# Integration of genomic and proteomic analyses in the classification of the *Siphoviridae* family

Evelien M. Adriaenssens[a,], Rob Edwards[b], John H.E. Nash[c], Padmanabhan Mahadevan[d], Donald Seto[e], Hans-Wolfgang Ackermann[f], Rob Lavigne[g], and Andrew M. Kropinski[c,h]

[a] Centre for Microbial Ecology and Genomics, Genomics Research Institute, University of Pretoria, Lynnwood Road, Pretoria 0028, South Africa

[b] Geology, Mathematics, and Computer Science, San Diego State University, 5500 Campanile Drive, San Diego, CA, 92182, USA

[c] Laboratory for Foodborne Zoonoses, Public Health Agency of Canada, 110 Stone Road West, Guelph, ON, N1G 3W4 Canada

[d] Department of Biology, University of Tampa, Tampa, FL, 33606, USA

[e] Bioinformatics and Computational Biology Program, School of Systems Biology, George Mason University,10900 University Blvd, Manassas, VA 20110, USA

[f] Département de Microbiologie-infectiologie et immunologie, Faculté de médecine, Université Laval, Québec, QC, G1K 7P4, Canada

[g] Department of Biosystems, Laboratory of Gene Technology, KU Leuven, Kasteelpark Arenberg 21 – b2462, 3001 Heverlee, Belgium

[h] Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, N1G 2A1, Canada

Corresponding authors:

Rob Lavigne: rob.lavigne@kuleuven.be

Andrew M. Kropinski: akropins@uoguelph.ca

## Abstract

Using a variety of genomic (BLASTN, ClustalW) and proteomic (Phage Proteomic Tree, CoreGenes) tools we have tackled the taxonomic status of members of the largest bacteriophage family, the *Siphoviridae*. In all over 400 phages were examined and we were able to propose 39 new genera, comprising 216 phage species, and add 62 species to two previously defined genera (*Phic3unalikevirus; L5likevirus*) grouping, in total, 390 fully sequenced phage isolates. Many of the remainder are orphans which the Bacterial and Archaeal Viruses Subcommittee of the International Committee on Taxonomy of Viruses (ICTV) chooses not to ascribe genus status at the time being.

## Introduction

The classification of bacteriophages has been the subject of discussion since their discovery in the beginning of the 20th century (Nelson, 2004). The use of electron microscopy (Ackermann, 2011; Luria et al., 1943; Ruska, 1940) and the discovery of the different forms of nucleic acid (Lwoff et al., 1962), brought together by Bradley in a classification scheme (Bradley, 1967), are still the basis of current phage classification (Ackermann, 2006). In this classification, there is one phage order of dsDNA phages, the *Caudovirales*, containing three families, *Myoviridae*, *Podoviridae* and *Siphoviridae*, the latter being the subject of this paper and, several other families which have not been assigned a higher taxon.

Making phage classification more complicated is the fact that there is no single gene present in all phages upon which a universal scheme could be based (Rohwer and Edwards, 2002). As a result, different research groups have proposed several classification schemes for the taxonomy of these viruses. One such scheme is the Phage Proteomic Tree, a grouping of completely sequenced phages based on protein distances with penalties (Rohwer and Edwards, 2002). Another one is based on the identification of mechanisms leading to cohesion in groups of viruses, with hierarchical levels at higher taxons and the possibility for groups to be reticulate, i.e., one virus can belong to more than one group, called 'modus' (Lawrence et al., 2002). This method was used for classification based on shared genes, resulting in a in a reticulate system in which each phage is characterized by its membership to a set of clusters, with the clusters a possible way to build modi (Lima-Mendez et al., 2008). Proux and colleagues suggested a taxonomy based on comparative genomics of a single structural gene module (Proux et al., 2002). Tetranucleotide usage deviations have also been proposed as a classification tool, but these predicted host range, rather than morphological similarities and were as such less compatible with the International Committee on Virus Taxonomy (ICTV) classification system (Pride et al., 2006). Lavigne and colleagues used BLASTP-related tools for the definition of genera and subfamilies, with cut-off values of respectively 40% and 20% shared proteins between phages. These BLASTP-related tools approaches have resulted in the creation of several proposed subfamilies, including the *Autographivirinae* and *Picovirinae* within the family *Podoviridae* (Lavigne et al., 2008), and the subfamilies *Tevenvirinae*, *Spounavirinae* and *Peduovirinae* in the family *Myoviridae* (Lavigne et al., 2009), which have now been ratified by ICTV.

Following these efforts, this paper attempts to bring order to the chaos currently present in the *Siphoviridae* family of phages. By mid-2014, the NCBI taxonomy browser included over 1200 entries in the *Siphoviridae* taxon, with 603 represented as complete genomes in the RefSeq

database. At the same time, only 31 siphoviruses had been classified in the 2013 ICTV taxonomy report (www.ictvdb.org/virusTaxonomy), grouped into 10 genera, *C2likevirus*, *L5likevirus*, *Lambdalikevirus*, *N15likevirus*, *Phic3unalikevirus*, *Psimunalikevirus*, *Spbetalikevirus*, *T5likevirus*, *Tunalikevirus* and *Yualikevirus*, all named after their type species (see Table 1 for more information). This left an enormous number of unclassified phages. The vast majority of these are currently grouped under the 'unclassified *Siphoviridae*' header at NCBI. Unfortunately, there are also a number of characterized siphoviruses present in the 'unclassified *Caudovirales*' grouping and the 'unclassified phages' group.

In this paper, we combine DNA and protein comparisons within and between groups of phages, in addition to physiological and morphological traits, to define new genera or add members to already existing, ICTV-ratified, genera. We propose 39 new genera, comprising 216 phage species, and add 62 species to two previously defined genera.

## Methods

The bacteriophages surveyed for this analysis were limited to isolates with fully sequenced genomes deposited in the GenBank/EMBL/DDBJ databases, excluding all prophages without a confirmed infectious virion stage. The steps followed in the Methods section are visualized as a flow chart in Supplementary Figure 1. In a first step, whole genome and whole proteome phylogenetic trees were constructed to identify groups of phages without inferring any taxonomic status. For the whole genome DNA trees, fasta-formatted files of the full length phage genomes were downloaded from the NCBI database and were aligned as is using ClustalW 2.0 at the default settings(Larkin et al., 2007), creating a guide-tree. For the proteome-based tree, the Phage Proteomic Tree approach was chosen, using a protein distance result with a penalty value of 10; for the tree-building method we refer to the original publication (Rohwer and Edwards, 2002). These trees were used in parallel in a qualitative manner (no distance cut-offs) to identify clusters of phages which were then investigated for their possible taxonomic grouping (Figure 1, Supplementary Figures 2 & 3). From each of these groups, a type phage was chosen – usually the first genome deposited in the database – and a BLASTN was performed (Altschul et al., 2005) to check for newly deposited genomes similar to the chosen phage. Phages belonging to the same group were also aligned with the visualization software progressiveMauve (Darling et al., 2004, 2010) to assess for genome colinearity. Phages with highly similar genomes (based on the previously collected data) were rendered, if necessary, collinear, and aligned with EMBOSS Stretcher (pairwise global alignment

with the Needlemann-Wunsch algorithm (Myers and Miller, 1988)) and phages showing more than 95% DNA identity were grouped in the same phage species.

With selected phages within and between groups, a CoreGenes 3.5 analysis was performed which compares the predicted proteomes of the phages from a GenBank file in a pairwise way (Mahadevan et al., 2009a, 2009b; Zafar et al., 2002). In accordance with the cut-off values for genera used for the families *Myoviridae* and *Podoviridae* (Lavigne et al., 2009, 2008), phages sharing over 40% proteins were grouped into the same genus, combined with other characteristics – where available - such as genome size and organization, morphology, and packaging and replication mechanism, meaning phages which differed radically in these features but shared 40% proteins will be grouped in separate genera.

Because of mosaicism within the *Siphoviridae* family (Born et al., 2011; Dorscht et al., 2009; Hatfull et al., 2008; Summer et al., 2006), at this time we have chosen not to define any subfamilies according to the criteria previously used in the classification of the *Myoviridae* and *Podoviridae* families.

At present, we have named all the genera tentatively, according to their type phage name, followed by "–like viruses". The quotation marks indicate that these genera have not yet been ratified by ICTV's Executive Committee and are subject to possible change.

# Results

## Phages infecting Gram-positive bacteria

### *Mycobacterium* phages

Bacteriophages infecting the genus *Mycobacterium* are presently the best-represented group within the *Siphoviridae* family, with over 300 sequenced phage genomes deposited in GenBank as of May 2014, including the recent deposition of 138 genomes (Hatfull, 2012a). The vast majority of these phages were isolated as part of the Phage Hunters Program of the University of Pittsburgh and associated universities, and have been grouped in 21 clusters and 37 subclusters in the Mycobacteriophage Database (Hatfull and Hendrix, phagesdb.org).

The division in clusters is based on whole genome comparisons, with phages belonging to the same cluster sharing a common genome organization with many orthologous genes and several distinct biophysical characteristics (Hatfull, 2012b). Our analyses corroborate this division and we are currently defining 14 new genera corresponding with the clusters B, D, E, F, G, H, I, J, K, L, M, N, O

and P (Table 1), while the existing, ICTV recognized genus *L5likevirus* will be expanded with the phages belonging to cluster A. Of interest is the presence in this genus of *Rhodococcus* phages RGL3 and RER2, which share 38.8% of their proteins with the type phage L5 (Petrovski et al., 2013). progressiveMauve (Darling et al., 2010) analysis of these three phages also showed DNA relatedness over the entire length of the genomes (data not shown). The recent clusters Q to T have not been included in this analysis. Phage Patience, the only deposited member of cluster U, was incorporated in the genus "Barnyard-like viruses" (cluster H) because of a shared protein content with phage Barnyard of over 40%.

Generally, *Mycobacterium* phage morphologies are similar, with head sizes of 50 to 60 nm and tail lengths of 150 – 200 nm (Mycobacteriophage Database). An exception is the proposed genus "Corndog-like viruses" (Cluster 0, Table 1), the members of which have extremely elongated heads with length to width ratios of 4:1 or greater.

Several of the clusters with a large number of members have been further subdivided in subclusters in the Mycobacteriophage database. Phages belonging to the same subcluster share a high percentage of DNA identity and are in the proteomic and DNA trees clearly recognizable as separate groups (Figure 1, Supplementary Figures 2 & 3). The taxonomic consequences of this subdivision will be attended to in the discussion section of this paper.

### "Andromeda-like viruses"

The six members of this genus (Andromeda, Eoghan, Curly, Taylor, Gemini and Finn) are *Bacillus pumilis*-infecting siphoviruses with genomes of ~49.5 kb, and a G+C content of approximately 42% (Lorenz et al., 2013). Their heads measure 41 to 60 nm in diameter and the flexible tails vary in length from 110 to 149 nm. All have ~820 bp terminal repeats, similar gene content, and a comparable host range.

### "C5-like viruses"

The genus "C5-like viruses" comprises only two phage species, *Lactobacillus* phages c5 and LL-Ku, both of which infect *L. delbrueckii* (Accolas and Spillmann, 1979; Riipinen et al., 2011). These phages were previously characterized lactic acid bacteriophages with isometric heads and a *cos*-type packaging mechanism, and were later found to be genetically closely related (Riipinen et al., 2011).

### "IEBH-like viruses"

Two *Bacillus cereus*-infecting phages are currently classified in this genus, IEBH and 250 (Lee and Park, 2010; Smeesters et al., 2011). The type phage IEBH has an isometric head of 55 nm and a long (150 nm) non-contractile tail displaying transverse tail discs. For phage 250 no dimensions have been reported. Both phages share 81.7% DNA identity, but only have 54.7% shared proteins. This is due to a lower number of ORFs annotated in phage 250.

