

Sequence analyses and a unifying system of virus taxonomy: consensus via consent

C. H. Calisher¹, M. C. Horzinek², M. A. Mayo³, H.-W. Ackermann⁴,
and J. Maniloff⁵

¹ Arthropod-borne and Infectious Disease Laboratory, Colorado State University, Fort Collins, Colorado, U.S.A.; ² Veterinary Faculty, Virology Unit, Utrecht University, Utrecht, The Netherlands;

³ Scottish Crop Research Institute, Invergowrie, Dundee, Scotland; ⁴ Department of Microbiology, Faculty of Medicine, Laval University, Quebec, Canada; ⁵ Department of Microbiology and Immunology, University of Rochester, Rochester, New York, U.S.A.

Introduction

Once a relationship, or lack of it, between viruses has been established, they can be classified and assigned to individual types. Virus classification and naming, collectively known as taxonomy, have been administered by the International Committee on Taxonomy of Viruses (ICTV) for the past 30 years (Fig. 1). Hundreds of virologists world-wide have been involved in this painstaking process. For many viruses, a broadly satisfactory and working taxonomy has been developed, though not without substantial, and sometimes lengthy, debates.

In recent years, nucleotide and amino acid sequences have been determined for many viruses (Murphy et al. [17]). Several of these studies have shown that there are similarities among certain gene products of distantly related viruses (e.g., Goldbach [10]; Strauss and Strauss [21]). These comparisons have raised questions about phylogenetic relationships among viruses and among virus taxa, such as genera and families. Our current understanding of the connections between evolution, natural selection, and molecular biology leads to the expectation that taxonomic relationships also reflect phylogenetic relationships (de Queiroz and Gauthier [5]). In particular, phylogenetic relationships derived from sequence analyses should be consistent with previously established taxonomic assignments and, when taxonomy has not been established, it should be possible to use phylogenetic relationships derived from sequence analyses to postulate taxonomic associations.

This paper discusses the influences of sequence alignment on taxonomic thinking and describes the interactive process that has evolved in the virology community for decision-making about virus taxonomy. Although some taxonomic relationships suggested by comparisons of viral genomic sequences might reflect phylogenetic relationships, this may not always be so. In many instances, such relationships are supported to an extent by similarities between viruses in features other than gene or gene product sequences, as for example in the order *Mononegavirales* (Pringle [19]). However, when there are no such

International Committee on Taxonomy of Viruses

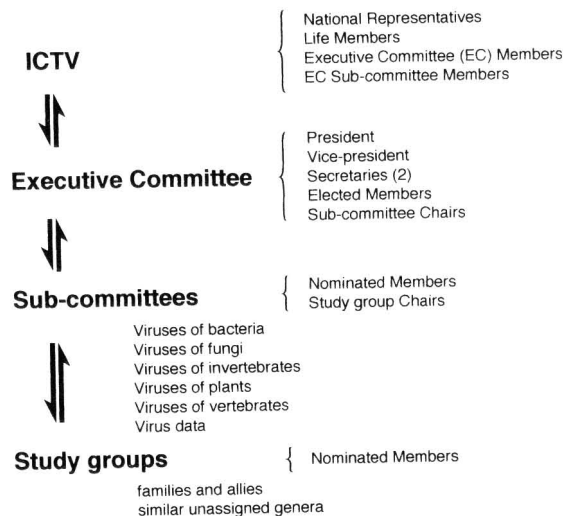


Fig. 1. Organization of the International Committee on Taxonomy of Viruses (ICTV)

supporting data, the process of identifying relationships between viruses based on sequence analyses alone may be misleading. It seems certain that virus genomes can recombine with other genomes to create new combinations of genes or even parts of genes, and this process of modular evolution (Botstein [3]; Gibbs [7]) makes the interpretation of comparisons between genome sequences very difficult.

