

Virus systematics: taxonomy for the tiny

The classification of viruses poses different problems from those of classifying more elaborate life forms as

Anne-Lise Haenni and **Mike Mayo** describe.

A question frequently asked of virologists by non-virologists is – are viruses living? Perhaps one answer to this chestnut is in the nature of virus systematics. Although molecular in scale (circoviruses have ~20 nm diameter virions and ~2 kb ssDNA genomes; viroid genomes consist of ssRNA of only ~0.3 kb), viruses and viroids are replicated by error-prone mechanisms that result in variations from which natural selection yields distinctive forms fitted to particular niches. Thus one can study and describe virus variation, investigate the causes and consequences of this variation and manipulate the data to produce a system of classification. This constitutes a classic definition of systematics and can be said to describe much of current virological research activity.

Background

Intrinsic to any biological research are **classification** – the production of a logical system of categories, each containing

any number of organisms, that allows easier reference to its components, **nomenclature** – the construction of a system of names and a formal guide to their use in taxonomy, and **taxonomy** – the process of naming organisms and classifying them hierarchically to express their mutual relationship in a simplified way and according to internationally agreed codes of practice. In virology, these activities are undertaken by the International Committee on the Taxonomy of Viruses (ICTV). ICTV consists of a network of some 500 virologists organized into Study Groups for particular groups of viruses (e.g. a family) that report to a subcommittee for each of the major groups of hosts (vertebrates, invertebrates, plants, prokaryotes, or fungi, protozoa and algae). The Chairs of these subcommittees form the backbone of the Executive Committee of ICTV.

It is feasible to administer virus taxonomy (including nomenclature) through such an organization because of the relatively small number of virus species currently recognized. In contrast to the approximately 17,000 new species

Genome	Family	Genera	Hosts	Genome size (kb)
dsDNA	Myoviridae	6	Pro	169
	Siphoviridae	8	Pro	43
	Podoviridae	4	Pro	40
	Tectiviridae	1	Pro	15
	Corticoviridae	1	Pro	10
	Plasmaviridae	1	Pro	12
	Lipothrixviridae	3	Pro	16
	Rudoviridae	1	Pro	33
	Fuselloviridae	1	Pro	15
	Guttaviridae	1	Pro	20
	Poxviridae	11	V	170
	Asfarviridae	1	VI	180
	Iridoviridae	5	I	150
	Phycodnaviridae	6	Algae	335
	Baculoviridae	2	I	134
	Nimaviridae	1	I	293
	Herpesviridae	10	V	130
	Adenoviridae	4	V	31
	Polyomaviridae	1	V	5
	Papillomaviridae	16	V	7
Polydnaviridae	2	I	245	
Ascoviridae	1	I	120	
Unassigned genera	3			
ssDNA	Inoviridae	2	Pro	6
	Microviridae	4	Pro	6
	Geminviridae	4	Pl	6
	Circoviridae	2	V	2
	Nanoviridae	1	Pl	6
	Parvoviridae	9	V	5
Unassigned genera	1			
RT	Hepadnaviridae	2	V	3
	Caulimoviridae	6	Pl	8
	Pseudoviridae	3	I,Pl,F	6
	Metaviridae	3	I,Pl,F	5
	Retroviridae	7	V	7
dsRNA	Cystoviridae	1	Pro	13
	Reoviridae	12	VI,Pl,F	24
	Birnaviridae	3	V	6
	Totiviridae	3	F	5
	Partitiviridae	3	FPl	4
	Chrysoviridae	1	F	13
	Hypoviridae	1	F	10
Unassigned genera	1			

Genome	Family	Genera	Hosts	Genome size (kb)
-ssRNA	Bornaviridae	1	V	9
	Rhabdoviridae	6	V,Pl,I	12
	Filoviridae	2	V	19
	Paramyxoviridae	7	V	15
	Orthomyxoviridae	5	VI	10
	Bunyviridae	5	V,Pl,I	12
	Arenaviridae	1	V	11
	Unassigned genera	4		
	+ssRNA	Leviviridae	2	Pro
Narnaviridae		1	F	3
Picornaviridae		9	V	7
Dicistroviridae		1	I	9
Marnaviridae		1	F	9
Sequiviridae		2	Pl	10
Comoviridae		3	Pl	10
Potyviridae		6	Pl	10
Caliciviridae		4	V	8
Astroviridae		2	V	7
Nodaviridae		2	I	4
Tetraviridae		2	I	8
Luteoviridae		3	Pl	6
Tombusviridae		8	Pl	4
Coronaviridae		2	V	28
Arteriviridae		1	V	13
Roniviridae		1	I	26
Flaviviridae		3	VI	11
Togaviridae		2	VI	10
Bromoviridae	5	Pl	8	
Tymoviridae	3	Pl,I	6	
Closteroviridae	3	Pl	16	
Flexiviridae	8	Pl	6	
Barnaviridae	1	F	4	
Unassigned genera	14			
Viroids	Pospiviroidae	5	Pl	0.4
	Avsunviroidae	2	Pl	0.3

Table 1. Virus families and some representative properties

Abbreviations: Pro, prokaryotes; V, vertebrates; I, invertebrates; Pl, plants; F, fungi. Colours indicate main hosts, gradients signify more than one type of host. Genome sizes are for a representative species.

of animals that are described every year, the virus world consists of a mere 1,950 species. As parasite diversity surely parallels host diversity, this number is clearly a massive under-representation, but is probably held in check because to establish a new virus species is a laborious research exercise and only ‘important’ hosts such as man are studied intensively.

