Res. Virol.
1993, 144, 419-453

REVIEW

Progress towards a higher taxonomy of viruses

C.W. Ward

CSIRO, Division of Biomolecular Engineering, Parkville, Victoria (Australia)

SUMMARY

The current consensus view is that a higher hierarchical taxonomy of viruses cannot be established for two reasons. Firstly, viruses appear to be polyphyletic in origin, with several sets of viruses arising by different, independent routes at different times. Secondly, subsequent virus adaptation for survival in different host/vector combinations has involved the selective acquisition of additional genes by a process of cassette or modular evolution, with these additional gene modules coming from other viruses or host genetic material. Thus, depending on the gene product used for comparison, different phylogenetic relationships can be deduced. Further virus adaptation can arise by reassortment of segmented genomes, gene duplication, deletions, frameshift mutations, point mutations or *de novo* development of new gene products from existing, unused reading frames.

The solution to the first objection is to place all viruses in a separate kingdom and assign the current viruses to several phyla that reflect these diverse origins. The solution to the second objection is to consider the core module of replication machinery as the major criterion on which to make the initial assignments to classes and orders. For RNA viruses, the major criterion is the sequence identity of the RNA-dependent RNA polymerase.

Using this criterion, the positive strand RNA viruses can be assigned to five classes that correspond to the recently recognized supergroups of RNA viruses. These five classes contain four, three, three, three and one order(s) respectively. These fourteen orders contain 31 virus families (including 17 families of plant viruses) and 48 genera (including 30 genera of plant viruses). This approach confirms the separation of the alphaviruses and flaviviruses into two families, the *Togaviridae* and *Flaviridae*, but suggests that several other current taxonomic assignments, such as the pestiviruses, hepatitis C virus, rubiviruses, hepatitis E virus and arteriviruses, may be wrong. The coronaviruses and toroviruses appear to be distinct families in distinct orders, not distinct genera of the same family as currently classified. In addition, the luteoviruses are split into two families and apple chlorotic leaf spot virus appears not to be a closterovirus but a new genus of the *Potexviridae*.

From an analysis of the polymerase dendrograms of the dsRNA viruses, it appears that they are not closely related to each other, but belong to four additional classes *(Partitiviridae, Reoviridae, Birnaviridae* and *Cystoviridae)* and one additional order *(Totiviri-*

Submitted July 1, 1993, accepted September 17, 1993.

dae) of one of the classes of positive ssRNA viruses in the same subphylum as the positive strand RNA viruses. The negative strand virus polymerase relationships confirm the assignment of the negative strand viruses to two orders in a single class in a separate subphylum of the RNA viruses.

This review includes preliminary data suggesting that the DNA viruses can also be assigned to higher taxa on the basis of the sequence identities of their highly conserved DNA polymerases. The suggested use of viral polymerases to establish higher order relationships is similar in principle to the use of highly conserved ribosomal RNA gene sequences in prokaryotic and eukaryotic taxonomy. This review also discusses the assignment of 33 of the 35 groups of plant viruses into genera of 25 families based on the nature of the genome and its arrangement, the level of sequence identity and, to a lesser extent, particle morphology.

Key-words: Virus, Taxonomy, Hierarchy; DNA viruses, RNA viruses; Review.

Historical background

At the turn of the century, the only physicochemical measurement available for comparing viruses was filterability. As a result, the earliest efforts to classify viruses were based on biological properties such as disease symptoms, ecological niches and transmission characteristics (Murphy and Kingsbury, 1990). By the 1930s, problems were already emerging with reliance on such criteria as it was becoming apparent that, in the same host, different viruses could cause very similar symptoms, while similar strains of the one virus could cause very different symptoms (Matthews, 1991). Consequently, the need for a unified approach to virus taxonomy and nomenclature was recognized at this time, and some of the early schemes have been reviewed by Matthews (1991).

By the 1950s, considerable biochemical and morphological data had been generated and the first groupings of viruses (myxoviruses, poxviruses, herpesviruses) on the basis of common virion properties were formulated (Murphy and Kingsbury, 1990). Around the same time, there was a rapid increase in the number of viruses discovered and the amount of information describing them. Between 1940 and 1966, various classification schemes were proposed independently by individuals and committees resulting in considerable confusion in the literature (Murphy and Kingsbury, 1990). None of these schemes were adopted by any significant number of virologists and the need for international cooperation and agreement to develop a single universal scheme for taxonomy and nomenclature became obvious (Matthews, 1991).

In 1966, at the International Congress of Microbiology in Moscow, the International Committee on Nomenclature of Viruses (ICNV) was formed. The history of the development of this organization, now known as the International Committee for Taxonomy of Viruses (ICTV), has been reviewed by Matthews (1983, 1985a, b). The reports of this committee (now triennial) provide the most up-to-date information on virus classification and nomenclature.

---

dsDNA = double-stranded DNA.
dsRNA = double-stranded RNA.
ICNV = International Committee on Nomenclature of Viruses.
ICTV = International Committee for Taxonomy of Viruses.
kb = kilobase.
kbp = kilobase pair.
ORF = open reading frame.
ssDNA = single-stranded DNA.
ssRNA = single-stranded RNA.

Note: acronyms for viruses are listed in legends to figures 1, 2, 3, 4 and 7.

The current position

The fifth ICTV report (Francki *et al.*, 1991) lists 73 families and groups of viruses infecting bacteria, algae, fungi, invertebrates, vertebrates and plants. It also reveals that the plant virologists had been reluctant, until recently, to abandon the concept of groups and accept the family and genus concepts so readily developed for viruses infecting other organisms. The fifth ICTV report listed three exceptions that had been accepted by ICTV at that time. These were the members of the *Reoviridae* (3 genera), *Rhabdoviridae* (2 genera) and *Bunyaviridae* (1 genus) that infect plants (Francki *et al.*, 1991) and which are so obviously similar to the other viruses in these families that infect vertebrates and invertebrates, that their inclusion in these families was inescapable (Matthews, 1991). As pointed out by Matthews (1985a, 1991), these developments reinforced the concept of the unity of virology and made it imperative that the current plant virus groups be assigned to families and genera since "the present situation where some groups are equivalent to families and others to genera has become confusing and anomolous, requiring urgent attention". This theme was also discussed by Van Regenmortel (1989), who suggested that the current plant virus groups be considered genera, and by Kingsbury (1987), who suggested they all be considered families.

During the last three years, considerable attention has been given by the ICTV plant virus subcommittee to the taxonomic assignment of these plant virus groups and these discussions have been concisely summarized by Martelli (1992). Following an international workshop in Braunschweig, Germany in September 1990, a proposal was developed for the family *Potyviridae.* At this meeting, the suggestion (Ward and Shukla, 1991) to use coat protein and gene sequence data as the basis for discriminating potyvirus species and strains and for assigning these viruses to four genera (at that time unnamed) was critically examined and compared with other approaches to potyvirus taxonomy. At the end of that meeting, the participants voted to establish the plant virus family, *Potyviridae* and the three genera equivalent to aphid- (Potyvirus genus), mite- (Ryemovirus genus) and fungus- (Baymovirus) transmitted viruses, with a fourth possible genus, Ipomovirus corresponding to the whitefly-transmitted potyvirus pending confirmation by appropriate sequence data. This proposal has been summarized by Barnett (1991), was accepted by the Plant Virus Subcommittee of ICTV with minor changes to two of the four suggested genus names, and was approved by the ICTV at their executive committee meeting in Glasgow 1993.

Other plant virus groups to be assigned to families were: the *Cryptoviridae* (two genera); the *Geminiviridae* (three genera); the *Tombusviridae* (tombusviruses and carmoviruses); the *Comoviridae* (comoviruses, nepoviruses and fabaviruses); the *Bromoviridae* (bromoviruses, cucumoviruses, ilarviruses and alfamoviruses); and the *Sequiviridae* (sequiviruses, e.g. parsnip yellow fleck virus, and waikaviruses, e.g. rice tungro spherical virus; Mayo *et al.*, 1993). These, along with the *Rhabdoviridae, Bunyaviridae, Reoviridae* and *Potyviridae,* make a total of 17 of the 35 plant virus groups that have been classified into 10 families (Martelli, 1992). The remaining 18 groups were left as unassigned genera (Martelli, 1992). These classifications are summarized in table I.

Assignment of other plant virus groups to genera and families

A tentative assignment of most of the remaining plant virus groups into families and genera is also shown in table I. These assignments are based primarily on three criteria: the nature of the genome, including the number of gene products and their organization; the level of sequence identity; and to a lesser extent, particle morphology. Excellent summaries of these viruses can be found in the fifth report of ICTV (Francki *et al.*, 1991). These families are arranged in table I under subheadings to conform with the way they are discussed in that volume (Francki *et al.*, 1991). This arrangement does not represent a higher taxonomy. Three of these families, *Reoviridae, Rhabdoviridae* and *Bunyaviridae* have long been recognized by

Table I. Assignment of plant virus groups into families and genera (a).

Family	Genus
dsDNA viruses	
Badnaviridae	Badnavirus (formerly the commelina yellow mottle virus group)
Caulimoviridae	Caulimovirus
ssDNA viruses	
Geminiviridae (b)	Two genera (subgroups I and II) Third genus (subgroup III) may belong to separate family or subfamily
dsRNA viruses	
Cryptoviridae (b)	Two genera (white clover subgroups 1 and 2)
Reoviridae (b)	Three genera (Phytoreovirus, Fijivirus and Oryzavirus (b))
Negative strand RNA enveloped viruses	
Bunyaviridae (b)	Tospovirus
Rhabdoviridae (b)	Two genera (plant rhabdovirus subgroup A and B)
Negative strand RNA filamentous viruses	
Tenuiviridae	Tenuivirus
Positive strand RNA viruses with isometric particles	
Bromoviridae (b)	Four genera (Bromovirus, Cucomovirus, Ilarvirus and Alfamovirus)
Comoviridae (b)	Three genera (Comovirus, Nepovirus, Fabavirus)
Dianthoviridae	Dianthovirus
Ortholuteoviridae (c)	Ortholuteovirus (BWY-subgroup)
Penamoviridae (d)	Penamovirus (d)
Paraluteoviridae (e)	Paraluteovirus (BYD-subgroup)
Sequiviridae	Two genera (Sequivirus, Waikavirus: formerly the parsnip yellow fleck virus and maize chlorotic dwarf virus groups)
Sobemoviridae	Sobemovirus
Tombusviridae (b)	Two genera (Tombusvirus, Carmovirus)
Tymoviridae	Tymovirus
Positive strand RNA viruses with rod-shaped particles	
Tobamoviridae	Tobamovirus
Tobraviridae	Tobravirus
Hordeiviridae	Hordeivirus
Furoviridae	Furovirus
Positive strand RNA viruses with filamentous particles	
Potexviridae	Three genera (Potexvirus, Carlavirus, Fibravirus (f)) and a possible genus (Capillovirus)
Potyviridae (b)	Three genera (Potyvirus, Rymovirus, Bymovirus) and a possible genus (Ipomovirus)
Closteroviridae	Closterovirus (SBY type (g)) and possibly a second genus (subgroup C)
Unassigned viruses	
Isometric viruses	Marafivirus, Necrovirus

(a) The listings of suggested families under each subheading do not represent higher taxonomic relationships. These are shown in tables II-IV.

(b) Names from Martelli (1992).

(c) Beet western yellows subgroup of luteoviruses.

(d) Suggested name for the pea enation mosaic virus group.

(e) Barley yellow dwarf subgroup of luteoviruses.

(f) Suggested name for the ACLSV type closteroviruses.

(g) Sugar beet yellows subgroup of closteroviruses.

ICTV (Francki *et al.*, 1991) and require no further comment.

The dsDNA viruses

The caulimoviruses and badnaviruses (formerly the commelina yellow mottle virus group) may belong to distinct families or represent different genera of a common family that also includes the animal *Hepadnaviridae* as suggested by Martelli (1992). They both have genomes of similar size (7.5-8.1 kb), and similar form (open circle dsDNA with single-strand discontinuities), and have been shown (caulimovirus) or are assumed (badnavirus) to replicate via reverse transcription. However, as summarized by Howard (1991), viruses of the animal *Hepadnaviridae* family have a significantly smaller genome (3.0-3.3 kb) than the plant viruses, and have already been assigned to two genera, Orthohepadnavirus (human, woodchuck and ground squirrel viruses) and Avihepadnavirus (duck and heron viruses), based on their relative sequence identities and other properties. For these reasons, it seems more appropriate to assign the two plant virus groups to one or two separate families depending on the relationship revealed when gene sequence data becomes available for the badnaviruses. Their very different particle morphology suggests they represent two distinct families.

The ssDNA viruses

The three subgroups of geminiviruses have been suggested to form a single family (Martelli, 1992; Matthews, 1985a) and appear to be distinct from the other ssDNA viruses that infect invertebrates and vertebrates *(Parvoviridae)* and bacteria (Francki *et al.*, 1991). The three subgroups have been assigned as three genera and reveal an interesting mixture of genome arrangements. Subgroups I and II have a monopartite genome (2.7-3.0 kb) and are transmitted by leafhoppers, whereas most of the subgroup III members have a bipartite genome of two molecules of ssDNA (each 2.4-2.8 kb) and are transmitted by whiteflies. The sequence data (Howarth and Goodman, 1986; Stanley *et al.*, 1986) reveals that the subgroup II genome organization resembles that of DNA 1 of the bipartite subgroup III viruses rather than the subgroup I viruses, with the exception of the coat protein, which is more closely related to subgroup I (25 %, particularly the C-terminal-end region) than subgroup III (15 %). The presence of a bipartite genome, with an extra gene segment which almost doubles the genome coding capacity, as well as the use of a different vector raises the possibility that the subgroup III geminiviruses could alternatively be considered a subfamily of the *Geminiviridae* or a separate family rather than just a separate genus.

The dsRNA cryptoviruses

The plant cryptovirus group contains two subgroups which appear to correspond with genera. Both have bipartite genomes of dsRNA with the subgroup I genome (estimated at 1.7 kbp and 1.4 kbp) and isometric particles (30 nm dia) being smaller than those of subgroup II (est 2.2 kbp and 1.9 kbp; 38 nm dia particles). Sequence data is required to establish whether these two genera of plant cryptoviruses constitute a separate family, the *Cryptoviridae,* as suggested by Martelli (1992), or are members of the fungal virus family *Partitiviridae,* which have similar particle dimensions and morphology and similar bipartite genomes (Buck and Ghabrial, 1991).

The filamentous negative strand RNA tenuiviruses

The tenuiviruses contain four or five gene segments, of which at least three contain an ambisense coding strategy (Kakutani *et al.*, 1990; Zhu *et al.*, 1991; Takahashi *et al.*, 1993; Hamamatsu *et al.*, 1993). They have some sequence identity with the *Arenaviridae* and *Bunyaviridae,* but are sufficiently different to warrant their classification as a distinct family *Tenuiviridae.* They contain a single capsid protein of 32 kDa and the virus particles are filamentous and appear folded, branched and supercoiled in the electron microscope (Zhu *et al.*, 1991).

The positive strand RNA viruses with isometric particles

Currently there are 18 groups of positive sense, ssRNA plant viruses with isometric particle morphology that are recognized by ICTV, and these are listed in table I. As will be discussed in this section, 16 of these can be assigned as genera of 10 families. At the present time, there is insufficient sequence data to allow the family/genus status of the other two (marafivirus group, necrovirus group), to be assessed.