Presentations such as this are a justification, albeit *post hoc*, for the way in which virus taxonomy has been developed. The process of decision-making in virus taxonomy always has been an essentially pragmatic process involving the establishment of a broad consensus for grouping virus species and genera into higher taxa (Murphy et al. [17]).

Taxonomic structures should be based on succinct data and a dose of opinions. Whereas good taxonomic ideas may come from anywhere, it is reasonable that the process extending from a good idea to taxonomic assignments proceeds by debate and the development of a consensus. The meticulous efforts required for establishment of a taxonomic scheme may be slow, but that is the price to be paid for constructing stable taxa. Therefore, while in no way wishing to inhibit speculative discussion, we urge that, when new ideas for virus taxonomic arrangements are developed, they be submitted to this process before they are disseminated in subsequent publications.

The problem

Viruses have been described as “mistletoe on the Tree of Life” (DJ McGeoch, pers. comm.). They parasitize distant branches of the tree, in sometimes similar ways, but relationships with, and origins from, their hosts are obscure. However, and more prosaically, whether viruses arose monophyletically (from a single ancestor, followed by divergence) or polyphyletically (from multiple ancestors, followed by divergence or convergence) there is

no doubt that viruses are heterogeneous. Current evidence suggests that viruses have arisen *de novo* from various sources at various times and, therefore, are not monophyletic and do not all have a common lineage (Rybicki [20]). However, some viruses with RNA genomes are thought to have evolved from common progenitors (Goldbach [10]; Strauss and Strauss [21]); this also could be so for some DNA viruses. Morse [16] suggests that Baltimore's grouping of viruses by nucleic acid type and strategy of replication [2] can be taken to hint at what we now call a "superfamily" arrangement. Such groupings emphasize common features of viruses and may allow the formation of much higher taxa than have been accepted to date.

Nevertheless, it is not, and probably never will be, known with certainty if viruses are monophyletic or polyphyletic in origin. Thus, any taxonomic scheme must, by definition, be contrived. Whether monophyletic with subsequent polydiversity, or polyphyletic with subsequent polydiversity, the origin of viruses undoubtedly is a complex issue. Therefore, devising a taxonomic scheme to suit presently available data must require considerable deliberation and discussion to achieve a panoramic view rather than a narrow one. Further, the occurrence of recombination and reassortment among parts of virus genomes in co-infected host cells, the exchange of genes between viruses and their hosts, and parallel evolution of viruses and hosts would seem to make it impossible to devise a single, simple explanation of such complex evolutionary topologies.

The ICTV approach

The organization charged by the international microbiological community with guiding and controlling virus taxonomy is the ICTV, a Committee of the Virology Division of the International Union of Microbiology Societies. Its objectives are (1) to develop an internationally agreed taxonomy for viruses; (2) to establish internationally agreed names for taxonomic groups of viruses; and (3) to communicate the latest results on the classification and nomenclature of viruses to virologists by holding meetings and publishing reports (e.g., Francki et al. [6]; Murphy et al. [17]). The ICTV has a hierarchical organization (Fig. 1) in which virologists contribute as national members, and/or as members of the Executive Committee, of its subcommittees, or of their Study Groups. Taxonomic questions are discussed at all levels until a consensus is reached. In this way, taxonomic decisions are made, revised, and up-dated at approximately 3-year intervals (Murphy et al. [17]).

The taxonomic schemes developed by the ICTV were not designed to have evolutionary implications, but they do not indicate otherwise. The problem is that while a universal virus taxonomic scheme is desirable, probably there is no single phylogenetic scheme applicable to all viruses. It has been the policy of the ICTV to allow the gradual evolution of a universal taxonomic scheme by accepting guidance from individual Study Groups composed of recognized experts in each area of virology. The evolution of such taxonomic formations is tempered by tests of their acceptability to virology in general.