### *Phic3unalikevirus*

This ICTV-ratified genus, originally only contained the type virus *Streptomyces* phage φC31 (Smith et al., 1999), is now supplemented with two other *Streptomyces*-infecting phages, φBT1 (Gregory et al., 2003) and TG1 (Foor et al., 1985). The genome organization of φBT1 is very similar to that of φC31, but there is evidence of mosaicism between the two genomes (Gregory et al., 2003). No TEM pictures are publically available for these phages, so no comparative morphological analysis could be performed.

### "phiFL-like viruses"

This genus represents a group of temperate siphoviruses active against *Enterococcus faecalis,* possessing 36.3 to 40.3 kb genomes (mol% G+C content of 34-37) which encode 59 - 74 proteins (Yasmin et al., 2010). Morphologically, their isometric heads measure 48 – 53 nm in diameter with tails of 205-229 nm in length. Based on a shared DNA identity of over 95 %, the phages described by Yasmin and colleagues were classified in three separate species, *Enterococcus phage Phifl1* comprising the isolates ΦFL1A, ΦFL1B and ΦFL1C, *Enterococcus phage Phifl2* with the isolates ΦFL2A and ΦFL2B, and *Enterococcus phage Phifl3* with ΦFL3A and ΦFL3B (Supplementary Table 1).

### "phiLJ1-like viruses"

This genus, named after *Lactobacillus* phage φJL-1, contains two other members, *Lactobacillus* phage ATCC 8014-B1 and *Pediococcus* phage clP1 (Briggiler Marcó et al., 2012; Kelly et al., 2012; Lu et al., 2005, 2003). The latter show a very high degree of DNA identity (92.7%) and shared protein content (96.5%), but are clearly distinct species according to their host range. Both phages share 63.0% of protein identity with φJL-1. Morphology of the type phage is an isometric head of about 59 nm diameter and a non-contractile tail of 182 nm long and 11 nm wide (Lu et al., 2003).

### "R4-like viruses"

The broad-host range temperate *Streptomyces* phage R4 (Chater and Carter, 1979) possesses a capsid of 64 nm in diameter and a laterally striated flexible tail of 190 nm length by 12.5 nm width. Its 51.1 kb genome, which is terminated by 11bp 3'-cohesive termini (CGCCGTGTCTT) (Mitsui and Takahashi, 1992), was until recently considered a genomic orphan. This changed with the isolation of *S. lividans* phage ELB20 and *S. hydroscopicus* phage φHau3 (Smith et al., 2013) with which R4 shares 97.2%, and 61.1% DNA sequence identity, respectively. At the protein level ELB20 and φHau3 share 90.7% and 61.6% proteins homologs.

### "Lika-like viruses"

Three other temperate viruses identified along with ELB20 (r4likevirus), and showing marginally larger genomes and a lower degree of DNA sequence identity to the genus *R4likevirus*, are *Streptomyces lividans* phages Lika, Zemiya and Sujidale (Smith et al., 2013). They all possess 51.0-51.5 kb genomes which are 65.7-65.8- mol%G+C and lack the tRNA genes found in the r4likeviruses. In comparison to phage Lika, the genomes of Sujidade and Zemiya are 95.1 and 89.8% identical and these three viruses share 97.3-98.7% protein homologs. Therefore, the *Likalikevirus* genus contains two species.

### "Sap6-like viruses"

The two members currently belonging to this genus, Sap6 and BC-611, both infect *Enterococcus faecalis* and were isolated as therapeutic agents against this bacterium (Horiuchi et al., 2012; Lee and Park, 2012). With 77.7 % sequence identity and 84.1 % shared proteins; these phages are clearly related, yet distinct. Both phages were reported as members of the family *Siphoviridae*, but no particle dimensions were published (Horiuchi et al., 2012; Lee and Park, 2012).

### The genera "Sfi11-like viruses" and "Sfi21-DT1-like viruses"

Phages belonging to these proposed genera all infect *Streptococcus thermophilus* (Supplementary Table 1). These two groups of phages were first recognized based on the presence of two ("Sfi21-DT1-like viruses") or three ("Sfi11-like viruses") major structural proteins on an SDS-PAGE profile, and also corresponding with the presence of *cos* and *pac* sites, respectively (Le Marrec et al., 1997). We have given the former genus a double name, since the term "Sfi21-like" was coined first, but phage DT1 was isolated first historically (Le Marrec et al., 1997; Proux et al., 2002; Tremblay and Moineau, 1999). Both groupings "Sfi11-like" and "Sfi21-like" were previously proposed, comprising many of the current *Streptococcus* phages, but also phages infecting other low GC-content Gram

positives (Brüssow and Desiere, 2001; Proux et al., 2002). While these proposals focused on similarity in the structural region of the genome, we are taking the whole genome into account and thus limit membership of these genera to the *Streptococcus* phages, which share more than 40% of their proteins within both genera.

### "SK1-like viruses"

This genus is derived from the 936-like phage group of dairy phages infecting *Lactococcus lactis*, a grouping currently without taxonomic validity. It is named after *Lactococcus* phage SK1, the type species and first fully sequenced isolate of the group (Chandry et al., 1997), since unfortunately phage 936 has not been sequenced completely. This genus also comprises a large group of Australian dairy phages (Castro-Nallar et al., 2012) which have been grouped into five species, with isolates belonging to the same species sharing over 95% DNA identity (Supplementary Table 1).

### "TP21-like viruses"

The type species of this proposed genus is *Bacillus* phage Tp21-l (Loessner et al., 1997), one of three *Bacillus* phages designated TP21 (Klumpp et al., 2010). Analysis of these phages by Klumpp and colleagues (2010) revealed that these are clearly separate phages and the authors suggest the use of suffixes to indicate the different isolates. The genomes of phages TP21-T (Walter and Aronson, 1991) and TP21-H (He et al., 1978) have not been deposited in a public database and are therefore not included in this analysis. The other member of this genus is *Bacillus* phage BMBtp2 (Dong et al., 2013), sharing 86.9% DNA identity with TP21-L and 82.1% shared proteins. Both phages share the same morphology, an isometric head and non-contractile flexible tail, with the dimensions for TP21-L 58.5 nm head diameter and a tail of 144.8 x 11.0 nm (Klumpp et al., 2010).

### "Wbeta-like viruses"

This genus comprises a set of *Bacillus anthracis* phages derived from the temperate phage Wβ (WBeta) (Schuch and Fischetti, 2006) which are classified in the same species because of their common origin. The lytic derivatives Fah, Cherry and γ (Gamma) have been reported previously to be in essence the same phage containing variations at three places in the genome (Fouts et al., 2006). The "Wbeta-like viruses" share a common morphology, an isometric head of 56 nm and a non-contractile tail of 200 nm with a small base plate and fibrous tail extension of about 63 nm (dimension as described for Wβ) (Schuch and Fischetti, 2006).

### The Staphylococcus-infecting genera "3A-like viruses", "77-like viruses" and "phiETA-like viruses"

The classification of *Staphylococcus*-infecting siphoviruses is described in detail in a previous publication which groups 40 phages in three new genera (Gutiérrez et al., 2014). The first two genera reflect a previous classification scheme of *Staphylococcus* phages by an anterior ICTV Bacterial Virus Subcommittee (Ackermann and DuBow, 1987) and summarize numerous studies of phage morphology, serology, and physicochemical properties of phages and their DNAs.

## Phages infecting Gram-negative bacteria

### "Chi-like viruses"

The type phage of this genus, *Salmonella* phage χ (Chi), was found to attack only motile *Salmonella* strains with its receptor at the base of the flagellum (Schade et al., 1967). Recently, its genome and that of several others (*Salmonella* phages iEPS5, SPN19, FSL SP-030, FSL SP-039, FSL SP-088, and FSL SP-124) were sequenced, with the isolates FSL SP-030 and FSL SP-039 grouped in species Fslp30 and the isolates FSL SP-088 and FSL SP-124 grouped into species Fslp88 based on above 95% DNA identity (Choi et al., 2013; Lee et al., 2013; Moreno Switt et al., 2013). The number of shared genes with the type species Chi ranges from 90 to 93%. The virion morphology is only known for Chi and iEPS5, with comparable dimensions, showing an isometric head and long flexible tail (Choi et al., 2013; Schade et al., 1967).

### "D3-like viruses"

Type species of this genus is temperate *Pseudomonas* phage D3, which has an isometric head 55 nm in diameter and a long flexible tail (113 nm x 7 nm) possessing six tail fibers with terminal knobs (Kropinski, 2000), which has the ability to seroconvert lysogens (Newton et al., 2004). The second phage in this genus, the virulent *Pseudomonas* phage phiPMG1, does not share this ability, but was grouped in the genus because of a high DNA identity and shared protein content of 73.7% (Krylov et al., 2012).

### "D3112-like viruses"

*Pseudomonas aeruginosa* phage D3112 (Wang et al., 2004) and its associated species (JD024, JBD5, JBD26 and LPB1) are all temperate transposable viruses possessing 34.6-37.8 kb genomes, showing DNA identities and shared protein content of over 80%. Originally mistakenly considered to be members of the "Mu-like virus", along with *Pseudomonas* phage B3, whole genome proteomic

analyses of this genus (data not shown) reveals that Mu and D3112 share only 11 homologs (20%), while B3 and D3112 share 19 homologs (34.5%) (Braid et al., 2004). These values are too low for these phages to be considered part of the same genus. In addition, Mu is a member of the *Myoviridae*, while the other phages are members of the *Siphoviridae*.

### "HK578-like viruses"

*Escherichia* phage HK578 was chosen as type species for this genus, as oldest isolate of the group (Dhillon et al., 1970). Members of this genus infect a range of *Enterobacteriaceae* such as *Escherichia, Sodalis* and *Shigella*, respectively phages HK578, JL1, and SSL-2009a, SO-1 and EP23 infecting both *E. coli* and *Shigella sonnei* (Chang and Kim, 2011; Li et al., 2010; Pan et al., 2013). Phages share a common genome organization and over 40% shared proteins, with some evidence of rearrangements in phage SSL-2009a when compared with the others (Chang and Kim, 2011).

### "Jersey-like viruses"

Bacteriophages belonging to this proposed genus have comparable genome sizes (40 – 44 kb) and morphology, with a head size of approximately 60 nm and tail of 119 nm for the type species *Salmonella phage Jersey* (Ackermann and Gershman, 1992; Ackermann et al., 1972). Phage SETP3, part of the *Salmonella* Enteritidis typing set, was previously recognized as a Jersey-like phage based on its morphology (De Lappe et al., 2009). The phages Sse3, SE2, vB_SenS-Ent1 have all been described as virulent phages infecting multiple *Salmonella enterica* serovars (Kang et al., 2013; S.-H. Kim et al., 2012; Tiwari et al., 2012; Turner et al., 2012). Relative to phage Jersey, the phages currently classified in this genus share at least 60% DNA identity and over 68% protein content. This genus will be the subject of a separate paper by H. Anany and colleagues (personal communication).

### "P23-45-like viruses"

The two current members of this genus are *Thermus thermophilus* lytic phages P23-45, appointed as type phage, and P74-26, which share 92.2% DNA identity and 95% protein content (Minakhin et al., 2008). The most interesting feature of these phages is their tail morphology, having extremely long (~800 nm) tails, which is reflected in the size of the tail tape measure protein, 5002 residues long for P23-45 (Minakhin et al., 2008). A similar morphology has been reported for *Thermus* phage TSP4, but no sequence data is available for this phage (Lin et al., 2010). Another unusual feature was, as reported by Minakhin and colleagues, the presence of homopolypurine-homopolypyrimidine mirror repeat sequences forming triplexes that affect DNA synthesis, but the biological function remains unclear.