1) *The* Bromoviridae

The bromovirus, cucumovirus, ilarvirus and alfamovirus groups constitute a single family as first suggested by Van Vloten-Doting *et al.* (1981). They referred to this family as the *Tricornaviridae*, while ICTV has now named them the *Bromoviridae* (Martelli, 1992). These viruses all have three genomic molecules plus an mRNA that codes for the coat protein, packaged in three or four (alfalfa mosaic virus) particles. There is a gradation in morphology from the isometric particles of bromovirus and cucumovirus, through the quasi-isometric and occasionally bacilliform particles of ilarvirus to the bacilliform particles of alfamovirus (see Francki *et al.*, 1991). Some workers had questioned the inclusion of the alfamoviruses in this family (Van Regenmortel, 1989) presumably because of; (i) the bacilliform particle morphology: (ii) the packaging of the coat protein mRNA in a fourth particle rather than in the third (along with RNA segment 3); (iii) the limited sequence identity in protein 3a, the putative movement protein coded for in the 5' region of RNA 3 (Davies and Symons, 1988); and (iv) the absence of sequence identity in the capsid protein coding region at the 3' end of RNA 3 (Dasgupta and Kaesberg, 1982; Davies and Symons, 1988). It is worth noting, however, that the sequence identities of the bromovirus and cucumovirus coat proteins are also low (Davies and Symons, 1988), and that the ilarvirus particles are quasi-isometric and occasionally bacilliform. The sequence data for RNA 1 and 2 of alfalfa mosaic virus and brome mosaic virus (Haseloff *et al.*, 1984) shows significant identity, suggesting that they are genera of a single family. The analyses of RNA-dependent RNA polymerases (see fig. 1) and helicases confirm this close relationship (Gorbalenya and Koonin, 1989; Koonin, 1991).

2) *The* Comoviridae

The comovirus, nepovirus and fabavirus groups have been assigned to a second family, the *Comoviridae* (Martelli, 1992). All have bipartite genomes of comparable size (RNA 1, 5.9-7.3 kb and RNA 2, 3.5-4.5 kb) with a 5' VPg and a 3' poly (A) tail, similar genome arrangements (Goldbach and Wellink, 1988), significant sequence identities in the non-structural proteins (Gorbalenya and Koonin, 1989; Koonin, 1991) as shown in figure 1, and three types of isometric particles.

3) *The luteoviruses*

The luteovirus group appears to consist of two distinct families, one of which appears to be related to the tombusviruses, carmoviruses and dianthroviruses and the other related to the sobemoviruses (Koonin, 1991; Martin *et al.*, 1990). The two families of luteoviruses have similar sized genomes (5.5-6.0 kb) with a genome-linked VPg at the 5' ends and probably no poly(A) tail or t-RNA-like structure at their 3' ends (Martin *et al.*, 1990). In general, the 3' halves of their genomes are similarly organized and code for: the coat protein (22-23 kDa); the VPg (17 kDa; in a different reading frame within the coat protein coding region); and a 50-56 kDa protein that is in the same reading frame as the coat protein (Martin *et al.*, 1990). However, there are major differences at the 5' half of the genomes of the two luteovirus families. This region of the genome

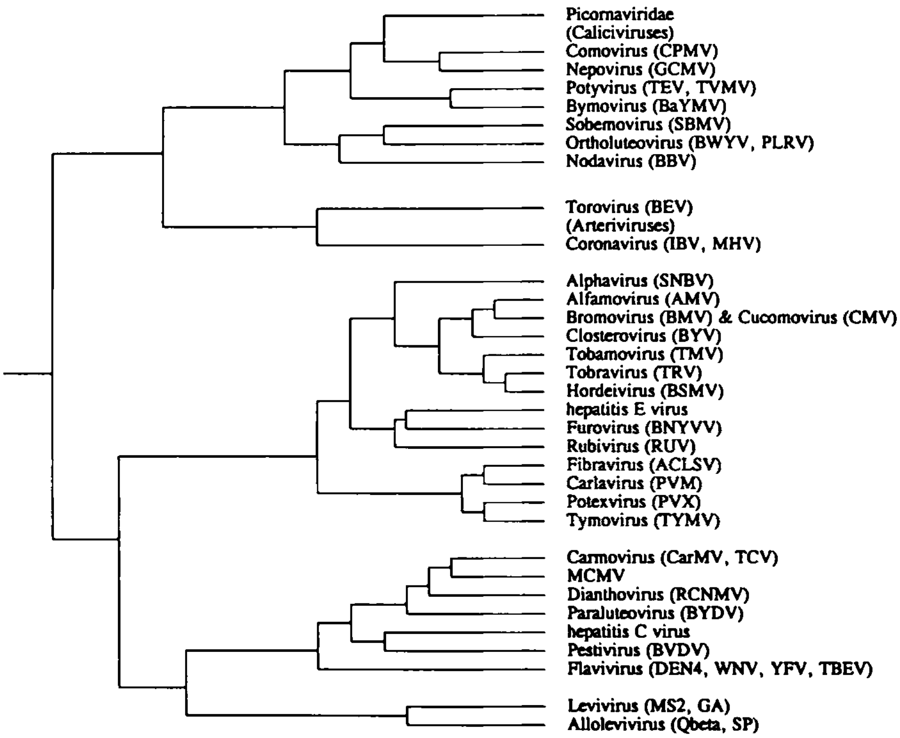


Fig. 1. Phylogenetic scheme for the RNA-dependent RNA polymerases.

The virus acronyms in alphabetical order are as follows: ACLSV, apple chlorotic leafspot virus; AMV, alfalfa mosaic virus; BaYMV, barley yellow mosaic virus; BBV, black beetle virus; BEV, Berne virus; BMV, brome mosaic virus; BNYVV, beet necrotic yellow vein virus; BSMV, barley stripe mosaic virus; BVDV, bovine viral diarrhoea virus; BWYV, beet western yellows virus; BYDV, barley yellow dwarf virus; BYV, beet yellow virus; CarMV, carnation mottle virus; CMV, cucumber mosaic virus; CPMV, cowpea mosaic virus; DEN4, dengue 4 virus; GA, bacteriophage GA; GCMV, Hungarian grapevine chrome mosaic virus; IBV, infectious bronchitis virus; MCMV, maize chlorotic mottle virus; MHV, mouse hepatitis virus; MS2, bacteriophage MS2; PLRV, potato leaf roll virus; PVM, potato virus M; PVX, potato virus X; Qbeta, bacteriophage Qbeta; RCNMV, red clover necrotic mottle virus; RUV, rubella virus; SBMV, southern bean mosaic virus; SNBV, sindbis virus; SP, bacteriophage SP; TBEV, tick-borne encephalitis virus; TCV, turnip crinkle virus; TEV, tobacco etch virus; TMV, tobacco mosaic virus; TRV, tobacco rattle virus; TVMV, tobacco vein mottling virus; TYMV, turnip yellow mosaic virus; WNV, West Nile virus; YFV, yellow fever virus. Redrawn from Dolja and Carrington (1992) and Koonin (1991).

codes for the RNA-dependent RNA polymerase in two overlapping open reading frames (ORF) that are translated as a fusion protein by a "-1" translational frameshift (Martin *et al.*, 1990) which is similar to the amber mutation read-through adopted by tombusviruses and carmoviruses. One family, represented by barley yellow dwarf virus, has a short (13 nt) overlap between ORF 1 and 2 and a polymerase with ~ 30 % sequence identity to that of the tombusviruses (Koonin, 1991), carmoviruses and dianthoviruses (Martin *et al.*, 1990) as shown in figure 1. The other luteovirus family, represented by beet western yellows virus and potato leaf roll virus, has an additional ORF (ORF 0) at the extreme 5' end of the genome, a larger (298 or 474 nt respectively) overlap between ORF 1 and 2 and a polymerase (fig. 1) that resembles that of the sobemovirus family (Martin *et al.*, 1990; Koonin, 1991).

The extent of these differences suggests that these two types of luteoviruses are members of distinct families, tentatively named *Ortholuteoviridae* (the beet western yellows subgroup) and *Paraluteoviridae* (the barley yellow dwarf subgroup) to retain the luteovirus connection in the same way that the myxoviruses were subsequently reclassified and renamed. The alternate possibility of assigning the two luteovirus subgroups as members of two subfamilies is less favoured given the major differences in the sequence identities of their RNA polymerases, the most important marker of higher taxonomic relationships. The third possibility of classifying them as genera of the *Sobemoviridae* and *Tombusviridae,* respectively, is not favoured given the differences in the 3' halves of their respective genomes. In addition, the tombusviruses, carmoviruses and dianthoviruses lack the VPg present in the paraluteoviruses (see Francki *et al.,* 1991). The 3' half of the sobemovirus genome is also quite different from that of the ortholuteoviruses, particularly in the location of VPg which is downstream in a second reading frame of the coat protein coding region in luteoviruses (Martin *et al.,* 1990) but upstream of the polymerase in the central region of ORF 2, the major translation product of sobemoviruses (Wu *et al.,* 1987).

The existence of two families of luteoviruses may appear confusing but is similar to the multiple families of viruses responsible for hepatitis in man (see Francki *et al.,* 1991). It is an interesting coincidence that both of these difficult sets of pathogens cause their infected hosts to go yellow.

4) *The pea enation mosaic virus subgroup*

The recently determined sequence of RNA 1 of the type member of the pea enation mosaic virus group shows that this virus resembles the beet western yellows family of ortholuteoviruses (Demler and de Zoeten, 1991). Its genome has a 5' VPg and no poly(A) tail and the genome arrangement of RNA 1 is similar to that of the ortholuteoviruses. It consists of: (i) an ORF 1 of unknown function, (ii) an overlapping ORF 2 that contains a proteinase motif, (iii) a third ORF that has a helicase-like motif, several RNA polymerase motifs, overlaps ORF 2 and is proposed to be translated by a frameshift fusion of ORF 2 and ORF 3 products; (iv) the fourth ORF codes for the coat protein and is immediately followed in frame by a 33-kDa ORF (Demler and de Zoeten, 1991). The sequence identities between these ORF of pea enation mosaic virus and the beet western yellows subgroup of ortholuteoviruses range from 40 % for ORF 3, ~ 30 % for the coat protein and 33-kDa protein and 17 % for ORF 2, although the central non-overlapping region of this ORF is much more conserved (Demler and de Zoeten, 1991).

This raises the question as to whether the pea enation mosaic virus group should be considered a genus or subfamily of the family of ortholuteoviruses that has developed a cell-to-cell movement function through acquisition of a second genome segment, RNA 2, enabling this luteo-like virus to infect its host systematically (Demler and de Zoeten, 1991). At this stage, I suggest that the pea enation mosaic subgroup of plant viruses be tentatively classified as a distinct family, the *Penamoviridae* (table I) given their possession of a bipartite genome (5.7 and ~ 3.7 kb), where the second gene segment contains substantial additional coding capacity. These alternative options need to be addressed by the Plant Virus Subcommittee of ICTV.

5) *The* Tombusviridae

The tombusvirus and carmovirus groups have been accepted to constitute a third family of RNA viruses, the *Tombusviridae* (Martelli, 1992), based on their similar genome size and organization (Goldbach and Wellink, 1988; Habili and Symons, 1989). The genome (4003 nt) for the carmovirus carnation mottle virus codes for three major proteins: a 27-kDa ORF 1 product: a 763-residue ORF 1/ORF 2, amber mutation readthrough product which is the RNA polymerase; and the p38 coat protein. The coat protein is coded by the + 1 reading frame, overlaps the polymerase by eight nucleotides and is synthesized from one or two encapsidated subgenomic RNA (Guilley *et al.,* 1985). The tombusvirus genome is larger in size (4776 nt) with five ORF, but is similarly organized. It also codes for its putative polymerase by readthrough of the ORF 1 (p33) amber terminator to give a p92, ORF 1/ORF 2 polymerase product (Hearne *et al.*, 1990). The p41 coat protein ORF 3 of tomato bushy stunt virus is downstream and starts 32 nucleotides after the polymerase terminator. Like the carmovirus coat protein, it is also translated from subgenomic RNA (Hearne *et al.*, 1990). The tombusvirus genome differs from that of the carmoviruses in having two extra ORF at the 3' end of the genome that are translated in overlapping reading frames to give p19 and p22 products which appeared to be unrelated to other ORF encoded by small RNA viruses (Hearne *et al.*, 1990).

The tombusvirus and carmovirus readthrough domains of their RNA polymerases are almost identical in size (518 and 522 amino acids) and exhibit sequence identities of 36-37 % (Hearne *et al.*, 1990), a value similar to that (31-34 %) found between the polymerase proteins of the Potyvirus and Bymovirus genera of *Potyviridae* (Kashiwazaki *et al.*, 1990). It is also similar to that (37 %) found between the two carmoviruses, carnation mottle virus and turnip crinkle virus (Hearne *et al.*, 1990), suggesting that the latter may also be a distinct genus. Some sequence identities could also be seen between the structural proteins of tombusviruses and carmoviruses, although these were highest (30 %) in the β-barrel S domain (Carrington *et al.*, 1987). It is interesting to note that the coat proteins of distinct species of tombusvirus showed modest sequence identity (37-44 %) when the total coat protein was compared, with considerable variation in their surface exposed P-domains. This is reminiscent of the situation with the N-terminal domains of the coat proteins of potyviruses (Shukla and Ward, 1989).

It is interesting to note that maize chlorotic mottle virus, which was originally classified as a possible member of the sobemovirus group (see Francki *et al.*, 1991), appears to be a carmovirus from its similar genome organization and sequence identity (Nutter *et al.*, 1989). Its RNA polymerase has 50 % sequence identity with those of carnation mottle virus and turnip crinkle virus and 38 % identity with three tombusviruses (Hearne *et al.*, 1990).

6) *The* Dianthoviridae

The dianthovirus group appears to represent a distinct family of plant viruses that is closely related to the carmoviruses and tombusviruses. Its RNA polymerase shows significant sequence identity (~ 35 %) with the corresponding proteins of the tombusviruses (Hearne *et al.*, 1990; Koonin, 1991), carmoviruses (Xiong and Lommel, 1989) and paraluteoviruses. Furthermore, significant sequence identity (27 %) is also seen between the coat proteins of dianthoviruses and carmoviruses (Xiong and Lommel, 1989).

The dianthoviruses, however, differ from the tombusviruses and carmoviruses in having a bipartite genome (3.9 and 1.4 kb) where the second gene segment increases the genome size by 40 %. The sequence of both gene segments from red clover necrotic mosaic dianthovirus has been completed. RNA 1 contains three ORF arranged in a pattern that is almost identical to that of the 5' region of the genome of the paraluteoviruses (the barley yellow dwarf subgroup of luteoviruses), including a putative translational frameshift between the first two ORF. These overlap by 7 nucleotides to give an 88-kDa fusion protein. RNA 2 codes for a 35 kDa protein with low sequence identity to the RNA 3a movement protein of bromoviruses (Lommel *et al.*, 1988). It has been suggested that this movement protein may be equivalent to the RNA 2 product of pea enation mosaic virus (Demler and de Zoeten, 1991).