The phylogenetic implications of viral taxonomy increase as viruses are grouped into higher order taxa: species into genera, genera into families, and families into orders. Because at present different amounts of data are available for different viruses, it would be unreasonable to organize more than a few viruses into higher order taxa. However, the ICTV has made progress towards a universal virus taxonomy by establishing families, placing many viruses in those families, establishing genera in which individual viruses have

been placed, accepting the definition of an order (Pringle [19]), and adopting the species concept (Van Regenmortel et al. [23]). It has been through this methodical approach to the chaotic picture presented by viruses that the ICTV has been most influential in the world community of virologists.

Most definitions of “viruses”, “species”, or “groups” do not take into consideration that expediency may be the principal factor in the creation of these taxa. Each working group of virologists, trying to elucidate the relationships of viruses within its area of expertise, sees its own charge in isolation from that of other groups. Recent advances in molecular techniques have provided detailed information about gene sequences for use in comparing viruses. It is tempting to take sequencing as the solution to all taxonomic questions. However, considerable historical and biological information is lost when obviously diverse entities are considered as a single entity because of sequence similarities. All taxonomy is opinion.

Alternative approaches

Although the key determinant of any biological taxonomic description is the genetic constitution of the “organism” (or virus), not all gene or genomic sequences are appropriate phylogenetic measures. Ward [24] has attempted to use gene sequences and viral polymerases as key phylogenetic characteristics. Although this was not the first attempt to devise higher taxa on the basis of similarities among viral polymerases (e.g., Kamer and Argos [12]; Argos [1]; Poch et al. [18]; Koonin [13]), it is a provocative scheme in that it assigns positive-strand RNA viruses to classes, which allows further subdivisions of these classes into orders, families, genera, and species. An advantage of this approach is that it establishes a possible framework on which to base a more extensive series of taxa at each level. One of the principal drawbacks is that it is not inclusive. For example, retroviruses are omitted, possibly because the polymerase data do not fit the hypothesis, and DNA viruses that do not encode a DNA polymerases are equally omitted.

Such a proposal illustrates the danger of a single-criterion approach to any taxonomic scheme. For example, on the basis of the presence of retroviral proviruses in the human genome, one could conclude that humans are retroviruses. Similarly, because a thymidylate synthetase gene is present in the genomes of T4 and certain tailed phages, but also in *Escherichia coli*, *Leishmania* sp., and humans (Hardy et al. [11]), are we to infer that all these species are relatives of humans? It is unlikely that this gene occurs in such varied life forms as a consequence of convergent evolution or serendipity. It is more likely that the thymidylate synthetase gene has moved horizontally between species and that this constitutes a prime example of “modular evolution” (Susskind and Botstein [22]).

Ward’s placement of the dsRNA totiviruses with the ssRNA carmoviruses is done on grounds that this is preferable to other suggested placements. His liberal movement of viruses, groups, genera, and families sometimes defies logic and uniformity, although this is justified by explaining that, because viruses clearly arose polyphyletically, their relations are polydiverse. While it increasingly appears that the origin of viruses is polyphyletic, it is also clear that their relationships and lack of relationships are far too complex to simply establish a taxon for each clade and move on.

It might have been possible for Ward to accommodate all DNA viruses in his proposal. He did not, suggesting that he recognizes the different organization of RNA and DNA

viruses. As he indicates, analyses of DNA polymerases of DNA viruses by others (Braithwaite and Ito [4]) suggest the possibility of establishing a higher taxonomy for these viruses. However, his arrangement of tailed phages is inconsistent, distributing only seven of them into six families and at least three classes and three orders based on structural relations of their DNA polymerases with DNA polymerases I and II of *E. coli*. Such a single-criterion approach seems made to order for proving one's hypothesis. Similarities between amino acid sequences among biological entities are common, so much so that one must be very careful when comparing such sequences and trying to answer questions such as "How long is long enough?". In fact, there are tailed phages that do not possess DNA polymerases at all, including, ironically, coliphage lambda, the phage that has given its name to the genus "lambda-like phages" (Murphy et al. [17]). We feel that Ward has overlooked the possibility of horizontal gene transfer, as witnessed by the existence of several types of DNA or RNA polymerases in a few tailed phages and their absence from others, suggesting that these polymerases are specified by "wandering genes". The occurrence of DNA polymerase types A and B in *E. coli* suggests that phages might have acquired them from bacteria.