Viruses are diverse in genome type and size and infect hosts in all major types of organisms. The table shows the currently recognized virus families clustered according to genome type. Taxonomy above the level of families is at present limited to the recognition of three orders; most families are not clustered into orders and some genera

are as yet unassigned to taxa above the level of genus. This reflects some of the uncertainties in current virus taxonomy.

The nomenclature of virus taxa, in particular species, differs fundamentally from that of the rest of biology. It is regulated by the International Code of Virus Classification and Nomenclature, which is published on the ICTV website (www.danforthcenter.org/iltab/ICTVnet/asp/_MainPage.asp) and in reports such as the current *8th ICTV Report*. This article is concerned only with the classification aspects of ICTV's work, in particular the factors that complicate virus classification: the intrinsic properties of viruses and the probable evolutionary origins of viruses.

Complicating factors – properties of viruses

As for many microbes, morphological characters are of limited use (although the advent of electron microscopy did lead to significant taxonomic advances) and small genetic changes can result in substantial changes in pathological impact. Virus genomes are small (see Table 1) and can often vary rapidly. However, it is now possible to obtain nucleotide sequences very quickly, sometimes without isolating the virus from its host. By using appropriate software, evolutionary distances between sequences can be deduced and phylogenetic trees obtained. This assumes that the sequences are derived from a single ancestral sequence and

that differences have arisen by mutation of single bases. Such a model seems most reasonable when sequences are clearly similar, such as when viruses in a genus are compared. When other sorts of sequence variation have happened, such as recombination, these simple comparison techniques are likely to lead to erroneous conclusions. This is vividly illustrated by the genomes of viruses that are classified in the family *Luteoviridae*. There is no dispute that viruses in this family are related both in sequence details and in their biology (all are transmitted by aphids in a circulative fashion that involves virions crossing several barriers inside aphid bodies). The RNA genomes of luteovirids consist largely of two gene blocks, one of structural genes and the other of replication-related (*pol*) genes (Fig. 1). The phylogenetic trees obtained by comparing luteovirid structural genes confirm the family cluster and show a separation of species into the constituent genera. But when *pol* genes are compared with those of other viruses, the genera *Luteovirus* and *Polerovirus* differ greatly and each seems more closely related to one or the other non-luteovirid genus. It seems that recom-

bination between gene blocks has happened during the evolution of these viruses (Fig. 1). Expecting sequence comparisons and the resulting trees to produce a durable classification is thus, at least in some instances, a naïve approach to virus classification. Whereas it works for some virus taxa, for others a more arbitrary and broadly based approach that uses multiple criteria is clearly necessary.

Complicating factors – evolutionary history

Biological taxonomy is well suited to classifying organisms that are related to one another via simple branched and diverging descent, with relatively long evolutionary distances between successive branch points. However, virus evolution differs from this simple paradigm in several ways.

1. Genomes of some viruses, such as luteovirids (see above; Fig. 1), have

been formed by recombination (joining of parts of separate genomes to produce a new genome) and/or reassortment (making new combinations among separate parts of the genomes of viruses with multipartite genomes, such as influenza viruses), resulting in chimeric viruses with polyphyletic (i.e. hybrid) genomes whose origins cannot be represented in a monophyletic scheme.

2. Genomes of some viruses seem to contain genes acquired from their hosts.
3. Viruses are unlikely to be descended from one original 'protovirus'; they probably arose more than once, but when and how many times is unclear.
4. Genomes of some viruses (e.g. retroviruses) integrate into the genomes of the hosts, where they are subject to selection pressures

different from those applying to independently replicating genomes.

5. Some viruses infect more than one type of host (one often serving as vector for transmission between other hosts) and therefore will have evolved in response to complex selection pressures.

Any or all of these factors greatly complicate, if not vitiate, attempts to deduce phylogeny from studies of current genome sequences and thus, whatever the taxonomic arrangement, some viruses and/or taxa will always seem to be misfits.

Conclusion

Notwithstanding the complications, some sort of classification is fundamental to most virological research, whether it be deducing the nature of a new disease agent by working out which of the known taxa it most resembles or developing a better understanding of the nature of a virus by predicting properties based on its assumed classification.