As discussed earlier for the pea enation mosaic virus group and the subgroup III geminiviruses, the presence of a bipartite genome and substantial additional coding capacity suggests that the dianthoviruses may be best classified as a distinct family rather than as a subfamily or genus of the *Tombusviridae.*

7) *The* Sobemoviridae

The sobemovirus group appears to correspond to a distinct family with an RNA polymerase that is more closely related to that of the ortholuteoviruses than to those of other RNA viruses (fig. 1). The sobemovirus genome (4.2kb), like that of the luteoviruses, contains a 5' VPg but no 3' poly(A) tail or tRNA-like structure. It contains 4 ORF spread across 3 reading frames (Wu *et al.*, 1987). The first ORF codes for a 21-kDa protein and overlaps the large central ORF 2. ORF 2 codes for a 105-kDa product that has similar organization to the VPg-protease-polymerase arrangement of the picornavirus supergroup, although the sequence identity is weak (Wu *et al.*, 1987). In this respect, it differs significantly from the luteoviruses, which code for their VPg in a second reading frame of the downstream coat protein coding region (Martin *et al.*, 1990). Thus, the sobemovirus group appears to be a distinct plant virus family. As mentioned above, the sequence data for the maize chlorotic mottle virus genome suggests that it is a carmovirus, not a possible member of the sobemovirus group as currently classified (see Francki *et al.*, 1991).

8) *The* Tymoviridae

The tymovirus group appears to be a distinct family. The tymovirus genome (6.3 kb) has a $5'm^7G^5ppp^5Gp$ cap and a 3' tRNA-like structure (Klein *et al.*, 1976; Mans *et al.*, 1991) and generates small amounts of subgenomic coat protein RNA. The genome of TYMV (turnip yellow mosaic virus), the type member, contains three conserved ORF. The first initiates at nucleotide 89 and codes for a 67-kDa protein which may be involved in intercellular movement (Keese *et al.*, 1989). The second ORF is in the +1 reading frame at nucleotide 96 and codes for a 206-kDa replicase protein, parts of which show similarity to the nucleotide binding proteins and replicases of viruses in the Sindbis virus supergroup (Goldbach and Wellink, 1988; Koonin, 1991) with their polymerases most closely related to those of the filamentous potex- and carlaviruses (figure 1). The third ORF codes for the coat protein, is at the 3' end of the genome and occurs in either the first, second or third reading frames depending on the virus species (Ding *et al.*, 1989; Keese *et al.*, 1989).

The ssRNA viruses with rod-shaped particles

The four groups of rod-shaped RNA plant viruses fall into two sets with the tobamoviruses, tobraviruses and hordeiviruses appearing to be more closely related than the furoviruses. The tobamovirus genome is monopartite, approximately 5.7 kb in size, has a $5'm^7G^5ppp^5Gp$ cap and a tRNA-like structure at the 3' end (Francki *et al.*, 1991; Mans *et al.*, 1991). The tobravirus genome is bipartite (with the smaller RNA segment coding for the coat protein), approximately 7.3 to 8.6 kb in size, has a $5'm^7G^5ppp^5Ap$ cap and a tRNA-like structure at the 3' end (Francki *et al.*, 1991; Mans *et al.*, 1991). The hordeivirus genome is tripartite and larger (10.3 kb) than those of the tobamoviruses and tobraviruses, but has similar 5' cap and 3' tRNA-like structures (Francki *et al.*, 1991; Mans *et al.*, 1991). The furovirus genome is bipartite and of similar size to that of the hordeiviruses, but does not have the 3' t-RNA-like structure or a poly(A) tail (Francki *et al.*, 1991; Mans *et al.*, 1991). The genome arrangement of tobamoviruses, tobraviruses, hordeiviruses and furoviruses is similar despite the differences in the numbers of gene segments and the size of the genomes (see fig. 2).

The RNA polymerases of tobamoviruses, tobraviruses and hordeiviruses show higher sequence identity than the polymerases of other accepted genera (fig. 1), suggesting that these viruses could be considered distinct genera of a single virus family. However, the coat proteins of tobamoviruses, tobraviruses and hordeiviruses exhibit relatively low sequence identity (Dolja *et al.*, 1991), as shown in figure 3, suggesting that their classification as either three distinct subfamilies or families is more appropriate.

The furovirus group appears to constitute a distinct family of rod-shaped plant viruses, the *Furoviridae*, since the polymerase of beet necrotic yellow vein virus (BNYVV), the furovirus representative analysed by Koonin (1991), is not closely related to those of the hordei-, tobamo- and tobraviruses (fig. 1).

The positive strand RNA viruses with filamentous particles

1) *The* Potyviridae

The taxonomy of the *Potyviridae* with its three genera, Potyvirus, Rymovirus, Bymovirus and possible fourth genus, Ipomovirus, has been discussed in detail previously (Ward and Shukla, 1991; Ward *et al.*, 1992; Barnett, 1992). The *Potyviridae* are quite distinct from the other filamentous plant viruses and have gene replication elements that more closely resemble some isometric viruses from animals, plants and insects (fig. 4).

2) *The* Potexviridae: *potexviruses, carlaviruses, the ACLSV closterovirus subgroup and possibly capilloviruses*

Genomic sequence information is clarifying the relationships between these filamentous viruses. The overall size and coding arrangements for potexviruses and carlaviruses are similar, but not identical, and the sequence data (Forster *et al.*, 1988; Huisman *et al.*, 1988; Rupasov *et al.*, 1989; Memelink *et al.*, 1990; Zavriev *et al.*, 1991) indicates that they are distinct genera of the one family. Their RNA-dependent RNA polymerases fall into the same phylogenetic subset (fig. 1) as do their helicases (Gorbalenya and Koonin, 1989) and their coat proteins (fig. 3). As shown in figure 5, the sequence identity between the coat proteins of the potexvirus PVX (potato virus X) and the carlavirus PVM (potato virus M) (29 %) is comparable to that between the two potexviruses PVX and WClMV (white clover mosaic virus) (35 %). The sizes of all three coat proteins vary markedly, from 188 to 304 residues, and the sequence identities are similar to those expected of distinct genera, suggesting that WClMV may even be a distinct genus of the potexvirus family.

Genome sequence data is required to establish the family/genus status of the capilloviruses. The data that indicates that the ACLSV subgroup of closteroviruses corresponds to a genus of the *Potexviridae* is discussed below.

3) *The closterovirus group — at least two distinct families*

The second filamentous group of plant viruses, the closteroviruses, can be divided into three subgroups based on the modal length of their particles and some biological properties (Agranovsky *et al.*, 1991). Subgroup A, represented by apple chlorotic leaf spot virus (ACLSV), has a particle length of 730 nm; subgroup B, represented by sugar beet yellows virus (SBYV), has a particle length of 1250 to 1450 nm; while subgroup C, represented by citrus tristeza virus has a particle length of 1650 to 2000 nm (Bar-Joseph and Murant, 1982). In the introduction to their sequence paper, Agranovsky *et al.* (1991) state that the current data on structure and expression of closterovirus genomes indicates that they should be reclassified into three separate virus groups, but do not indicate whether those groups are equivalent to distinct genera or distinct families. The genome of the subgroup A virus ACLSV is 7555 nucleotides long and very different in size, arrangement and sequence identity (German *et al.*, 1990) to the larger genome (approximately 14,000 nucleotides) of the closterovirus subgroup B type member SBYV (Agranovsky *et al.*, 1991). Their RNA-dependent RNA polymerases fall into different phylogeny groups (fig. 1) as do their coat proteins (fig. 3). Furthermore, the 3' end of the ACLSV genome is polyadenylated, while that of SBYV is not.

These data suggest that these two viruses belong to different families. SBYV, the type member is representative of a true *Closteroviridae* family, while the genome size, arrangement and sequence of ACLSV (except for its coat protein) suggest that it belongs to the potexvirus family as a separate genus or subfamily. Its genome organization is as similar to the potexviruses and carlaviruses as the bymovirus genome is to the potyviruses and rymoviruses. ACLSV also resembles the potexviruses and carlaviruses in particle size (730 nm by 12 nm). The coat protein dendrogram shown in figure 3, however, shows that the coat protein of ACLSV is very different from those of the other suggested genera (potexviruses, carlaviruses) of the *Potex-*

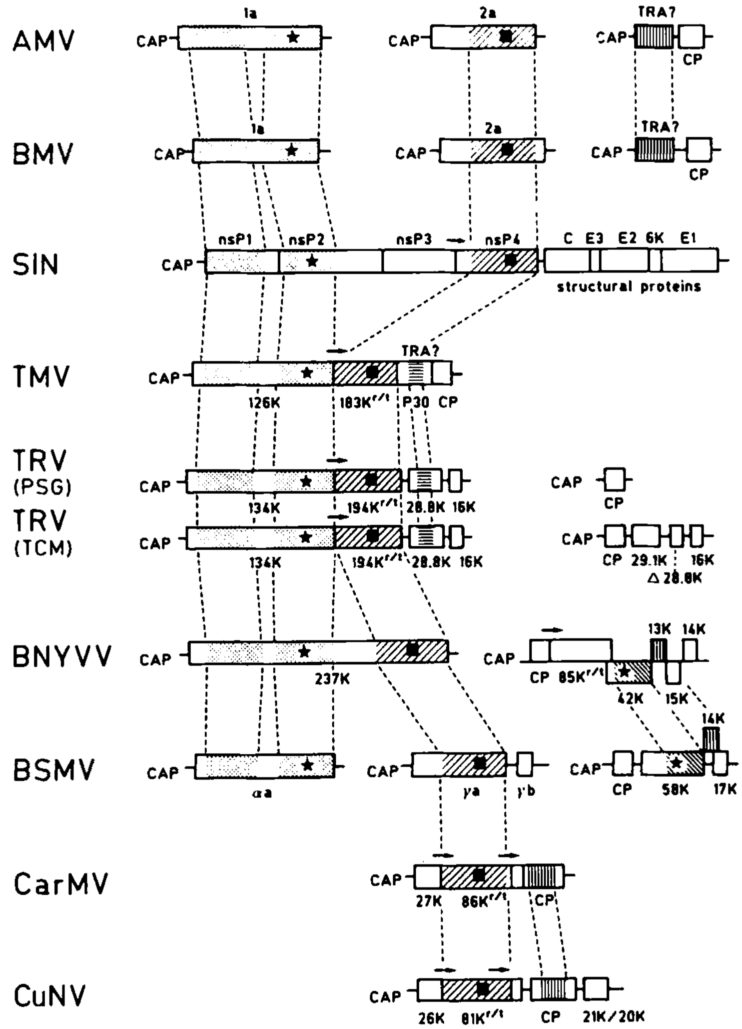
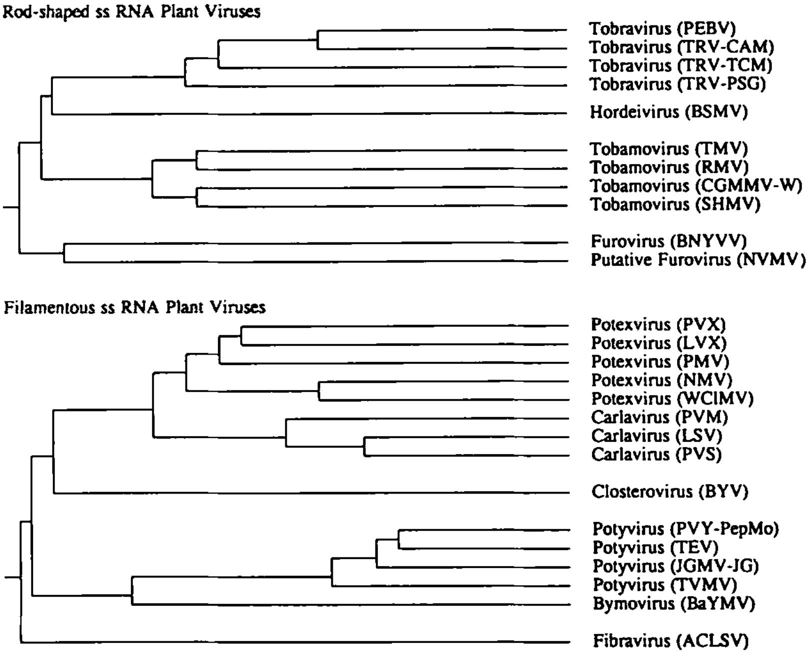


Fig. 2. Schematic representation of the genomes of the togavirus, Sindbis virus and the Sindbis-like plant viruses.

From Goldbach and Wellink (1988) with permission. The virus acronyms in alphabetical order are: AMV, alfalfa mosaic virus (alfamovirus); BMV, brome mosaic virus (bromovirus); BNYVV, beet necrotic yellow vein virus (furovirus); BSMV, barley stripe mosaic virus (hordeivirus); CarMV, carnation mottle virus (carmovirus); CuNV, cucumber necrosis virus (tombusvirus); SIN, sindbis virus (alphavirus); TMV, tobacco mosaic virus (tobamovirus); TRV-PSG and TRV-TCM, the PSG and TCM strains of tobacco rattle virus (tobravirus). Subsequent analyses have added rubiviruses, potexviruses, carlaviruses and closteroviruses to this supergroup and placed the carmoviruses and tombusviruses in a different supergroup with the dianthoviruses, luteoviruses, flaviviruses, pestiviruses and hepatitis C virus (Habili and Symons, 1989; Dolja and Carrington, 1992).

viridae. However, this does not justify placing ACLSV into a distinct family, as it is similar to the situation seen with the negative stranded RNA orthomyxoviruses, Thogoto virus and Dhori virus. There, the two gene segments coding for the surface glycoproteins, haemagglutinin and neuraminidase, have been replaced by an appropriately modified gene segment that codes for a surface fusion glycoprotein related to gp64 of the DNA-containing baculovirus


Rod-shaped ss RNA Plant Viruses

Filamentous ss RNA Plant Viruses

Fig. 3. Cluster dendrograms showing the relationships between the sequences of virus capsid proteins of rod-shaped viruses (panel A) and filamentous viruses (panel B).

The virus acronyms in alphabetical order are: ACLSV, apple chlorotic leafspot virus; BaYMV, barley yellow mosaic virus; BNYVV, beet necrotic yellow vein virus; BSMV, barley stripe mosaic virus; BYV, beet yellows virus; CGMMV-W, cucumber green mottle mosaic virus W strain; JGMV-JG, Johnson grass mosaic virus JG strain; LSV, lilia symptomless virus; LVX, lilia virus X; NMV, narcissus mosaic virus; NVMV, *Nicotinia velutina* mosaic virus; PEBV, pea early browning virus; PMV, papaya mosaic virus; PVM, potato virus M; PVS, potato virus S; PVX, potato virus X; PVY-PepMo, the pepper mottle strain of PVY; RMV, ribgrass mosaic virus; TEV, tobacco etch virus; TMV, tobacco mosaic virus; TVMV, tobacco vein mottling virus; SHMV, sun-hemp mosaic virus; TRV-CAM, TRV-PSG and TRV-TCM, the CAM, PSG and TCM strains of tobacco rattle virus; WClMV, white clover mosaic virus. Redrawn from (Dolja *et al.,* 1991).

(Morse *et al.,* 1992), producing a virus with seven gene segments that retains the replication strategies typical of the family of orthomyxoviruses.

