Bacteriophage taxonomy

Hans-W Ackermann

Department of Microbiology
Immunology and Infectiology
Faculty of Medicine, Laval
University, Vandry Pavilion
Avenue de la Médecine
Quebec, QC, G1K 7P4
Canada
Tel 418-656-2131 # 2558
Fax 418-656-7898
Email ackermann@mcb.ulaval.ca

Bacteriophages or "phages" are viruses of prokaryotes. At least 5,360 tailed and 179 cubic, filamentous, and pleomorphic bacterial viruses have been examined in the electron microscope since the introduction of negative staining in 1959. Since at least 100 novel bacterial viruses are described every year[1], the approximate number of viruses under consideration is over 6,000. Numerically, this makes bacteriophages the largest virus group known. Phages are presently classified in a hierarchical and holistic system with one order and 10 families. Over 96% of phages are tailed and contain dsDNA. The seven families of cubic, filamentous and pleomorphic phages are small and well defined. They contain ds or ss DNA or RNA. The most important developments are reclassifications of the Podoviridae and Myoviridae families of tailed phages.

A. Phages in nature

Bacteriophages were discovered twice at the beginning of the 20th century. In 1915, the English bacteriologist FW Twort described a transmissible lysis in a "micrococcus" and, in 1917, the Canadian Felix d'Herelle, then at the Pasteur Institute in Paris, described the lysis of Shigella cultures[2,3]. Twort abandoned his discovery and tried instead to propagate vertebrate viruses, such as the cowpox virus, on inert media. D'Herelle, however, devoted the rest of his scientific life to bacteriophages and the phage therapy of infectious diseases. He coined the term "bacteriophage" and stated that there was only one phage, the "bacteriophagum intestinale", with many races[4].

Bacteriophages occur everywhere in the biosphere and have colonised even such forbidding habitats as volcanic hot springs. Their main habitats are the oceans and topsoil. Lysogenic bacteria seem to be the main reservoir. From counts of marine phages, the total number of phages in the biosphere has been estimated at over 10^{30} particles[5-7]. The number of phage species in nature has been evaluated at several 100,000 or even millions[8,9]. This is to some extent confirmed by metagenomics, that is the culture-independent identification of phage genomes. The immense majority of viral sequences are not found in databases and only a few can be related to known phages such as T4 and T7[5,8]. Moreover, most phages are from North America and Europe, while we know almost nothing of phages in the environment of vast regions such as Black Africa or South America. Our knowledge of the phage world is evidently incomplete and we have barely scratched its surface.

B. History of phage classification

The forerunners of phage classification were the great Australian microbiologist, Sir Macfarlane Burnet, who proved in 1937 that phages differed in size and resistance against physicochemical agents[10] and H Ruska, who proved that phages were morphologically diverse. In 1943, Ruska proposed a classification of viruses by electron microscopy[11]. In 1948, Holmes classified viruses into three families. Phages constituted the family Phagineae. The Holmes classification was based on host range and symptoms of disease[12]. For example, herpesviruses and poxviruses were lumped together because they produced pustules on the skin. This classification is of historical interest only.

In 1962, Lwoff, Horne and Tournier stated that virus classification should be based on the properties of the virion and its nucleic acid and proposed a system with a latinised nomenclature[13] that included several phages. A Provisional Committee on Nomenclature of Viruses (PCNV) was founded in 1965, later to become the International Committee on Taxonomy of Viruses (ICTV)[14]. In 1971, the ICTV issued its first report which included six phage "genera": T-even phages, I , lipid phage PM2, the f X group, "filamentous phage", and the "ribophage group". Groups were listed with type species and properties[14]. This may be considered as the starting point of phage classification.