### "phiCBK-like viruses"

The genus "phiCBK-like viruses" represents a group of *Caulobacter* phages with large, prolate heads and correspondingly large genomes of over 200 kb (Gill et al., 2012). The type phage, φCbK, has a head of 195 - 205 by 56 -64 nm and a tail of 275 - 300 nm, with measurements varying slightly between publications (Agabian-Keshishian and Shapiro, 1970; Gill et al., 2012). Gill and colleagues (2012) reported phages φCbK, Karma, Magneto and Swift as a cohesive group showing over 88% DNA identity and Rogue more distantly related at 62.6% DNA identity with φCbK. Overall, the phages belonging to this genus share genome size (205 – 225 kb), organization, gene synteny, the presence of a large number of tRNAs and particle morphology (Gill et al., 2012).

### "phiE125-like viruses"

This genus contains three phages infecting *Burkholderia* showing similar genome sizes and organization, with the type phage φE125 specific for *B. mallei* (Woods et al., 2002). The morphology associated with this genus is an isometric head (63 nm for φE125) and long, flexible tail of about 200nm length (DeShazer, 2004; Woods et al., 2002). Deshazer and colleagues (2004) recognized both the relatedness between φE125 and phage φ1026b, sharing a large number of proteins, and also the mosaic nature of the latter phage, with the head morphogenesis genes resembling more those of *Pseudomonas* phages. Despite this, the shared protein content of 80.6% warrants the grouping in the same genus. In addition, phage φ644-2 (unpublished) is also a member of this genus with 72.2% of proteins in common with φE125.

### "Xp10-like viruses"

This genus comprises five viruses infecting *Xanthomonas* species, with phage Xp10 being the type species (Yuzenkova et al., 2003). The main distinguishing feature of the five phages, Xp10, OP1, φXo411, phiL7 and CP1, is the presence of a single-subunit DNA-dependent RNA polymerase, like the members of the *Autographivirinae* subfamily within the *Podoviridae* (Ahmad et al., 2014; Yasuhiro Inoue et al., 2006; Lavigne et al., 2008; Lee et al., 2009, 2006; Yuzenkova et al., 2003). A second feature is the presence in the genome of several HNH-type homing endonucleases (Lee et al., 2007). Furthermore, the phages share a common morphology (isometric head of 53 nm and tail of 173 nm for Xp10), genome size and genome organization, all sharing over 58% of proteins with the type phage.

## Discussion

It has been ten years since Daniel Nelson wrote the famous quote "we agree to disagree" about the taxonomy of bacteriophages, reflecting different views in the phage community (Nelson, 2004). In the meantime, the number of phage genomes sequenced has exploded and the majority of them remain unclassified. Previous efforts by Lavigne and colleagues (2008, 2009) resulted in the creation of new ICTV-ratified subfamilies and genera in the families *Myoviridae* and *Podoviridae*, paving the road for further phage taxonomy. The proteome-based genus demarcation criteria used in these publications were included in this study.

Our results are based upon comparative genomics and proteomics, backed up by phage morphology and physiology. Combining these approaches, it becomes clear that while mosaicism is a defining characteristic of phages it does not necessarily complicate the creation of a logical taxonomy for many viruses, particularly those that are virulent. Nevertheless, the presence of mosaicism in certain siphovirus groups prevented the creation of consistent subfamilies. The criterion of 20% similar proteins could in this case occur within a single exchange module without overall gene content identity between the phages in question. Therefore, we have at this time only defined genera, not subfamilies, in this virus family with pervasive mosaicism.

We have deliberately chosen not to tackle the *Lambalikevirus* genus since preliminary proteomic data suggest that the archetypal lambdoid coliphages HK022 and HK97 are more closely related to each other than to coliphage λ (Dhillon et al., 1970; Kameyama et al., 1999). This is also visible on the proteomic tree (Figure 1, supplementary Figure 2) where HK022 and HK97 cluster together, but seem unrelated to Lambda which stands alone. In addition, D. Refardt has recently sequenced a number of lambdoid phages and the initiative to fill out this genus depends on the incorporation of these newly emerging data.

Transposable phages have been an important tool for bacterial genetics and a broadly present as prophages in bacterial genomes. From a genome organization perspective, Mu-related phages have been shown to have either long-contractile or non-contractile tails (sipho- and myoviruses). Hulo and colleagues propose to group these phages into two subfamilies, "Siphosaltovirinae" and "Myosaltovirinae" (Salto = "to jump") which could/should be joined into a single family ("Saltoviridae") if/when the current *Siphoviridae*/*Myoviridae* taxonomy level is abandoned (discussed further in this section) (Hulo et al., this issue). A similar solution can be envisioned for the lambdoid phages across the *Siphoviridae* and *Podoviridae* family. In a recent study by Grose and Casjens, investigating phages infecting the host family *Enterobacteriaceae*, they found a "lambda-

like supercluster" which was a large transitive, closed set formed of temperate lambda-like phages with syntenic gene functions and a similar transcription pattern (Grose and Casjens, 2014). This could be the basis of a new "Lambdaviridae" family containing "Myolambdavirinae" and "Podolambdavirinae" subfamilies, but this falls beyond the scope of this paper.

For species demarcation, we used pairwise DNA similarities using a global alignment algorithm (Needleman-Wunsch). This type of global alignment in combination with BLASTN has been used for the classification of eukaryotic viruses, PASC (PAirwise Sequence Comparison, http://www.ncbi.nlm.nih.gov/sutils/pasc/viridty.cgi) (Bao et al., 2012). In the case of filoviruses similarities of 50-64% (BLASTN) and 64-72% (Needleman–Wunsch) has been suggested for membership of the same genus; whereas 30-41%, and 52-58% respectively suggested membership in different genera. For the plant-pathogenic ssDNA viruses of the genus *Mastrevirus* (family *Geminiviridae*), a similar PASC classification system has been used with a cut-off of 78% to group isolates within one species and 94% to group them into the same strain (Muhire et al., 2013). The chosen percentage of DNA identity for species demarcation is highly dependent on the order and family of viruses. Given the phenotypic differences associated with phages showing a high degree of DNA similarity, we chose a threshold 95% DNA identity over the whole length of the genomes for grouping isolates within the same species.

During the genome analyses, distinct clades within several genera were discovered (see Figure 1), usually showing a DNA identity greater than 60% as calculated with EMBOSS Stretcher (data not shown). This has been observed before for the *Mycobacterium* phages, corresponding with the subclusters in the Mycobacteriophage Database (Hatfull and Hendrix,phagesdb.org). A similar DNA sequence-based grouping was also noted for the *Streptomyces* phages (Smith et al., 2013). Based on these data, we acknowledge the existence of potential "subgenera". These groupings hold no taxonomic relevance as of yet, but might form the basis of possible future genome-based taxonomic rearrangements, i.e. the reconfiguration of the morphology-based families *Myoviridae, Siphoviridae* and *Podoviridae* to better incorporate genomic data, taking into account mosaicism and chimeric phages, as proposed with the introduction of the "Saltoviridae" mentioned earlier. However, before such a massive reorganization can take place, there is a need for established and well-characterized genera in each of the families concerned. Additionally, by suggesting such a rearrangement of the order *Caudovirales*, we hope to get this specific taxonomy discussion started and get feedback from the community.

Members of the phage community want a simple method by which they can determine whether their recently completed genome fits within a given phage genus or subfamily. Comparative proteomic analyses using the CoreGenes suite of programs is useful but care must be taken that homologs exist over the length of the genome and are not selectively grouped. In addition, a pervasive error in the use of comparative proteomics to classify phages is that in certain cases phages have been severely over or under-annotated, DNA sequence identity employing EMBOSS Stretcher or similar programs only works for co-linear genomes, and will be problematic with the larger phages in which genomic rearrangements have occurred. The Prokaryote Virus Subcommittee of the ICTV intends to prepare readily implementable molecular taxonomic tools to aid phage scientists correctly address the taxonomic position of their newly sequenced phages. At this time, to classify a newly isolated phage, the same approach as in the methods section can be used, starting from the BLASTN step, this to exclude the need for the high computational power necessary for generation of the trees.

## Acknowledgements.

This research was supported by Grant G.0323.09 from the 'Fonds voor Wetenschappelijk Onderzoek' (FWO, Belgium). EMA holds a Vice-Chancellor's Postdoctoral Fellowship at the University of Pretoria.

## Figure titles

**Figure 1.** Phage proteomic tree (Rohwer & Edwards, 2002) of the phage genomes belonging to the family *Siphoviridae* classified in this study. The proposed genera are colored randomly in different colors. The tree was rooted at the node of the "phiCBK-like viruses". No nodes were collapsed and taxon names were replaced with their respective proposed genera. The full version of this tree is Supplementary Figure 2.

## References


Accolas, J.-P., Spillmann, H., 1979. Morphology of bacteriophages of *Lactobacillus bulgaricus*, *L. lactis* and *L. helveticus*. J Appl Microbiol 47, 309–319. doi:10.1111/j.1365-2672.1979.tb01760.x

Ackermann, H.-W., 2006. Classification of bacteriophages, in: Calendar, R. (Ed.), The Bacteriophages. Oxford University Press, New York, NY, USA, pp. 8–17.

Ackermann, H.-W., 2011. Visualization of bacteriophage lysis in the hypermicroscope. Bacteriophage 1, 183–187.

Ackermann, H.-W., Berthiaume, L., Kasatiya, S.S., 1972. Morphologie des phages de lysotypie de *Salmonella paratyphi* B (schéma de Felix et Callow). Can J Microbiol 18, 77–81. doi:10.1139/m72-012

Ackermann, H.-W., DuBow, M.S., 1987. Viruses of prokaryotes. CRC Press, Boca Raton, FL, USA.

Ackermann, H.-W., Gershman, M., 1992. Morphology of phages of a general *Salmonella* typing set. Res Virol 143, 303–310. doi:10.1016/S0923-2516(06)80118-4

Agabian-Keshishian, N., Shapiro, L., 1970. Stalked bacteria: Properties of deoxyribonucleic acid bacteriophage phiCbK. J Virol 5, 795–800.

Ahmad, A.A., Ogawa, M., Kawasaki, T., Fujie, M., Yamada, T., 2014. Characterization of bacteriophages Cp1 and Cp2, the strain-typing agents for *Xanthomonas axonopodis* pv. citri. Appl Environ Microbiol 80, 77–85. doi:10.1128/AEM.02310-13

Altschul, S.F., Wootton, J.C., Gertz, E.M., Agarwala, R., Morgulis, A., Schäffer, A.A., Yu, Y.-K., 2005. Protein database searches using compositionally adjusted substitution matrices. FEBS J 272, 5101–5109. doi:10.1111/j.1742-4658.2005.04945.x

Bae, T., Baba, T., Hiramatsu, K., Schneewind, O., 2006. Prophages of *Staphylococcus aureus* Newman and their contribution to virulence. Mol Microbiol 62, 1035–1047. doi:10.1111/j.1365-2958.2006.05441.x

Bao, Y., Chetvernin, V., Tatusova, T., 2012. PAirwise Sequence Comparison (PASC) and its application in the classification of filoviruses. Viruses 4, 1318–1327. doi:10.3390/v4081318

Battaglioli, E.J., Baisa, G.A., Weeks, A.E., Schroll, R.A., Hryckowian, A.J., Welch, R.A., 2011. Isolation of generalized transducing bacteriophages for uropathogenic strains of *Escherichia coli*. Appl Environ Microbiol 77, 6630–6635. doi:10.1128/AEM.05307-11

Born, Y., Fieseler, L., Marazzi, J., Lurz, R., Duffy, B., Loessner, M.J., 2011. Novel virulent and broad-host-range *Erwinia amylovora* bacteriophages reveal a high degree of mosaicism and a relationship to *Enterobacteriaceae* phages. Appl Environ Microbiol 77, 5945–5954. doi:10.1128/AEM.03022-10

Bradley, D.E., 1967. Ultrastructure of bacteriophage and bacteriocins. Bacteriol Rev 31, 230–314.

