A new bacteriophage subfamily - "Jerseyvirinae"

Hany Anany • Andrea I Moreno Switt• Niall De Lappe• Hans-Wolfgang Ackermann • Darren M. Reynolds• Andrew M. Kropinski • Martin Wiedmann• Mansel W. Griffiths • Denise Tremblay •Sylvain Moineau • John H.E. Nash • Dann Turner

H. Anany (*) (Corresponding author) • M.W. Griffiths
University of Guelph, Canadian Research Institute for Food Safety, ON; N1G 2W1, Canada,
(*) Microbiology Department, Faculty of Science, Ain Shams University, Abbassia, Cairo, Egypt

A. I. Moreno Switt

Universidad Andres Bello, Escuela de Medicina Veterinaria, Facultad de Ecología y Recursos Naturales, Republica 440,8370251 Santiago, Chile

N. De Lappe
National *Salmonella, Shigella* & *Listeria* Reference Laboratory, Medical Microbiology Department, University Hospital Galway, Galway, Ireland.

H.-W. Ackermann
Université Laval, Department of Microbiology, Immunology, and Infectiology, Faculty of Medicine, Quebec, QC; G1X 4C6, Canada

Andrew M. Kropinski (Corresponding author)
University of Guelph, Department of Molecular & Cellular Biology, ON; N1G 2W1, Canada. Email: akropins@uoguelph.ca

M. Wiedmann
Cornell University, Department of Food Science, Ithaca, NY 14850, USA

D.M. Reynolds • D. Turner
University of the West of England, Centre for Research in Biosciences, Faculty of Health and Applied Sciences, Bristol, BS16 1QY, UK

S. Moineau • D. Tremblay
Université Laval, Département de biochimie, de microbiologie et de bio-informatique, Faculté des sciences et de génie, QC, G1V 0A6, Canada

J.H.E. Nash
Public Health Agency of Canada, Laboratory for Foodborne Zoonoses, Guelph, ON; N1G 3W4; &, Department of Pathobiology, Ontario Veterinary College, University of Guelph, Guelph; ON, N1G 2W1, Canada

Abstract

Based upon morphology, comparative nucleotide and protein sequence analysis, a new subfamily of the family *Siphoviridae* is proposed, named the "Jerseyvirinae" and consisting of three genera, the "Jerseylikevirus", the "Sp3unalikevirus" and the "K1glikevirus". To date, this subfamily consists of 18 phages, for which the genomes have been sequenced. *Salmonella* phages Jersey, vB_SenS_AG11, vB_SenS-Ent1, vB_SenS-Ent2, vB_SenS-Ent3, FSL SP-101, SETP3, SETP7, SETP13, SE2, SS3e and wksl3 form the "Jerseylikevirus". The "K1glikevirus" consist of *Escherichia* phages K1G, K1H, K1ind1, K1ind2 and K1ind3. The Sp3unalikevirus contains one member so far. Jersey-like phages appear to be widely distributed, as the above phages were isolated in the UK, Canada, USA and South Korea between 1970 and the present day. The distinguishing features of this subfamily include a distinct siphovirus morphotype, terminally redundant, circularly permuted genomes of 40.7-43.6 kb (49.6-51.4 mol% G+C), syntenic genome organisation and a high degree of nucleotide sequence identity and shared proteins. All known members of the proposed subfamily are strictly lytic.

Introduction

The advent of affordable whole genome sequencing and the renewed interest in using bacteriophages as alternatives to antibiotics or to decontaminate food substances has led to a marked increase in submissions of complete phage genomes to public databases such as GenBank. While this process has led to comparative genome analyses and resulted in the recognition of relationships between newly submitted phage genomes and those already deposited, official recognition of these new taxonomic units has lagged. This is particularly true for members of the *Siphoviridae* family (double-stranded dsDNA genome, non-contractile tail), for which the present International Committee on Virus Taxonomy (ICTV) only recognizes ten genera and thirty-one species (www.ictvonline.org/virusTaxonomy.org). These species account for less than 10% of the fully sequenced genomes from members of this family, a situation that must be addressed. A significant number of these unclassified *Siphoviridae* infect *Salmonella* hosts.

Serovars of *Salmonella* are widespread aetiological agents of food and waterborne diseases in humans and livestock. *Salmonella enterica* is classified into six subspecies by biochemical tests. Based upon serology of the lipopolysaccharide (O) and flagella (H) antigens, over 2,600 serovars of *Salmonella* have been described [23]. Not surprisingly, *Salmonella*phages are also numerous and varied. The morphology of 177 *Salmonella* phages was reviewed in 2007 [2] (updated in [56]) with most representatives belonging to the *Caudovirales* order (dsDNA genome, tailed-phages) and its three families, the *Siphoviridae*, *Myoviridae* and *Podoviridae*. A small number of phages representing the *Inoviridae*, *Leviviridae*, *Microviridae* and *Tectiviridae* families have also been documented.

A number of *Salmonella* phages were originally isolated for the purposes of phage typing. Phage typing represents a useful epidemiological tool whereby strains of a particular serovar may be differentiated into phage types based upon susceptibility to a panel of bacteriophages. Phage Jersey was originally isolated and used by Felix and Callow in the development of a typing scheme for *Salmonella paratyphi* B [1].

Jersey presents a morphotype recognizable by electron microscopy with a tail terminal disk-like structure with 6 club-shaped spikes [1, 2, 15]. To date, almost all phages examined which exhibit the Jersey-type morphotype have been restricted to *Salmonella* and *Escherichia*, the sole exception to this rule being *Serratia* phage η [16]. No phages with this morphotype have been observed to infect other Gram-negative genera including *Aeromonas, Pseudomonas, Vibrio,* or *Rhizobium* (Table 1). The genome of phage Jersey has recently been sequenced and discontinuous MEGABLAST revealed similarity to a number of siphovirus genomes deposited in GenBank.

On the basis of an in-depth bioinformatics analysis, this work proposes the creation of three new genera within a novel subfamily, the “Jerseyvirinae”.

Results

Morphological characteristics

Members of this subfamily which have been examined by transmission electron microscopy are typical siphoviruses with isometric heads of 60 nm in diameter between opposite apices and long, non-contractile tails of 120 x 8 nm (Table 1). Capsids appear as hexagonal or pentagonal and are thus icosahedral in shape. Tails have about 27 transverse striations with a thin, 17 nm wide baseplate with 5-6 spikes of 10 x 3 nm. Capsid and tail dimensions differ slightly from those of phosphotungstate-stained phage Jersey (head = 68 nm, tail length 116 nm), which were obtained in 1970 without the benefit of rigorous electron magnification calibration. A morphological diagram of Jersey-like phages is presented in Figure 1. Phages exhibiting a Jersey-morphology have been described many times in the literature (Table 1).

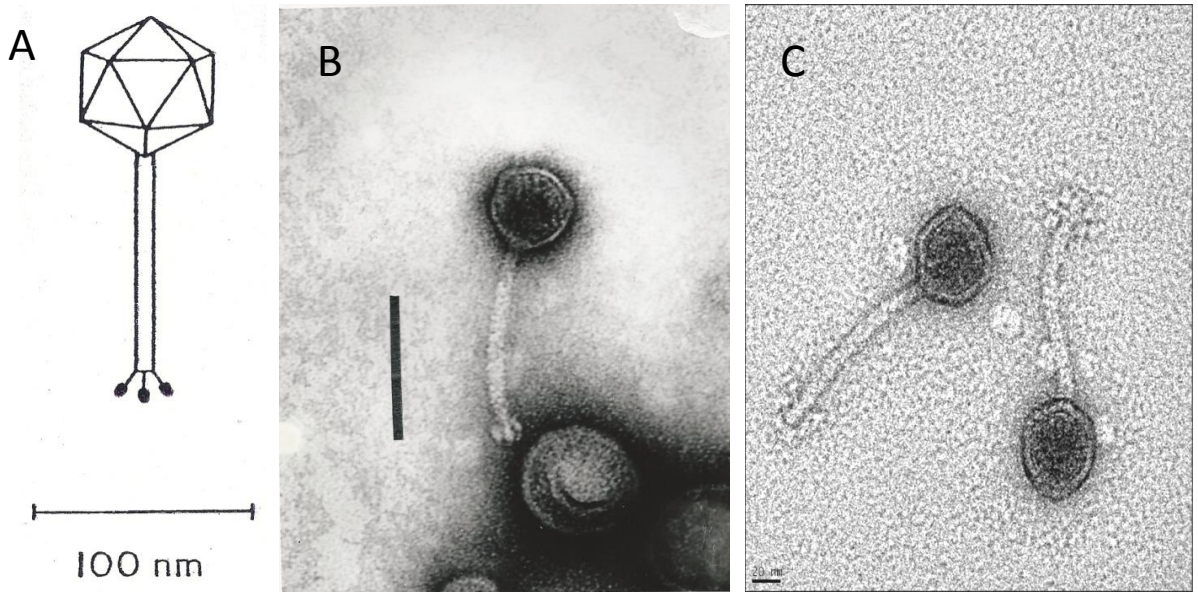


Fig 1: Scale drawing of Jersey-like phage (A). Representative negatively stained TEM images of phage Jersey (B) and K1H (C). Scale bars are 100 nm and 20 nm, respectively

Comparative genomics and proteomics of bacteriophages belonging to the subfamily "Jerseyvirinae"

Phages considered in this manuscript are summarised in Table 2. Each was initially linked by similar morphology and discontinuous MEGABLAST analysis.