The ICTV is the only international body concerned with virus taxonomy. It has subcommittees for vertebrate, invertebrate, bacterial, plant, protozoal and fungal viruses. About 400 virologists are members of the ICTV. Taxonomical proposals should be submitted to the relevant subcommittee. The ICTV issues reports, ideally after each International Congress of Virology. The IXth Report is in print and will hopefully reach the scientific community this year. It includes six orders, 87 families, 19 subfamilies and 348 genera[15]. The families are the most stable parts of the system. The ICTV uses, in principle, every available criterium. In phages, this amounts to some 70 properties[16]. For practical purposes, the most important properties are the nature of nucleic acid and morphology and physicochemical properties of the virion, now increasingly completed by genomic data. The

ICTV classifies virions and not isolated genes or proteins. In the past, the ICTV has seen spirited battles, more on nomenclature than classification, for example, the use of English vernacular names versus a latinised nomenclature. Among others, it was proposed that viral properties should be indicated by a system of eight descriptors, including nature and molecular weight of nucleic acid. This was called a "Cryptogram"[17], but it is not used any more, despite its high descriptive value.

Phage classification started in earnest in 1967 with a seminal paper by Bradley[18]. He proposed six basic morphological types, corresponding respectively to tailed phages (with contractile tails, long and noncontractile tails, and short tails), small isometric ssDNA viruses, filamentous phages and small ssRNA phages. This scheme was adopted by the ICTV[14]. At that time, only 111 phages were known to any extent[19]. In 1974, the tailed phages of the Bradley scheme were subdivided into morphotypes, but this was purely for better identification by electron microscopy[20].

C. The present state of phage classification

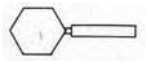
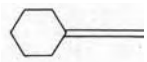
1. Orders and families

Phages have double-stranded or single-stranded DNA or RNA. Particles are tailed or polyhedral, filamentous or pleomorphic. Morphology, physicochemical and physiological properties of phage families have been reviewed several times and the reader is referred to these publications[16,21-23]. Detailed descriptions of some phage taxa may be found in reference 24.

Tailed phages constitute the order Caudovirales with three families, characterised by contractile, long and noncontractile, or short tails and named respectively Myoviridae, Siphoviridae, and Podoviridae (Table 1). They represent over 96% of phages. Their heads are icosahedra or closely related bodies. Most problems of phage classification are linked to tailed phages because of their extraordinary numbers and an enormous amount of data (often of low quality). The VIIIth ICTV Report includes 17 genera of tailed phages[24].

The seven families of polyhedral, filamentous, and pleomorphic families are separated by profound differences in nucleic acid content and structure. All families are small, sometimes have a single member and are taxonomically unproblematic. The virions of four groups contain lipids and two of them have lipoprotein envelopes (Table 1).

Table 1. Overview of phage families.

Shape	Order or family	Nucleic acid, particulars, size	Member	Number[a]
	Caudovirales	dsDNA (L), no envelope		
	Myoviridae	Tail contractile	T4	1312
	Siphoviridae	Tail long, noncontractile	l	3262
	Podoviridae	Tail short	T7	771
	Microviridae	ssDNA (C), 27 nm, 12 knoblike capsomers	f X174	38
	Corticoviridae	dsDNA (C), complex capsid, lipids, 63 nm	PM2	3?
	Tectiviridae	dsDNA (L), inner lipid vesicle, pseudo-tail, 60 nm	PRD1	19
	Leviviridae	ssRNA (L), 23 nm, like poliovirus	MS2	38
	Cystoviridae	dsRNA (L), segmented, lipidic envelope, 70–80 nm	f 6	3
	Inoviridae	ssDNA (C), filaments or rods, 85–1950 x 7 nm	fd	66
	Plasmaviridae	dsDNA (C), lipidic envelope, no capsid, 80 nm	MVL2	5

[a] From reference 1. C, circular; L, linear.