The reclassification of subgroup A closteroviruses (ACLSV subgroup) requires a new name. Since the name closterovirus was derived from the greek *kloster,* "spindle or thread" and the name capillovirus from the Latin *capillus,* "a hair", the name fibravirus from the Latin *fibra* meaning "fibre" seems appropriate and consistent. However, before naming this new genus of the *Potexviridae,* its relationship to the capilloviruses needs to be established.

A decision regarding the family/genus status of the subgroup C closteroviruses must await the availability of genome sequence data which is in progress (Boyko *et al.,* 1992).

Higher taxa

There has been considerable reluctance in the past to consider taxonomic categories above the level of family. The first serious attempt was made by Lwoff (1967), who used two main features of the virion, the type of nucleic acid and subsequently the architecture of the virus capsid (symmetry, then presence or absence of an envelope, then dimensions) to arrange viruses

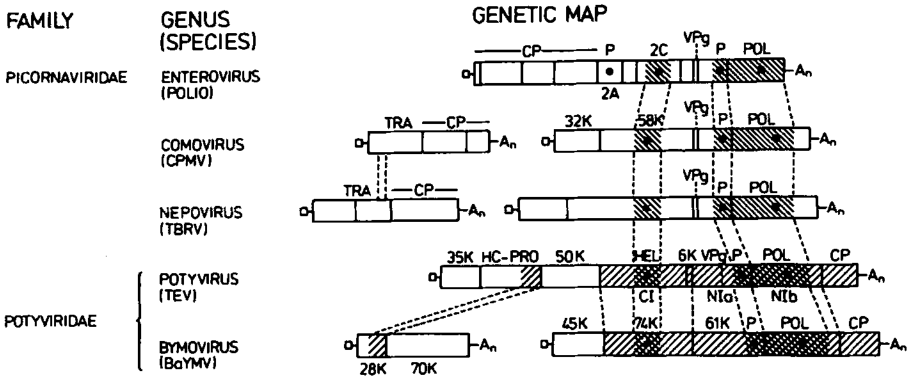


Fig. 4. Schematic representation of the genomes of the aphid-transmitted potyviruses, the fungal-transmitted bymoviruses and other members of the picorna-like supergroup of viruses.

The virus acronyms are: CPMV, cowpea mosaic virus; TBRV, tomato black ring virus; TEV, tobacco etch virus; BaYMV, barley yellow mosaic virus. Coding regions in the genomes are indicated as open bars; regions of amino acid sequence identity in the gene products are indicated by similar shading. The symbols are: CP, coat protein; TRA, transport protein; HEL, helicase; P, proteinase; POL, polymerase; A_n, poly(A) tail; VPg is denoted by the small open square at the N-terminal end of each polyprotein. From Goldbach (1992) with permission.

```
WClMV MA---------------TTTAT---TPPSL----------TDIRALK-----------------------------------------------------  19
      *                ***  *   **              *
PVX   MSAPASTTQ-ATGSTTSTTTKTAGATPATA--SGLFT--IPDGDF-------------------------------------------------------  40
      *         *    **     *   *  *   *  *    **
PVM   MGDSTKKAETAKDEGTSQERREARPLPTAADFEGKDTSENTDGRAADADGEMSLERRLDSLREFLRERRGAIRVTNPGLETGRPRLQLAENMRPDPTNPY 100

WClMV -------YTSSTVSVAS--PAEIEAITK---TWAETFKIPNDVLPLACWDL-RAFADVGASSKSELTGDSAALAG-VS-RKQLAQAIKIHCTIRQFCNYF 104
              ***   *  *   *    *    * * *    * *** *  **** *   *         *  * *  ** ***  ** *****
PVX   -------FSTARAIVASNAVATNEDLSKIEAIW-KDMKVPTDTMAQAAWDLVRHCADVGSSAQTEMIDTGPYSNG-IS-RARLAAAIKEVCTLRQFCNKY 130
              *    *  **  *  **   *    .    ***    *     *  * *  **             * *   * **   *   ***  *  *
PVM   NRPSIEALSRIKPIAISNNMATSEDMMRIYVNL-EGLGVPTEHVQQVVIQAVLFCKDASSSVFLDPRGSFEWPRGAITADAVLAVLKKDAETLRRVCRLY 199

WClMV ANIVWNIMLDTKTPPASWSKLGYKEESKFAGFDFFDGVNHPAALMPADGLIRGPSDAEILAHQTAKQVALHRDAKPTWHKRCQLC*               188
      *  *** **    *** *   * * * *** **** **  *** **  **** ** **  * ***  *
PVX   APVVWNWMLTNNSPPANWQAQGFKPEHKFAAFDFFNGVTNPAAIMPKEGLIRPPSEAEMNAAQTAAFVKITKA-RAQSNDFASLDAAVTRGRITGTTTAE 230
      *** ** *** * *** * * **  *  ***** *  * * **  * ***** *   *  *  *        * * *   * ** * ** *      *
PVM   APVTWNHMLTHNAPPADWAAMGFQYEDRFAAFDCFDYVENTAAVQPLEGLIRRPTPREKVAHNTHKDIALRGANRNQ--VFSSLNAEVTGGMNGPELTRD 298

PVX   AVVTLPPP    237
       *
PVM   YVKSNRK     304
```

Fig. 5. Comparison of the amino acid sequences of the coat proteins of the potexviruses PVX and WClMV and the carlavirus PVM.

into descending hierharchical divisions. The main objections to this scheme were: (i) the likelihood that the qualification that it did not reflect phylogenetic relationships would be forgotten; (ii) the arbitrary selection and weighting of the available criteria; and (iii) the need to collect a lot more data about individual viruses and their relationships to other virus groups (see Lwoff, 1967). The latter is now well advanced, As Matthews (1985a) and Strauss *et al.* (1990) point out, the amount of new information appearing on virus replication strategies, genome sequences and 3D structures is staggering, revealing interrelationships between virus families that previously were thought to be unrelated. As long ago as 1985, Matthews (1985a) pointed out that the original view, that viral taxonomy has no evolutionary implications, was no longer tenable.

Despite these developments, there is still much resistance to the establishment of higher taxa. The first major objection (Rybicki, 1990) is that viruses are probably polyphyletic in origin (to be discussed below), and arise by different routes at different times requiring the construction of separate evolutionary trees, not a single tree. The solution to this objection is to place all viruses in a separate kingdom and assign the current viruses to several phyla that reflect these diverse origins.

The second major objection is the recombinative character of current virus families where members contain gene cassettes that have come from diverse origins, thus compromising the establishment of a hierarchical evolutionary history, since different genes give different trees (Rybicki, 1990; Goldbach, 1992). The solution to this problem is also at hand. The observed amino acid sequence similarities between the non-structural proteins of diverse groups of viruses infecting plants and animals (Ahlquist *et al.*, 1985; Argos *et al.*, 1984; Franssen *et al.*, 1984; Haseloff *et al.*, 1984) led to the observation that many plus-stranded RNA plant viruses could be classified into two major supergroups (Goldbach, 1987; Goldbach and Wellink, 1988; Zimmern, 1987). The viruses within these supergroups may have a common evolutionary origin and there are some sequence similarities between the two major supergroups. This concept appears to have gained widespread acceptance. It has been further refined by the inclusion of additional viruses (see Dolja and Carrington, 1992) and the establishment of a further supergroup (Habili and Symons, 1989). It has been referred to, and the figures reproduced, in most major reviews and texts on the subject since. I propose that these supergroups provide the elusive connection required for a higher taxonomy of viruses and are equivalent to classes.

In the remaining sections of this review, the theories on the multiphyletic origin of viruses will be summarized briefly, followed by a discussion of how this and the wealth of molecular information can be used to to make a tentative start to assigning the 73 currently recognized virus families and groups to higher taxa.

Theories of origin

Currently described viruses are very diverse, and there is no compelling reason to suppose that all viruses arose in the same way (Matthews, 1991). Not only does the nature of their genome vary (DNA or RNA), but the size of their genomes also varies enormously, ranging over almost three orders of magnitude, with the largest (poxviruses) having genomes comparable to those of the simplest cells (Matthews, 1991).

There are three different theories regarding the origin of viruses and these are reviewed in detail in Strauss *et al.* (1990) and Matthews (1991). One theory is that viruses evolved from autonomous, self-replicating host cell molecules such as plasmids or transposons, by acquiring appropriate genes that code for packaging proteins. In prokaryotes, there are strong parallels between some members of the bacteriophage families and bacterial plasmids. The phages contain large dsDNA genomes that, like some plasmids, can exist either in an integrated state in the host chromosome or as an autonomously replicating form. Their evolution as viruses is associated with their capacity to package their genome in virus particles which protect the genetic material while outside the host cell and provide a mechanism for attachment to new host cells and DNA injection. The origin of the bacteriophage coat proteins is not known, but it is interesting to note that they have some similarities in their structure and complex assembly patterns (Georgopoulos *et al.*, 1983; Katsura, 1983) with bacterial self assembly proteins such as pilin and flagellin (Uhlin *et al.*, 1985; Parge *et al.*, 1987).

Similarly, in eukaryotes, there are strong parallels between retrotransposons and the viruses that use reverse transcriptase. For example, the genome of the Ty element in yeast is similar to that of the retroviruses. It is about 5900 nucleotides in length and contains two genes, TYA and TYB. TYA codes for a protein P1, which is analogous to the retroviral *gag* proteins and assembles into virus-like particles that resemble retrovirus cores; while TYB codes for the enzyme activities protease, integrase and reverse transcriptase and is analogous to the retroviral *pol* genes (Kingsman *et al.*, 1991). The major difference between the retrotransposons and the retroviruses is the absence of an *env* analogue in Ty, resulting in its inability to bud from infected cells or to bind to and infect new target hosts.

More recently the transmissable hypovirulence element of the chestnut blight fungus *Cryphonectria parasitica* has been sequenced and shown to be a dsRNA molecule with some similarities to a viral genome (Koonin *et al.*, 1991). Its polymerase appeared more closely related to those of the positive strand RNA viruses than the dsRNA viruses and with two out of three algorithms appeared distantly related to potyviruses (Koonin *et al.*, 1991). Its helicase was also related to those of the positive strand RNA viruses being equidistant from those of the potyviruses on the one hand (although in a different position in the genome) and the flaviviruses, pestiviruses and hepatitis C virus on the other (Koonin *et al.*, 1991). It also coded for a cysteine-like protease with weak identity to the HC-protease of potyviruses. The major difference between this hypovirulence element and the RNA viruses is its lack of a capsid protein. The distant relationship between this fungal hypovirulence element and potyviruses is intriguing given the belief that potyviruses first arose in fungi (Ward *et al.*, 1994).

Whether other RNA viruses arose in a similar manner or by *de novo* assembly from a polymerase core is not known. Many types of uninfected plant cells contain RNA-dependent RNA polymerases (Strauss *et al.*, 1990) and the replicase function of many RNA viruses is an ancient evolutionary core feature that predates the cassette assembly process that generated the curent RNA virus groups (Rybicki, 1990). With regard to the origin of the capsid proteins of the small icosahedral positive strand RNA viruses of plants, insects and animals, it is interesting to note that the plant storage protein phaseolin, which has a packaging function, also has a domain containing the same 8-stranded, antiparallel β-barrel motif (Lawrence *et al.*, 1990) found in the viral proteins (Harrison, 1990). Thus it is tempting to speculate that these isometric viruses arose first in plants and then subsequently evolved in insects before being transferred to mammals. Rybicki (1990) has pointed out that the picorna-like viruses of insects differ more than the picornaviruses of mammals, and suggested that the latter may have originated in insects.

A second theory is that some viruses arose by degeneration from primitive cells in a manner similar to that proposed for the evolution of cellular organelles such as mitochondria and chloroplasts from bacteria (Strauss *et al.*, 1990). This process would entail: (i) the loss of a bounding membrane that separates the replication of the primitive parasitic cell from that of the host cell, thus prohibiting binary fission which is characteristic of cell division and (ii) the use of host cell protein-synthesizing and metabolic machinery (Matthews, 1991). There are some problems with this type of explanation for the origin of all viruses (Strauss *et al.*, 1990) but it has been suggested to be a possible mechanism by which the poxviruses arose (Fenner, 1979). Their very large genome size, their complex structure, the presence of many enzymes within the virus particle and their ability to replicate in the cytoplasm independently of host nuclear functions suggest that these large, enveloped DNA viruses may have arisen from simple cellular parasites (Matthews, 1991).

A third theory is that some RNA viruses are descendants of prebiotic RNA polymers. RNA molecules can carry out nucleolytic cleavage, self-splicing reactions, ligations and even polymerization in a template-dependent fashion (Strauss *et al.*, 1990; Matthews, 1991). In addition, the tRNA structures found at the 3' end of some RNA viruses (tymo-, tobamo-, tobra-, bromo-, cucumo- and hordeiviruses (Mans *et al.*, 1991)) have been suggested to represent molecular fossils of the original RNA world which tagged genomic RNA for the initiation of replication and functioned as primitive telomeres to prevent loss of terminal nucleotides during the replication process (Weiner and Maizels, 1987). This theory suggests that some RNA viruses might have evolved from the prebiotic RNA world and parasitized the earliest cells.

A fourth possibility is that some viruses evolved from viroids or virusoids, although it is equally possible that these small RNA, rather than being progenitors of viruses, are recent degenerative products of the more complex self-replicating systems (Strauss *et al.,* 1990). Neither code for any proteins.

How many phyla of viruses?

Simple division into two phyla based on whether the genome is DNA or RNA may be too simplistic, as it ignores the current recognition of the multiphyletic origin of viruses. However, with regard to the RNA viruses, it has been postulated that their replicase function is an ancient common evolutionary feature (Rybicki, 1990) and that these viruses may have arisen only once, with all current RNA viruses derived from this single protovirus by linear divergence, recombination and gene duplication (Strauss *et al.,* 1990). Thus it seems sensible to suggest that there be a single phylum of RNA viruses (excluding the retroviruses) that contains the positive strand RNA viruses, the double-stranded RNA viruses and the negative strand RNA viruses.

The viruses using reverse transcription could form a second phylum rather than be grouped with the RNA viruses because of their possible origin from retrotransposons and the profound differences between these viruses and the RNA viruses (Strauss *et al.,* 1990). This phylum may also include the DNA-containing hepadnaviruses that infect vertebrates and the caulimovirus and badnavirus families of plant viruses which differ from the true retroviruses by encapsidating the DNA phase of their replication cycle rather than the earlier RNA phase. The higher taxonomy of these viruses is not discussed in this review, but analysis of their relationships should prove rewarding.

The remaining phyla would accommodate the DNA viruses and detailed analyses of these viruses is required to establish the number of phyla and their phylogenetic relationships. A start to such analyses has recently been made (Braithwaite and Ito, 1993) and will be discussed at the end of this review after the RNA viruses.

RNA virus supergroups constitute classes

In establishing phylogenetic networks among the RNA viruses, Rybicki (1990) inclined to the view that only the "core module" of polymerase and associated replication machinery should be considered. Everything else in the genomes of specific virus families may have been acquired from diverse sources to enable their survival in different biological environments. Thus it is the polymerases and associated proteins that constitute the essence of a particular genome strategy (Rybicki, 1990), and these can be used to construct phylogenetic dendrograms (Koonin, 1991) or to assign the RNA viruses to supergroups (Matthews, 1985a; Goldbach and Wellink, 1988; Strauss and Strauss, 1988; Habili and Symons, 1989; Strauss *et al.,* 1990; Dolja and Carrington, 1992) as depicted in figures 1, 2 and 4.