Braid, M.D., Silhavy, J.L., Kitts, C.L., Cano, R.J., Howe, M.M., 2004. Complete genomic sequence of bacteriophage B3, a Mu-like phage of *Pseudomonas aeruginosa*. J Bacteriol 186, 6560–6574. doi:10.1128/JB.186.19.6560-6574.2004

Briggiler Marcó, M., Garneau, J.E., Tremblay, D., Quiberoni, A., Moineau, S., 2012. Characterization of two virulent phages of *Lactobacillus plantarum*. Appl Environ Microbiol 78, 8719–8734. doi:10.1128/AEM.02565-12

Brüssow, H., Desiere, F., 2001. Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. Mol Microbiol 39, 213–223. doi:10.1046/j.1365-2958.2001.02228.x

Castro-Nallar, E., Chen, H., Gladman, S., Moore, S.C., Seemann, T., Powell, I.B., Hillier, A., Crandall, K.A., Chandry, P.S., 2012. Population genomics and phylogeography of an Australian dairy factory derived lytic bacteriophage. Genome Biol Evol 4, 382–393. doi:10.1093/gbe/evs017

Ceyssens, P.-J., Mesyanzhinov, V. V, Sykilinda, N., Briers, Y., Roucourt, B., Lavigne, R., Robben, J., Domashin, A., Miroshnikov, K.A., Volckaert, G., Hertveldt, K., 2008. The genome and structural proteome of YuA, a new *Pseudomonas aeruginosa* phage resembling M6. J Bacteriol 190, 1429–1435. doi:10.1128/JB.01441-07

Chandry, P.S., Moore, S.C., Boyce, J.D., Davidson, B.E., Hillier, A.J., 1997. Analysis of the DNA sequence, gene expression, origin of replication and modular structure of the *Lactococcus lactis* lytic bacteriophage sk1. Mol Microbiol 26, 49–64. doi:10.1046/j.1365-2958.1997.5491926.x

Chang, H.-W., Kim, K.-H., 2011. Comparative genomic analysis of bacteriophage EP23 infecting *Shigella sonnei* and *Escherichia coli*. J Microbiol 49, 927–934. doi:10.1007/s12275-011-1577-0

Chater, K.F., Carter, A.T., 1979. A new, wide host-range, temperate bacteriophage (R4) of Streptomyces and its interaction with some restriction-modification systems. J Gen Microbiol 115, 431–442. doi:10.1099/00221287-115-2-431

Choi, Y., Shin, H., Lee, J.-H., Ryu, S., 2013. Identification and characterization of a novel flagellum-dependent *Salmonella*-infecting bacteriophage, iEPS5. Appl Environ Microbiol 79, 4829–4837. doi:10.1128/AEM.00706-13

Christie, G.E., Matthews, A.M., King, D.G., Lane, K.D., Olivarez, N.P., Tallent, S.M., Gill, S.R., Novick, R.P., 2010. The complete genomes of *Staphylococcus aureus bacteriophages* 80 and 80α--implications for the specificity of SaPI mobilization. Virology 407, 381–390. doi:10.1016/j.virol.2010.08.036

Chung, I.-Y., Cho, Y.-H., 2012. Complete genome sequences of two *Pseudomonas aeruginosa* temperate phages, MP29 and MP42, which lack the phage-host CRISPR interaction. J Virol 86, 8336. doi:10.1128/JVI.01127-12

Crutz-Le Coq, A.-M., Cesselin, B., Commissaire, J., Anba, J., 2002. Sequence analysis of the lactococcal bacteriophage bIL170: insights into structural proteins and HNH endonucleases in dairy phages. Microbiology 148, 985–1001.

Daniel, A., Bonnen, P.E., Fischetti, V.A., 2007. First complete genome sequence of two *Staphylococcus epidermidis* bacteriophages. J Bacteriol 189, 2086–2100. doi:10.1128/JB.01637-06

Darling, A.C.E., Mau, B., Blattner, F.R., Perna, N.T., 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14, 1394–1403. doi:10.1101/gr.2289704

Darling, A.E., Mau, B., Perna, N.T., 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5, e11147. doi:10.1371/journal.pone.0011147

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. J Med Microbiol 58, 86–93. doi:10.1099/jmm.0.000034-0

Demerec, M., Fano, U., 1945. Bacteriophage-resistant mutants in *Escherichia coli*. Genetics 30, 119–136.

Depew, J., Zhou, B., McCorrison, J.M., Wentworth, D.E., Purushe, J., Koroleva, G., Fouts, D.E., 2013. Sequencing viral genomes from a single isolated plaque. Virol J 10, 181. doi:10.1186/1743-422X-10-181

DeShazer, D., 2004. Genomic diversity of *Burkholderia pseudomallei* clinical isolates: subtractive hybridization reveals a *Burkholderia mallei*-specific prophage in *B. pseudomallei* 1026b. J Bacteriol 186, 3938–3950. doi:10.1128/JB.186.12.3938-3950.2004

Desiere, F., Lucchini, S., Brüssow, H., 1998. Evolution of *Streptococcus thermophilus* bacteriophage genomes by modular exchanges followed by point mutations and small deletions and insertions. Virology 241, 345–356. doi:10.1006/viro.1997.8959

Deveau, H., Barrangou, R., Garneau, J.E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D. a, Horvath, P., Moineau, S., 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J Bacteriol 190, 1390–400. doi:10.1128/JB.01412-07

Dhillon, T.S., Chan, Y.S., Sun, S.M., Chau, W.S., 1970. Distribution of coliphages in Hong Kong sewage. Appl Environ Microbiol 20, 187–191.

Dong, Z., Peng, D., Wang, Y., Zhu, L., Ruan, L., Sun, M., 2013. Complete genome sequence of *Bacillus thuringiensis* bacteriophage BMBtp2. Genome Announc 1, e00011–12. doi:10.1128/genomeA.00011-12

Dorscht, J., Klumpp, J., Bielmann, R., Schmelcher, M., Born, Y., Zimmer, M., Calendar, R., Loessner, M.J., 2009. Comparative genome analysis of *Listeria* bacteriophages reveals extensive mosaicism, programmed translational frameshifting, and a novel prophage insertion site. J Bacteriol 191, 7206–7615. doi:10.1128/JB.01041-09

Fan, X., Teng, T., Wang, H., Xie, J., 2012. Biology of a novel mycobacteriophage, SWU1, isolated from Chinese soil as revealed by genomic characteristics. J Virol 86, 10230–10231. doi:10.1128/JVI.01568-12

Foor, F., Roberts, G.P., Morin, N., Snyder, L., Hwang, M., Gibbons, P.H., Paradiso, M.J., Stotish, R.L., Ruby, C.L., Wolanski, B., Streicher, S.L., 1985. Isolation and characterization of the *Streptomyces cattleya* temperate phage TG1. Gene 39, 11–16. doi:10.1016/0378-1119(85)90101-5

Ford, M.E., Sarkis, G.J., Belanger, A.E., Hendrix, R.W., Hatfull, G.F., 1998a. Genome structure of mycobacteriophage D29: implications for phage evolution. J Mol Biol 279, 143–164. doi:10.1006/jmbi.1997.1610

Ford, M.E., Stenstrom, C., Hendrix, R.W., Hatfull, G.F., 1998b. Mycobacteriophage TM4: genome structure and gene expression. Tuber lung Dis 79, 63–73. doi:10.1054/tuld.1998.0007

Fouts, D.E., Rasko, D.A., Cer, R.Z., Jiang, L., Fedorova, N.B., Shvartsbeyn, A., Vamathevan, J.J., Tallon, L., Althoff, R., Arbogast, T.S., Fadrosh, D.W., Read, T.D., Gill, S.R., 2006. Sequencing *Bacillus anthracis* typing phages gamma and cherry reveals a common ancestry. J Bacteriol 188, 3402–3408. doi:10.1128/JB.188.9.3402-3408.2006

García, P., Martínez, B., Obeso, J.M., Lavigne, R., Lurz, R., Rodríguez, A., 2009. Functional genomic analysis of two *Staphylococcus aureus* phages isolated from the dairy environment. Appl Environ Microbiol 75, 7663–7673. doi:10.1128/AEM.01864-09

Gill, J.J., Berry, J.D., Russell, W.K., Lessor, L., Escobar Garcia, D.A., Hernandez, D., Kane, A., Keene, J., Maddox, M., Martin, R., Mohan, S., Thorn, A.M., Russell, D.H., Young, R., 2012. The *Caulobacter crescentus* phage phiCbK: genomics of a canonical phage. BMC Genomics 13, 542. doi:10.1186/1471-2164-13-542

Gregory, M.A., Till, R., Smith, M.C.M., 2003. Integration site for *Streptomyces* phage phiBT1 and development of site-specific integrating vectors. J Bacteriol 185, 5320–5323.

Grose, J.H., Casjens, S.R., 2014. Understanding the enormous diversity of bacteriophages: The tailed phages that infect the bacterial family Enterobacteriaceae. Virology 468-470, 421–443. doi:10.1016/j.virol.2014.08.024

Guglielmotti, D.M., Deveau, H., Binetti, A.G., Reinheimer, J.A., Moineau, S., Quiberoni, A., 2009. Genome analysis of two virulent *Streptococcus thermophilus* phages isolated in Argentina. Int J Food Microbiol 136, 101–109. doi:10.1016/j.ijfoodmicro.2009.09.005

Gutiérrez, D., Adriaenssens, E.M., Martínez, B., Rodríguez, A., Lavigne, R., Kropinski, A.M., García, P., 2014. Three proposed new bacteriophage genera of staphylococcal phages: “3alikevirus”, “77likevirus” and “Phietalikevirus”. Arch Virol 159, 389–398. doi:10.1007/s00705-013-1833-1

Gutiérrez, D., Martínez, B., Rodríguez, A., García, P., 2010. Isolation and characterization of bacteriophages infecting *Staphylococcus epidermidis*. Curr Microbiol 61, 601–608. doi:10.1007/s00284-010-9659-5

Hatfull, G.F., 2010. Mycobacteriophages: genes and genomes. Annu Rev Microbiol 64, 331–356. doi:10.1146/annurev.micro.112408.134233

Hatfull, G.F., 2012a. Complete genome sequences of 138 mycobacteriophages. J Virol 86, 2382–2384. doi:10.1128/JVI.06870-11

Hatfull, G.F., 2012b. The secret lives of Mycobacteriophages, in: Lobocka, M.B., Szybalski, W.T. (Eds.), Advances in Virus Research, Vol 82. Academic Press, Burlington, pp. 179–288.

Hatfull, G.F., 2013. Complete genome sequences of 63 mycobacteriophages. Genome Announc 1. doi:10.1128/genomeA.00847-13

Hatfull, G.F., Cresawn, S.G., Hendrix, R.W., 2008. Comparative genomics of the mycobacteriophages: insights into bacteriophage evolution. Res Microbiol 159, 332–339. doi:10.1016/j.resmic.2008.04.008

Hatfull, G.F., Hendrix, R.W., n.d. The Mycobacteriophage Database [WWW Document]. URL www.phagesdb.org

Hatfull, G.F., Jacobs-Sera, D., Lawrence, J.G., Pope, W.H., Russell, D.A., Ko, C.-C., Weber, R.J., Patel, M.C., Germane, K.L., Edgar, R.H., Hoyte, N.N., Bowman, C.A., Tantoco, A.T., Paladin, E.C., Myers, M.S., Smith, A.L., Grace, M.S., Pham, T.T., O'Brien, M.B., Vogelsberger, A.M., Hryckowian, A.J., Wynalek, J.L., Donis-Keller, H., Bogel, M.W., Peebles, C.L., Cresawn, S.G., Hendrix, R.W., 2010. Comparative genomic analysis of 60 Mycobacteriophage genomes: Genome clustering, gene acquisition, and gene size. J Mol Biol 397, 119–143.