2. Subdivision of the *Podoviridae* and *Myoviridae* families

The fully sequenced genomes of 55 Podoviridae and later 102 Myoviridae were compared by the CoreGenes and CoreExtractor programs[25,26]. Taxa were defined by the number of shared homologous/orthologous proteins. ICTV phage genera were generally confirmed and often extended and subdivided. The results are summarised in Tables 2 and 3. The very large T7, f29, P2, SPO1, and T4 "supergroups" were subdivided into subfamilies and many new "genera" were set up. In addition, both the Podoviridae and Myoviridae groups included some 20 viruses which, apparently, stood alone, were unrelated to other phages, and seemingly represented independent genera. This approach must now be extended to the Siphoviridae family. We found very few cross-reactions between phages of different families, the most notable being those between lambda-like siphoviruses and P22-like podoviruses. The finding of a swarm of apparently unrelated "orphan" viruses is consistent with the extreme diversity of bacteriophages indicated by metagenomics.

Our approach was a development of the Phage Proteomic Tree[27,28], but went considerably farther. There is an impressive consensus between ICTV phage genera, our schemes, and the Phage Proteomic Tree. Many more confirmations are seen in other genomic approaches, for example, a phylogenetic approach based on terminase subunits[29], the Phage Finder program for prophages[30] and also in the clusters of related proteins in the ACLAME database[31] and of orthologous genes in completely sequenced dsDNA phages[32]. It appears that horizontal gene transfer does not totally obliterate evolutionary relationships between phages[25].

3. Species

The ICTV is moving toward species definitions; however, in my opinion, no biologist can certify what a species is. There are 23 species definitions in the literature, including one for dinosaurs and other fossils[33]. Many biologists would like to content themselves with the "biological species definition" by Mayr[34]. It postulates that a species is "a group of interbreeding

Table 2. Reclassification of Podoviridae phages.

Subfamily	Genus	Example	Members	Host
Autographivirinae	T7-like	T7	8	Enterics, Pseudomonas, Vibrio
	SP6-like	SP6	4	Enterics
	f KMV-like	f KMV	3	Pseudomonas
	P60-like	P60	3	Prochlorococcus, Synechococcus
Nanovirinae	f29	f29	4	Bacillus
	44AHJD	44AHJD	7	Staphylococcus
(P22-like)	P22-like	P22	7	Enterics
---	BPP-1-like	BPP1	4	Bordetella, Burkholderia
---	e15-like	e15	2	Enterics
---	N4-like	N4	1	Enterics
---	119-like	119	2	Pseudomonas
---	VP2-like	VP2	2	Vibrio

Table 3. Reclassification of Myoviridae phages.

Subfamily	Genus	Example	Members	Host
Teequatrovirinae	T4-like	T4	15	Enterics, Acinetobacter, Aeromonas
	KVP40-like	KVP40	5	Aeromonas, Vibrio
	(Cyanophages)	S-PM2	4	Synechococcus, Prochlorococcus
Peduovirinae	P2-like	P2	13	Enterics, Burkholderia, Mannheimia, Pseudomonas, Ralstonia
	HP1-like	HP1	6	Aeromonas, Haemophilus, Pasteurella, Vibrio
Spounavirinae	SPO1-like	SP01	1	Bacillus
	Twort-like	Twort	7	Staphylococcus, Listeria
---	Mu-like	Mu	2	Enterics
---	P1-like	P1	2	Enterics
---	Bcep781-like	Bcep781	5	Burkholdera, Xanhomonas
---	BcepMu-like	BcepMu	2	Burkholderia
---	Felix O1-like	Felix O1	3	Enterics
---	HAP1-like	HAP1	2	Halomonas, Vibrio
---	I3-like *	I3	7	Mycobacterium
---	f CD119-like	f CD119	3	Clostridium
---	f KZ-like	f KZ	2	Pseudomonas
---	PB1-like	PB1	7	Pseudomonas

* Renamed I3-like after Mycobacterium phage I3.

natural populations that are reproductively isolated for other groups". However, this definition was created for songbirds, is totally inapplicable to haploid entities like viruses, and already fails when it comes to dogs and wolves; for example, when the Eskimos decide that their dogs must be improved, they leave a bitch outside and the wolves oblige. Presently, the ICTV has adopted the "polythetical species definition", meaning that a virus species is a polythetic class of individuals that constitute a replicating lineage and share a particular biotic niche[24]. Unfortunately, this definition is of no practical help. Classification into species is thus left to the intuition of individual taxonomists and remains very much an art.