The positive strand RNA viruses

I have used the sequence relationships between the replicase proteins to assign the positive ssRNA virus families into five classes with four, three, three, three and one order(s), respectively, as shown in table II. The *Tetraviridae* family of isometric, positive strand, RNA insect viruses was not included in these analyses (Koonin, 1991; Dolja and Carrington, 1992) and has not been assigned.

1) *Class 1*

The first class corresponds to the picorna-like supergroup (fig. 4) which had been shown (Goldbach and Wellink, 1988) to contain the icosahedral animal *Picornaviridae* (five genera), the

Table II. Higher taxa of positive ssRNA and dsRNA viruses.

Phylum	Subphylum	Class	Order	Suborder	Family	Genus	Species
RNA viruses	positive ssRNA or dsRNA	class 1 (Picorna-like supergroup)	order 1		*Picornaviridae*	Enterovirus	human poliovirus 1
						Hepatovirus	human hepatitis A virus
						Cardiovirus	encephalomyocarditis virus
						Rhinovirus	human rhinovirus 1A
						Aphthovirus	aphthovirus O
					Sequiviridae	Sequivirus	parsnip yellow fleck virus
						Waikavirus	rice tungro spherical virus
					Caliciviridae	Calicivirus	vesicular exanthema of swine virus
					Comoviridae	Comovirus	cowpea mosaic virus
						Nepovirus	tobacco ringspot virus
						Fabavirus	broad bean wilt virus
			order 2		*Potyviridae*	Potyvirus	potato virus Y
						Rymovirus	ryegrass mosaic virus
						Bymovirus	barley yellow mosaic virus
						Ipomovirus	sweet potato mild mottle virus
			order 3		*Sobemoviridae*	Sobemovirus	southern bean mosaic virus
					Ortholuteoviridae	Ortholuteovirus	beet western yellows virus
					Penamoviridae	Penamovirus	pea enation mosaic virus
			order 4		*Nodaviridae*	Nodavirus	nodamura virus
			order 5		*Totiviridae*	Totivirus	*Saccaromyces cerevisiae* virus L1
						Giardiavirus	*Giardia lamblia* virus
		class 2	order 1		*Coronaviridae*	Coronavirus	avian infectious bronchitis virus
			order 2		*Toroviridae*	Torovirus	Berne virus
			order 3		*Arteriviridae*	Arterivirus	equine arteritis virus
		class 3 (Sindbis-like supergroup)	order 1	suborder 1	*Togaviridae*	Alphavirus	Sindbis virus
				suborder 2	*Tobamoviridae*	Tobamovirus	tobacco mosaic virus
					Tobraviridae	Tobravirus	tobacco rattle virus
					Hordeiviridae	Hordeivirus	barley stripe mosaic virus
				suborder 3	*Closteroviridae*	Closterovirus	sugar beet yellows virus
				suborder 4	*Bromoviridae*	Bromovirus	brome mosaic virus
						Cucomovirus	cucumber mosaic virus
						Ilarvirus	tobacco streak virus
						Alfamovirus	alfalfa mosaic virus
			order 2	suborder 1	*Rubiviridae*	Rubivirus	rubella virus
				suborder 2	*Hepeviridae*	Hepatitis E	hepatitis E virus
				suborder 3	*Furoviridae*	Furovirus	beet necrotic yellow vein virus
			order 3	suborder 1	*Potexviridae*	Potexvirus	potato virus X
						Carlavirus	carnation latent virus
						Capillovirus	apple stem grooving virus
						Fibravirus	apple chlorotic leaf spot virus
				suborder 2	*Tymoviridae*	Tymovirus	turnip yellow mosaic virus

Table II (continued). Higher taxa of positive ssRNA and dsRNA viruses.

Phylum	Subphylum	Class	Order	Suborder	Family	Genus	Species
		class 4 (carmo-like supergroup)	order 1		*Tombusviridae*	Tombusvirus	tomato bushy stunt virus
						Carmovirus	carnation mottle virus
					Dianthoviridae	Dianthovirus	carnation ringspot virus
					Paraluteoviridae	Paraluteovirus	barley yellow dwarf virus
			order 2		*Pestiviridae*	Pestivirus	bovine diarrhoea virus
					Hepciviridae	Hepatitis C	hepatitis C virus
			order 3		*Flaviviridae*	Flavivirus	yellow fever virus
		class 5 (ss phages)	order 1		*Leviviridae*	Levivirus	phage MS2
					Alloleviviridae	Allolevivirus	phage Qβ
		class 6	order 1		*Partitiviridae*	Partitivirus	*Gaeumannomyces graminis* virus
						Chrysovirus	*Penicillium chrysogenum* virus
						Alphacryptovirus	white clover cryptic virus I
						Betacryptovirus	white clover cryptic virus II
		class 7	order 1		*Reoviridae*	Orthoreovirus	reovirus type 1
						Orbivirus	bluetongue virus
						Cypovirus	cytoplasmic polyhedrosis vir. *B. mori*
						Fijivirus	Fiji disease virus
						Oryzavirus	rice ragged stunt virus
						Rotavirus	human rotavirus
						Aquareovirus	golden shiner virus
						Coltivirus	Colorado tick fever virus
						Phytoreovirus	wound tumour virus
		class 8	order 1		*Birnaviridae*	Aquabirnavirus	fish pancreatic necrosis virus
						Avibirnavirus	infectious bursal disease virus
						Entomobirnavirus	*Drosophila X* virus
		class 9 (ds phages)	order 1		*Cystoviridae*	Cystovirus	*Pseudomonas* phage Φ6

isometric plant virus family the *Comoviridae* (3 genera) and the filamentous *Potyviridae* (3, possibly 4 genera). The RNA polymerase-based dendrogram, shown in figure 1, confirms this arrangement and indicates that the plant virus families *Sobemoviridae* and *Ortholuteoviridae* (beet western yellows subgroup) as well as the invertebrate virus family *Nodaviridae* are included in this class. The sobemoviruses and ortholuteoviruses had been placed previously in a separate supergroup with the tombusviruses, carmoviruses and paraluteoviruses (Habili and Symons, 1989). The more recent report by Dolja and Carrington (1992) shows that the animal icosahedral *Caliciviridae* family (one genus) is also a member of this supergroup. Consequently, it is placed in this class as shown in table II. The newly recognized *Sequiviridae* (former plant picornaviruses) have not been included in such analyses but presumably belong to this class as they exhibit distant relationships to the picornaviruses (Mayo *et al.*, 1993).

The data in figure 1 show that the extent of sequence divergence between the RNA-dependent RNA polymerases of comoviruses and nepoviruses is of similar magnitude as that of the two genera of *Potyviridae* examined, while that between the sobemoviruses and ortholuteoviruses is more substantial and equivalent to that between the *Picornaviridae* and *Comoviridae*. This confirms the assignment of the comoviruses and nepoviruses as genera of the one family *Comoviridae*, as suggested by Martelli (1992), while the sobemoviruses and ortholuteoviruses represent distinct families. Pea enation mosaic virus was not included in the RNA polymerase analysis of Koonin (1991), but is included in this class because of its close relationship to the *Ortholuteoviridae* and *Sobemoviridae* as discussed earlier.

The members of this class have been divided into four orders on the basis of the dendrogram in figure 1. These orders also parallel major differences in accessory genes and in the case of the *Potyviridae* (order 2) particle morphology but not necessarily major host differences. It is also interesting to note that the isometric plant viruses (comoviruses, nepoviruses and fabaviruses) in class 1, order 1, have genome arrangements that much more closely resemble the animal viruses than the other isometric plant viruses (sobemoviruses, ortholuteoviruses and penamoviruses) which are placed in class 1, order 3 (table II). The polymerase dendrogram places the insect viruses, the *Nodaviridae*, into a separate order, order 4.

2) *Class 2*

The polymerase-based analysis placed the two pleomorphic, enveloped animal viruses, coronaviruses and toroviruses, into a second class. By comparison with the sequence relationships between other accepted genera (*e.g.*, potyviruses *vs* bymoviruses; comoviruses *vs* nepoviruses), other accepted families (*Sobemoviridae vs Ortholuteoviridae*) and suggested orders, it would appear that the coronaviruses and toroviruses are not genera of the same family, *Coronaviridae*, as currently assigned (see Francki *et al.*, 1991), but are distinct families which may belong to distinct orders as shown in table II. Although not included in the analysis by Koonin (1991), the arteriviruses have been shown to be more closely related to the coronaviruses than the togaviruses on the basis of genome organization (see Dolja and Carrington, 1992). According to the fifth report of ICTV, they will almost certainly be reclassified either as a genus of the family *Coronaviridae* or as a new family *Arteriviridae* (Strauss, 1991). Since the spherical particle morphology of the arteriviruses is so different from the other two families in this class, one would predict that the arteriviruses constitute a new family, *Arteriviridae*, in a new order, order 3, as tentatively suggested in table II. Thus another member of the original *Togaviridae* family has been reclassified (see Strauss *et al.*, 1990).

3) *Class 3*

The third class indicated by the polymerase-based dendrogram in figure 1 contains the *Togaviridae* family of animal viruses and the plant virus families, *Bromoviridae* (Bromovirus, Cucumovirus, Ilarvirus and Alfamovirus genera); *Tobamoviridae, Tobraviridae, Furoviridae* and *Hordeiviridae* initially assigned to the Sindbis-like superfamily (Goldbach and Wellink, 1988; Strauss and Strauss, 1988). It also contains those viruses subsequently assigned to this supergroup (Dolja and Carrington, 1992; Habili and Symons, 1989) such as the plant *Tymoviridae, Potexviridae* (potexvirus, carlavirus, and the ACLSV group of closteroviruses), *Closteroviridae,* the animal rubiviruses which appear to be a distinct family *(Rubiviridae)* and the hepatitis E virus family which I suggest be named *Hepeviridae.*

The polymerase-based analysis places the viruses in this class into three orders which have been split further into suborders on the basis of particle morphology (enveloped, rod-shaped, filamentous, isometric). This appears to be sensible from a practical point of view despite the very close relationship between the polymerases of: the *Closteroviridae* and *Bromoviridae* in order 1; the *Furoviridae* (as represented by beet necrotic yellow vein virus) and the two animal virus families *Rubiviridae* and *Hepeviridae* in order 2; and the *Tymoviridae* and *Potexviridae* in order 3 (fig. 1).

The polymerase-based analysis confirms the placement of the former closterovirus ACLSV into the same family *(Potexviridae)* as the other filamentous plant viruses (potexviruses and carlaviruses) which, as discussed earlier, are all thought to correspond to genera. The relationship of the capilloviruses to these filamentous plant viruses is not known at this stage.

What is fascinating is that the polymerase dendrogram in figure 1 splits the rubiviruses from the alphaviruses where they are currently classified (Strauss, 1991) and places them in a separate family *(Rubiviridae)* in a separate order (along with the *Hepeviridae*), but the same class as the *Togaviridae.* This separation is supported by the gene organization data (Dolja and Carrington, 1992), and makes the disintegration of the original *Togaviridae* family complete (Strauss *et al.,* 1990). The *Togaviridae* family initially contained four genera, the alphaviruses, the flaviviruses, the rubiviruses and the pestiviruses. Gene sequence data showed that although morphologically similar and arthropod-borne, the flaviviruses belonged to a different family, the *Flaviviridae* (Strauss *et al.,* 1990). Subsequently the pestiviruses were also assigned to the *Flaviviridae* and are currently classified there along with the hepatitis C virus (Strauss, 1991). The data in figure 1 suggests that even this is wrong and that the pestiviruses and the hepatitis C virus group constitute genera in separate new families (tentatively named *Pestiviridae* and *Hepciviridae*) in a separate order (order 2), but the same class (class 4) as the *Flaviviridae* as shown in table II.

4) *Class 4*

Other members of the fourth class in addition to the *Flaviviridae,* the *Pestiviridae* and the *Hepciviridae* (hepatitis C virus group) are all plant viruses. These viruses have been assigned to three orders (table II) on the basis of the sequence identities of their RNA-dependent RNA polymerases (fig. 1). These assignments also correlate with major differences in particle morphology between the isometric plant viruses (order 1) and the spherical enveloped animal viruses (orders 2 and 3).

5) *Class 5*

The fifth class shown in table II consists only of the RNA phages. In the original analysis (Koonin, 1991), the phages were recorded as a distant branch of the class 4 supergroup (figure 1) when the similar clustering and maximum parsimony methods were employed, but not when the maximum topological similarity method was used. Koonin (1991) discussed the relative merits of the three approaches and concluded that the final decisions regarding the phage polymerases were uncertain. For this reason and the practical advantage of keeping the RNA bacteriophages separate, I have placed the RNA phages in a single order in a class of their own. The degree of difference between their

RNA polymerases (fig. 1) suggests the leviviruses and alloleviviruses are distinct families, not genera of the same family, as currently classified (Ackermann, 1991b).

The double-stranded RNA viruses

The phylogenetic relationships of the dsRNA viruses have recently been examined (Bruenn, 1991; Koonin *et al.*, 1991, 1992). A comparative analysis of amino acid sequences of the complete genomes of members of the *Cystoviridae, Reoviridae, Birnaviridae, Totiviridae* and the plant cryptoviruses failed to show significant similarities betwen viruses from the different groups (Koonin, 1992). The analysis also showed that there was minimal similarity with proteins from related genera of the *Reoviridae* (Orbivirus, Rotavirus, Orthoreovirus), suggesting that some of these genera may, in fact, represent distinct families.

The only common denominator for all the dsRNA viruses examined was their RNA-dependent RNA polymerase which contained sequence motifs that are also found in the positive stranded RNA viruses (Koonin, 1992). Since the positive strand and dsRNA viruses appear to be recognizably similar in polymerase sequence, whereas the negative strand RNA viruses are distinctly different (Bruenn, 1991), it seems preferable to assign the positive strand and dsRNA viruses to a single subphylum rather than to two distinct subphyla.

Tentative phylogenetic trees have been constructed which place the *Cystoviridae* in a distinct class (class 9) of its own, group the plant cryptoviruses with the supergroup II positive strand RNA viruses and place the *Reoviridae, Birnaviridae* and *Totiviridae* with the supergroup I positive strand RNA viruses (Koonin *et al.*, 1991, 1992). Given that these latter associations with the positive strand RNA virus polymerases are very tentative and were not observed by Bruenn (1991), it seems preferable to: (i) assign the plant cryptoviruses with the *Partitiviridae* to a distinct class (class 6) as was done with the leviviruses (class 5); (ii) assign the *Reoviridae* and *Birnaviridae* to two distinct classes (classes 7 and 8); and (iii) only assign the *Totiviridae* as a distinct order (order 5) of one of the classes (class 1) of positive ssRNA viruses as shown in table II.