To develop and maintain such a classification, ICTV communicates with virologists in several ways. Via its website, ICTV provides many of the details of its organization as well as current issues in virus taxonomy. ICTV also publishes regular formal reports (such as the recent *8th ICTV Report*) and occasional notes in the Virology Division News section of *Archives of Virology*. Also, workshops, such as the recent meeting on virus evolution and taxonomy which has been summarized by U. Desselberger, are held and, as work progresses, ICTV expects to organize further such workshops.

New viruses are being discovered all the time. Some can be fitted into existing taxonomic structures (e.g. *Severe*

acute respiratory syndrome coronavirus; SARS-CoV), making it possible to predict their properties, and others are unlike anything seen previously (e.g. *Acanthamoeba polyphaga mimivirus*). Thus, virus taxonomy must have both an internal logical justification and a capability of expansion to accommodate novelty. Whatever the virological endeavour, virus taxonomy should provide a provocative framework to which all virologists can contribute. Current ICTV practice aims to achieve this.

Anne-Lise Haenni

Current ICTV Secretary

Institut Jacques Monod, CNRS and Universities VI and VII, 2 Place Jussieu – Tour 43, 75251 Paris Cedex 05, France

(e haenni@ijm.jussieu.fr)

Mike Mayo

ICTV Secretary, 1999–2005

Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK

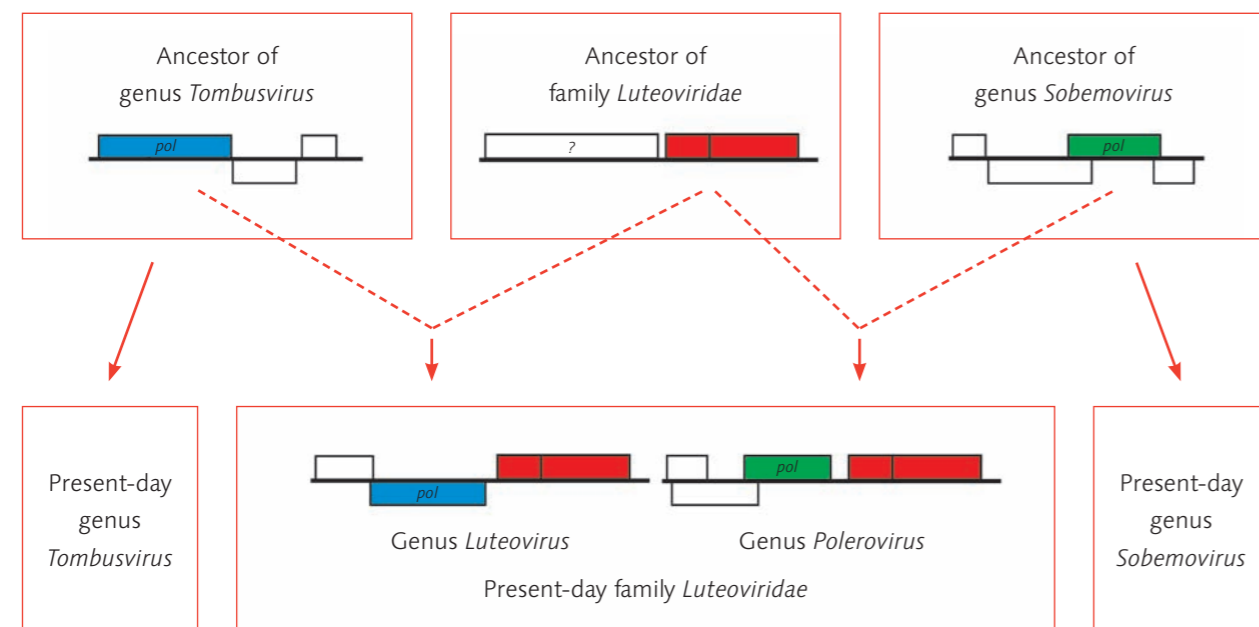
(e m.mayo@scri.ac.uk)

Further reading

Fauquet, C.M., Mayo, M.A., Malinoff, J., Desselberger, U. & Ball, L.A. (editors) (2005). *Virus taxonomy, VIIIth Report of the ICTV*. London: Elsevier/Academic Press.

Desselberger, U. (2005). Report on an ICTV-sponsored symposium on virus evolution. *Arch Virol* 150, 629–635.

▼ Fig. 1. Diagram of possible origins of the genomes of viruses in different genera in the family *Luteoviridae*. The upper boxes contain hypothetical genomes and assume that the ancestors of tombusviruses and sobemoviruses had genomes like those of their current-day counterparts. Red boxes represent blocks of structural protein genes. *pol* signifies replication-related genes. ? signifies a gene block of unknown type. Further details can be found in the *8th ICTV Report*.



Virus taxonomy must have both an internal logical justification and a capability of expansion to accommodate novelty. Whatever the virological endeavour, virus taxonomy should provide a provocative framework to which all virologists can contribute.