D. Nomenclature

Nomenclature is inseparable from classification. The ICTV uses latinised terms for order, family, subfamily and genus names. Families are characterised by the suffix, -viridae. Species epithets are not latinised; for example, phage T4 is and will remain T4. The family names have been proven to be very useful. Indeed, it is much more elegant to say "Siphoviridae" instead of "phages with long, noncontractile tails". The ICTV has now banned hyphens and Greek letters in virus names which, unfortunately, are very frequent in phages (for example, phage f X174). I believe that this was not the right decision and that virus names should never be modified. In recent times, a system for naming phages has been devised that recalls the Cryptogram[35].

E. Problems of classification

Classification is defined as the act of classifying and the edifice resulting from this. Humans classify all the time, while the human mind tends to simplify by screening out data and criteria. Biological classification should ideally reflect evolutionary relationships. The problems of classification are both virus-related and man-related.

1. Viral problems

a. The viral properties themselves may be inappropriate for classification or are, as genomic sequences, only determined in specialised laboratories and at great cost of money and time. As an example of the former, I remember the heady days of protein sequencing, when some people believed that viruses should be classified by their amino acids. It was also believed that all illnesses of bacteriological classification could be cured by determining G+ C percentages. Fortunately, this is now history.

b. A completely hierarchical virus classification appears as an impossible dream since viruses, including bacteriophages, are clearly polyphyletic.

c. Phages (and any viruses) evolved vertically and by horizontal gene transfer from other phages and a variety of other organisms. The latter is probably the main mechanism of phage evolution and gives rise to reticulate groups. Indeed, many phage genomes, especially of siphoviruses, appear as genetic mosaics composed of "modules", that is single genes or groups of genes that are exchangeable. The modules include head and tail and possibly other genes. A reticulate group of phages has been called a "modus". Phenetic properties are seen as unreliable and classification should be based entirely on genome sequences[36]. It has also been proposed to base tailed phage taxonomy on a single structural module of head or tail genes[37]. These proposals are limited to a few phages and there has been no follow-up.

A modular classification is attractive as it reflects evolutionary relationships. Unfortunately, the number of possible cross-links[38] may be enormous (imagine 5,000 phages with, say, 50 genes) and it is unclear whether all genes are to be counted. For example, the genome of P. aeruginosa phage f KZ is a collection of genes of the most diverse origin (worms, the Drosophila, the rat, Bacillus phages)[39]. It seems that "mosaicism" and reticulate evolution are general features of the living world and not specific to phages. For example, the human genome contains a T4-type lysozyme[21] gene and some 100,000 defective endogenous retrovirus genomes from monkeys, birds, and cats[40]. The feasibility of a reticulate classification is not evident.

2. Man-made problems

a. The major problem is poor electron microscopy, namely unsharp, low-contrast pictures with unreliable dimensions. Standards clearly fell in recent years and many pictures in the newer literature are far inferior to those obtained in 1959 at the moment of introduction of negative staining[41].

b. Valuable viral properties, such as the complete base composition (ATGC), the presence of sugars and modified bases in phage DNA, particle weight, or DNA-DNA homology are no longer determined[16] because of the present emphasis on genomics.

c. Databases. For example, the important GenBank database is user-driven and accepts data from unpublished papers that may never see print. The reason is that many journals require new sequences to be deposited in a database prior to acceptance of the papers describing them. Since at least 50% of papers are rejected, this leads to the accumulation of possibly worthless material. Furthermore, GenBank makes no difference between "phages" and "prophages", however defective.

d. Classification by a single criterium. This is a very dangerous undertaking. It went well with phage classification by terminases[29]. On the other hand, a classification by RNA and DNA polymerases worked well with RNA plant viruses, but backfired when seven tailed bacteriophages were sorted into two phyla, two classes, and six orders according to their DNA polymerases[42].