Table1. Jersey-like phages in the literature: (I) phages for which morphology and genome sequence are known, (II) phages with sequence only, and (III) phages with morphology only.

Group	Phages	Host	Origin	Capsid, nm	Tail, nm	References
I	**Jersey,** 1, 2, 3a, 3aI, 1010	*S*. Paratyphi B	England	68	116	[1]
	AG11	*S*. Enteritidis	Canada	61	117 x 8	[5]
	Ent1	*S*. Enteritidis	England	64	116 x 9	[58]
	MB78	*S*. Typhimurium	India	60	90-100	[29]
	SE2	*S*. Enteritidis	South Korea	---	---	[57]
	SEPT3, SEPT5, **SEPT7**, SEPT11, SEPT12, **SEPT13**	*S*. Enteritidis	England	63	120 x 7	[14]
	wksl3	*S*. Enteritidis	South Korea	63	121 x 8	[30]
II	**FSL SP-031**	*S*. Cerro	USA			[43]
	FSL SP-038⌘	*S*. Cerro	USA			[43]
	FSL SP-049⌘	*S*. Cerro	USA			[43]
	FSL SP-101⌘	*S*. Dublin	USA			[43]
	K1G	*E. coli*	USA			[9]
	K1H	*E. coli*	USA			[9]
	K1ind1	*E. coli*	USA			[9]
	K1ind2	*E. coli*	USA			[9]
	K1ind3	*E. coli*	USA			[9]
	SSe3	Broad range (*)	South Korea			[31]
	L13	Broad range (**)	South Korea			[26]
	ST4	Broad range(***)	South Korea			[26]
III	FGCSSa2	*S*. Typhimurium	New Zealand	66	122 x 9	[11]
	JS77.1, JS85.2	*E. coli* O127:K63	Bangladesh	---	---	[12]
	San3, San7, San8	*Salmonella* spp.	USA	57 x 60	113-119 x 7-11	[2], [3]
	ΦSH17, ΦSH18	*S*. Typhimurium	England	---	---	[27]
	2, 3	*S*. Heidelberg	Canada	58	115 x 8	[15]

Bold face: complete or partial (⌘) sequence. Dimensions in italics indicate size determination after calibration with catalase crystals or T4 phage tails.

106 * Hosts: *E. coli*, *Shigella sonnei*, *Enterobacter cloacae*, *Serratia marcescens*.
107 ** Hosts: *E. coli, S*. Gallinarum, *S*. Pullorum, *S*. Enteritidis, and *S*. Typhimurium.
108 *** Hosts: *S*. Gallinarum, *S*. Pullorum, *S*. Enteritidis.
109

110 **Table 2.** Bacteriophages with fully sequenced genomes belonging to the proposed subfamily
111 "Jerseyvirinae".

Phage name	Proposed genera	Genome size (bp)	Mol % G+C	No. of annotated CDSs	Reference	GenBank Accession Number
vB_SenS_AG11	Jerseylikevirus	41,546	49.91	66	-	JX297445
SETP3	Jerseylikevirus	42,572	49.85	63	[14]	EF177456
SETP7	Jerseylikevirus	42,749	49.90	68	[14]	KF562865
SETP13	Jerseylikevirus	42,665	49.79	68	[14]	KF562864
SS3e syn. KS7	Jerseylikevirus	40,793	50.08	58	[31]	AY730274
SE2	Jerseylikevirus	43,221	49.64	61	[57]	JQ007353
wksl3	Jerseylikevirus	42,633	49.80	64	[7]	JX202565
vB_SenS-Ent1	Jerseylikevirus	42,391	49.79	58	[58]	HE775250
vB_SenS-Ent2	Jerseylikevirus	42,093	49.92	56	-	HG934469
vB_SenS-Ent3	Jerseylikevirus	42,764	49.79	60	-	HG934470
FSLSP-101	Jerseylikevirus	41,873	50.27	57	[43]	KC139511
Jersey	Jerseylikevirus	43,447	49.97	69	-	KF148055
K1G syn. K1-dep(4)	K1glikevirus	43,587	51.07	52	[9]	GU196277
K1H syn. K1-dep(1)	K1glikevirus	41,632	51.17	50	[9]	GU196278
Kind1	K1glikevirus	42,292	51.27	51	[9]	GU196279
Kind2	K1glikevirus	42,765	51.35	48	[9]	GU196280
Kind3	K1glikevirus	43,461	51.15	49	[9]	GU196281
FSLSP-031	Sp3unalikevirus	42,215	51.07	59	[43]	KC139518

112

113 In addition to phages with completely sequenced genomes, several phages with partial sequences in
114 GenBank were also identified using discontigous MEGABLAST; ST4 [JX233783], L13 [KC832325] and
115 MB78 [AY040866, AF156970, AF349435, AJ277754, AJ249347, AJ245858, AJ245537, X87092,
116 X86562, Y19202, Y19203, Y18133]. These phages are considered tentative members, subject to
117 morphological examination and complete sequencing. Two *Salmonella* phages, FSL SP-038 and FSL
118 SP-049, are also represented by partial genomic sequences in GenBank. Analyses suggest that they
119 are part of the "Sp3unalikevirus" genus.

120

121 Each of the phages listed in **Table 2** was colinearized and then examined for overall sequence
122 similarity using progressiveMauve [13] (data not shown); and, DNA sequence identity using EMBOSS
123 Stretcher [45] and CLUSTALW [34], which have been widely used for nucleotide sequence alignment
124 of viruses [4, 42, 59]. Using the latter methodology, three clades were clearly defined (**Figure 2**).