The classification of luteoviruses is a further example of problems with this approach. All viruses in this genus have a similar arrangement of genes in the 3'-half of the genome, which encodes structural proteins (Martin et al. [14]; Mayo and Ziegler-Graff [15]). A characteristic of luteoviruses is that they are transmitted by aphids in a circulative, non-propagative fashion, probably reflecting similarities of 50% or more among luteovirus coat protein amino acid sequences. However, the 5'-portion of the genomes of luteoviruses encodes non-structural proteins including the viral polymerase, and the genus can be divided into two subgroups which have polymerases related to either of two completely dissimilar types with almost no sequence similarity. In some proposed taxonomies (Martin et al. [14]; Ward [24]) the distinct polymerases have led to the subgroups being classified in very distant branches of the phylogenetic tree. It seems certain that recombination has transferred different polymerase genes into a genome containing the luteovirus structural protein genes (or *vice versa*). Indeed, it seems likely that there have been several recombination events during the evolution of luteoviruses and other similar viruses of plants (Gibbs [8]; Gibbs and Cooper [9]). Such alternative recombinational histories for manifestly related viruses call into question the feasibility of trying to deduce virus phylogeny from sequence comparisons among particular genes.

The Ward approach to virus taxonomy and phylogeny represents a "top down" strategy, with higher order taxa (i.e., phyla) being defined for all viruses and then subdivided until the species level is reached. In contrast, the ICTV is following a data-driven "bottom up" strategy, with individual virus species being defined by groups of international experts based on experimental data. Clustering into higher taxa is then driven by the accumulation of data showing relationships between lower taxa. In some instances, such as the *Mononegavirales*, there are enough data to establish an order, the highest taxonomic level the ICTV has, thus far, certified. For other viruses, only limited data are available and only lower level taxa have been defined.

In summary, because the major evolutionary patterns of viruses and of cellular organisms appear to be quite different, their taxonomies cannot be made parallel. It may be that progenitor viruses of most higher taxa obtained their genes by recombination between viral genomes or from their hosts, whereas the more hierarchical phylogeny of cellular

organisms appears to have been by mutation and genomic rearrangement. Higher taxa, such as that which Ward advocates, can only be achieved by assuming or arbitrarily deciding that some genes are taxonomically more important than are others, although the basis of such a decision must be arbitrary. In essence, Ward's proposal applies the taxonomic paradigm used for cellular organisms to the taxonomic organization of viruses. But it is precisely because the ICTV emphasizes the differences as well as the similarities between viruses and cellular organisms that we thus far have been unable to include all viruses in higher taxa. We feel that to do otherwise would simply be creating artificial taxa.

Conclusions

Virology is now such a huge field that no single person, however articulate and widely read, can possibly address all the issues necessary for classifying all viruses. However, the ICTV encourages contributions by interested individuals to the collective insight of groups that not only attempt to represent the best interests of the scientific community but also have the wide-ranging expertise to handle very complex matters. As we recognize, the "collective insight of committees" can become parochial. This is among the many reasons that ICTV is structured so as to be as inclusive as possible of different opinions (Fig. 1). Virus taxonomy, as do so many other controversial issues, requires open debate among knowledgeable scientists.

Far-reaching taxonomic proposals have implications not simply for committees and publishers but also for the formation of concepts about virus evolution. We therefore repeat that the international community of virologists will continue to be best served if taxonomic proposals are reviewed and approved by the ICTV, the organization established by virologists for this purpose. In this way, taxonomic schemes are assured of equitable and critical reviews by experts in each area of consideration and individuals who submit such proposals can be given feedback with respect to consensus opinions. We trust this paper will serve as a stimulus for others to submit their opinions.