E. Why phage classification?

The main purposes of classification are generalisation and simplification. It is impossible and pointless to memorise the properties of 5000 individual tailed phages, but it is much more rewarding to study tailed phages as a group. Classification facilitates comparisons and thus virus research and understanding of viruses. It is also indispensable for teaching, textbooks, doctoral theses, phylogenetic studies and databases. Classification is also necessary for identification of novel and therapeutic phages, of harmful phages in biotechnology and industrial fermentations, and of industrially important phages in patent applications. Finally, it is a valuable research aid as it allows for the control of the accuracy of data by comparison with known phages.

References

1. Ackermann, H.-W. (2007) 5500 Phages examined in the electron microscope. Arch. Virol. 152, 277–243.
2. Twort, F.W. (1915) An investigation on the nature of ultra-microscopic viruses. Lancet ii, 1241–1243.
3. D'Herelle, F. (1917) Sur un microbe invisible antagoniste des bacilles dysentériques. C.R. Hebd. Seances Acad. Sci. D 165, 373–375.
4. D'Herelle, F. (1918) Technique de la recherche du microbe filtrant bacteriophage (Bacteriophagum intestinale). C. R. Seances Soc. Biol. Filiales 81, 1160–1162.
5. Breitbart, M. and Rohwer, F. (2005) Here a virus, there a virus, everywhere the same virus? Trends Microbiol. 13, 278–284.
6. Brüssow, H. and Hendrix, R.W. (2002) Phage genomics: small is beautiful. Cell 108, 13–16.
7. Suttle, C.A. (2005) Viruses in the sea. Nature 437, 356–361.
8. Angly, F.E. et al. (2006) The marine viromes of four oceanic regions. PLoS Biol. 4, 2121–2131 (e368).
9. Rohwer, F. (2003) Global phage diversity. Cell 113, 141.
10. Burnet, F.M. (1933) The classification of dysentery-coli bacteriophages. III. A correlation of the serological classification with certain biochemical tests. J Pathol. Bacteriol. 37, 179–184.
11. Ruska, H. (1943) Versuch zu einer Ordnung der Virusarten. Arch. Ges. Virusforsch. 2, 480–498.
12. Holmes, F.O. (1948) Order Virales. In Bergey's Manual of Determinative Bacteriology (6th edn) (Breed, R.S. et al., eds), pp. 1126–1286. Williams & Wilkins, Baltimore, MD.
13. Lwoff, A. et al. (1962) A system of viruses. In Basic Mechanisms in Animal Virology. Cold Spring Harb. Symp. Quant. Biol. 27, 51–62.
14. Wildy, P. (1971) Classification and nomenclature of viruses. First Report of the International Committee on Nomenclature of Viruses. S. Karger, Basel.
15. Anonymous (2011) Virus classification. Wikipedia, http://en.wikipedia.org/wiki/Virus_classification. Accessed 1 February 2011
16. Ackermann, H-W. (2009) Phage classification and characterization. In Bacteriophages. Methods and Protocols, Vol. I, Isolation, Characterization, and Interactions (Clokie M.R.J. and Kropinski A.M., eds), Methods in Molecular Biology, 501, 127–140, Humana Press, Clifton, NJ.
17. Gibbs, A.J. et al. (1966) What's in a virus name.? Nature 209, 450–454.
18. Bradley, D.E. (1967) Ultrastructure of bacteriophages and bacteriocins. Bacteriol. Rev. 31, 230–314.
19. Eisenstark, A. (1967) Bacteriophage techniques. In Methods in Virology, Vol. 1 (Maramorosch K. and Koprowski, H., eds), pp. 449–524. Academic Press, London.
20. Ackermann, H.-W. and Eisenstark, A. (1974) The present state of phage taxonomy. Intervirology 3, 201–219.
21. Ackermann, H.-W. (2005) Bacteriophage classification. In Bacteriophages: Biology and Applications (Kutter, E. and Sulakvelidze, A., eds), pp. 169–187, CRC Press, Boca Raton, FL.