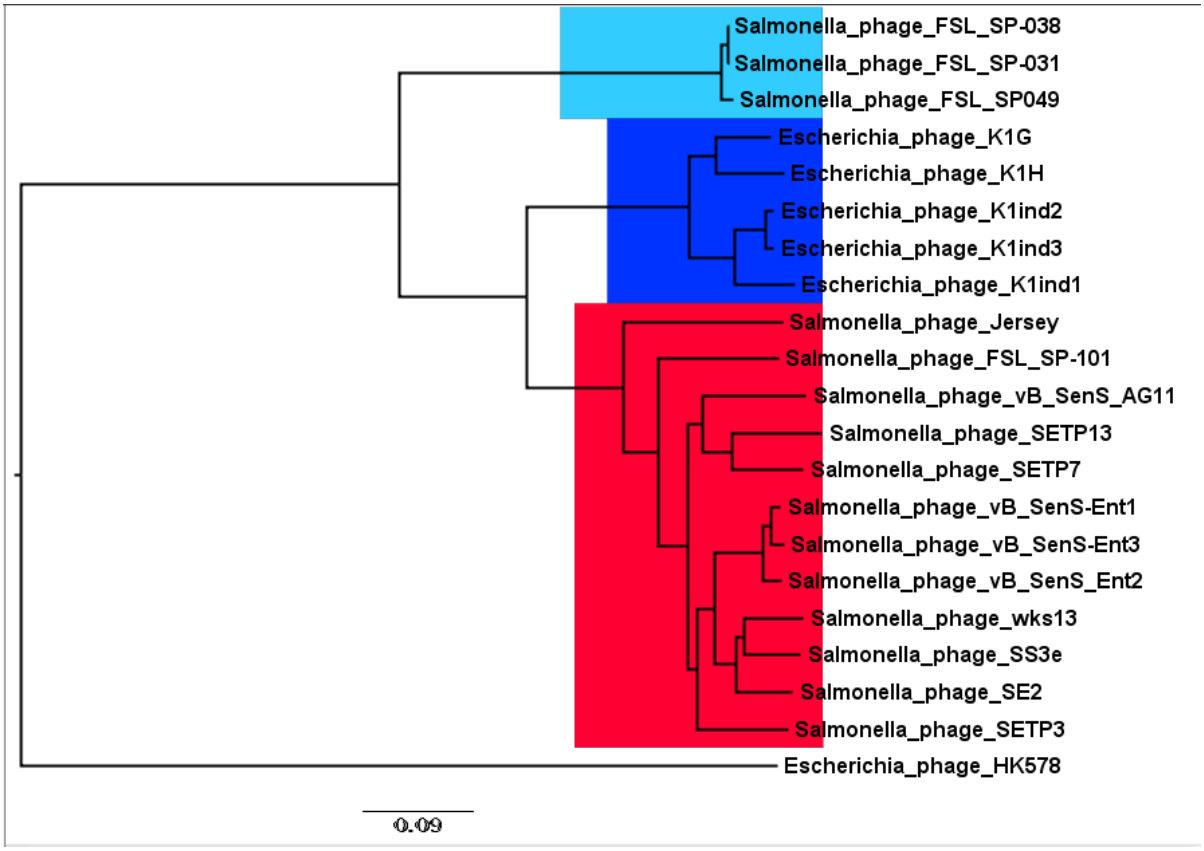


Fig 2. Clustal analysis reveals that the Jersey-like phages fall into three distinct groups, for which we proposed three genera "Jerseylikevirus" (red), "K1glikevirus" (dark blue), and "Sp3unalikevirus" (light blue) within a proposed subfamily, the "Jerseyvirinae." The genomes were colinearized before analysis. *E. coli* phage HK578 was included as an outlier. The scale bar represents 0.09 substitutions per site.

In addition, the viral proteomes were subjected to pairwise comparisons using CoreGenes 3.0 which provided a measure of the total similarity at the protein level based upon pairs of proteins scoring above a pre-defined BLASTP bit score threshold [65]. The CoreGenes and Stretcher results are presented in **Table 3**. Lastly, using the Markov clustering algorithm OrthoMCL [38] with an e-value threshold of $1e^{-5}$ and an inflation value of 1.15, all the proteins encoded by members of the proposed subfamily were examined (data not shown). Functional annotations of proteins encoded by Jersey-like phages were obtained using BLAST tools and HHpred [51, 52]. Transmembrane domains were identified using TMHMM [32] and conserved domains using Pfam [17] and InterProScan [28]. In line with the 95% DNA sequence identity threshold used to delineate phage species specified by the Bacterial and Archaeal Viruses Subcommittee of the ICTV, the genus "Jerseylikevirus" is comprised of 10 species whose sequence identities range from 71.6% (FSL SP-101) to 67.6% (SETP13) and share between 79.7% and 68.1% proteins relative to Jersey (Table 3).

Table 3. Proteome and nucleotide sequence similarity of members of the "Jerseyvirinae". Protein homology represents the percentage of shared proteins related to phage Jersey as determined using CoreGenes 3.0. Nucleotide sequence identity relative to phage Jersey was determined using EMBOSS Stretcher.

Genus	Phage name	GenBank Accession	Nucleotide sequence identity	Protein Homology
"Jerseylikevirus"	Jersey	KF148055	100%	100.0%
	AG11	JX297445	70.9%	79.7%
	wksI3	JX202565	71.2%	78.3%
	SE2	JQ007353	69.5%	71.0%
	Ent1	HE775250	69.5%	75.4%
	Ent2	HG934469	69.7%	73.9%
	Ent3	HG934470	69.4%	76.8%
	SS3e	AY730274.2	68.1%	68.1%
	SETP3	EF177456	71.1%	68.1%
	SETP7	KF562865	69.7%	79.7%
	SETP13	KF562864	67.6%	73.9%
	FSL SP-101*	KC139511	71.6%	72.5%
"K1glikevirus"	K1G	GU196277	59.5%	59.4%
	K1H	GU196278	59.0%	59.4%
	K1ind2	GU196280	58.0%	55.1%
	K1ind3	GU196281	59.5%	56.5%
	K1ind1	GU196279	59.7%	58.0%
"Sp3unalikevirus"	FSL SP-031	KC139518	53.4%	66.7%

* Partially sequenced

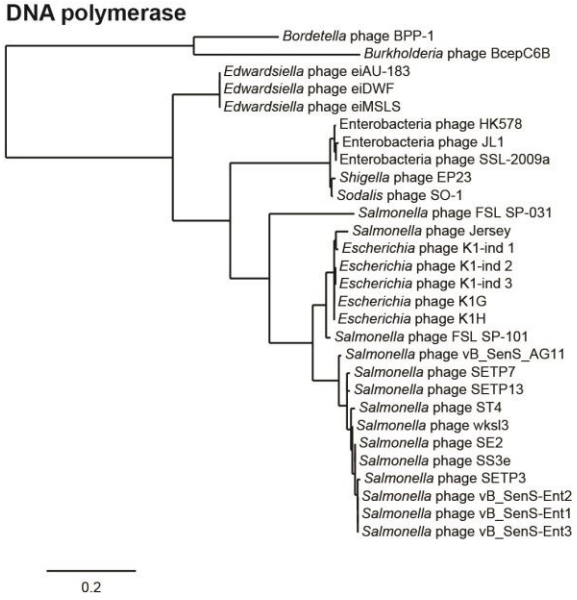
The "K1glikevirus" genus consists of four species - K1H, K1G, K1ind1 and K1ind2. They share between 79% and 97.1 % DNA sequence identity and a minimum of 84% homologous proteins. Relative to phage Jersey, the "K1glikevirus" exhibit at minimum 58% nucleotide sequence identity. Despite exhibiting significant protein homology (66.7%), the DNA sequence of FSL SP-031 shows only 53.4% identity to Jersey, a difference sufficient to indicate that this phage is distinct from other members and warrant the creation of a separate genus, the "Sp3unalikevirus" within the subfamily.

The results of DNA sequence alignment, homologous proteins, morphology, and genome organisation indicate that the Jersey-like phages form three distinct groups, united by 53% DNA sequence identity, which are proposed to represent three genera; the "Jerseylikevirus", "Sp3unalikevirus" and "K1glikevirus" (Fig. 2). In the following sections the common properties of viruses, which belong to the "Jerseyvirinae", are discussed alongside specifics of the three genera.

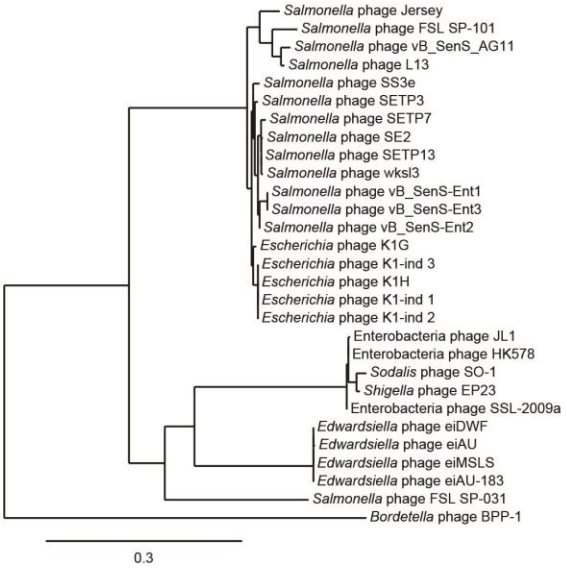
Protein phylogeny

Phylogenetic trees were constructed to investigate common proteomic features of the "Jerseyvirinae" for the large terminase subunit (TerL), portal protein, DNA polymerase (DpoI), helicase, major capsid, and major tail (MTP) proteins (Fig. 3). Analysis of the helicase proteins indicates that this group of phages are phylogenetically related and distinct from the "Hk574likevirus". The trees constructed using the TerL, portal and DpoI sequences clearly indicate that this group of viruses can be subdivided into three clades. Analysis of the major capsid and MTPs reveals that the K1G clade is significantly different from the "Jerseylikevirus" group. Interestingly, the K1H major capsid protein is distinct from other members of the subfamily, a feature corroborated by the OrthoMCL groupings.