Hatfull, G.F., Pedulla, M.L., Jacobs-Sera, D., Cichon, P.M., Foley, A., Ford, M.E., Gonda, R.M., Houtz, J.M., Hryckowian, A.J., Kelchner, V.A., Namburi, S., Pajcini, K. V, Popovich, M.G., Schleicher, D.T., Simanek, B.Z., Smith, A.L., Zdanowicz, G.M., Kumar, V., Peebles, C.L., Jacobs, W.R., Lawrence, J.G., Hendrix, R.W., 2006. Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform. PLoS Genet 2, e92. doi:10.1371/journal.pgen.0020092

Hatfull, G.F., Sarkis, G.J., 1993. DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. Mol Microbiol 7, 395–405.

He, N.-B., Chen, J.-Z., Lin, C.-C., 1978. Six distinct types of bacteriophage attacking *Bacillus thuringiensis*. Acta Microbiol Sin 18, 220–224.

Hejnowicz, M.S., Gołebiewski, M., Bardowski, J., 2009. Analysis of the complete genome sequence of the lactococcal bacteriophage bIBB29. Int J Food Microbiol 131, 52–61. doi:10.1016/j.ijfoodmicro.2008.06.010

Henry, M., O'Sullivan, O., Sleator, R.D., Coffey, A., Ross, R.P., McAuliffe, O., O'Mahony, J.M., 2010. In silico analysis of Ardmore, a novel mycobacteriophage isolated from soil. Gene 453, 9–23. doi:10.1016/j.gene.2009.12.007

Heo, Y.-J., Chung, I.-Y., Choi, K.B., Lau, G.W., Cho, Y.-H., 2007. Genome sequence comparison and superinfection between two related *Pseudomonas aeruginosa* phages, D3112 and MP22. Microbiology 153, 2885–2895. doi:10.1099/mic.0.2007/007260-0

Hong, J., Kim, K.-P., Heu, S., Lee, S.J., Adhya, S., Ryu, S., 2008. Identification of host receptor and receptor-binding module of a newly sequenced T5-like phage EPS7. FEMS Microbiol Lett 289, 202–9. doi:10.1111/j.1574-6968.2008.01397.x

Horiuchi, T., Sakka, M., Hayashi, A., Shimada, T., Kimura, T., Sakka, K., 2012. Complete genome sequence of bacteriophage BC-611 specifically infecting *Enterococcus faecalis* strain NP-10011. J Virol 86, 9538–9539. doi:10.1128/JVI.01424-12

Hoshiba, H., Uchiyama, J., Kato, S., Ujihara, T., Muraoka, A., Daibata, M., Wakiguchi, H., Matsuzaki, S., 2010. Isolation and characterization of a novel *Staphylococcus aureus* bacteriophage, phiMR25, and its therapeutic potential. Arch Virol 155, 545–552. doi:10.1007/s00705-010-0623-2

Hulo, C., Masson, P., Le Mercier, P., Toussaint, A., n.d. A structured annotation frame for the transposable phages, a new Caudovirales group, the Saltoviridae. Virology.

Iandolo, J.J., Worrell, V., Groicher, K.H., Qian, Y., Tian, R., Kenton, S., Dorman, A., Ji, H., Lin, S., Loh, P., Qi, S., Zhu, H., Roe, B.A., 2002. Comparative analysis of the genomes of the temperate bacteriophages φ11, φ12 and φ13 of *Staphylococcus aureus* 8325. Gene 289, 109–118. doi:10.1016/S0378-1119(02)00481-X

Inoue, Y., Matsuura, T., Ohara, T., Azegami, K., 2006. Bacteriophage OP1, lytic for *Xanthomonas oryzae* pv. *oryzae*, changes its host range by duplication and deletion of the small domain in the deduced tail fiber gene. J Gen Plant Pathol 72, 111–118. doi:10.1007/s10327-005-0252-x

Inoue, Y., Matsuura, T., Ohara, T., Azegami, K., 2006. Bacteriophage OP1, lytic for Xanthomonas oryzae pv. oryzae, changes its host range by duplication and deletion of the small domain in the deduced tail fiber gene. J Gen Plant Pathol 72, 111–118. doi:DOI: 10.1007/s10327-005-0252-x

Juhala, R.J., Ford, M.E., Duda, R.L., Youlton, A., Hatfull, G.F., Hendrix, R.W., 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J Mol Biol 299, 27–51. doi:10.1006/jmbi.2000.3729

Kameyama, L., Fernández, L., Calderón, J., Ortiz-Rojas, A., Patterson, T.A., 1999. Characterization of wild lambdoid bacteriophages: detection of a wide distribution of phage immunity groups and identification of a nus-dependent, nonlambdoid phage group. Virology 263, 100–111. doi:10.1006/viro.1999.9888

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. Appl Environ Microbiol 79, 1956–1968. doi:10.1128/AEM.02793-12

Kelly, D., O'Sullivan, O., Mills, S., McAuliffe, O., Ross, R.P., Neve, H., Coffey, A., 2012. Genome sequence of the phage clP1, which infects the beer spoilage bacterium *Pediococcus damnosus*. Gene 504, 53–63. doi:10.1016/j.gene.2012.04.085

Kim, M., Ryu, S., 2011. Characterization of a T5-like coliphage, SPC35, and differential development of resistance to SPC35 in *Salmonella enterica* serovar Typhimurium and *Escherichia coli*. Appl Environ Microbiol 77, 2042–2050. doi:10.1128/AEM.02504-10

Kim, S., Rahman, M., Seol, S.Y., Yoon, S.S., Kim, J., 2012. *Pseudomonas aeruginosa* bacteriophage PA1Ø requires type IV pili for infection and shows broad bactericidal and biofilm removal activities. Appl Environ Microbiol 78, 6380–6385. doi:10.1128/AEM.00648-12

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. J Virol 86, 10253–10254. doi:10.1128/JVI.01550-12

Klumpp, J., Calendar, R., Loessner, M.J., 2010. Complete nucleotide sequence and molecular characterization of *Bacillus* phage TP21 and its relatedness to other phages with the same name. Viruses 2, 961–971. doi:10.3390/v2040961

Kropinski, A.M., 2000. Sequence of the genome of the temperate, serotype-converting, *Pseudomonas aeruginosa* bacteriophage D3. J Bacteriol 182, 6066–6074.

Krylov, S. V., Kropinski, A.M., Pleteneva, E.A., Shaburova, O. V., Burkal'tseva, M. V., Mirosnnikov, K.A., Krylov, V.N., 2012. Properties of the new D3-like *Pseudomonas aeruginosa* bacteriophage phiPMG1: Genome structure and prospects for the use in phage therapy. Russ J Genet 48, 902–911. doi:10.1134/S1022795412060087

Kwan, T., Liu, J., DuBow, M., Gros, P., Pelletier, J., 2005. The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. Proc Natl Acad Sci U S A 102, 5174–5179. doi:10.1073/pnas.0501140102

Kwan, T., Liu, J., Dubow, M., Gros, P., Pelletier, J., 2006. Comparative genomic analysis of 18 *Pseudomonas aeruginosa* bacteriophages. J Bacteriol 188, 1184–1187. doi:10.1128/JB.188.3.1184

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948. doi:10.1093/bioinformatics/btm404

Lavigne, R., Darius, P., Summer, E.J., Seto, D., Mahadevan, P., Nilsson, A.S., Ackermann, H.-W., Kropinski, A.M., 2009. Classification of *Myoviridae* bacteriophages using protein sequence similarity. BMC Microbiol 9, 224. doi:10.1186/1471-2180-9-224

Lavigne, R., Seto, D., Mahadevan, P., Ackermann, H.-W., Kropinski, A.M., 2008. Unifying classical and molecular taxonomic classification: analysis of the *Podoviridae* using BLASTP-based tools. Res Microbiol 159, 406–414. doi:10.1016/j.resmic.2008.03.005

Lawrence, J.G., Hatfull, G.F., Hendrix, R.W., 2002. Imbroglios of viral taxonomy: Genetic exchange and failings of phenetic approaches. J Bacteriol 184, 4891–4905. doi:10.1128/JB.184.17.4891-4905.2002

Lazarevic, V., Soldo, B., Düsterhöft, A., Hilbert, H., Mauël, C., Karamata, D., 1998. Introns and intein coding sequence in the ribonucleotide reductase genes of *Bacillus subtilis* temperate bacteriophage SPbeta. Proc Natl Acad Sci U S A 95, 1692–1697.

Le Marrec, C., van Sinderen, D., Walsh, L., Stanley, E., Vlegels, E., Moineau, S., Heinze, P., Fitzgerald, G., Fayard, B., 1997. Two groups of bacteriophages infecting *Streptococcus thermophilus* can be distinguished on the basis of mode of packaging and genetic determinants for major structural proteins. Appl Environ Microbiol 63, 3246–3253.

Lee, C.-N., Hu, R.-M., Chow, T.-Y., Lin, J.-W., Chen, H.-Y., Tseng, Y.-H., Weng, S.-F., 2007. Comparison of genomes of three *Xanthomonas oryzae* bacteriophages. BMC Genomics 8, 442. doi:10.1186/1471-2164-8-442

Lee, C.-N., Lin, J.-W., Chow, T.-Y., Tseng, Y.-H., Weng, S.-F., 2006. A novel lysozyme from *Xanthomonas oryzae* phage varphiXo411 active against *Xanthomonas* and *Stenotrophomonas*. Protein Expr Purif 50, 229–237. doi:10.1016/j.pep.2006.06.013

Lee, C.-N., Lin, J.-W., Weng, S.-F., Tseng, Y.-H., 2009. Genomic characterization of the intron-containing T7-like phage phiL7 of *Xanthomonas campestris*. Appl Environ Microbiol 75, 7828–7837. doi:10.1128/AEM.01214-09

Lee, J.-H., Shin, H., Choi, Y., Ryu, S., 2013. Complete genome sequence analysis of bacterial-flagellum-targeting bacteriophage chi. Arch Virol 158, 2179–2183. doi:10.1007/s00705-013-1700-0

Lee, Y.-D., Kim, J.-Y., Park, J.-H., Chang, H., 2012. Genomic analysis of bacteriophage ESP2949-1, which is virulent for *Cronobacter sakazakii*. Arch Virol 157, 199–202. doi:10.1007/s00705-011-1147-0

Lee, Y.-D., Park, J.-H., 2010. Genomic sequence of temperate phage 250 isolated from emetic *B. cereus* and cloning of putative endolysin. Food Sci Biotechnol 19, 1643–1648. doi:10.1007/s10068-010-0232-6

Lee, Y.-D., Park, J.-H., 2012. Complete genome sequence of enterococcal bacteriophage SAP6. J Virol 86, 5402–5403. doi:10.1128/JVI.00321-12

Lévesque, C., Duplessis, M., Labonté, J., Labrie, S., Fremaux, C., Tremblay, D., Moineau, S., 2005. Genomic organization and molecular analysis of virulent bacteriophage 2972 infecting an exopolysaccharide-producing *Streptococcus thermophilus* strain. Appl Environ Microbiol 71, 4057–4068. doi:10.1128/AEM.71.7.4057-4068.2005

Li, S., Liu, L., Zhu, J., Zou, L., Li, M., Cong, Y., Rao, X., Hu, X., Zhou, Y., Chen, Z., Hu, F., 2010. Characterization and genome sequencing of a novel coliphage isolated from engineered *Escherichia coli*. Intervirology 53, 211–220. doi:10.1159/000299063

Lima-Mendez, G., Van Helden, J., Toussaint, A., Leplae, R., 2008. Reticulate representation of evolutionary and functional relationships between phage genomes. Mol Biol Evol 25, 762–777. doi:10.1093/molbev/msn023

Lin, L., Hong, W., Ji, X., Han, J., Huang, L., Wei, Y., 2010. Isolation and characterization of an extremely long tail *Thermus* bacteriophage from Tengchong hot springs in China. J Basic mMcrobiology 50, 452–456. doi:10.1002/jobm.201000116

Liu, J., Dehbi, M., Moeck, G., Arhin, F., Bauda, P., Bergeron, D., Callejo, M., Ferretti, V., Ha, N., Kwan, T., McCarty, J., Srikumar, R., Williams, D., Wu, J.J., Gros, P., Pelletier, J., DuBow, M., 2004. Antimicrobial drug discovery through bacteriophage genomics. Nat Biotechnol 22, 185–191. doi:10.1038/nbt932

Loessner, M.J., Maier, S.K., Daubek-Puza, H., Wendlinger, G., Scherer, S., 1997. Three *Bacillus cereus* bacteriophage endolysins are unrelated but reveal high homology to cell wall hydrolases from different bacilli. J Bacteriol 179, 2845–2851.

