ARTICLE IN PRESS

Virology ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

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,*], 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, Canada N1G 3W4*
[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, Canada G1K 7P4*
[g] *Department of Biosystems, Laboratory of Gene Technology, KU Leuven, KasteelparkArenberg 21 – b2462, Heverlee 3001, Belgium*
[h] *Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada N1G 2A1*

ARTICLE INFO

Article history:
Received 10 July 2014
Returned to author for revisions
8 September 2014
Accepted 17 October 2014

Keywords:
Siphoviridae
Taxonomy
Phage Proteomic Tree
CoreGenes
ClustalW
ICTV

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 remainders 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, 1967 in a classification scheme, 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 et al. (2002) suggested a taxonomy based on comparative genomics of a single structural gene module. 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

---

[*] Corresponding author.
[**] Corresponding author at: Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada N1G 2A1.
E-mail addresses: rob.lavigne@kuleuven.be (R. Lavigne), akropins@uoguelph.ca (A.M. Kropinski).

http://dx.doi.org/10.1016/j.virol.2014.10.016
0042-6822/

respectively 40% and 20% shared proteins between phages. These BLASTP-related 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.

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)

Please cite this article as: Adriaenssens, E.M., et al., Integration of genomic and proteomic analyses in the classification of the *Siphoviridae* family. Virology (2014), http://dx.doi.org/10.1016/j.virol.2014.10.016