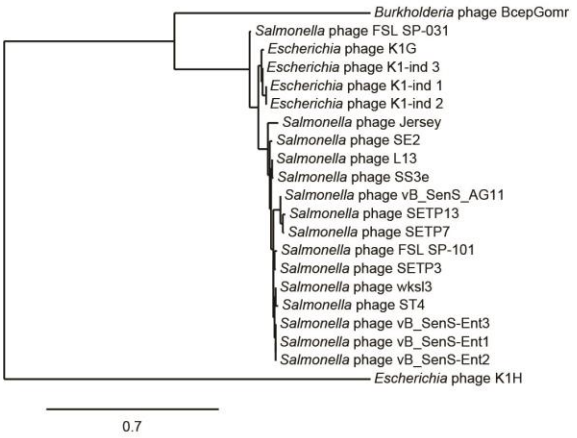
DNA polymerase



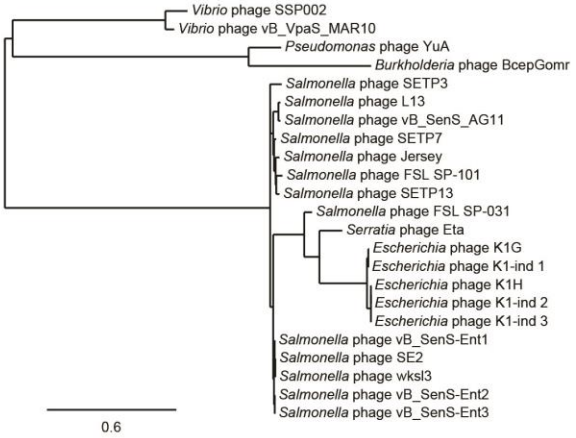
Helicase



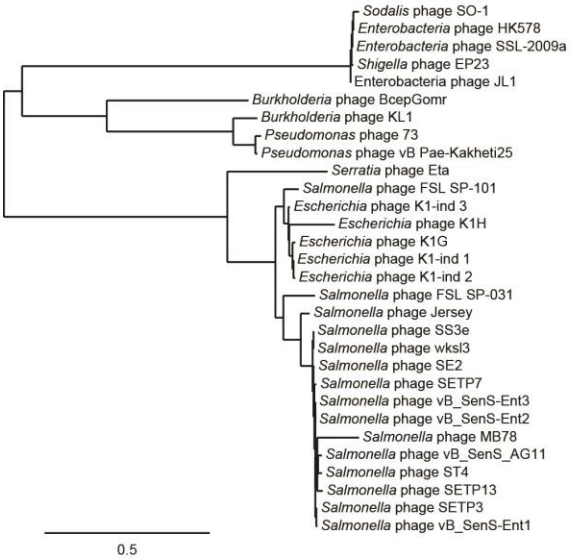
Major capsid protein



Major tail protein



Portal protein



Large terminase subunit

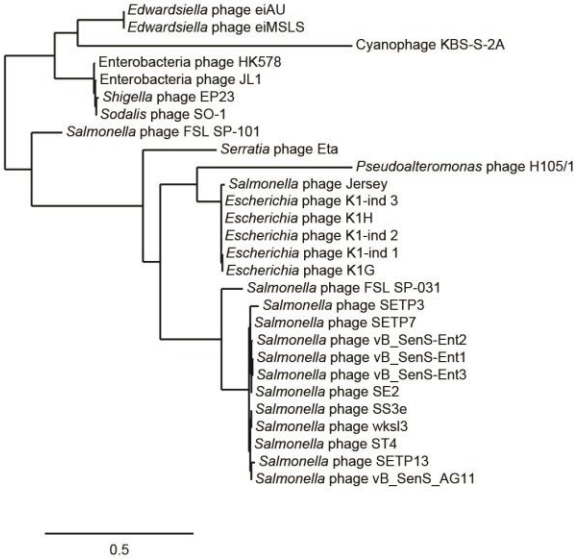


Fig 3. Phylogenetic analysis of the "Jerseyvirinae" common proteins produced using www.phylogeny.fr. Branch length is proportional to the number of substitutions per site.

Common features of the "Jerseyvirinae"

The genome of the proposed "Jerseyvirinae" phages is ranging in size from 40.7 to 43.6 kb with a G+C content from 49.6 to 51 %. They also encode between 48 to 69 proteins but no tRNAs. These phages appear to be strictly lytic/virulent as none have been shown capable of lysogeny or harbour homologues of known integrases, recombinases or excisionases. As with most phages, their genomes exhibit a modular structure and the organisation of genes shows a high degree of synteny across all members (Fig. 4). The genome may be divided into four modules on the basis of the predicted function of component genes; (i) virion structure and assembly, (ii) regulation/immunity, (iii) genome replication and (iv) host lysis. Despite the presence of a number of deduced proteins of unknown function, the roles of some gene products can be predicted on the basis of BLAST and HHpred searches or by the presence of conserved domains.

Using OrthoMCL, the 1,057 proteins encoded by the 18 members of the three proposed genera formed 90 clusters of orthologous proteins and 51 singletons. Of these, 25 clusters were present in all members of the three proposed genera representing conserved proteins in the structural, replicative and lysis gene modules. In addition to these 'core' genes/proteins, a further 1, 5 and 12 clusters were present in all members of the "Jerseylikevirus", "K1glikevirus", and "Sp3unalikevirus", respectively. The structural gene module represents the most highly conserved region with 17 ortholog clusters. Almost half of the genes encoded by the "Jerseyvirinae" phages are devoted to genome packaging, assembly and structure of mature virions. The structural module follows a strongly conserved gene order, encoding genes involved in DNA packaging followed by the virion capsid, tail and adsorption apparatus in an arrangement reminiscent of that observed in many phages.

Several features are of interest within the morphogenesis module. Using phage Jersey as reference, the major capsid protein (gp12) shows structural similarity to *Bacillus* phage SPP1 and coliphage HK97 (Protein Data Bank/PDB accession numbers: 4an5 and 3p8q, respectively) suggesting that the capsid protein shares the evolutionary conserved HK97-like fold [18]. Three gene products each return structural similarity to Ig-set domain proteins with HHpred. Two of these proteins, Jersey gp7 (AGP24895) and SETP13 gp12 (AGX84616), return matches to the Hoc protein of Enterobacteria phage RB49 (PDB 3shs) while a third, Jersey gp13 (AGP24902), returned a match to the head fibre of *Bacillus* phage Φ29 (PDB 3qc7). Ig-like domains are found widely within the order *Caudovirales* and

are predominantly associated with structural proteins suggesting that these proteins play a role within capsid assembly or completion and may also be involved in non-specific binding to host cells [19]. The presence of a weak match to an Ig-like I-set domain (pfam: PF07679) in gp07 and the proximity of its gene to the gene coding for the major capsid protein might suggest that this gene encodes a capsid decoration protein [53]. Head decoration proteins have been described in a number of phages such as L, λ and ES18 and are thought to aid stabilisation of the capsid structure [21]. However, to date none of the structural proteins present in mature Jersey-like virions have been identified using mass spectrometry.