Lohr, J.E., Chen, F., Hill, R.T., 2005. Genomic analysis of bacteriophage PhiJL001: insights into its interaction with a sponge-associated alpha-proteobacterium. Appl Environ Microbiol 71, 1598–1609. doi:10.1128/AEM.71.3.1598-1609.2005

Lorenz, L., Lins, B., Barrett, J., Montgomery, A., Trapani, S., Schindler, A., Christie, G.E., Cresawn, S.G., Temple, L., 2013. Genomic characterization of six novel *Bacillus pumilus* bacteriophages. Virology 444, 374–383.

Lu, Z., Altermann, E., Breidt, F., Predki, P., Fleming, H.P., Klaenhammer, T.R., 2005. Sequence analysis of the *Lactobacillus plantarum* bacteriophage PhiJL-1. Gene 348, 45–54. doi:10.1016/j.gene.2004.12.052

Lu, Z., Breidt, F., Fleming, H.P., Altermann, E., Klaenhammer, T.R., 2003. Isolation and characterization of a *Lactobacillus plantarum* bacteriophage, ΦJL-1, from a cucumber fermentation. Int J Food Microbiol 84, 225–235. doi:10.1016/S0168-1605(03)00111-9

Lubbers, M.W., Waterfield, N.R., Beresford, T.P., Le Page, R.W., Jarvis, A.W., 1995. Sequencing and analysis of the prolate-headed lactococcal bacteriophage c2 genome and identification of the structural genes. Appl Environ Microbiol 61, 4348–4356.

Lucchini, S., Desiere, F., Brüssow, H., 1999. Comparative genomics of *Streptococcus thermophilus* phage species supports a modular evolution theory. J Virol 73, 8647–8656.

Luria, S.E., Delbrück, M., Anderson, T.F., 1943. Electron microscope studies of bacterial viruses. J Bacteriol 46, 57–77.

Lwoff, A., Horne, R., Tournier, P., 1962. A System of Viruses. Cold Spring Harb Symp Quant Biol 27, 51–55. doi:10.1101/SQB.1962.027.001.008

Ma, X.X., Ito, T., Chongtrakool, P., Hiramatsu, K., 2006. Predominance of clones carrying Panton-Valentine leukocidin genes among methicillin-resistant *Staphylococcus aureus* strains isolated in Japanese hospitals from 1979 to 1985. J Clin Microbiol 44, 4515–4127. doi:10.1128/JCM.00985-06

Mahadevan, P., King, J.F., Seto, 2009a. Data mining pathogen genomes using GeneOrder and CoreGenes and CGUG: gene order, synteny and in silico proteomes. Int J Comput Biol Drug Des 2, 100–114.

Mahadevan, P., King, J.F., Seto, D., 2009b. CGUG: in silico proteome and genome parsing tool for the determination of "core" and unique genes in the analysis of genomes up to ca. 1.9 Mb. BMC Res Notes 2, 168. doi:10.1186/1756-0500-2-168

Mahony, J., Deveau, H., Mc Grath, S., Ventura, M., Canchaya, C., Moineau, S., Fitzgerald, G.F., van Sinderen, D., 2006. Sequence and comparative genomic analysis of lactococcal bacteriophages jj50, 712 and P008: evolutionary insights into the 936 phage species. FEMS Microbiol Lett 261, 253–261. doi:10.1111/j.1574-6968.2006.00372.x

Matsuzaki, S., Yasuda, M., Nishikawa, H., Kuroda, M., Ujihara, T., Shuin, T., Shen, Y., Jin, Z., Fujimoto, S., Nasimuzzaman, M.D., Wakiguchi, H., Sugihara, S., Sugiura, T., Koda, S., Muraoka, A., Imai, S., 2003. Experimental protection of mice against lethal *Staphylococcus aureus* infection by novel bacteriophage phi MR11. J Infect Dis 187, 613–624. doi:10.1086/374001

Mediavilla, J., Jain, S., Kriakov, J., Ford, M.E., Duda, R.L., Jacobs, W.R., Hendrix, R.W., Hatfull, G.F., 2000. Genome organization and characterization of mycobacteriophage Bxb1. Mol Microbiol 38, 955–970.

Minakhin, L., Goel, M., Berdygulova, Z., Ramanculov, E., Florens, L., Glazko, G., Karamychev, V.N., Slesarev, A.I., Kozyavkin, S.A., Khromov, I., Ackermann, H.-W., Washburn, M., Mushegian, A., Severinov, K., 2008. Genome comparison and proteomic characterization of *Thermus thermophilus* bacteriophages P23-45 and P74-26: siphoviruses with triplex-forming sequences and the longest known tails. J Mol Biol 378, 468–480. doi:10.1016/j.jmb.2008.02.018

Minakhin, L., Semenova, E., Liu, J., Vasilov, A., Severinova, E., Gabisonia, T., Inman, R., Mushegian, A., Severinov, K., 2005. Genome sequence and gene expression of *Bacillus anthracis* bacteriophage Fah. J Mol Biol 354, 1–15. doi:10.1016/j.jmb.2005.09.052

Mishra, C.K., Choi, T.J., Kang, S.C., 2012. Isolation and characterization of a bacteriophage F20 virulent to *Enterobacter aerogenes*. J Gen Virol 93, 2310–2314. doi:10.1099/vir.0.043562-0

Mitsui, H., Takahashi, H., 1992. Cohesive single-stranded ends of *Streptomyces* temperate bacteriophage R4. Mol Gen Genet 231, 360–362. doi:10.1007/BF00292703

Mondigler, M., Holz, T., Heller, K.J., 1996. Identification of the receptor-binding regions of pb5 proteins of bacteriophages T5 and BF23. Virology 219, 19–28. doi:10.1006/viro.1996.0218

Moreno Switt, A.I., Orsi, R.H., den Bakker, H.C., Vongkamjan, K., Altier, C., Wiedmann, M., 2013. Genomic characterization provides new insight into *Salmonella* phage diversity. BMC Genomics 14, 481. doi:10.1186/1471-2164-14-481

Muhire, B., Martin, D.P., Brown, J.K., Navas-Castillo, J., Moriones, E., Zerbini, F.M., Rivera-Bustamante, R., Malathi, V.G., Briddon, R.W., Varsani, A., 2013. A genome-wide pairwise-identity-based proposal for the classification of viruses in the genus *Mastrevirus* (family *Geminiviridae*). Arch Virol 158, 1411–1424. doi:10.1007/s00705-012-1601-7

Myers, E.W., Miller, W., 1988. Optimal alignments in linear space. Bioinformatics 4, 11–17. doi:10.1093/bioinformatics/4.1.11

Narita, S., Kaneko, J., Chiba, J., Piémont, Y., Jarraud, S., Etienne, J., Kamio, Y., 2001. Phage conversion of Panton-Valentine leukocidin in *Staphylococcus aureus*: molecular analysis of a PVL-converting phage, φSLT. Gene 268, 195–206. doi:10.1016/S0378-1119(01)00390-0

Nelson, D., 2004. Phage taxonomy: we agree to disagree. J Bacteriol 186, 7029–7031. doi:10.1128/JB.186.21.7029-7031.2004

Newton, G.J., Daniels, C., Burrows, L.L., Kropinski, A.M., Clarke, A.J., Lam, J.S., 2004. Three-component-mediated serotype conversion in *Pseudomonas aeruginosa* by bacteriophage D3. Mol Microbiol 39, 1237–1247. doi:10.1111/j.1365-2958.2001.02311.x

Niu, Y.D., Stanford, K., Kropinski, A.M., Ackermann, H.-W., Johnson, R.P., She, Y.-M., Ahmed, R., Villegas, A., McAllister, T.A., 2012. Genomic, proteomic and physiological characterization of a T5-like bacteriophage for control of Shiga toxin-producing *Escherichia coli* O157:H7. PLoS One 7, e34585. doi:10.1371/journal.pone.0034585

Pan, F., Wu, H., Liu, J., Ai, Y., Meng, X., Meng, R., Meng, Q., 2013. Complete genome sequence of *Escherichia coli* O157:H7 lytic phage JL1. Arch Virol 158, 2429–2432. doi:10.1007/s00705-013-1727-2

Panis, G., Lambert, C., Viollier, P.H., 2012. Complete genome sequence of *Caulobacter crescentus* bacteriophage φCbK. J Virol 86, 10234–10235. doi:10.1128/JVI.01579-12

Pedulla, M.L., Ford, M.E., Houtz, J.M., Karthikeyan, T., Wadsworth, C., Lewis, J.A., Jacobs-Sera, D., Falbo, J., Gross, J., Pannunzio, N.R., Brucker, W., Kumar, V., Kandasamy, J., Keenan, L., Bardarov, S., Kriakov, J., Lawrence, J.G., Jacobs, W.R., Hendrix, R.W., Hatfull, G.F., 2003. Origins of highly mosaic mycobacteriophage genomes. Cell 113, 171–182.

Petrovski, S., Seviour, R.J., Tillett, D., 2013. Characterization and whole genome sequences of the *Rhodococcus* bacteriophages RGL3 and RER2. Arch Virol 158, 601–609. doi:10.1007/s00705-012-1530-5

Pham, T.T., Jacobs-Sera, D., Pedulla, M.L., Hendrix, R.W., Hatfull, G.F., 2007. Comparative genomic analysis of mycobacteriophage Tweety: evolutionary insights and construction of compatible site-specific integration vectors for mycobacteria. Microbiology 153, 2711–2723. doi:10.1099/mic.0.2007/008904-0

Pope, W.H., Ferreira, C.M., Jacobs-Sera, D., Benjamin, R.C., Davis, A.J., DeJong, R.J., Elgin, S.C.R., Guilfoile, F.R., Forsyth, M.H., Harris, A.D., Harvey, S.E., Hughes, L.E., Hynes, P.M., Jackson, A.S., Jalal, M.D., MacMurray, E.A., Manley, C.M., McDonough, M.J., Mosier, J.L., Osterbann, L.J., Rabinowitz, H.S., Rhyan, C.N., Russell, D.A., Saha, M.S., Shaffer, C.D., Simon, S.E., Sims, E.F., Tovar, I.G., Weisser, E.G., Wertz, J.T., Weston-Hafer, K.A., Williamson, K.E., Zhang, B., Cresawn, S.G., Jain, P., Piuri, M., Jacobs, W.R., Hendrix, R.W., Hatfull, G.F., 2011a. Cluster K mycobacteriophages: insights into the evolutionary origins of mycobacteriophage TM4. PLoS One 6, e26750. doi:10.1371/journal.pone.0026750

Pope, W.H., Jacobs-Sera, D., Best, A.A., Broussard, G.W., Connerly, P.L., Dedrick, R.M., Kremer, T.A., Offner, S., Ogiefo, A.H., Pizzorno, M.C., Rockenbach, K., Russell, D.A., Stowe, E.L., Stukey, J., Thibault, S.A., Conway, J.F., Hendrix, R.W., Hatfull, G.F., 2013. Cluster J mycobacteriophages: intron splicing in capsid and tail genes. PLoS One 8, e69273. doi:10.1371/journal.pone.0069273

