PLOS | ONE

RESEARCH ARTICLE

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

Marine Henry[1], Louis-Marie Bobay[2,3,4], Anne Chevallereau[1,5], Emilie Saussereau[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

CrossMark
click for updates

OPEN ACCESS

Citation: Henry M, Bobay L-M, Chevallereau A, Saussereau 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, 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

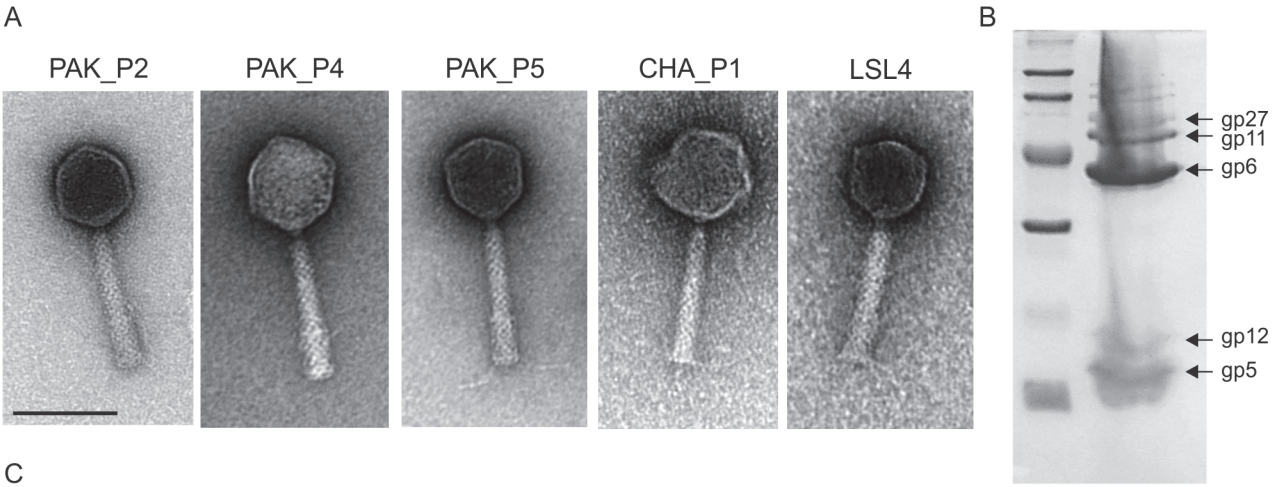
Isolation of six new bacteriophages infecting *P. aeruginosa*

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



C

Gp	MW(kDa)	Nb Pept.	Nb Spec.	Nb Spec./Mw	Sequence coverage (%)	Predicted function
2	56.3	21	199	3.5	39.6	Portal protein
4	31.8	9	34	1.1	29.8	Unknown
5	13.5	10	265	19.6	76.6	Decoration protein
6	40.2	20	527	13.1	61.3	Major capsid protein
7	18.6	6	91	4.9	34.8	Unknown
8	17.5	2	2	0.1	22.6	Unknown
9	14.3	3	11	0.8	41.5	Head-tail adaptor
10	23.4	3	5	0.2	20.9	Unknown
11	47.9	14	220	4.6	35	Tail sheath protein
12	18.8	6	107	5.7	52.6	Structural protein
13	18.3	5	6	0.3	41.3	Unknown
14	18.7	5	42	2.2	36.5	Unknown
17	84.8	16	40	0.5	24	Tape measure protein
18	27.9	2	10	0.4	9	Unknown
19	14.6	3	10	0.7	26.4	Unknown
20	32.6	9	33	1.0	44	Tail protein
21	27.2	3	16	0.6	23.8	Baseplate assembly protein
22	14.1	2	6	0.4	20.7	Tail-associated lysozyme
23	53.3	12	85	1.6	35	Baseplate protein
24	28.1	3	13	0.5	16.1	Unknown
25	70.2	18	88	1.3	44	Fiber protein
27	52.9	14	229	4.3	46.5	Fiber protein
72	8.1	2	2	0.2	30.3	Unknown
160	20.7	3	7	0.3	18.1	Unknown
166	13.8	6	101	7.3	67.4	Fiber protein

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-N_{17}-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