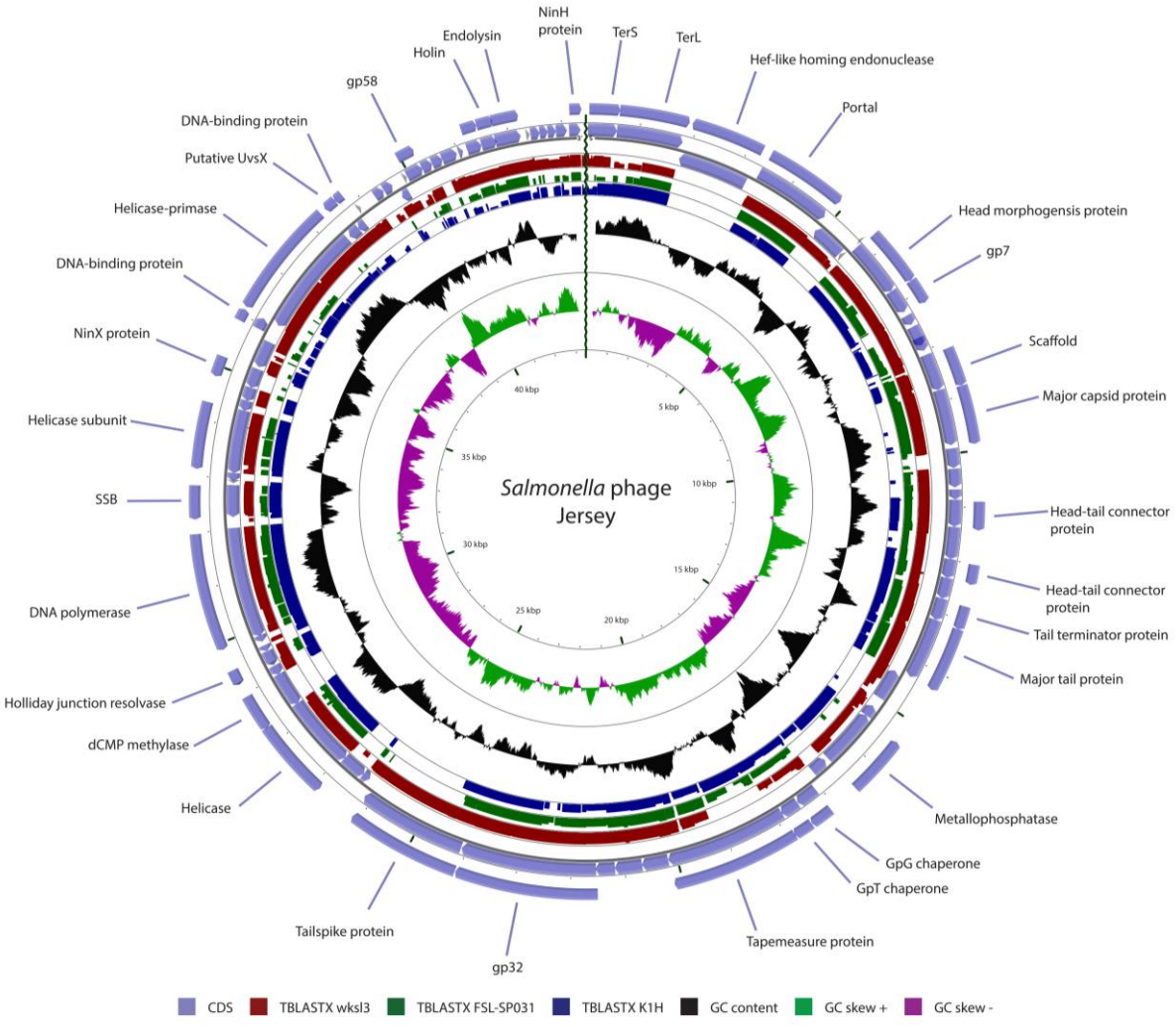


Fig 4. Genetic and physical map of *Salmonella* phage Jersey prepared using CGView [22]. For sequence similarity comparison, TBLASTX was used versus wks13 (red), FSL-SP031 (dark green) and K1H (dark blue). GC content is depicted in black while positive and negative GC skew is denoted by green and purple, respectively.

Two coding sequences positioned immediately upstream of the gene coding for the putative tape measure form a single cluster using OrthoMCL, which is present in all members of the three proposed genera. The positioning of these genes is similar to the ones of λ tail assembly chaperones gpG and gpGT. In λ, gpG and the translational frameshift product gpGT interact with the tape measure and major tail proteins and are required for correct tail formation [61-63]. Analysis of the putative gpG and gpGT genes in phage Jersey (AGP24914 and AGP24915) with MFOLD and HPKNOTTER provided evidence for the presence of a stem-loop structure and a pseudoknot, respectively, suggesting that the Jersey-like phages produce a gpGT-like fusion product.

The putative tail fibre (Jersey gp32, AGP24920) shows distant similarity to the central tail fibre gpJ of phage λ and p33 of phage T1 using PSI-BLAST and is conserved among all members of the "Jerseyvirinae". However, a central tail tip fibre has not been observed in electron micrographs so the precise role and location of this protein remains unclear. This protein may in fact form the virion baseplate in conjunction with a gene (Jersey gp31, AGP24919) encoded immediately upstream which exhibits similarity to the endolysin of *Pseudomonas* phage ϕKZ (PBD: 3bkh), suggesting a role in cell wall degradation.

In all "Jerseyvirinae", the morphogenesis gene module is interrupted after the gene coding for the major tail protein by a cassette of between 1 to 5 genes encoded in the opposite orientation where no single protein is present across the subfamily. With the exception of phages SS3e and SP-031, all members of the "Jerseyvirinae" encode a putative serine/threonine protease in this region (pfam: PF12850, metallophos_2), which is linked by HHpred to the NinI protein phosphatase in λ (PDB 1g5b).

Four critical units involved in genome replication, a replicative family A DNA polymerase (pfam: PF00476), DEAD-box helicase-primase (pfam: PF13481) and helicase (pfam: PF00176) are also conserved across all members of the proposed subfamily. Analysis of the replication gene cluster using HHpred provides additional evidence for the presence of a Holliday junction resolvase (PDB 1hh1), helicase subunit (PDB 3h4r) and a helix-destabilising ssDNA-binding protein similar to gp2.5 of phage T7 (PDB 1je5).

Finally, the lysis gene cluster codes for proteins facilitating lysis of the infected host cell, allowing egress of newly formed virions into the surrounding environment. The lysis or late gene module represents an area of significant divergence among the three genera and between individual phages. The module is replete with ORFs of unknown function and only three gene products, the endolysin,

holin and a protein of unknown function are conserved across the subfamily. While the holin formed a single cluster using OrthoMCL, the number of predicted transmembrane domains differed between phage. On the basis of predicted transmembrane domains, phages AG11, Jersey, SE2, SETP3, wksl3, Ent2 and SP-031 are presumed to encode class I holins while class II holins are encoded by Ent1, Ent3, SP-101, SETP7, SETP13 and SS3e [64].The endolysin belongs to the glycoside hydrolase 24 family (muraminidase, pfam:PF00959). Many of the lysis gene products are shared between only some members of each genus, while others are found between limited representatives of one or more genera, suggesting that this region has been the site of frequent genetic exchange.

Description of individual genera

"Jerseylikevirus"

This proposed genus is named after the first characterized phage of this morphotype, *Salmonella* phage Jersey [1]. The genus is distinguished by a distinct morphology, a similar genome size among its members, conserved gene organisation and the use of a P22-like tailspike to facilitate host recognition and adsorption. The latter distinguished them from the "K1glikevirus" and "Sp3unalikevirus" as well as by an additional 25 accessory genes shared between two or more species. To date, 12 members, isolated from four different countries have been fully sequenced and annotated (**Table 2**). Members of the "Jerseylikevirus" have an average G+C content of 49.79 %, slightly lower than the average of 52 % reported for serovars of *Salmonella enterica* [41].

Host specificity is conferred by six tailspikes, observed as short clubs attached to the tail terminus. Like the tailspikes of podophages P22, Sf6 and HK620, they exhibit a modular design consisting of a conserved N-terminal binding domain and a P22-like C-terminal catalytic domain (data not shown). In phage P22, the tailspike facilitates adsorption to *Salmonella* O-antigen 12 of the cell surface LPS, expressed by members of White-Kauffmann-Le Minor serogroups A, B and D1 [23]. The high degree of conservation of the P22-like domain and catalytic residues combined with host range data available [15, 30, 57] for SETP3, SETP7, SETP13, vB_SenS-Ent1 and wksl3 suggests that "Jerseylikevirus" isolated to date are limited to these *Salmonella* serogroups. However, SS3e appears to be an exception, being reported as capable of lysing enterobacterial genera other than *Salmonella*, including *E. coli*, *Enterobacter cloacae*, *Shigella sonnei* and *Serratia marcescens* [57].

Several members of the "Jerseylikevirus" (Jersey, Ent1, Ent2, Ent3, SE2, wksl3 and SP-101) encode a putative DNA binding protein containing one or two Pfam family domains; ANT (PF03374) and pRha (PF09669). In phage P22, *ant* encodes an anti-repressor which inhibits binding of the c2 repressor to

the PL and PR operators enabling the expression of genes necessary for the lytic development [10]. The pRha domain represents a family of proteins whose expression is detrimental for lytic growth in the absence of integration host factor function [55].

