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.
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.

**Table 1.** 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.

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

Bold face: complete or partial (*) sequence. Dimensions in italics indicate size determination after calibration with catalase crystals or T4 phage tails.
Hosts: *E. coli, Shigella sonnei, Enterobacter cloacae, Serratia marcescens.*

Hosts: *E. coli, S. Gallinarum, S. Pullorum, S. Enteritidis, and S. Typhimurium.*

Hosts: *S. Gallinarum, S. Pullorum, S. Enteritidis.*

<table>
<thead>
<tr>
<th>Phage name</th>
<th>Proposed genera</th>
<th>Genome size (bp)</th>
<th>Mol % G+C</th>
<th>No. of annotated CDSs</th>
<th>Reference</th>
<th>GenBank Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>vB_SenS_AG11</td>
<td>Jerseylikevirus</td>
<td>41,546</td>
<td>49.91</td>
<td>66</td>
<td>-</td>
<td>JX297445</td>
</tr>
<tr>
<td>SETP3</td>
<td>Jerseylikevirus</td>
<td>42,572</td>
<td>49.85</td>
<td>63</td>
<td>[14]</td>
<td>EF177456</td>
</tr>
<tr>
<td>SETP7</td>
<td>Jerseylikevirus</td>
<td>42,749</td>
<td>49.90</td>
<td>68</td>
<td>[14]</td>
<td>KF562865</td>
</tr>
<tr>
<td>SETP13</td>
<td>Jerseylikevirus</td>
<td>42,665</td>
<td>49.79</td>
<td>68</td>
<td>[14]</td>
<td>KF562864</td>
</tr>
<tr>
<td>SS3e syn. KS7</td>
<td>Jerseylikevirus</td>
<td>40,793</td>
<td>50.08</td>
<td>58</td>
<td>[31]</td>
<td>AJ277754</td>
</tr>
<tr>
<td>SE2</td>
<td>Jerseylikevirus</td>
<td>43,221</td>
<td>49.64</td>
<td>61</td>
<td>[57]</td>
<td>JQ007353</td>
</tr>
<tr>
<td>wksl3</td>
<td>Jerseylikevirus</td>
<td>42,633</td>
<td>49.80</td>
<td>64</td>
<td>[7]</td>
<td>JC202565</td>
</tr>
<tr>
<td>vB_SenS-Ent1</td>
<td>Jerseylikevirus</td>
<td>42,391</td>
<td>49.79</td>
<td>58</td>
<td>[58]</td>
<td>HE775250</td>
</tr>
<tr>
<td>vB_SenS-Ent2</td>
<td>Jerseylikevirus</td>
<td>42,093</td>
<td>49.92</td>
<td>56</td>
<td>-</td>
<td>HG934469</td>
</tr>
<tr>
<td>vB_SenS-Ent3</td>
<td>Jerseylikevirus</td>
<td>42,764</td>
<td>49.79</td>
<td>60</td>
<td>-</td>
<td>HG934470</td>
</tr>
<tr>
<td>FSLSP-101</td>
<td>Jerseylikevirus</td>
<td>41,873</td>
<td>50.27</td>
<td>57</td>
<td>[43]</td>
<td>KF148055</td>
</tr>
<tr>
<td>Jersey</td>
<td>Jerseylikevirus</td>
<td>43,447</td>
<td>49.97</td>
<td>69</td>
<td>-</td>
<td>GU196277</td>
</tr>
<tr>
<td>K1H syn. K1-dep(1)</td>
<td>K1glikevirus</td>
<td>41,632</td>
<td>51.17</td>
<td>50</td>
<td>[9]</td>
<td>GU196279</td>
</tr>
<tr>
<td>Kind1</td>
<td>K1glikevirus</td>
<td>42,292</td>
<td>51.27</td>
<td>51</td>
<td>[9]</td>
<td>GU196280</td>
</tr>
<tr>
<td>Kind2</td>
<td>K1glikevirus</td>
<td>42,765</td>
<td>51.35</td>
<td>48</td>
<td>[9]</td>
<td>GU196281</td>
</tr>
<tr>
<td>Kind3</td>
<td>K1glikevirus</td>
<td>43,461</td>
<td>51.15</td>
<td>49</td>
<td>[9]</td>
<td>GU196281</td>
</tr>
<tr>
<td>FSLSP-031</td>
<td>Sp3unalikevirus</td>
<td>42,215</td>
<td>51.07</td>
<td>59</td>
<td>[43]</td>
<td>KC139518</td>
</tr>
</tbody>
</table>

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

Each of the phages listed in Table 2 was colinearized and then examined for overall sequence similarity using progressiveMauve [13] (data not shown); and, DNA sequence identity using EMBOSS Stretcher [45] and CLUSTALW [34], which have been widely used for nucleotide sequence alignment of viruses [4, 42, 59]. Using the latter methodology, three clades were clearly defined (Figure 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 “Sp3unlikevirus” (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 $10^{-5}$ 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.

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

* 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 (Dpol), 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 Dpol 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.
**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.

![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.

Fig 4.
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 (PDB: 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).

“K1glkevirus”

The Escherichia phages K1G, K1H, K1ind1, K1ind2 and K1ind3, comprising the “K1glkevirus” 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-acetylg glucosaminidase 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-acetyl-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-like phages 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 “Sp3unalikevirus” 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 SSSM7 (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 Sp3unalikeviruses 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 taxonomic 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

to paper network.

Conflict of interest

The authors declare that they have no competing interests.
References


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