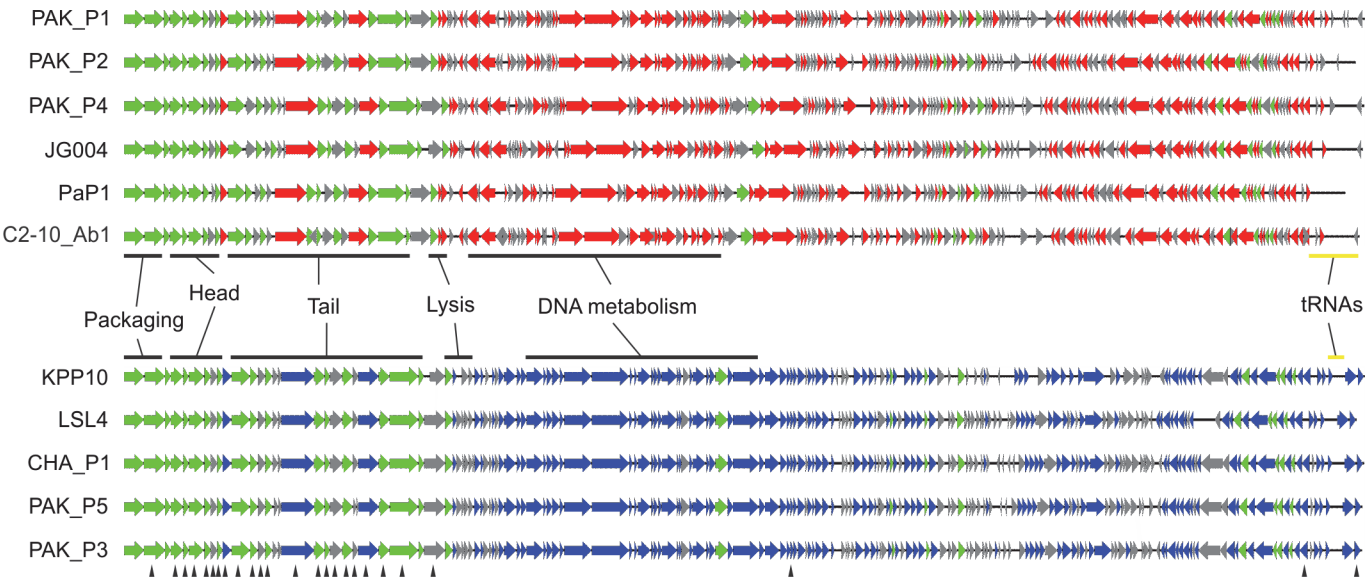


Figure 2. Maps of the 11 genomes involved in this study. In green, the core ORFs homologous in the 11 bacteriophages; red and blue ORFs are common to the PAK_P1-like and KPP10-like bacteriophages, respectively. Arrows (▲) designate the sequences corresponding to proteins identified by mass spectrometry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