A gene product with similarity to inner membrane immunity (Imm) proteins, which protect against superinfecting phages [33], is also found in all Jersey-like phages, except Jersey and AG11. This gene product contains an Imm_superinfect motif (pfam: PF14373), is predicted to localize at the cytoplasmic membrane (PSORTb) and contains two transmembrane domains (TMHMM). Only the putative immunity protein [AAZ41745] is annotated in SS3e, although this appears to be due to an incomplete genome sequence rather than an absence of further proteins in this region.

An interesting feature of the "Jerseylikevirus" is that intein insertion is evident in the DNA helicase of phages vB_SenS-Ent1, Ent2, Ent3, SETP3 and SETP7 and also within the DNA polymerase of phages FSL SP-101, Ent1, Ent2, Ent3, SE2 and SETP3 (data not shown). Inteins are defined as protein sequences embedded within a precursor sequence which, upon translation, catalyzes self-excision from the host polypeptide and ligation of the flanking sequences to yield two stable products; the mature protein (extein) and the intein [48].

In SE2 and SS3e the DNA polymerase appears to be encoded by more than one gene. For SE2, two gene products, gp05 (AEX56144) and gp06 (AEX56145) have predicted DNA polymerase activity. The large subunit, gp06, is predicted to possess an intein similar to that found in SETP3 and vB_SenS-Ent1. Like SE2, the DNA polymerase of SS3e also appears to be split into two coding sequences (gp41 and gp43) although in this case, the subunits are interrupted by an additional gene (gp42; AAW51247) with a predicted C-terminal HNH endonuclease domain. Notably all three gene products show short matches to either the N- or C-terminal ends of the SETP3 intein.

A total of 15 accessory genes are encoded within the lysis/late gene cassette of which only 3 have homologues with predicted functions; a putative protease (9 members), a HNH homing endonuclease (7 members) and a putative RNA-binding protein (2 members).

"K1glikevirus"

The *Escherichia* phages K1G, K1H, K1ind1, K1ind2 and K1ind3, comprising the "K1glikevirus"are described as K1-dependant or -independent, denoting the requirement for the K1 capsule for productive infection [9]. The GC content of their genome is slightly higher than the "Jerseylikevirus" ranging between 51.1 to 51.5%, a value closer to that of their host *E. coli* (50.8 %). Moreover, they encode fewer proteins than the "Jerseylikevirus".

Members of the “K1glikevirus” possess tailspikes, which are of similar size to the “Jerseylikevirus”. These proteins exhibit similarity to the conserved N-terminal domain identified within the “Jerseylikevirus” but possess divergent C-terminal domains, indicative of their different host specificity. Tailspikes from the K1-independant phages K1ind1, K1ind2 and K1ind3 exhibit high similarity to the HK620 tailspike [PDB 2x6w] and appear to belong to the pectate lysase 3 family (pfam: PF12708). The HK620 tailspike possess endo-*N*-acetylglucosaminidase activity that degrades the O-antigen of *E. coli* serotype O18A1 [7]. In contrast, the tailspike encoded by the K1-dependant phages K1G and K1H exhibit endo-*N*-acyl-neuraminidase (endosialidase) activity [9] and are nearly identical to the tailspike of the T7-like phage K1F (PDB 3ju4). Endosialidases bind to and degrade the K1 capsular polysaccharide, a homopolymer made up of α2,8-linked sialic acid residues [6]. Each of the K1-dependant tailspikes are predicted to contain a C-terminal Peptidase_S74 domain (PF13884) which shows homology to the protease domains of K1F, K1E, K1-5 tailspikes as well as the long tail fibre of T5. In these phages this domain has been shown to function as an intramolecular chaperone, whose presence and subsequent auto-cleavage is essential for folding and assembly of mature proteins [50]. These data indicate that the K1-dependant tailspikes undergo a distinct maturation process to their counterparts in the “Jerseylikevirus” and K1-independent members of the “K1glikevirus”.

With one or two exceptions, members of the “K1glikevirus” share the complete structural module with phages of the “Jerseylikevirus” genus. Phage K1ind3 has a HNH homing endonuclease gene immediately downstream of a gene coding for a gp7 family morphogenesis protein (gp04) related to *Bacillus* phage SPO1 (PBD 1u3e). K1G encodes another homing endonuclease (gp12) with 31 % identity to MobE of coliphage T6. Notably, the K1H major capsid sequence differs substantially to other “Jerseyvirinae” phages, but is predicted to have a similar structure to SPP1 and HK97 using HHpred. A further gene product of unknown function, K1H gp09, not found in other members of the proposed subfamily is encoded immediately downstream of the gene coding for the major capsid protein (K1H gp08, ADA82303).

Three gene products are conserved across all K1G-like phages in the immunity gene module, a serine/threonine protein phosphatase (pfam: PF12850), an acid-phosphatase B domain protein (pfam: PF03767) and a protein with inferred ATPase activity (pfam: P13207). With the exception of K1H, all members of the “K1glikevirus” encode a super-infection immunity protein. Neither Jersey, AG11 nor K1H appear to harbour an immunity protein, instead these phages encode a hypothetical protein forming a single OrthoMCL cluster.

In addition to the core replication proteins of the “Jerseyvirinae”, each of the K1G-likephages encode a C-5 cytosine specific methylase (pfam: PF00145) which bears little sequence similarity to the methylase of phage Jersey or FSL SP-101 and falls under a separate cluster using OrthoMCL. All K1G-like phages except K1ind1 encode a putative UvsX-like protein. In coliphage T4, UvsX functions as a RecA-like recombinase interacting with the UvsY helicase to promote strand-exchange during genome replication [20].

Finally, a total of seven proteins are conserved across all K1G-like phages in the lysis or late gene module, three of which are conserved across the subfamily; the holin, lysin and a hypothetical protein. Unlike the “Jerseylikevirus”, each of the K1G-like phages are predicted to encode a separate class I and class II holin, with three and two transmembrane domains, respectively. Each K1G-like phage has a NinH–like domain protein (pfam: PF06322) in addition to two gene products of unknown function. Lastly, a gene product found only in the K1-independent phages, returns a DUF3850 family motif (pfam: PF12961), suggested to be involved in RNA recognition.

“Sp3unalikevirus”

This genus was recently proposed by Moreno Switt et al. (2014). To date, three members (FSL SP-031, FSL SP-038, and FSL SP-049) have been reported, though only the genome of phage FLS SP-031 is fully sequenced, hence the proposed genus “Sp3unalikevirus” [43]. Phages in the “Sp03unalikevirus”genus were isolated from dairy farms in the state of New York, with history of *Salmonella* isolation [44]. Sp031-like viruses have a GC content of 51.1%, closer to the *Salmonella* G+C of approximately 52% [47].

Unique characteristics of “Sp3unalikevirus” are non-essential “cargo” genes and genes associated with host-specificity. The genes coding for a number of unique hypothetical proteins with unknown function are located in the replication module of SP-031-likeviruses. Two annotated proteins that are absent in the “Jerseylikevirus” and “K1glikevirus” are homing endonucleases. These homing endonucleases are inserted in the replication module of FLS SP-031, one endonuclease shows 41% identity over the 93% of the protein with an endonuclease in *Synechococcus* phage S-SSM7 (GenBank YP_004324330.1), and the other endonuclease, with the homing endonuclease domain PF13392, showed 42% identity over the 89% with an endonuclease in the genome of *Salmonella* phage FLS SP-126 (GenBank AGF87903.1). Another difference of Sp03unalikeviruses is the presence of a gene encoding a putative phosphoadenosine phosphosulfate (PAPS) reductase (pfam: PF01507) in the virion assembly module. PAPS reductases are involved in the reduction of sulfate to hydrogen sulfide, which is generated by *Salmonella* in the gut of humans and animals [60]. While PAPs

reductases have previously been reported in temperate phages (e.g., BcepB1A [54]), there is no current evidence to indicate whether this enzyme confers a fitness advantage to the host.

Sp031-like phages have a very narrow host range. When 23 different *Salmonella* serovars were tested, only *Salmonella* Cerro was lysed [44]. This host specificity is likely related with the tail spikes in the "Sp3unalikevirus". BLAST of the tail spike amino acid sequence of FSL SP-031 showed as the best hit (78% over 80 % of the protein) a tail spike of a P22-like prophage found in the genome of *S.* Cerro str. 818 (GenBank ESH26034), with the divergence at N-terminal. This prophage shows no further identity to any other CDSs of Sp031-like phages. . To further investigate the potential genetic mechanism involved in this host range, an alignment of the tail spike amino acid sequences of representative Jersey-likeviruses (i.e., Jersey, L3, SETP7, SETP3, SE2, Ent2, SETP12, and FSL SP-101) and Sp031-like viruses (FSL SP-031) showed a conserved N-terminal (approx. 87 % aa identity) and a divergent C-terminal (approx. 15% aa identity). This finding corresponds with the host specificity reported; while other *Salmonella* phages in the "Jerseyvirinae" subfamily infect serogroups A, B, and D; Sp031-likeviruses only infect serogroup K.