By analogy with the negative strand viruses, the *Reoviridae* with their multipartite genomes would be expected to be placed in a different order than that of the other dsRNA virus families, which have only one to three segments. Their larger size and more complex structure of successive layers of proteins in the virus particle (Prasad *et al.*, 1992) further support their placement in a separate class. The differing particle sizes (see Francki *et al.*, 1991) and/or genome arrangements and replication strategies for the *Cystoviridae, Birnaviridae* (Hudson *et al.*, 1986; Jagadish *et al.*, 1988; Morgan *et al.*, 1988), *Totiviridae* (Fujimura and Wickner, 1988) and *Partitiviridae* suggest they too should each be assigned to distinct classes and orders. These virus families also infect very different host organisms (see Francki *et al.*, 1991). Comparative sequence data is required to confirm the relationship between the plant cryptoviruses and the fungal *Partitiviridae.*

The negative strand RNA viruses

The negative strand viruses of vertebrates and plants form a fairly homogeneous group structurally and share a number of common features of genome organization and replication strategy, as summarized in Strauss *et al.* (1990) and Tordo *et al.* (1992). In the fifth report of ICTV (Francki *et al.*, 1991) it was agreed to combine the three families of monopartite, negative strand viruses, *Paramyxoviridae, Filoviridae,* and *Rhabdoviridae,* into the order *Mononegavirales* as outlined by Pringle (1990). This is the first classification of order of viruses that has been accepted by ICTV. It was also suggested that consideration be given to combining the multisegmented negative strand virus families, *Orthomyxoviridae, Arenaviridae* and *Bunyaviridae* into a second order (Pringle, 1990) that, for consistency, could be called the *Multinegavirales.* The individual gene segments of the *Orthomyx-*

oviridae can be arranged to match the functions of the ordered gene set in the *Paramyxoviridae* and *Rhabdoviridae* (Strauss *et al.*, 1990; Tordo *et al.*, 1992), suggesting that these two orders belong to a single class, the *Negavirata,* as summarized in table III. Phylogentic analyses of the negative strand virus polymerases confirm these assignments of the negative strand viruses into two orders of a single class (Tordo *et al.*, 1992). The data further indicate that the pneumovirus, RSV (respiratory syncytial virus), appears to be a member of a distinct virus family *(Pneumoviridae)* in the order *Mononegavirales,* not the subfamily *(Pneumovirinae)* of the *Paramyxoviridae* as currently classified. There are two genera of plant viruses in the *Rhabdoviridae* and one genus of plant virus (Tospovirus) in the *Bunyaviridae.* It has recently been shown that the segmented genome of the tenuivirus family of plant viruses is ambisense and resembles that of the *Arenaviridae* and *Bunyaviridae* (Kakutani *et al.*, 1990; Zhu *et al.*, 1991; Takahashi *et al.*, 1993; Hamamatsu *et al.*, 1993), but is sufficiently different to warrant its classification as a distinct family *(Tenuiviridae)* in the order *Multinegavirales* as shown in table III.

The RNA-dependent RNA polymerases of the negative strand viruses are distinctly different from those of the positive strand and dsRNA viruses (Bruenn, 1991), confirming their assignment to a single class and raising the question whether they should be considered as a separate subphylum.

The DNA viruses

Given the relationships established for the positive strand RNA viruses from an analysis of their RNA-dependent RNA polymerases, it is tempting to speculate whether an analysis of the DNA polymerases of the 24 families of DNA viruses (see Francki *et al.*, 1991) can be used to establish a higher taxonomy for these viruses. Such an analysis has recently been initiated by Braithwaite and Ito (1993) and reveals some intriguing results. They examined the DNA polymerases from 26 viruses of 10 DNA virus families, and the results are summarized in table IV and figures 6 and 7.

These results suggest that the T5, T7, Spo1 and Spo2 phages belong to one phylum. They all have type A DNA polymerases that are structurally related to the *Escherichia coli,* polA gene product, DNA polymerase I (Braithwaite and Ito, 1993). The phylogenetic tree for these polymerases (fig. 6) suggests that the T5 phages and the Spo2 phages belong to different families, not the same family *(Siphoviridae)* as currently classified (Ackermann, 1991a). The name *Allosiphoviridae* is a possible name for the Spo2 family of tailed phages, as it is consistent with the nomenclature used to distinguish the leviviruses and alloleviviruses (Ackermann, 1991b).

The remaining DNA viruses examined have type B polymerases that are structurally related to the *E. coli* polB gene product, DNA polymerase II (Braithwaite and Ito, 1993). These can be assigned to a second phylum that contains three subphyla (table IV). The first subphylum contains the *Adenoviridae* and the phages PRD1 *(Tectiviridae),* phi29 *(Podoviridae)* and M2 *(Podoviridae).* Their polymerases are protein-primed DNA polymerases and are similar to those from bacterial and fungal plasmids supporting their possible evolution from these self-replicating molecules. The sequence identity between the DNA polymerases of these four families of viruses (fig. 7) suggests the three phages belong to a different order but probably the same class as the *Adenoviridae* (table IV). The analysis also reveals that the phi29, M2 and T7 phages, which are currently classified as members of the same family *Podoviridae* (Ackermann, 1991a), have very different DNA polymerases and belong to different families in different phyla (table IV). As discussed above, the name *Allopodoviridae* is suggested for the phi29/M2 phages family to distinguish them from the well characterized T7 *Podoviridae.*

The second subphylum comprises those viruses with RNA-primed type B DNA polymerases (table IV). These are the T4 family of tailed phages *(Myoviridae),* the *Phycodnaviridae* (algal viruses), and the *Poxviridae, Baculoviridae*

Table III. Higher taxa of negative ssRNA viruses.

Phylum	Subphylum	Class	Order	Family	Genus	Species
RNA viruses	negative ssRNA	Negavirata	*Mononegavirales*	*Paramyxoviridae*	Paramyxovirus	Newcastle disease virus
					Morbillivirus	measles virus
				Pneumoviridae	Pneumovirus	human respiratory syncytial virus
				Filoviridae	Filovirus	Marburg virus
				Rhabdoviridae	Vesiculovirus	vesicular stomatitis Indiana virus
					Lyssavirus	rabies virus
					Plant rhabdo A	lettuce necrotic yellows virus
					Plant rhabdo B	potato yellow dwarf virus
			Multinegavirales	*Orthomyxoviridae*	Influenza A	influenza A/PR/8/34
					Influenza B	influenza B/Lee/40
					Influenza C	influenza C/Taylor/1233/47
					Acarivirus	Dhori virus
				Bunyaviridae	Bunyavirus	Bunyamwera virus
					Phlebovirus	sandfly fever Sicilian virus
					Nairovirus	Crimean-Congo haemorrhagic fever
					Hantavirus	Hantaan virus
					Tospovirus	tomato spotted wilt virus
				Arenaviridae	Arenavirus	lymphocytic choriomeningitis virus
				Tenuiviridae	Tenuivirus	rice stripe virus

Table IV. Higher taxa of some DNA viruses.

Phylum	Subphylum	Class	Order	Family	Subfamily	Genus	Species
Phylum 1 (type A polymerase)		?	?	*Siphoviridae*		Coliphage lambda group	phage T5
		?	?	*Allomyoviridae* (*)		*B. subtilis* hmU phage group	phage Spo1
		?	?	*Podoviridae*		Coliphage T7 group	phage T7
		?	?	*Allosiphoviridae* (*)		*B. subtilis* temperate group	phage Spo2
Phylum 2 (type B polymerase)	Subphylum 1 (protein-primed)	class 1	order 1	*Adenoviridae*		Mastadenovirus	human adenovirus 2
			order 2	*Tectiviridae*		Tectivirus	phage PRD1
				Allopodoviridae (*)		*Bacillus* spp. podovirus group	phage phi29 and M2
	Subphylum 2 (RNA-primed)	class 1	order 1	*Myoviridae*		T4 phage group	phage T4
		class 2	order 1	*Poxviridae*	Chordopoxvirinae	Orthopoxvirus	vaccinia virus
						Avipoxvirus	fowlpox virus
					Entomopoxvirinae	Entomopoxvirus A	*Choristoneura biennis* poxvirus
				Baculoviridae	Eubaculovirinae	Nuclear polyhedrosis	*Autographa californica* nuclear polyhedrosis virus
							Lymantra dispar nuclear polyhedrosis virus
		class 3	order 1	*Herpesviridae*	Alphaherpesvirinae	Simplexvirus	human (alpha) herpesvirus 1
						Varicellovirus	equid (alpha) herpesvirus 1
							varicella-zoster virus
					Gammaherpesvirinae	Lymphocryptovirus	Epstein-Barr virus
						Rhadinovirus	saimirine herpesvirus 1
					Betaherpesvirinae	Cytomegalovirus	human cytomegalovirus
						Muromegalovirus	murine cytomegalovirus
						human herpes type 6	human herpesvirus 6
			order 2	*Phycodnaviridae*		Phycodnavirus	Chlorella virus NY-2A
	Subphylum 3	class 1	order 1	Channel catfish virus family		Channel catfish virus	Channel catfish virus

Not all DNA virus families or genera are listed here, only those corresponding to the 26 viruses whose DNA polymerases were compared by Braithwaite and Ito (1993).

(*) Allo prefix to distinguish these viruses from the *Myoviridae* (T4 phage group), *Siphoviridae* (coliphage lambda group) or the *Podoviridae* (T7 colophage group) respectively.

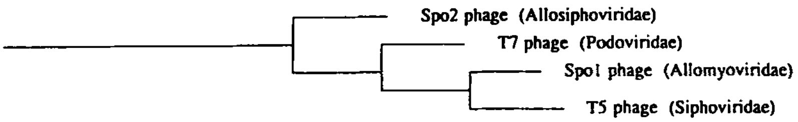


Fig. 6. Phylogenetic phenogram tree of the type A viral DNA polymerases.

The allo prefix is used to distinguish the Spo1 phage and the Spo2 phage from the *Myoviridae* (T4 phage group) and *Siphoviridae* (coliphage lambda group) respectively. Redrawn from Braithwaite and Ito (1993).

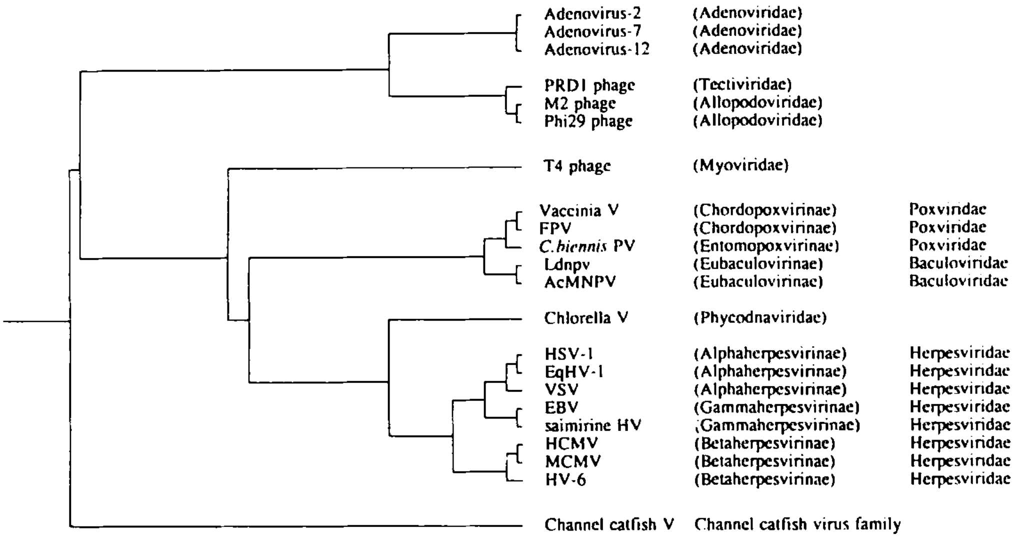


Fig. 7. Phylogenetic phenogram tree of the type B viral DNA polymerases.

The virus acronyms in alphabetical order are: AcMNPV, *Autographa californica* nuclear polyhedrosis virus; *C. biennis* PV, *Choristoneura biennis* pox virus; EBV, Epstein-Barr virus; EqHV-1, equine herpes virus type-1; VSV, varicella-zoster virus; FPV, fowlpox virus; HCMV, human cytomegalovirus; HSV-1, herpes simplex virus type-1; HV-6, human herpes virus type-6; LdNPV, *Lymantria dispar* nuclear polyhedrosis virus; MCMV, murine cytomegalovirus; saimirine HV, *Herpesvirus saimiri;* the allo prefix is used to distinguish the phi29 and M2 phage from the *Podoviridae* (T7 coliphage group). Redrawn from Braithwaite and Ito (1993).

and *Herpesviridae* (animal and insect viruses). The sequence identities of these DNA polymerases (fig. 7) suggest that: (i) the T4 tailed phages *(Myoviridae)* belong to one class and are quite distinct from the *Bacillus subtilis,* Spo1, hmU phage group (Rabussay and Geiduschek, 1977) that are also currently classified as *Myoviridae* (Fraenkel-Conrat, 1985) but have a type A DNA polymerase (table IV) and appear to belong to a different phylum; (ii) the *Herpesviridae* and *Phycodnaviridae* belong to different orders of a second class; and (iii) the *Poxviridae* and *Baculoviridae* are closely related and belong to the same order of a third class in this subphylum of DNA viruses. As shown in figure 7, the DNA polymerases of the *Poxviridae* and *Baculoviridae* are as closely related to each other as the Alphaherpesvirinae and Gammaherpesvirinae and more closely related than the third *Herpesviridae* subfamily, the Betaherpesvirinae. It is also interesting to note that the latter has a considerably larger genome size: ~ 235-250 kbp for human cytomegalovirus (Betaherpesvirinae) compared to 125 kbp for the varicella-zoster virus (Alphaherpesvirinae) and the 175 kbp for EBV (Gammaherpesvirinae; Francki *et al.*, 1991).

Channel catfish virus appears to belong to a third subphylum, as its DNA polymerase is the most divergent of the type B enzymes (fig. 7). It is very different from those of the alphaherpesviruses with which it is currently classified (Fraenkel-Conrat, 1985).

Concluding remarks

In this review, two long-standing problems in viral taxonomy have been addressed. The first is the assignment of the remaining plant virus groups to families and genera, which had been resisted by plant virologists for a long time, but which has recently gained momentum. The second is the reluctance to consider higher taxa, given the polyphyletic origin of viruses and the recombinative process of cassette assembly, where gene acquisitions from a variety of sources have accompanied the evolution of different virus families. Both problems can now be approached given the wealth of information that is currently accumulating on virus genome sequences, replication strategies and structural relationships.

With regard to the first problem, plant virus taxonomy, considerable progress has been made with the recognition that the current 35 plant virus groups represent a mixture of families and genera. The current state of deliberation by the Plant Virus Subcommittee of ICTV has been concisely summarized by Martelli (1992). Three plant virus families, *Rhabdoviridae, Bunyaviridae* and *Reoviridae,* have been recognized for some time since the plant viruses are obviously members of these accepted larger families that include viruses infecting vertebrates and invertebrates. More recently proposals for the *Geminiviridae,* the *Cryptoviridae,* the *Bromoviridae,* the *Comoviridae,* the *Tombusviridae,* the *Potyviridae* and the *Sequiviridae* have been developed (Martelli, 1992; Mayo *et al.*, 1993). In assigning these plant virus groups to seven families no new names were coined. Rather, the families were named after their best characterized genus, as done with many of the animal virus families, to minimize the proliferation of unnecessary new names (Martelli, 1992). The remaining 18 plant virus groups were provisionally classified as genera by Martelli (1992), but were not assigned to families.