doi:10.1371/journal.pone.0117163.g002

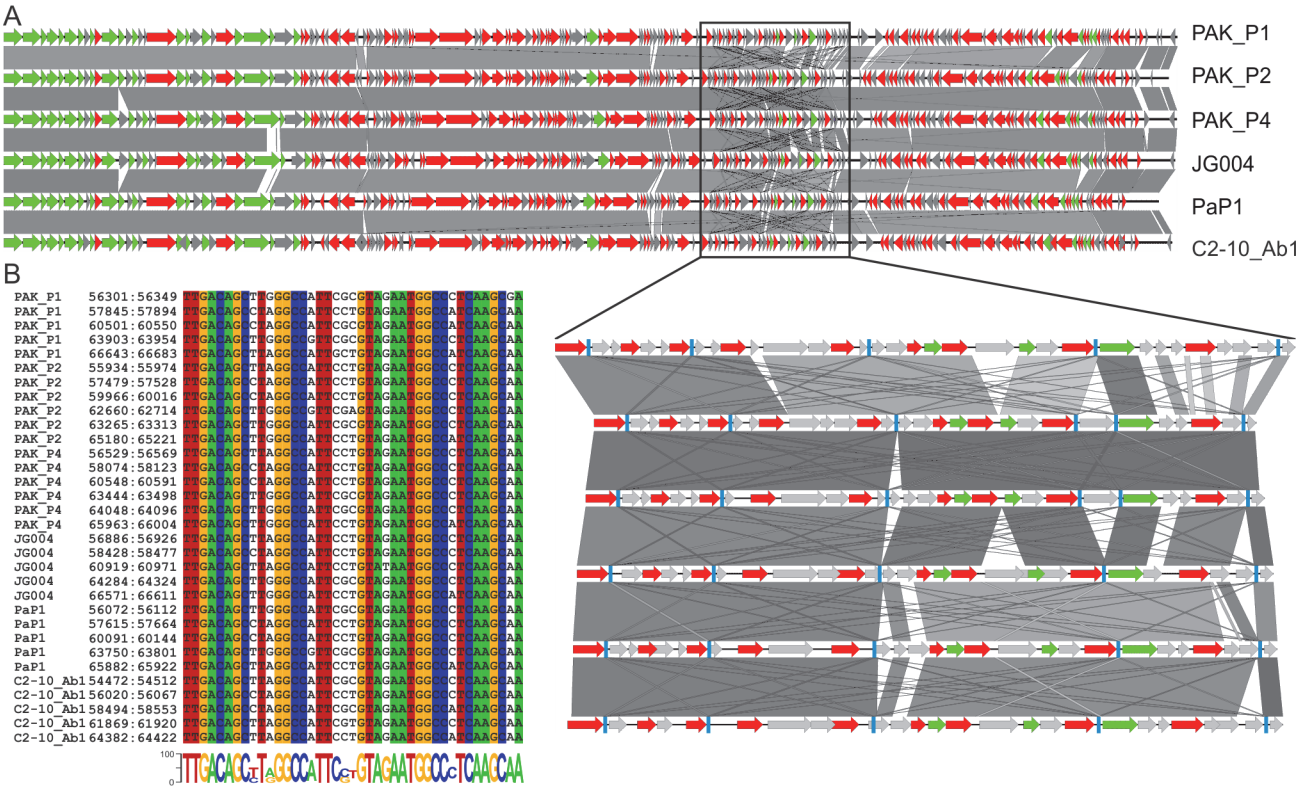


Figure 3. Identification of repeated promoter-like regions of the PAK_P1-like clade. A. Representation of the homology (blastn) between the six PAK_P1-like bacteriophages, with an enlargement of the region in which 41-nt repeats were identified (represented as light blue boxes). B. Alignment of the total 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 $1x10^7$ 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