Discussion

The availability of an increasing number of genome sequences and improvements in gene prediction methods has resulted in a sizeable shift towards the inclusion of genomic and proteomic data for the taxonomic classification of bacteriophages [37]. However, a number of different, and sometimes conflicting, approaches have been reported in the literature for the purposes of delineating evolutionary relationships between phages including the proteomic tree [49], numbers of shared homologous/orthologous proteins [35, 36] and reticulate classification based on gene content [39, 40]

Our analysis reveals that the classification criteria introduced for members of the *Podoviridae* and *Myoviridae* [35, 36], based upon the existence of protein homologs, tend to "lump" taxa rather than represent the true taxonomic relationship within a genus. Using an alternative approach employing NCBI BLASTN and TBLASTX, genera can be defined a possessing ≥ 65% DNA sequence identity while subfamilies show ≥40% protein homologs (Kropinski, Edwards and Mahadevan, unpublished results). These values are those derived by Niu et al. in their assessment of the T1-likeviruses [46].

Phylogenetic analysis also demonstrates that the *Salmonella* and *Escherichia* phages described here fall into three clusters, substantiating the establishment of three genera. Analysis using EMBOSS Stretcher demonstrated that while phages within each genera were closely related, a significant

relationship based upon shared homologous proteins identified using CoreGenes and OrthoMCL existed between genera. Based upon these data we propose the establishment of a subfamily of *Siphoviridae*, the "Jerseyvirinae" comprised of three genera, the "Jerseylikevirus", "K1Glikevirus" and "Sp3unalikevirus". The three genera of closely related phages were isolated from geographically disparate locations. This suggests that phages of this subfamily are widely distributed and have persisted and evolved in various environments. Considering the length of time between the isolation of Jersey and the other members, horizontal gene transfer does not always mask the identification of taxomic relationships between phages. Grose and Casjens have recently clustered 337 phages infecting various members of family *Enterobacteriaceae* [24] based upon average nucleotide identity, conserved gene product content in addition to whole genome nucleotide and amino acid dotplots. They report a "SETP3 supercluster", formed of 5 clusters of lytic phages; SETP3-like, SO-1-like, ECO1230-10-like, Gj1-like and PY100-like phages. Our independent findings are in broad agreement with those reported by Grose and Casjens but provide a more detailed analysis of the phages comprising the SETP3-like subcluster.

All Jersey-like phages isolated to date exhibit a strictly lytic lifestyle and encode no proteins with homology to known toxins or allergens. As such members of this subfamily appear suited to biocontrol applications [8, 25], particularly the "Jerseylikevirus" which exhibit a broad host range encompassing important serogroups of *Salmonella*.

Undoubtedly more Jersey-like bacteriophages will be isolated in the future. The authors hope the data provided here will act as a starting point for the annotation of these future isolates.

Author's contributions

DT, AIMS, HA, AK, SM and HWA wrote the manuscript. DT, AIMS, HA, AMK, JEN, HWA, NDL and SM contributed towards the analysis of data. All authors read and approved the final manuscript.

Acknowledgements

S.M. acknowledges funding from NSERC of Canada (Discovery program). S.M. holds a Tier 1 Canada Research Chair in Bacteriophages. HA and MWG acknowledges funding from Sentinel Bioactive paper network.

Conflict of interest

The authors declare that they have no competing interests.

435

References

1. Ackermann H-W, Berthiaume L, Kasatiya SS (1972) Morphologie des phages de lysotypie de *Salmonella paratyphi* B (schéma de Felix et Callow). Canadian Journal of Microbiology 18:77-81
2. Ackermann HW, Gershman M (1992) Morphology of phages of a general *Salmonella* typing set. Research in Virology 143:303-310
3. Ackermann HW, DuBow MS, Gershman M, Karska-Wysocki B, Kasatiya SS, Loessner MJ, Mamet-Bratley MD, Regué M (1997) Taxonomic changes in tailed phages of enterobacteria. Arch Virol 142:1381-1390
4. Alonso C, Murtaugh MP, Dee SA, Davies PR (2013) Epidemiological study of air filtration systems for preventing PRRSV infection in large sow herds. Preventive veterinary medicine 112:109-117
5. Anany H (2010) Biocontrol of foodborne bacterial pathogens using immobilized bacteriophages. Food Science. University of Guelph, Guelph
6. Barbirz S, Muller JJ, Uetrecht C, Clark AJ, Heinemann U, Seckler R (2008) Crystal structure of Escherichia coli phage HK620 tailspike: podoviral tailspike endoglycosidase modules are evolutionarily related. Mol Microbiol 69:303-316
7. Barbirz S, Müller JJ, Uetrecht C, Clark AJ, Heinemann U, Seckler R (2008) Crystal structure of *Escherichia coli* phage HK620 tailspike: podoviral tailspike endoglycosidase modules are evolutionarily related. Molecular Microbiology 69:303-316
8. Brovko LY, Anany H, Griffiths MW (2012) Bacteriophages for detection and control of bacterial pathogens in food and food-processing environment. Adv Food Nutr Res 67:241-288
9. Bull JJ, Vimr ER, Molineux IJ (2010) A tale of tails: Sialidase is key to success in a model of phage therapy against K1-capsulated *Escherichia coli*. Virology 398:79-86
10. Byl CV, Kropinski AM (2000) Sequence of the Genome of *Salmonella* Bacteriophage P22. Journal of Bacteriology 182:6472-6481
11. Carey-Smith GV, Billington C, Cornelius AJ, Hudson JA, Heinemann JA (2006) Isolation and characterization of bacteriophages infecting *Salmonella* spp. FEMS Microbiology Letters 258:182-186
12. Chibani-Chennoufi S, Sidoti J, Bruttin A, Dillmann M-L, Kutter E, Qadri F, Sarker SA, Brüssow H (2004) Isolation of *Escherichia coli* Bacteriophages from the Stool of Pediatric Diarrhea Patients in Bangladesh. Journal of Bacteriology 186:8287-8294
13. Darling AE, Mau B, Perna NT (2010) progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. PLoS ONE 5:e11147
14. De Lappe N, Doran G, O'Connor J, O'Hare C, Cormican M (2009) Characterization of bacteriophages used in the *Salmonella enterica* serovar Enteritidis phage-typing scheme. Journal of Medical Microbiology 58:86-93
15. Demczuk W, Ahmed R, Ackermann H-W (2004) Morphology of *Salmonella enterica* serovar Heidelberg typing phages. Canadian Journal of Microbiology 50:873-875
16. Denyes J, Krell P, Manderville R, Ackermann H-W, She Y-M, Kropinski A (2014) The genome and proteome of *Serratia* bacteriophage eta which forms unstable lysogens. Virology Journal 11:6
17. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M (2014) Pfam: the protein families database. Nucleic Acids Research 42:D222-D230
18. Fokine A, Leiman PG, Shneider MM, Ahvazi B, Boeshans KM, Steven AC, Black LW, Mesyanzhinov VV, Rossmann MG (2005) Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry. Proceedings of the National Academy of Sciences of the United States of America 102:7163-7168