Pope, W.H., Jacobs-Sera, D., Russell, D.A., Peebles, C.L., Al-Atrache, Z., Alcoser, T.A., Alexander, L.M., Alfano, M.B., Alford, S.T., Amy, N.E., Anderson, M.D., Anderson, A.G., Ang, A.A.S., Ares, M., Barber, A.J., Barker, L.P., Barrett, J.M., Barshop, W.D., Bauerle, C.M., Bayles, I.M., Belfield, K.L., Best, A.A., Borjon, A., Bowman, C.A., Boyer, C.A., Bradley, K.W., Bradley, V.A., Broadway, L.N., Budwal, K., Busby, K.N., Campbell, I.W., Campbell, A.M., Carey, A., Caruso, S.M., Chew, R.D., Cockburn, C.L., Cohen, L.B., Corajod, J.M., Cresawn, S.G., Davis, K.R., Deng, L., Denver, D.R., Dixon, B.R., Ekram, S., Elgin, S.C.R., Engelsen, A.E., English, B.E. V, Erb, M.L., Estrada, C., Filliger, L.Z., Findley, A.M., Forbes, L., Forsyth, M.H., Fox, T.M., Fritz, M.J., Garcia, R., George, Z.D., Georges, A.E., Gissendanner, C.R., Goff, S., Goldstein, R., Gordon, K.C., Green, R.D., Guerra, S.L., Guiney-Olsen, K.R., Guiza, B.G., Haghighat, L., Hagopian, G. V, Harmon, C.J., Harmson, J.S., Hartzog, G.A., Harvey, S.E., He, S., He, K.J., Healy, K.E., Higinbotham, E.R., Hildebrandt, E.N., Ho, J.H., Hogan, G.M., Hohenstein, V.G., Holz, N.A., Huang, V.J., Hufford, E.L., Hynes, P.M., Jackson, A.S., Jansen, E.C., Jarvik, J., Jasinto, P.G., Jordan, T.C., Kasza, T., Katelyn, M.A., Kelsey, J.S., Kerrigan, L.A., Khaw, D., Kim, J., Knutter, J.Z., Ko, C.-C., Larkin, G. V, Laroche, J.R., Latif, A., Leuba, K.D., Leuba, S.I., Lewis, L.O., Loesser-Casey, K.E., Long, C.A., Lopez, A.J., Lowery, N., Lu, T.Q., Mac, V., Masters, I.R., McCloud, J.J., McDonough, M.J., Medenbach, A.J., Menon, A., Miller, R., Morgan, B.K., Ng, P.C., Nguyen, E., Nguyen, K.T., Nguyen, E.T., Nicholson, K.M., Parnell, L.A., Peirce, C.E., Perz, A.M., Peterson, L.J., Pferdehirt, R.E., Philip, S. V, Pogliano, K., Pogliano, J., Polley, T., Puopolo, E.J., Rabinowitz, H.S., Resiss, M.J., Rhyan, C.N., Robinson, Y.M., Rodriguez, L.L., Rose, A.C., Rubin, J.D., Ruby, J.A., Saha, M.S., Sandoz, J.W., Savitskaya, J., Schipper, D.J., Schnitzler, C.E., Schott, A.R., Segal, J.B., Shaffer, C.D., Sheldon, K.E., Shepard, E.M., Shepardson, J.W., Shroff, M.K., Simmons, J.M., Simms, E.F., Simpson, B.M., Sinclair, K.M., Sjoholm, R.L., Slette, I.J., Spaulding, B.C., Straub, C.L., Stukey, J., Sughrue, T., Tang, T.-Y., Tatyana, L.M., Taylor, S.B., Taylor, B.J., Temple, L.M., Thompson, J. V, Tokarz, M.P., Trapani, S.E., Troum, A.P., Tsay, J., Tubbs, A.T., Walton, J.M., Wang, D.H., Wang, H., Warner, J.R., Weisser, E.G., Wendler, S.C., Weston-Hafer, K.A., Whelan, H.M., Williamson, K.E., Willis, A.N., Wirtshafter, H.S., Wong, T.W., Wu, P., Yang, Y. jeong, Yee, B.C., Zaidins, D.A., Zhang, B., Zúniga, M.Y., Hendrix, R.W., Hatfull, G.F., 2011b. Expanding the diversity of mycobacteriophages: insights into genome architecture and evolution. PLoS One 6, e16329. doi:10.1371/journal.pone.0016329

Pride, D.T., Wassenaar, T.M., Ghose, C., Blaser, M.J., 2006. Evidence of host-virus co-evolution in tetranucleotide usage patterns of bacteriophages and eukaryotic viruses. BMC Genomics 7, 8. doi:10.1186/1471-2164-7-8

Proux, C., van Sinderen, D., Suarez, J., Garcia, P., Ladero, V., Fitzgerald, G.F., Desiere, F., Brussow, H., 2002. The dilemma of phage taxonomy illustrated by comparative genomics of Sfi21-like *Siphoviridae* in lactic acid bacteria. J Bacteriol 184, 6026–6036. doi:10.1128/JB.184.21.6026-6036.2002

Rabsch, W., Ma, L., Wiley, G., Najar, F.Z., Kaserer, W., Schuerch, D.W., Klebba, J.E., Roe, B.A., Laverde Gomez, J.A., Schallmey, M., Newton, S.M.C., Klebba, P.E., 2007. FepA- and TonB-dependent bacteriophage H8: receptor binding and genomic sequence. J Bacteriol 189, 5658–74. doi:10.1128/JB.00437-07

Riipinen, K.-A., Forsman, P., Alatossava, T., 2011. The genomes and comparative genomics of *Lactobacillus delbrueckii* phages. Arch Virol 156, 1217–1233. doi:10.1007/s00705-011-0980-5

Roberts, M.D., Martin, N.L., Kropinski, A.M., 2004. The genome and proteome of coliphage T1. Virology 318, 245–266. doi:10.1016/j.virol.2003.09.020

Rohwer, F., Edwards, R., 2002. The Phage Proteomic Tree: a genome-based taxonomy for phage. J Bacteriol 184, 4529–4535. doi:10.1128/JB.184.16.4529

Rousseau, G.M., Moineau, S., 2009. Evolution of *Lactococcus lactis* phages within a cheese factory. Appl Environ Microbiol 75, 5336–5344. doi:10.1128/AEM.00761-09

Ruska, H., 1940. Die Sichtbarmachung der bakteriophagen Lyse im Übermikroskop. Naturwissenschaften 28, 45–46.

Sampson, T., Broussard, G.W., Marinelli, L.J., Jacobs-Sera, D., Ray, M., Ko, C.-C., Russell, D., Hendrix, R.W., Hatfull, G.F., 2009. Mycobacteriophages BPs, Angel and Halo: comparative genomics reveals a novel class of ultra-small mobile genetic elements. Microbiology 155, 2962–2977. doi:10.1099/mic.0.030486-0

Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F., Petersen, G.B., 1982. Nucleotide sequence of bacteriophage lambda DNA. J Mol Biol 162, 729–773.

Scaltriti, E., Tegoni, M., Rivetti, C., Launay, H., Masson, J.-Y., Magadan, A.H., Tremblay, D., Moineau, S., Ramoni, R., Lichière, J., Campanacci, V., Cambillau, C., Ortiz-Lombardía, M., 2009. Structure and function of phage p2 ORF34(p2), a new type of single-stranded DNA binding protein. Mol Microbiol 73, 1156–1170. doi:10.1111/j.1365-2958.2009.06844.x

Schade, S.Z., Adler, J., Ris, H., 1967. How bacteriophage chi attacks motile bacteria. J Virol 1, 599–609.

Schouler, C., Ehrlich, S.D., Chopin, M.C., 1994. Sequence and organization of the lactococcal prolate-headed bIL67 phage genome. Microbiology 140, 3061–3069.

Schuch, R., Fischetti, V.A., 2006. Detailed genomic analysis of the Wbeta and gamma phages infecting *Bacillus anthracis*: implications for evolution of environmental fitness and antibiotic resistance. J Bacteriol 188, 3037–3051. doi:10.1128/JB.188.8.3037-3051.2006

Smeesters, P.R., Drèze, P.-A., Bousbata, S., Parikka, K.J., Timmery, S., Hu, X., Perez-Morga, D., Deghorain, M., Toussaint, A., Mahillon, J., Van Melderen, L., 2011. Characterization of a novel temperate phage originating from a cereulide-producing *Bacillus cereus* strain. Res Microbiol 162, 446–459. doi:10.1016/j.resmic.2011.02.009

Smith, M.C., Burns, R.N., Wilson, S.E., Gregory, M.A., 1999. The complete genome sequence of the *Streptomyces* temperate phage straight phiC31: evolutionary relationships to other viruses. Nucleic Acids Res 27, 2145–2155.

Smith, M.C.M., Hendrix, R.W., Dedrick, R., Mitchell, K., Ko, C.-C., Russell, D., Bell, E., Gregory, M., Bibb, M.J., Pethick, F., Jacobs-Sera, D., Herron, P., Buttner, M.J., Hatfull, G.F., 2013. Evolutionary

relationships among actinophages and a putative adaptation for growth in *Streptomyces* spp. J Bacteriol 195, 4924–4935. doi:10.1128/JB.00618-13

Stanley, E., Fitzgerald, G.F., Le Marrec, C., Fayard, B., van Sinderen, D., 1997. Sequence analysis and characterization of phi O1205, a temperate bacteriophage infecting *Streptococcus thermophilus* CNRZ1205. Microbiology 143, 3417–3429.

Summer, E.J., Gonzalez, C.F., Bomer, M., Carlile, T., Embry, A., Kucherka, A.M., Lee, J., Mebane, L., Morrison, W.C., Mark, L., King, M.D., LiPuma, J.J., Vidaver, A.K., Young, R., 2006. Divergence and mosaicism among virulent soil phages of the *Burkholderia cepacia* complex. J Bacteriol 188, 255–268. doi:10.1128/JB.188.1.255-268.2006

Tiwari, B.R., Kim, S., Kim, J., 2012. Complete genomic sequence of *Salmonella enterica* serovar Enteritidis phage SE2. J Virol 86, 7712. doi:10.1128/JVI.00999-12

Tremblay, D.M., Moineau, S., 1999. Complete genomic sequence of the lytic bacteriophage DT1 of *Streptococcus thermophilus*. Virology 255, 63–76. doi:10.1006/viro.1998.9525

Turner, D., Hezwani, M., Nelson, S., Salisbury, V., Reynolds, D., 2012. Characterization of the *Salmonella* bacteriophage vB_SenS-Ent1. J Gen Virol 93, 2046–2056. doi:10.1099/vir.0.043331-0

Vostrov, A.A., Vostrukhina, O.A., Svarchevsky, A.N., Rybchin, V.N., 1996. Proteins responsible for lysogenic conversion caused by coliphages N15 and phi80 are highly homologous. J Bacteriol 178, 1484–1486.

Walter, T.M., Aronson, A.I., 1991. Transduction of certain genes by an autonomously replicating *Bacillus thuringiensis* phage. Appl. Envir. Microbiol. 57, 1000–1005.

Wang, P.W., Chu, L., Guttman, D.S., 2004. Complete sequence and evolutionary genomic analysis of the *Pseudomonas aeruginosa* transposable bacteriophage D3112. J Bacteriol 186, 400–410. doi:10.1128/JB.186.2.400-410.2004

Wietzorrek, A., Schwarz, H., Herrmann, C., Braun, V., 2006. The genome of the novel phage Rtp, with a rosette-like tail tip, is homologous to the genome of phage T1. J Bacteriol 188, 1419–1436. doi:10.1128/JB.188.4.1419-1436.2006

Woods, D.E., Jeddeloh, J.A., Fritz, D.L., DeShazer, D., 2002. *Burkholderia thailandensis* E125 harbors a temperate bacteriophage specific for *Burkholderia mallei*. J Bacteriol 184, 4003–4017.