References

1. Argos P (1988) A sequence motif in many polymerases. *Nucleic Acids Res* 16: 9909–9916
2. Baltimore D (1971) Expression of animal virus genomes. *Bacteriol Rev* 35: 235–2417
3. Botstein D (1980) A theory of modular evolution for bacteriophages. *Ann NY Acad Sci* 354: 484–491
4. Braithwaite DK, Ito J (1993) Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res* 21: 787–802
5. de Queiroz K, Gauthier J (1994) Toward a phylogenetic system of biological nomenclature. *Trends Ecol Evol* 9: 27–31
6. Francki RIB, Fauquet CM, Knudson DL, Brown F (eds) (1991) Statutes of the ICTV. In: *Classification and Nomenclature of Viruses: Fifth Report of the International Committee on Taxonomy of Viruses*. Springer, Wien New York, pp 48–52 (*Arch Virol [Suppl]* 2)
7. Gibbs A (1987) Molecular evolution of viruses; 'trees', 'clocks' and 'modules'. *J Cell Sci* 7 [Suppl]: 319–337
8. Gibbs MJ (1995) The genome of carrot mottle mimic umbravirus and the evolution of the carmo- and sobemovirus families. Ph.D. Thesis, University of Oxford
9. Gibbs MJ, Cooper JI (1995) A recombinational event in the history of luteoviruses probably induced by base-pairing between the genomes of two distinct viruses. *Virology* 206: 1129–1132
10. Goldbach R (1987) Genomic similarities between plant and animal RNA viruses. *Microbiol Sci* 4: 197–202

11. Hardy LW, Finer-Moore JS, Montfort WR, Jones MO, Santi DV, Stroud RM (1987) Atomic structure of thymidylate synthetase: target for rational drug design. *Science* 235: 448–455
12. Kamer G, Argos P (1984) Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Res* 12: 7269–7282
13. Koonin EV (1991) The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J Gen Virol* 72: 2197–2206
14. Martin RR, Keese PK, Young MJ, Waterhouse PM, Gerlach WL (1990) Evolution and molecular biology of luteoviruses. *Ann Rev Phytopathol* 28: 341–363
15. Mayo MA, Ziegler-Graff V (1995) Recent developments in the molecular biology of luteoviruses. *Adv Virus Res* (in press)
16. Morse S (1994) Toward an evolutionary biology of viruses. In: Morse SS (ed) *The evolutionary biology of viruses*. Raven Press, New York, pp 1–28
17. Murphy FA, Fauquet CM, Bishop DHL, Ghabrial SA, Jarvis AW, Martelli GP, Mayo MA, Summers MD (eds) (1995) *Virus Taxonomy. Classification and Nomenclature of Viruses. Sixth Report of the International Committee on Taxonomy of Viruses*. Springer, Wien New York (Arch Virol [Suppl] 10)
18. Poch O, Blumberg BM, Bougueleret L, Tordo N (1990) Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. *J Gen Virol* 71: 1153–1162
19. Pringle CR (1991) The order *Mononegavirales*. *Arch Virol* 117: 137–140
20. Rybicki EP (1990) The classification of organisms at the edge of life, or problems with virus systematics. *S Afr J Sci* 86: 182–186
21. Strauss JH, Strauss EG (1988) Evolution of RNA viruses. *Annu Rev Microbiol* 42: 657–683
22. Susskind MM, Botstein D (1978) Molecular genetics of bacteriophage P22. *Microbiol Rev* 42: 385–413
23. Van Regenmortel MH, Maniloff J, Calisher CH (1991) The concept of virus species. *Arch Virol* 120: 313–314
24. Ward CW (1993) Progress towards a higher taxonomy of viruses. *Res Virol* 144: 419–433

Corresponding author: Prof. Dr. Dr. h.c. M. C. Horzinek, Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, Yalelaan 1, De Uithof, 3508 TD Utrecht, The Netherlands.