In this review, 16 of these remaining 18 plant virus groups have been tentatively assigned to new families (table I) primarily on the basis of genome structure and organization, gene sequence identity and particle morphology. The suggested possible names for these families are based on the names of the best characterized genus following the principles discussed by Martelli (1992). Many of these plant virus families contain only one genus but appear to be sufficiently distinct to warrant their classification as separate families. The most contentious conclusions are: (i) the separation of the luteovirus group into two families tentatively named *Ortholuteoviridae* and *Paraluteoviridae* to retain the luteovirus connecting for both families; (ii) the separation of the closteovirus group of filamentous viruses into two families, the *Closteroviridae* (sugar beet yellows subgroup) and the ACLSV subgroup, which appears to be a genus of the *Potexviridae.* The assignment of the potexvirus and carlavirus groups of filamentous plant viruses to a single family, the *Potexviridae,* is less contentious.

The tentative assignment of the plant virus groups to families and genera discussed here requires further verification by the expert working groups, and it would seem to be timely for the plant virus subcommittee of ICTV to form intergroup committees to examine such proposals. In considering these assignments, one must keep in mind the need to accommodate biological variation in any ordered system of classification. Thus the number of gene segments, the presence of additional unique coding regions and the use of alternate vectors are all properties that can vary between genera of the same family if the gene sequence data shows that most of the other gene products are appropriately related. This is important in the final decision regarding the classification of the subgroup III geminiviruses, the dianthoviruses, the pea enation mosaic virus group, the bymoviruses, the tobamoviruses, the tobraviruses and the hordeiviruses. The classification of the pea enation mosaic virus group as a distinct family rather than a genus of the *Ortholuteoviridae* and the assignment of the dianthovirus group as a distinct family rather than a genus of the *Tombusviridae* is based primarily on the presence of a bipartite genome where the second RNA segment has considerable coding capacity. These groups could just as easily be classified as subfamilies of the *Ortholuteoviridae* and *Tombusviridae,* respectively. The final decision about the assignment of the tobamovirus, tobravirus and hordeivirus groups of rod-shaped viruses to separate families rather than to subfamilies or genera of a single family will be based on similar arguments given their very similar gene organization, close polymerase sequence identity, 5' capped structure and 3' tRNA-like structures. This raises the question whether the third subgroup of the *Geminiviridae,* with their additional gene segment (Howarth and Goodman, 1986; Stanley *et al.,* 1986) should also be classified as a distinct family or subfamily rather than a distinct genus.

With regard to the question of higher taxa, the two major objections, multiple origins and recombinative cassette evolution, are not sufficient justification to conclude that the problem is insoluble. The first objection can be overcome by assigning all viruses to a separate kingdom and having multiple phyla to reflect the diverse origins. The second problem is overcome by basing the higher taxa on a restricted set of primitive characteristics. For the RNA viruses, this is the replicase function, as it is postulated to be an ancient evolutionary feature that predates cassette assembly and probably evolved only once (Rybicki, 1990).

Using the phylogenetic dendrograms (Koonin, 1991; Dolja and Carrington, 1992) obtained for the RNA-dependent RNA polymerases of the positive strand RNA viruses, it is possible to assign all the positive strand RNA viruses examined into classes, orders, families and genera. The validity of these assignments is strengthened by the observation that the five classes correspond to the supergroups previously recognized (Goldbach and Wellink, 1988; Habili and Symons, 1989; Strauss *et al.,* 1990; Dolja and Carrington, 1992) and widely accepted by plant and animal virologists. The analysis places the RNA bacteriophages into a single class and is consistent with the decisions to split the luteoviruses into two families and to assign the ACLSV closterovirus to the *Potexviridae.*

The use of the RNA polymerase as the primary parameter to establish higher relationships is further validated by the fact that the taxonomy obtained (table II) is consistent with recent trends in the classification of the *Togaviridae.* The original *Togaviridae* family contained four genera (Alphavirus, Flavivirus, Rubivirus and Pestivirus) but was subsequently split into two families: the *Togaviridae* (with three genera, Alphavirus, Rubivirus and Arterivirus), and the *Flaviviridae* (with three genera, Flavivirus, Pestivirus and hepatitis C virus). However the taxonomy indicates that these current assignments are still incomplete as: (i) the rubiviruses appear to belong to a separate family, the *Rubiviridae,* in a different order but the same class as the *Togaviridae,* (ii) the pestiviruses and the hepatitis C group are not genera of the *Flaviviridae* but belong to separate families, the *Pestiviridae* and *Hepciviridae* in a separate order but the same class as the *Flaviviridae.* The taxonomy also accounts for the intended decision to reclassify the arteriviruses as a genus of the *Coronaviridae* or a new closely related family (Strauss, 1991), since it shows the *Arteriviridae* as a distinct family in the same class as the coronaviruses and toroviruses. It also indicates that the toroviruses are not a genus of the *Coronaviridae* as currently classified, but are a separate family, *Toroviridae,* in a separate order from the *Coronaviridae* but the same class as the *Arteriviridae.*

The dsRNA polymerases have also been compared and shown to have only distant relationships to each other and to the positive strand RNA virus polymerases. It is concluded that they correspond to an additional order in class 1 and four additional classes (classes 6-9) in the same subphylum as the positive strand RNA viruses (table II).

A higher taxonomy of negative strand RNA viruses is already being addressed with the acceptance by ICTV of the order *Mononegavirales* that contains the *Paramyxoviridae, Filoviridae* and *Rhabdoviridae* (Pringle, 1990). Analyses of the RNA polymerases of the negative strand viruses has confirmed this assignment and indicated that the multisegmented *Orthomyxoviridae, Bunyaviridae, Arenaviridae* and *Tenuiviridae* constitute a second order *(Multinegavirales),* in the same class *(Negavirata)* as the *Mononegavirales.* These negative strand viruses have been placed in a second subphylum.

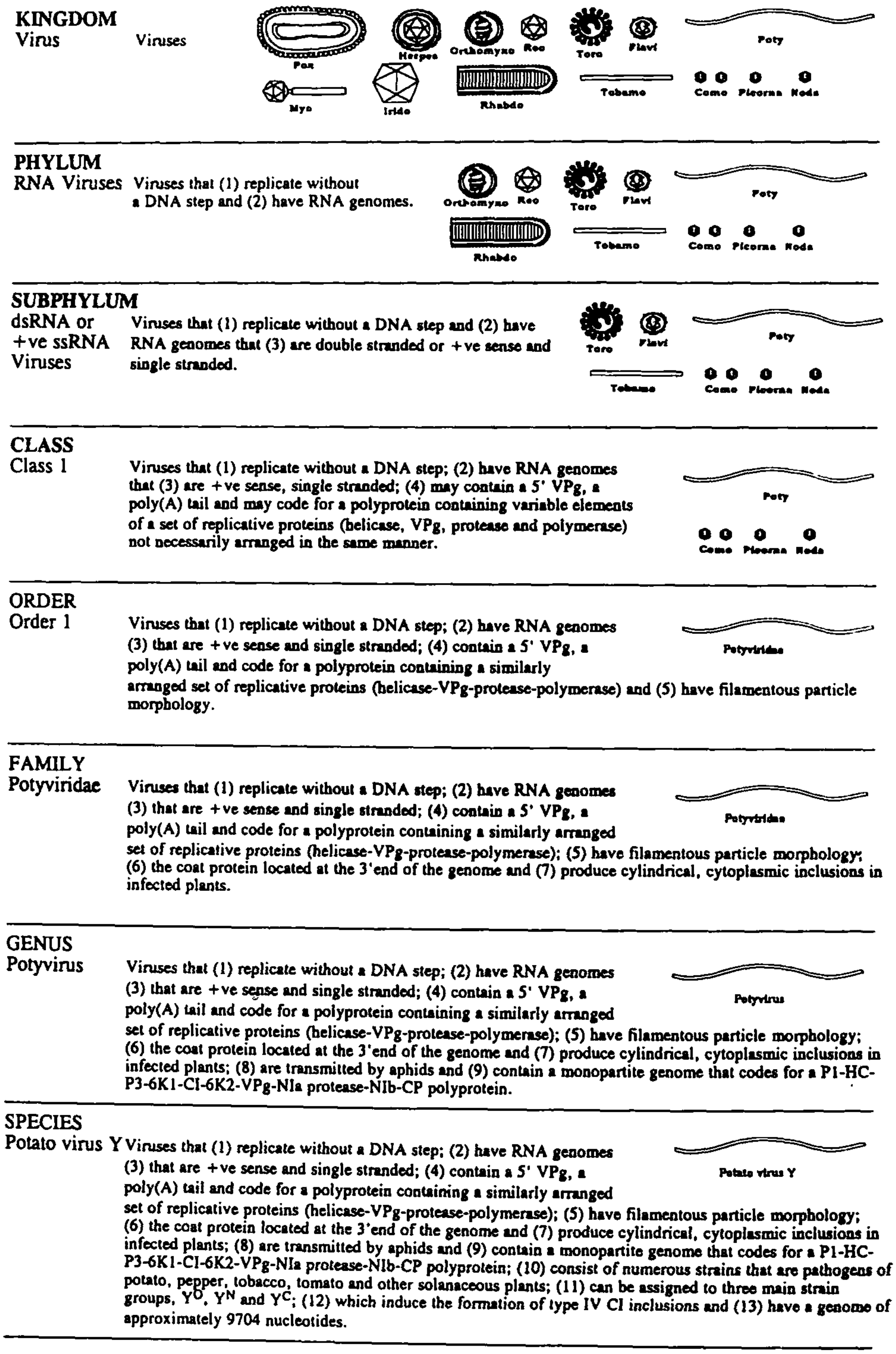
Finally, the preliminary analysis of the DNA polymerases of the 10 families of DNA viruses (Braithwaite and Ito, 1993) needs to be extended to include the remaining genera of the families examined as well as representatives of the genera of the other 14 families/groups of DNA viruses to see if the DNA polymerase proteins can be used to establish higher taxonomic relationships. The results suggest that, despite the large size and complexity of DNA virus genomes, the sequences of just the highly conserved DNA polymerase can be used to establish higher order relationships in the same way that the highly conserved ribosomal RNA gene sequences have been used in prokaryotic and eukaryotic taxonomy (Milinkovitch *et al.,* 1993; Stackebrandt, 1994). The results also indicate that the current taxonomy, where 95 % of the known viruses infecting eubacteria are assigned to one of the three families of tailed bacteriophages (Matthews, 1985a), is inadequate. The data presented in table IV and figures 6 and 7 suggests there are at least six families of tailed phages in two different phyla.

One should not be concerned that the taxonomy presented in tables II, III and IV results in the presence of several one-member classes and orders since, as Matthews (1985a) points out, there are undoubtedly many more viruses that infect archebacteria, marine bacteria and invertebrates awaiting discovery. Invertebrates account for 97 % of living animal species.

The combination of these two sets of observations, polymerase dendrograms and virus supergroups, has provided a higher taxonomy of the positive strand RNA viruses that appears to be workable and consistent with other findings. As depicted in table V and figure 8, the viruses in each level have more and more features in common as you move down the classification table from kingdom to species. In the scheme described here, the nature of the viral genome and the viral polymerase are the criteria that define phyla and subphyla, while the relative sequence identity of the polymerase and the arrangement of the core replication module are the criteria that define class. Sequence identity of the polymerase gene is also the major criterion that defines order and suborder, which in turn correlate well with particle morphology and the presence or absence of an envelope. All members of the same family would be expected to have similar particle morphology and share most gene coding regions but do not necessarily have equivalent numbers of gene segments or identical gene sets. Members of the same genera would be expected to possess the same gene set and show moderate sequence identity (35-85 %) be-

Table V. Criteria defining virus taxonomy.

Taxon level	Criteria
Phylum subphylum	(1) nature of genome (2) nature of polymerase
Class	(1) and (2) and (3) sequence identity of the polymerase and (4) replication strategy and nature of gene set of non-structural replication proteins
Order	(1) to (4) and (5) particle morphology
Family	(1) to (5) and (6) specific gene set
Genus	(1) to (6) and (7) nature of vector and (8) sequence identity of other gene products
Species	(1) to (8) and (9) biological and serological properties



KINGDOM	Virus	Viruses
PHYLUM	RNA Viruses	Viruses that (1) replicate without a DNA step and (2) have RNA genomes.
SUBPHYLUM	dsRNA or +ve ssRNA Viruses	Viruses that (1) replicate without a DNA step and (2) have RNA genomes that (3) are double stranded or +ve sense and single stranded.
CLASS	Class 1	Viruses that (1) replicate without a DNA step; (2) have RNA genomes that (3) are +ve sense, single stranded; (4) may contain a 5' VPg, a poly(A) tail and may code for a polyprotein containing variable elements of a set of replicative proteins (helicase, VPg, protease and polymerase) not necessarily arranged in the same manner.
ORDER	Order 1	Viruses that (1) replicate without a DNA step; (2) have RNA genomes (3) that are +ve sense and single stranded; (4) contain a 5' VPg, a poly(A) tail and code for a polyprotein containing a similarly arranged set of replicative proteins (helicase-VPg-protease-polymerase) and (5) have filamentous particle morphology.
FAMILY	Potyviridae	Viruses that (1) replicate without a DNA step; (2) have RNA genomes (3) that are +ve sense and single stranded; (4) contain a 5' VPg, a poly(A) tail and code for a polyprotein containing a similarly arranged set of replicative proteins (helicase-VPg-protease-polymerase); (5) have filamentous particle morphology; (6) the coat protein located at the 3'end of the genome and (7) produce cylindrical, cytoplasmic inclusions in infected plants.
GENUS	Potyvirus	Viruses that (1) replicate without a DNA step; (2) have RNA genomes (3) that are +ve sense and single stranded; (4) contain a 5' VPg, a poly(A) tail and code for a polyprotein containing a similarly arranged set of replicative proteins (helicase-VPg-protease-polymerase); (5) have filamentous particle morphology; (6) the coat protein located at the 3'end of the genome and (7) produce cylindrical, cytoplasmic inclusions in infected plants; (8) are transmitted by aphids and (9) contain a monopartite genome that codes for a P1-HC-P3-6K1-CI-6K2-VPg-NIa protease-NIb-CP polyprotein.
SPECIES	Potato virus Y	Viruses that (1) replicate without a DNA step; (2) have RNA genomes (3) that are +ve sense and single stranded; (4) contain a 5' VPg, a poly(A) tail and code for a polyprotein containing a similarly arranged set of replicative proteins (helicase-VPg-protease-polymerase); (5) have filamentous particle morphology; (6) the coat protein located at the 3'end of the genome and (7) produce cylindrical, cytoplasmic inclusions in infected plants; (8) are transmitted by aphids and (9) contain a monopartite genome that codes for a P1-HC-P3-6K1-CI-6K2-VPg-NIa protease-NIb-CP polyprotein; (10) consist of numerous strains that are pathogens of potato, pepper, tobacco, tomato and other solanaceous plants; (11) can be assigned to three main strain groups, Y^O, Y^N and Y^C; (12) which induce the formation of type IV CI inclusions and (13) have a genome of approximately 9704 nucleotides.

Fig. 8. Illustrated summary of plant virus taxonomy.