Yamaguchi, T., Hayashi, T., Takami, H., Nakasone, K., Ohnishi, M., Nakayama, K., Yamada, S., Komatsuzawa, H., Sugai, M., 2000. Phage conversion of exfoliative toxin A production in *Staphylococcus aureus*. Mol Microbiol 38, 694–705. doi:10.1046/j.1365-2958.2000.02169.x

Yasmin, A., Kenny, J.G., Shankar, J., Darby, A.C., Hall, N., Edwards, C., Horsburgh, M.J., 2010. Comparative genomics and transduction potential of *Enterococcus faecalis* temperate bacteriophages. J Bacteriol 192, 1122–1130. doi:10.1128/JB.01293-09

Yuzenkova, J., Nechaev, S., Berlin, J., Rogulja, D., Kuznedelov, K., Inman, R., Mushegian, A., Severinov, K., 2003. Genome of *Xanthomonas oryzae* bacteriophage Xp10: an odd T-odd phage. J Mol Biol 330, 735–748.

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. doi:10.1186/1471-2105-3-12

Zegans, M.E., Wagner, J.C., Cady, K.C., Murphy, D.M., Hammond, J.H., O'Toole, G.A., 2009. Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviors of *Pseudomonas aeruginosa*. J Bacteriol 191, 210–219. doi:10.1128/JB.00797-08

## Table 1: Characteristics of the proposed and existing genera in the *Siphoviridae* family.

| **Proposed genus name** | **Number of proposed species** | **Genome size range (kb)** | **Features** | **Infecting** | **References** |
|---|---|---|---|---|---|
| "Barnyard-like viruses" | 4 | 68 – 71 | Cluster H + U | *Mycobacterium* | (Hatfull et al., 2010, 2006; Pedulla et al., 2003) |
| "Bignuz-like viruses" | 2 | 45 – 49 | Cluster P | *Mycobacterium* | (Hatfull, 2012a) |
| "Charlie-like viruses" | 2 | 42 – 43 | Cluster N | *Mycobacterium* | (Hatfull, 2012a) |
| "Che8-like viruses" | 28 | 52 – 61 | Cluster F | *Mycobacterium* | (Hatfull, 2013, 2012a; Hatfull et al., 2010, 2006; Henry et al., 2010; Pedulla et al., 2003; Pham et al., 2007) |
| "Che9c-like viruses" | 3 | 47 – 57 | Cluster I | *Mycobacterium* | (Hatfull, 2012a; Hatfull et al., 2010; Pedulla et al., 2003) |
| "CJW1-like viruses" | 9 | 74 – 76 | Cluster E | *Mycobacterium* | (Hatfull, 2012a; Hatfull et al., 2010, 2006; Pedulla et al., 2003; Pope et al., 2011b) |
| "Corndog-like viruses" | 2 | 69 – 71 | Cluster O, extremely elongated heads (4:1-7:1, length:width) | *Mycobacterium* | (Hatfull, 2012a; Hatfull et al., 2006) |
| "Halo-like viruses" | 2 | 41 – 42 | Cluster G | *Mycobacterium* | (Hatfull, 2012a; Hatfull et al., 2006; Pope et al., 2011b; Sampson et al., 2009) |
| "Lebron-like viruses" | 5 | 69 – 76 | Cluster L, ~9 tRNAs | *Mycobacterium* | (Hatfull, 2012a; Pope et al., 2011b) |
| "Omega-like viruses" | 6 | 106 – 112 | Cluster J, contains introns and homing endonucleases | *Mycobacterium* | (Hatfull, 2012a; Pedulla et al., 2003; Pope et al., 2013) |
| "PBI1-like viruses" | 1 | 59 – 60 | Cluster D | *Mycobacterium* | (Hatfull, 2010; Hatfull et al., 2006) |
| "PG1-like viruses" | 13 | 67 – 71 | Cluster B | *Mycobacterium* | (Hatfull, 2012a; Hatfull et al., 2010, 2006; Pedulla et al., 2003) |
| "Rey-like viruses" | 2 | 80 – 84 | Cluster M, ~20 tRNAs | *Mycobacterium* | (Hatfull, 2012a; Pope et al., 2011a) |
| "TM4-like viruses" | 9 | 52 – 63 | Cluster K | *Mycobacterium* | (Ford et al., 1998b; Hatfull, 2013, 2012a; Pope et al., 2011a) |
| "Andromeda-like viruses" | 5 | 49 – 50 | | *Bacillus* | (Lorenz et al., 2013) |

| | | | | | |
|---|---|---|---|---|---|
| "C5-like viruses" | 2 | 31 – 32 | Group b lactic acid phages | *Lactobacillus* | (Accolas and Spillmann, 1979; Riipinen et al., 2011) |
| "IEBH-like viruses" | 2 | 53 – 57 | Transverse tail discs | *Bacillus* | (Lee and Park, 2010; Smeesters et al., 2011) |
| "phiFL-like viruses" | 3 | 36 – 40 | | *Enterococcus* | (Yasmin et al., 2010) |
| "phiJL1-like viruses" | 3 | 36 – 38 | | *Pediococcus, Lactobacillus* | (Briggiler Marcó et al., 2012; Kelly et al., 2012; Lu et al., 2005, 2003) |
| "R4-like viruses" | 2 | 51 | Broad host range | *Streptomyces* | (Chater and Carter, 1979) |
| "Lika-like viruses" | 2 | 51 | Related to the r4likeviruses | *Streptomyces* | (Smith et al., 2013) |
| "Sap6-like viruses" | 2 | 54 – 59 | | *Enterococcus* | (Horiuchi et al., 2012; Lee and Park, 2012) |
| "Sfi21-DT1-like viruses" | 5 | 34 – 41 | | *Streptococcus* | (Desiere et al., 1998; Guglielmotti et al., 2009; Le Marrec et al., 1997; Tremblay and Moineau, 1999) |
| "Sfi11-like viruses" | 5 | 34 – 43 | | *Streptococcus* | (Deveau et al., 2008; Guglielmotti et al., 2009; Lévesque et al., 2005; Lucchini et al., 1999; Stanley et al., 1997) |
| "SK1-like viruses" | 17 | 27 – 32 | 936-like group of dairy phages | *Lactococcus* | (Castro-Nallar et al., 2012; Chandry et al., 1997; Crutz-Le Coq et al., 2002; Hejnowicz et al., 2009; Mahony et al., 2006; Rousseau and Moineau, 2009; Scaltriti et al., 2009) |
| "TP21-like viruses" | 2 | 36 – 38 | | *Bacillus* | (Dong et al., 2013; Klumpp et al., 2010; Loessner et al., 1997) |
| "Wbeta-like viruses" | 1 | 36 – 41 | Multiple virulent derivatives from same temperate phage | *Bacillus* | (Fouts et al., 2006; Minakhin et al., 2005; Schuch and Fischetti, 2006) |
| "3a-like viruses" | 6 | 43 – 46 | | *Staphylococcus* | (García et al., 2009; Iandolo et al., 2002; Kwan et al., 2005; Narita et al., 2001) |
| "77-like viruses" | 3 | 41 – 45 | | *Staphylococcus* | (Iandolo et al., 2002; Liu et al., 2004; Ma et al., 2006) |
| "phiETA-lika viruses" | 31 | 39 – 45 | 3 subclusters | *Staphylococcus* | (Bae et al., 2006; Christie et al., 2010; Daniel et al., 2007; Gutiérrez et al., 2010; Hoshiba et al., 2010; Iandolo et al., 2002; Kwan et al., 2005; Matsuzaki et al., 2003; Yamaguchi et al., 2000) |
| "Chi-like viruses" | 5 | ~59 | Infect motile strains | *Salmonella, Serratia, Escherichia* | (Choi et al., 2013; Lee et al., 2013) |
| "D3-like viruses" | 2 | 54 – 57 | 1 – 4 tRNAs, tail | *Pseudomonas* | (Kropinski, 2000; Krylov et al., 2012) |

| | | | | | |
|---|---|---|---|---|---|
| | | | with terminal knobs | | |
| "D3112-like viruses" | 4 | 34 – 38 | Originally presumed Mu-like | *Pseudomonas* | (Chung and Cho, 2012; Depew et al., 2013; Heo et al., 2007; S. Kim et al., 2012; Wang et al., 2004; Zegans et al., 2009) |
| "HK578-like viruses" | 5 | 39 – 45 | Broad host range | *Escherichia, Shigella, Sodalis* | (Chang and Kim, 2011; Li et al., 2010; Pan et al., 2013) |
| "Jersey-like viruses" | 6 | 40 – 44 | Virulent, infecting multiple serovars | *Salmonella* | (Ackermann and Gershman, 1992; De Lappe et al., 2009; Kang et al., 2013; S.-H. Kim et al., 2012; Tiwari et al., 2012; Turner et al., 2012) |
| "P23-45-like viruses" | 2 | 83 – 84 | Extremely long tail, R-Y mirror repeats | *Thermus* | (Minakhin et al., 2008) |
| "phiCBK-like viruses" | 5 | 215 – 223 | Extremely large genomes | *Caulobacter* | (Gill et al., 2012; Panis et al., 2012) |
| "phiE125-like viruses" | 3 | 48 – 55 | Mosaicism present | *Burkholderia* | (DeShazer, 2004; Woods et al., 2002) |
| "Xp10-like viruses" | 5 | 43 - 45 | RNA polymerase, HNH endonucleases | *Xanthomonas* | (Ahmad et al., 2014; Y Inoue et al., 2006; Lee et al., 2009, 2007, 2006; Yuzenkova et al., 2003) |
| **In ICTV** | | | | | |
| *C2likevirus* | 2 | ~22 | | *Lactococcus* | (Lubbers et al., 1995; Schouler et al., 1994) |
| *L5likevirus* | 2 → 62 | 48 – 54 | | *Mycobacterium, Rhodococcus* | (Fan et al., 2012; Ford et al., 1998a, 1998b; Hatfull and Sarkis, 1993; Hatfull, 2013, 2012a, 2010; Hatfull et al., 2006; Mediavilla et al., 2000; Pedulla et al., 2003; Petrovski et al., 2013; Pope et al., 2011b) |
| *Lambdalikevirus* | 3 | | | *Escherichia* | (Juhala et al., 2000; Sanger et al., 1982) |
| *N15likevirus* | 1 | ~46 | | *Enterobacteria* | (Vostrov et al., 1996) |
| *Phic3unalikevirus* | 1 → 3 | ~41 | | *Streptomyces* | (Foor et al., 1985; Gregory et al., 2003; Smith et al., 1999) |
| *Psimunalikevirus*[a] | 1 | | DNA sequence incomplete | *Methanobacterium* | |
| *Spbetalikevirus* | 1 | ~134 | | *Bacillus* | (Lazarevic et al., 1998) |
| *T5likevirus* | 6 | 108 - 121 | | *Escherichia, Vibrio, Salmonella* | (Demerec and Fano, 1945; Hong et al., 2008; Kim and Ryu, 2011; Mondigler et al., 1996; Niu et al., 2012; Rabsch et al., 2007) |
| *Tunalikevirus* | 9 | 46 - 51 | | *Escherichia, Cronobacter, Enterobacter, Shigella* | (Battaglioli et al., 2011; Lee et al., 2012; Mishra et al., 2012; Roberts et al., 2004; Wietzorrek et al., 2006) |

| | | | | | |
|---|---|---|---|---|---|
| *Yualikevirus* | 3 | ~60 | | *Pseudomonas* | (Ceyssens et al., 2008; Kwan et al., 2006; Lohr et al., 2005) |
| **Genera 49** | **Species 307** | | | | |

[a] The genus *Psimunalikevirus* is a historical genus containing one species, *Methanobacterium phage psiM1*, for which the deposited DNA sequence is incomplete.