19. Fraser JS, Yu Z, Maxwell KL, Davidson AR (2006) Ig-Like Domains on Bacteriophages: A Tale of Promiscuity and Deceit. Journal of Molecular Biology 359:496-507
20. Gajewski S, Webb MR, Galkin V, Egelman EH, Kreuzer KN, White SW (2011) Crystal Structure of the Phage T4 Recombinase UvsX and Its Functional Interaction with the T4 SF2 Helicase UvsW. Journal of Molecular Biology 405:65-76
21. Gilcrease EB, Winn-Stapley DA, Hewitt FC, Joss L, Casjens SR (2005) Nucleotide Sequence of the Head Assembly Gene Cluster of Bacteriophage L and Decoration Protein Characterization. Journal of Bacteriology 187:2050-2057
22. Grant J, Arantes A, Stothard P (2012) Comparing thousands of circular genomes using the CGView Comparison Tool. BMC Genomics 13:202
23. Grimont PAD, Weill F-X (2007) Antigenic formulae of the *Salmonella* serovars. WHO Collaborating Centre for Reference and Research on *Salmonella*, p 166
24. 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
25. Hagens S, Loessner MJ (2010) Bacteriophage for biocontrol of foodborne pathogens: calculations and considerations. Curr Pharm Biotechnol 11:58-68
26. Hong S, Jeong J, Lee J, Kim S, Min W, Myung H (2013) Therapeutic Effects of Bacteriophages Against *Salmonella* Gallinarum Infection in Chickens. Journal of Microbiology and Biotechnology 23:1478-1483
27. Hooton SPT, Atterbury RJ, Connerton IF (2011) Application of a bacteriophage cocktail to reduce *Salmonella* Typhimurium U288 contamination on pig skin. International Journal of Food Microbiology 151:157-163
28. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236-1240
29. Joshi A, Siddiqi JZ, Rao GR, Chakravorty M (1982) MB78, a virulent bacteriophage of *Salmonella typhimurium*. Journal of Virology 41:1038-1043
30. Kang H-W, Kim J-W, Jung T-S, Woo G-J (2013) wksl3, a New Biocontrol Agent for *Salmonella enterica* Serovars Enteritidis and Typhimurium in Foods: Characterization, Application, Sequence Analysis, and Oral Acute Toxicity Study. Applied and Environmental Microbiology 79:1956-1968
31. Kim S-H, Park J-H, Lee B-K, Kwon H-J, Shin J-H, Kim J, Kim S (2012) Complete Genome Sequence of *Salmonella* Bacteriophage SS3e. Journal of Virology 86:10253-10254
32. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. Journal of Molecular Biology 305:567-580
33. Labrie SJ, Samson JE, Moineau S (2010) Bacteriophage resistance mechanisms. Nat Rev Microbiol 8:317-327
34. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947-2948
35. Lavigne R, Seto D, Mahadevan P, Ackermann H-W, Kropinski AM (2008) Unifying classical and molecular taxonomic classification: analysis of the Podoviridae using BLASTP-based tools. Research in Microbiology 159:406-414
36. Lavigne R, Darius P, Summer E, Seto D, Mahadevan P, Nilsson A, Ackermann H, Kropinski A (2009) Classification of *Myoviridae* bacteriophages using protein sequence similarity. BMC Microbiology 9:224
37. Lawrence JG, Hatfull GF, Hendrix RW (2002) Imbroglios of Viral Taxonomy: Genetic Exchange and Failings of Phenetic Approaches. Journal of Bacteriology 184:4891-4905

38. Li L, Stoeckert CJ, Roos DS (2003) OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. Genome Research 13:2178-2189
39. Lima-Mendez G, Van Helden J, Toussaint A, Leplae R (2008) Reticulate Representation of Evolutionary and Functional Relationships between Phage Genomes. Molecular Biology and Evolution 25:762-777
40. Lima-Mendez G, Toussaint A, Leplae R (2011) A modular view of the bacteriophage genomic space: identification of host and lifestyle marker modules. Research in Microbiology 162:737-746
41. McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, Porwollik S, Ali J, Dante M, Du F, Hou S, Layman D, Leonard S, Nguyen C, Scott K, Holmes A, Grewal N, Mulvaney E, Ryan E, Sun H, Florea L, Miller W, Stoneking T, Nhan M, Waterston R, Wilson RK (2001) Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. Nature 413:852-856
42. Mollov D, Lockhart B, Zlesak D (2013) Complete nucleotide sequence of rose yellow mosaic virus, a novel member of the family Potyviridae. Arch Virol 158:1917-1923
43. Moreno Switt A, Orsi R, den Bakker H, Vongkamjan K, Altier C, Wiedmann M (2013) Genomic characterization provides new insight into *Salmonella* phage diversity. BMC Genomics 14:481
44. Moreno Switt AI, den Bakker HC, Vongkamjan K, Hoelzer K, Warnick LD, Cummings KJ, Wiedmann M (2013) *Salmonella* bacteriophage diversity reflects host diversity on dairy farms. Food Microbiology 36:275-285
45. Myers EW, Miller W (1988) Optimal alignments in linear space. Computer applications in the biosciences : CABIOS 4:11-17
46. Niu YD, McAllister TA, Nash JH, Kropinski AM, Stanford K (2014) Four Escherichia coli O157: H7 Phages: A New Bacteriophage Genus and Taxonomic Classification of T1-Like Phages. PLoS ONE 9:e100426
47. Papanikolaou N, Trachana K, Theodosiou T, Promponas V, Iliopoulos I (2009) Gene socialization: gene order, GC content and gene silencing in *Salmonella*. BMC Genomics 10:597
48. Perler FB, Olsen GJ, Adam E (1997) Compilation and analysis of intein sequences. Nucleic Acids Research 25:1087-1093
49. Rohwer F, Edwards R (2002) The Phage Proteomic Tree: a Genome-Based Taxonomy for Phage. Journal of Bacteriology 184:4529-4535
50. Schwarzer D, Stummeyer K, Gerardy-Schahn R, Mühlenhoff M (2007) Characterization of a Novel Intramolecular Chaperone Domain Conserved in Endosialidases and Other Bacteriophage Tail Spike and Fiber Proteins. Journal of Biological Chemistry 282:2821-2831
51. Söding J (2005) Protein homology detection by HMM–HMM comparison. Bioinformatics 21:951-960
52. Söding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Research 33:W244-W248
53. Spinelli S, Bebeacua C, Orlov I, Tremblay D, Klaholz BP, Moineau S, Cambillau C (2014) Cryo-electron microscopy structure of lactococcal siphophage 1358 virion. J Virol 88:8900-8910
54. Summer EJ, Gill JJ, Upton C, Gonzalez CF, Young R (2007) Role of phages in the pathogenesis of *Burkholderia*, or 'Where are the toxin genes in *Burkholderia* phages?'. Current Opinion in Microbiology 10:410-417
55. Susskind MM, Botstein D (1975) Mechanism of action of Salmonella phage P22 antirepressor. Journal of Molecular Biology 98:413-424
56. Switt AIM, Sulakvelidze A, Wiedmann M, Kropinski AM, Wishart DS, Poppe C, Liang Y (2014) Salmonella phages and prophages – genomics, taxonomy and applied aspects. In: Schatten H, Eisenstark A (eds) Salmonella: Methods and Protocols (Methods in Molecular Biology). Humana Press

57. Tiwari BR, Kim S, Kim J (2012) Complete Genomic Sequence of *Salmonella enterica* Serovar Enteritidis Phage SE2. Journal of Virology 86:7712
58. Turner D, Hezwani M, Nelson S, Salisbury V, Reynolds D (2012) Characterization of the *Salmonella* bacteriophage vB_SenS-Ent1. Journal of General Virology 93:2046-2056
59. Westover KM, Rusinko JP, Hoin J, Neal M (2013) Rogue taxa phenomenon: A biological companion to simulation analysis. Molecular phylogenetics and evolution 69:1-3
60. Winter SE, Thiennimitr P, Winter MG, Butler BP, Huseby DL, Crawford RW, Russell JM, Bevins CL, Adams LG, Tsolis RM, Roth JR, Baumler AJ (2010) Gut inflammation provides a respiratory electron acceptor for *Salmonella*. Nature 467:426-429
61. Xu J, Hendrix RW, Duda RL (2004) Conserved Translational Frameshift in dsDNA Bacteriophage Tail Assembly Genes. Molecular Cell 16:11-21
62. Xu J, Hendrix RW, Duda RL (2013) A Balanced Ratio of Proteins from Gene G and Frameshift-Extended Gene GT Is Required for Phage Lambda Tail Assembly. Journal of Molecular Biology 425:3476-3487
63. Xu J, Hendrix RW, Duda RL (2014) Chaperone–Protein Interactions That Mediate Assembly of the Bacteriophage Lambda Tail to the Correct Length. Journal of Molecular Biology 426:1004-1018
64. Young R, Bläsi U (1995) Holins: form and function in bacteriophage lysis. FEMS Microbiology Reviews 17:195-205
65. Zafar N, Mazumder R, Seto D (2002) CoreGenes: A computational tool for identifying and cataloging "core" genes in a set of small genomes. BMC Bioinformatics 3:12