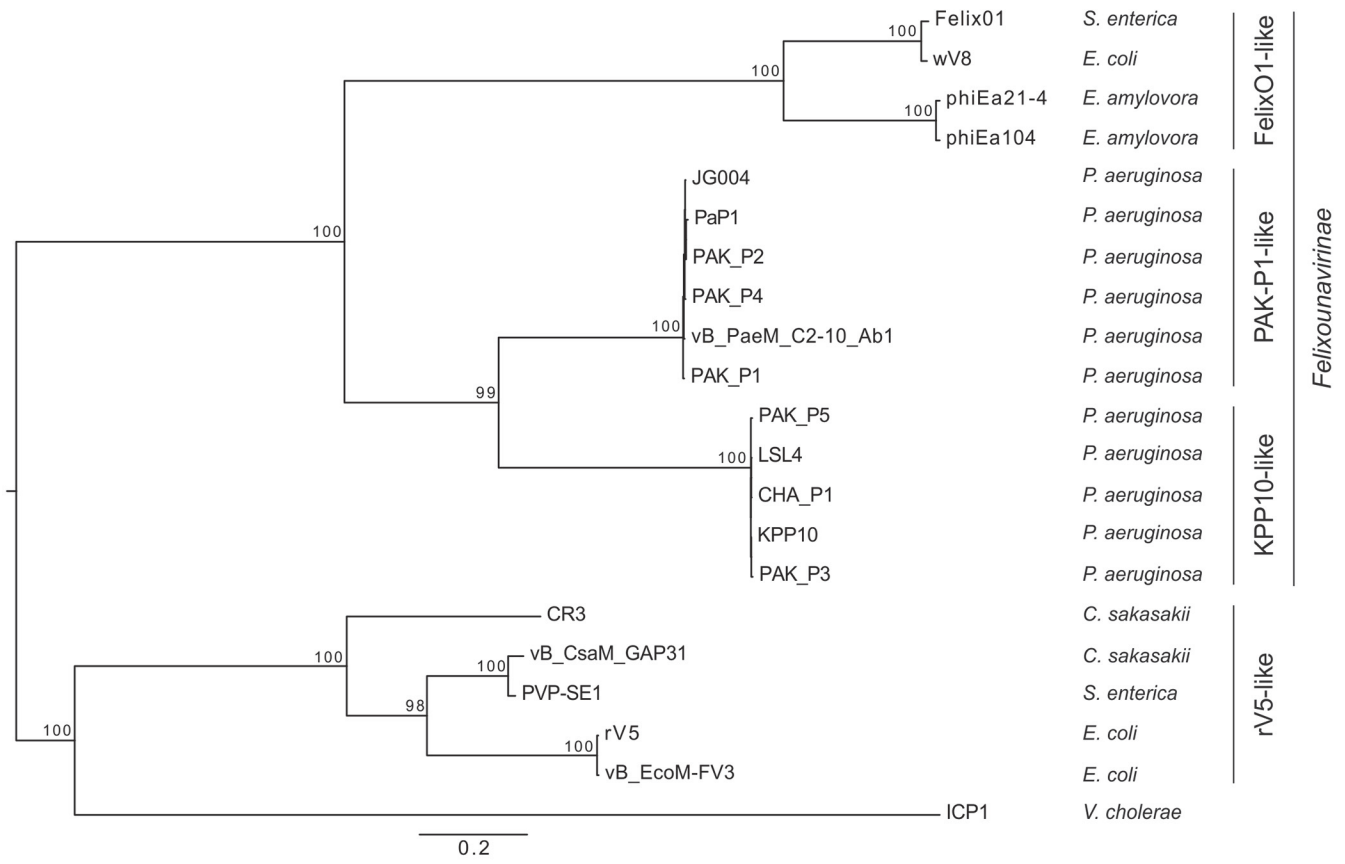


Figure 4. Phylogenetic tree of PAK_P1-like and KPP10-like bacteriophages and their closest relatives. The maximum likelihood tree was built from a concatenated alignment of three core proteins (predicted to encode the portal protein, the major capsid protein and the tail sheath protein) common to the 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 nonredundant bacteriophages, downloaded from GenBank (ftp://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-denaturated 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 PMID: 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 PMID: 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 PMID: 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 PMID: 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 PMID: 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 PMID: 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 PMID: 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 PMID: 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 PMID: 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 PMID: 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 PMID: 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 PMID: 21569567
14. Essoh 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 PMID: 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 PMID: 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 PMID: 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
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 PMID: 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 PMID: 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 PMID: 23050599
21. Jiang X, Jiang H, Li C, Wang S, Mi Z, et al. (2011) Sequence characteristics of T4-like bacteriophage IME08 benome termini revealed by high throughput sequencing. Virol J 8: 194. doi: 10.1186/1743-422X-8-194 PMID: 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
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
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
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
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 PMID: 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 PMID: 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 PMID: 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 PMID: 23497209
31. Juhala RJ, Ford ME, Duda RL, Youlton A, Hatfull GF, et al. (2000) Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J Mol Biol 299: 27–51. PMID: 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 PMID: 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
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 PMID: 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 PMID: 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 PMID: 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 PMID: 24603854

38. Eddy SR (2011) Accelerated Profile HMM Searches. PLoS Comput Biol 7: e1002195. doi: 10.1371/journal.pcbi.1002195 PMID: 22039361

39. Rohwer F, Edwards R (2002) The Phage Proteomic Tree: a genome-based taxonomy for phage. J Bacteriol 184: 4529–4535. PMID: 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 PMID: 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 PMID: 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 PMID: 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 PMID: 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 PMID: 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 PMID: 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 PMID: 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 PMID: 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

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

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

51. Sullivan MJ, Petty NK, Beatson SA (2011) Easyfig: a genome comparison visualizer. Bioinformatics 27: 1009–1010. doi: 10.1093/bioinformatics/btr039 PMID: 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 PMID: 22445902

53. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797. PMID: 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 PMID: 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 PMID: 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 PMID: 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