22. Ackermann, H.-W. (2006) Classification of bacteriophages. In The Bacteriophages, (2nd edn) (Calendar, R., ed.), pp. 8–16. Oxford University Press.
23. Ackermann, H.-W and DuBow, M.S. (1987) Viruses of Prokaryotes. Vol. II. General Properties of Bacteriophages, pp. 1–54 and 171–218. CRC Press, Boca Raton, FL.
24. Fauquet CM et al., eds (2005) Virus Taxonomy. VIIIth Report of the International Committee on Taxonomy of Viruses, pp. 7, 43–94, 279–299, 443–446, 741–749, Elsevier Academic Press.
25. Lavigne, R. et al. (2008) Unifying classical and molecular taxonomic classification: analysis of the Podoviridae using BLASTP-based tools. Res. Microbiol. 159, 406–414.
26. Lavigne, R. et al. (2009) Classification of Myoviridae bacteriophages using protein sequence similarity. BMC Microbiol. 9, 224.
27. Edwards, R.A. and Rohwer, F. (2005) Viral metagenomics. Nature Rev. Microbiol. 3, 504–510.
28. Rohwer, F. and Edwards, R. (2002) The Phage Proteomic Tree: a genome-based taxonomy for phage. J. Bacteriol. 184, 4529–4535.
29. Serwer, P. et al. (2004) Improved isolation of undersampled bacteriophages: finding of distant terminase genes. Virology 329, 412–424.
30. Fouts, D.E. (2006) Phage-Finder: automated identification and classification of prophage regions in complete bacterial genome sequences. Nucleic Acids Res. 34, 5839–5851.
31. Lima-Mendez, G. et al. (2008) Reticulate representation of evolutionary and functional relationships between phage genomes. Mol. Biol. Evol. 25, 762–777.
32. Glazko, G. et al. (2007) Evolutionary history of bacteriophages with double-stranded DNA genomes. Biol. Direct 2 art 36.
33. Ackermann, H.-W et al. (1992) The species concept and its application to bacterial viruses. Arch. Virol. 124, 69–82.
34. Mayr, E. (1969) Principles of Systematic Zoology, p. 42. McGraw-Hill, New York.
35. Kropinski, A. et al. (2009) Position paper: the creation of a rational scheme for the nomenclature of viruses of Bacteria and Archaea. Environ. Microbiol. 11, 2771–2953.
36. Lawrence, J.G. et al. (2002) Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. J. Bacteriol. 184, 4891–4905.
37. Proux, C. et al. (2002) The dilemma of phage taxonomy illustrated by comparative genomics of Sfi21-like Siphoviridae in lactic acid bacteria. J. Bacteriol. 184, 6026–6036.
38. Hendrix, R.W. et al. (1999) Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc. Natl. Acad. Sci. USA 96, 2192–2197.
39. Mesyanzhinov, V.V. et al. (2002) The genome of bacteriophage fKZ of Pseudomonas aeruginosa. J. Mol. Biol. 317, 1–19.
40. Belshaw, R. et al. (2004) Long-term reinfection of the human genome by endogenous retroviruses. Proc. Natl. Acad. Sci. USA 104, 4894–4899.
41. Brenner, S. et al. (1959) Structural components of bacteriophage. J. Mol. Biol. 1, 281–292.
42. Ward, C.W. (1993) Progress towards a higher taxonomy of viruses. Res. Virol. 144, 419–453.

Biography

Hans-W Ackermann MD, is of German origin and a retired professor of microbiology. He has been living in Canada for over 40 years. He was a vice-president of the International Committee on Taxonomy of Viruses (ICTV) and for some 20 years chairman of the Bacterial Virus Subcommittee of the ICTV. His research interests are electron microscopy and phage taxonomy. He has published over 200 articles and several books on these subjects. Although retired, he continues to work in his electron microscopy laboratory.