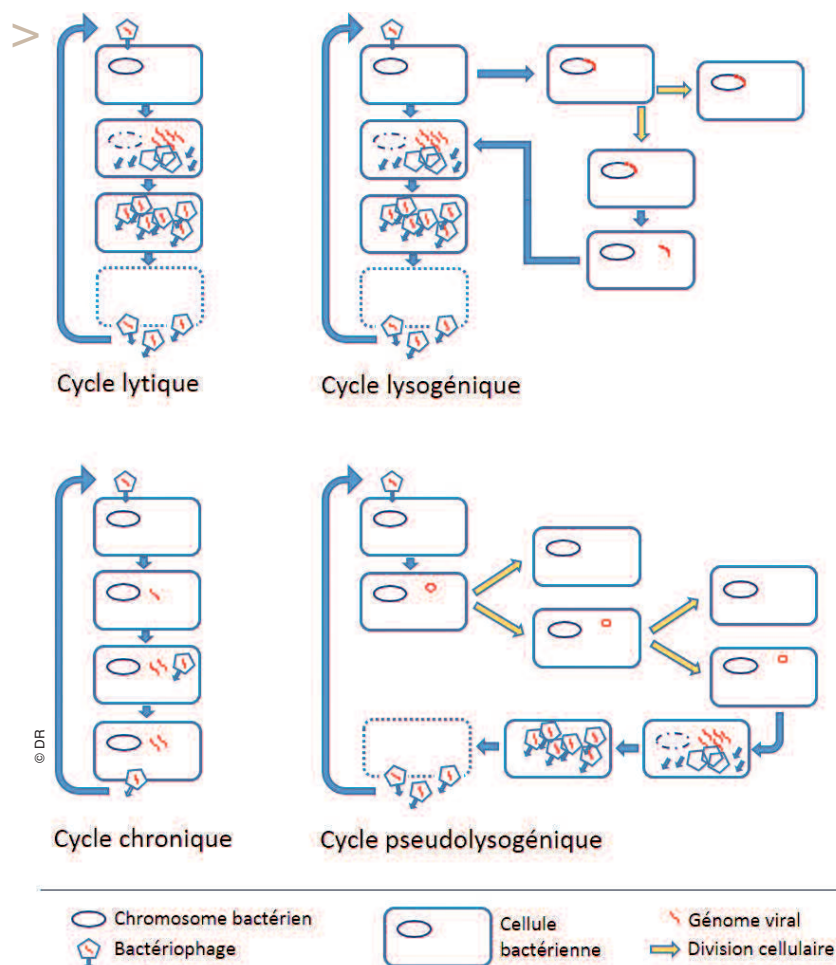
+ Pour en savoir plus

tinyurl.com/life-in-our-phage-world
Morello E, Debarbieux L
(2010) *Biofutur* 309, 33-6

* Technique reposant sur l'expression d'un répertoire protéique à la surface de bactériophages à la surface de bactériophages, sélectionnés pour leur capacité de liaison à une cible (anticorps, enzyme, récepteur, acide nucléique...).



Schémas des différents cycles infectieux des bactériophages. Une description détaillée de chacun des cycles est donnée dans le texte.



sélectif à la bactérie qui les reçoit. Elle contribue également au brassage génétique entre virus et bactéries, participant ainsi à l'évolution bactérienne.

On parle de « transduction généralisée » quand l'encapsulation erronée d'un fragment d'ADN bactérien – chromosomique ou plasmidique – provenant d'une première bactérie hôte n'entraîne pas le développement d'un cycle infectieux lorsqu'il atteint une deuxième bactérie. Une fois injecté, l'ADN bactérien étranger doit s'intégrer, par des mécanismes de recombinaison, au chromosome de la nouvelle bactérie infectée pour être conservé et transmis. Il peut également rester libre dans le cytoplasme et être perdu au fur et à mesure des divisions cellulaires.

La transduction spécialisée ne concerne que les bactériophages dotés d'un cycle lysogénique. Lors de l'excision du prophage, des erreurs peuvent survenir à faible fréquence – en moyenne une erreur par million d'excisions pour le bactériophage λ – et aboutir à l'excision d'une molécule d'ADN

hybride, contenant un fragment d'ADN viral et un fragment d'ADN bactérien adjacent à la zone d'intégration du prophage. S'il est fonctionnel sur le plan de l'adsorption et de l'éjection de son matériel génétique, le virus contenant ce matériel hybride est rarement en mesure d'aboutir à la formation de nouveaux virus autonomes car souvent privés de l'intégralité de leur génome.

UN RÉSERVOIR ILLIMITÉ DE FONCTIONS POTENTIELLEMENT THÉRAPEUTIQUES

Bien qu'étudiés depuis plus d'un siècle, les bactériophages recèlent encore un nombre très important de secrets quant à leurs modalités précises de fonctionnement, dont les détails sont vraisemblablement très variables d'un virus à l'autre. Grâce aux nouvelles méthodes d'analyse (études transcriptomiques, protéomiques, microfluidiques...) et à la diminution des coûts de séquençage à haut débit,

il devient possible d'accéder aux mécanismes intimes développés par les bactériophages pour pirater le plus efficacement possible leurs hôtes et contrer les mécanismes de résistances mis en place par ces derniers. Ces approches permettent notamment d'identifier des molécules virales exerçant une activité antibactérienne, à l'instar des antibiotiques. Compte tenu de leur immense diversité et du très grand nombre de gènes dont les fonctions sont actuellement inconnues, les bactériophages constituent une source très prometteuse de nouvelles découvertes ayant des applications thérapeutiques. Parmi les molécules phagiques séduisantes, les lysines sont des enzymes à fort potentiel pour lesquelles les premiers essais thérapeutiques chez l'homme sont en cours. Produites par les bactériophages lors de la phase finale du cycle infectieux, elles provoquent la lyse de la bactérie. Elles sont chargées d'hydrolyser le peptidoglycane, un composant majeur de la paroi bactérienne – notamment des bactéries à Gram+ –, et peuvent ainsi lyser en quelques secondes une culture de *Staphylococcus aureus*.

Avec les progrès de la biologie moléculaire et de la bio-ingénierie, les bactériophages peuvent également être détournés de leur fonction initiale et servir, par exemple, de sonde de détection des organismes pathogènes tels *Listeria monocytogenes* dans l'industrie agro-alimentaire (la société américaine Sample6 commercialise un tel outil diagnostique, supérieure aux outils conventionnels). De même, ils peuvent assurer le ciblage de cellules tumorales et délivrer *in situ* une molécule anti-tumorale, constituant ainsi des nanovecteurs thérapeutiques.

Plus simplement, l'utilisation directe des bactériophages représente actuellement l'espoir le plus prometteur en pathologie infectieuse. Effet conjugué de la progression inéluctable de l'antibiorésistance et de la pénurie de nouveaux antibiotiques, les rapports de force avec les micro-organismes pathogènes communs ont changé. Les bactériophages, de par leur diversité et leur mode d'action spécifique, sont sans aucun doute des partenaires qu'il faut dès à présent associer dans nos développements thérapeutiques. ■

(1) Rohwer F (2003) *Cell* 113, 141

(2) Pietila MK et al. (2014) *Trend Microbiol* 22, 334-44

(3) Hulo C et al. (2011) *Nucleic Acids Res* 39, D576-82

(4) Salmond GP, Fineran PC (2015) *Nat Rev Microbiol* 13, 777-86

(5) Los M, Wegrzyn G (2012) *Adv Virus Res* 82, 339-49