As you move down the tables from kingdom to species, the viruses at each level have more and more features in common.

tween most of their genes, but to vary in biological properties in some cases. Most virus species have been described in terms of their biological properties. With strains of individual species, the sequence identity of all gene coding regions is relatively high (85-90 %) and biological properties such as host range, symptomatology, resistance genes and serology are of increasing importance in strain characterization.

It is intriguing to note that the few parameters listed in table V that define orders (*e.g.* genome type, genome strandedness, polymerase sequence identity, particle morphology and the presence or absence of an envelope) are, except for polymerase sequence identity, the very same restricted set of key differential parameters that Francki (1983) suggested should be used to assign families of viruses to orders. As Martelli (1992) points out, there was no follow-through of Francki's proposal, as there was determined opposition among plant virologists to the use of traditional taxonomic systems at that time. Given the present level of information on most families of viruses, it seems overly pessimistic to avoid addressing the question of the higher taxonomy of viruses for fear of not getting it correct at the first attempt. The taxonomies of bacteria, plants, invertebrates and higher animals are constantly being reviewed and adjusted as new data becomes available (Milinkovitch *et al.*, 1993; Stackebrandt, 1994). As the philosopher Karl Popper advocated, the original approach to scientific investigation of attempting to verify hypotheses formulated from initial observations is inadequate. Popper suggests that one should attempt to *disprove* one's hypotheses as these refutations, combined with the original theory, will yield a better one (White and Gribbin, 1992). This was true for influenza virus research where the provocative theory of Fazekas de St Groth (1970, 1978), although wrong, precipitated a large international effort that elucidated the molecular basis of antigenic variation in this family of viruses (Ward, 1981; Colman and Ward, 1985). It is also clearly happening in viral taxonomy, where the reassignment of the *Togaviridae* is but one example. It is hoped that this process will continue as more and more information becomes available for each group of viruses, and that increasing attention will be directed to critically examining the higher order relationships discussed here.

Vers une taxonomie hiérarchique des virus

L'opinion consensuelle actuelle est qu'une taxonomie hiérarchique des virus ne peut être établie, cela pour deux raisons. La première est que les virus ont des origines variées car divers groupes de virus sont apparus de façon indépendante à différents moments. La seconde est que l'adaptation secondaire des virus, pour leur survie dans différentes combinaisons hôte/vecteur, a entraîné l'acquisition sélective de modules géniques supplémentaires issus d'autres virus ou du matériel génétique de l'hôte. Ainsi, selon le produit génique utilisé pour la comparaison, différentes relations peuvent être établies.

La solution à la première objection est de placer tous les virus dans un règne séparé et de ranger les virus actuels dans divers phylums reflétant cette variété des origines. La solution à la deuxième objection est de considérer le module au cœur du mécanisme de la réplication comme critère majeur pour faire une première répartition des virus en classes et en ordres. Pour les virus à ARN le critère majeur est l'identité de la séquence de l'ARN polymérase ARN-dépendante.

Sur la base de ce critère, les virus ARN à brin positif peuvent être répartis dans 5 classes qui correspondent aux supergroupes des virus ARN récemment reconnus. Les dendrogrammes de la polymérase montrent également que les virus double-brin appartiennent au même sous-phylum que celui des virus ARN à brin positif. Les virus ARN double-brin ne sont pas étroitement reliés les uns aux autres et peuvent être répartis dans quatre classes supplémentaires et dans un ordre supplémentaire d'une des classes des virus ARN à simple-brin positif. Les dendrogrammes de la polymérase des virus à brin négatif permettent de confirmer le classement des virus à brin négatif dans deux ordres d'une seule classe d'un sous-phylum séparé parmi les virus à ARN. Dans cette revue sont incluses des données qui laissent à penser que les virus à ADN peuvent être aussi mieux répartis sur la base des identités de séquence de leurs ADN polymérases fortement conservées.

L'idée d'utiliser les polymérases virales pour l'établissement de relations mieux ordonnées, vient de ce qui se fait pour les études taxonomiques des procaryotes et eucaryotes qui utilisent les séquences géniques ARNr fortement conservées. Dans cette revue, nous incluons un essai de classement de 33/35 groupes de virus végétaux en tant que genres de

25 familles, basé sur la nature et l'agencement du génome, sur le niveau d'identité de séquence et, pour une moindre mesure, sur la morphologie des particules.

Mots-clés: Taxonomie, Virus, Hiérarchie; Virus ADN, Virus ARN; Revue.

References

Ackermann, H.-W. (1991a), Tailed phages, *in* "Classification and Nomenclature of Viruses: Fifth Report of the International Committee on Taxonomy of Viruses" (Francki, R.I.B., Fauquet, C.M., Knudson, D.L. & Brown, F.) (pp. 159-166). Springer-Verlag, Heidelberg, Berlin.

Ackermann, H.-W. (1991b), Leviviridae, *in* "Classification and Nomenclature of Viruses: Fifth Report of the International Committee on Taxonomy of Viruses" (Francki, R.I.B., Fauquet, C.M., Knudson, D.L. & Brown, F.) (pp. 306-308). Springer-Verlag, Heidelberg, Berlin.

Agranovsky, A.A., Boyki, A.P., Karasev, A.V., Lunina, N.A., Koonin, E.V. & Dolja, V.V. (1991), Nucleotide sequence of the 3'-terminal half of beet yellow closterovirus RNA genome: unique arrangement of eight virus genes. *J. Gen. Virol.*, 72, 15-23.

Ahlquist, B., Strauss, E.G., Rice, C.M., Strauss, J.H., Haseloff, J. & Zimmern, D. (1985), Sindbis virus proteins nsP1 and nsP2 contain homology to nonstructural proteins from several RNA plant viruses. *J. Virol.*, 53, 536-542.

Argos, P., Kamer, G., Nicklin, M.J.H. & Wimmer, E. (1984), Similarity in gene organization and homology between proteins of animal picornaviruses and a plant comovirus suggest common ancestry of these virus families. *Nucl. Acids Res.*, 122, 7251-7267.

Bar-Joseph, M. & Murant, A.F. (1982), Closterovirus group. CMI/AAB Descriptions of Plant Viruses, No. 260.

Barnett, O.W. (1991), Potyviridae, a proposed family of plant viruses. *Arch. Virol.*, 118, 139-141.

Barnett, O.W. (1992), Potyvirus Taxonomy. Springer-Verlag, Heidelberg, Berlin.

Boyko, P.V., Karasev, A.V., Agranovsky, A.A., Koonin, E.V. & Dolja, V.V. (1992), Coat protein gene duplication in a filamentous RNA virus of plants. *Proc. Natl. Acad. Sci.* (Wash.), 89, 9156-9160.

Braithwaite, D.K. & Ito, J. (1993), Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucl. Acids Res.*, 21, 787-802.

Bruenn, J.A. (1991), Relationships among the positive strand and double-strand RNA viruses as viewed through their RNA-dependent RNA polymerases. *Nucl. Acids Res.*, 19, 217-226.

Buck, K.W. & Ghabrial, S.A. (1991), Totiviridae, *in* "Classification and Nomenclature of Viruses: Fifth Report of the International Committee on Taxonomy of Viruses" (Francki, R.I.B., Fauquet, C.M., Knudson, D.L. & Brown, F.) (pp. 203-205). Springer-Verlag, Heidelberg, Berlin.

Carrington, J.C., Morris, T.J., Stockley, P.G. & Harrison, S.C. (1987), Structure and assembly of turnip crinkle virus. Analysis of the coat protein gene and implications of the subunit primary structure. *J. Mol. Biol.*, 194, 265-276.

Colman, P.M. & Ward, C.W. (1985), Structure and diversity of influenza virus neuraminidase. *Curr. Topics Microbiol. Immunol.*, 114, 177-255.

Dasgupta, R. & Kaesberg, P. (1982), Complete nucleotide sequences of the coat protein messenger RNAs of brome mosaic virus and cowpea chlorotic mottle virus. *Nucl. Acids Res.*, 10, 703-713.

Davies, C. & Symons, R.H. (1988), Further implications for the evolutionary relationships between tripartite plant viruses based on cucumber mosaic virus RNA 3. *Virology*, 165, 216-224.

Demler, S.A. & de Zeoten, G.A. (1991), The nucleotide sequence and luteovirus-like nature of RNA 1 of an aphid non-transmissible strain of pea enation mosaic virus. *J. Gen. Virol.*, 72, 1819-1834.

Ding, S., Keese, P. & Gibbs, A. (1989), Nucleotide sequence of the ononis yellow mosaic tymovirus genome. *Virology*, 172, 555-563.

Dolja, V.V. & Carrington, J.C. (1992), Evolution of positive-strand RNA viruses. *Sem. Virol.*, 3, 315-326.

Dolja, V.V., Boyko, V.P., Agranovsky, A.A. & Koonin, E.V. (1991), Phylogeny of capsid proteins of rod-shaped and filamentous RNA plant viruses: two families with distinct patterns of sequence and probably structure conservation. *Virology*, 184, 79-86.

Fazekas de St Groth, S. (1970), Evolution and hierarchy of influenza viruses. *Arch. Environ. Health*, 21, 293-303.

Fazekas de St Groth, S. (1978), Antigenic, adaptive and adsorptive variants of the influenza A hemagglutinin, *in* "The Influenza Virus Hemagglutinin" (Laver, W.G., Bachmayer, H. & Weil, R.) (pp. 25-48). Springer-Verlag, Heidelberg, Berlin.

Fenner, F. (1979), Portraits of viruses: the poxviruses. *Intervirology*, 11, 137-157.

Forster, R.L.S., Bevan, M.W., Harbison, S.-A. & Gardner, R.C. (1988), The complete nucleotide sequences of the potexvirus white clover mosaic virus. *Nucl. Acids Res.*, 16, 291-303.

Fraenkel-Conrat, H. (1985), The Viruses: catalogue, characterization and classification. Plenum Press, New York.

Francki, R.I.B. (1983), Current problems in plant virus taxonomy, *in* "A Critical Appraisal of Viral Taxonomy" (Matthews, R.E.F.) (pp. 63-104). CRC Press, Boca Raton, FL.

Francki, R.I.B., Fauquet, C.M., Knudson, D.L. & Brown, F. (1991), Classification and Nomenclature of Viruses: Fifth Report of the International Committee on Taxonomy of Viruses. Springer-Verlag, Heidelberg, Berlin.

Franssen, H., Leunissen, J., Goldbach, R., Lomonosoff, G. & Zimmern, D. (1984), Homologous sequences in non-structural proteins from cowpea mosaic virus and picornaviruses. *EMBO J.*, 3, 855-861.

Fujimura, T. & Wickner, R.B. (1988), Gene overlap results in a viral protein having an RNA binding domain and a major coat protein domain. *Cell*, 55, 663-671.

Georgopoulos, C., Tilly, K. & Casjens, S. (1983), Lambdoid phage head assembly, *in* "Lambda II" (Hendrix, R.W., Roberts, J.W., Stahl, F.W. & Weisberg, R.A.)

(pp. 279-304). Cold Spring Harbor Laboratory, New York.

German, S., Candresse, T., Lanneau, M., Huet, J.C., Pernollet, J.C. & Dunez, J. (1990), Nucleotide sequence and genomic organisation of apple chlorotic leaf spot closterovirus. *Virology,* 179, 104-112.

Goldbach, R. (1987), Genome similarities between plant and animal RNA viruses. *Microbiol. Sci.,* 4, 197-202.

Goldbach, R. (1992), The recombinative nature of potyviruses: implications for setting up true phylogenetic taxonomy, *in* "Potyvirus Taxonomy" (Barnett, O.W.) (pp. 299-304). Springer-Verlag, Heidelberg, Berlin.

Goldbach, R. & Wellink, J. (1988), Evolution of plus-stranded RNA viruses. *Intervirology,* 29, 260-267.

Gorbalenya, A.E. & Koonin, E.V. (1989), Viral proteins containing the purine NTP-binding sequence pattern. *Nucl. Acids Res.,* 17, 8413-8440.

Guilley, H., Carrington, J.C., Balazs, E., Jonard, G., Richards, K. & Morris, T.J. (1985), Nucleotide sequence and genome organization of carnation mottle virus RNA. *Nucl. Acids Res.,* 13, 6663-6677.

Habili, N. & Symons, R.H. (1989), Evolutionary relationship between luteoviruses and other RNA plant viruses based on sequence motifs in their putative RNA polymerases and nucleic acid helicases. *Nucl. Acids Res.,* 17, 9543-9555.

Hamamatsu, C., Toriyama, S., Toyoda, T. & Ishihama, A. (1993), Ambisense coding strategy of the rice stripe virus genome: *in vitro* translation studies. *J. Gen. Virol.,* 74, 1125-1131.

Harrison, S.C. (1990), Principles of virus structure, *in* "Virology" (Fields, B.N., Knipe, D.M., Chanock, R.M., Hirsch, M.S., Melnick, J.L., Monath, T.P. & Roizman, B.) 2nd Edition (pp. 37-61). Raven Press, New York.

Haseloff, J., Goelet, P., Zimmern, D., Ahlquist, P., Dasgupta, R. & Kaesberg, P. (1984), Striking similarities in amino acid sequence among nonstructural proteins encoded by RNA viruses that have dissimilar genomic organisation. *Proc. Natl. Acad. Sci.* (Wash.), 81, 4358-4362.

Hearne, P.Q., Knorr, D.A., Hillman, B.I. & Morris, T.J. (1990), The complete genome structure and synthesis of infectious RNA from clones of tomato bushy stunt virus. *Virology,* 177, 141-151.

Howard, C.R. (1991), Hepadnaviridae, *in* "Classification and Nomenclature of Viruses: Fifth Report of the International Committee on Taxonomy of Virues" (Francki, R.I.B., Fauquet, C.M., Knudson, D.L. & Brown, F.) (pp. 111-116). Springer-Verlag, Heidelberg, Berlin.

Howarth, A.J. & Goodman, R.M. (1986), Divergence and evolution of geminivirus genomes. *J. Mol. Evol.,* 23, 313-319.

Hudson, P.J., McKern, N.M., Power, B.E. & Azad, A.A. (1986), Genomic structure of the large RNA segment of infectious bursal disease virus. *Nucl. Acids Res.,* 14, 5001-5012.

Huisman, M.J., Linthorst, H.J.M., Bol, J.F. & Cornelissen, B.J.C. (1988), The complete nucleotide sequence of potato virus X and its homologies at the amino acid level with various plus-stranded RNA viruses. *J. Gen. Virol.,* 69, 1789-1798.

Jagadish, M.N., Staton, V.J., Hudson, P.J. & Azad, A.A. (1988), Birnavirus precursor polyprotein is processed in *Escherichia coli* by its own virus-encoded polypeptide. *J. Virol.,* 62, 1084-1087.

Kakutani, T., Hayano, Y., Hayashi, T. & Minobe, Y. (1990), Ambisense segment 4 of rice stripe virus: possible evolutionary relationship with phleboviruses and uukuviruses *(Bunyaviridae). J. Gen. Virol.,* 71, 1427-1432.

Kashiwazaki, S., Minobe, Y., Minobe, T. & Hibino, H. (1990), Nucleotide sequence of barley yellow mosaic virus RNA 1: a close evolutionary relationship with potyviruses. *J. Gen. Virol.,* 71, 2781-2790.

Katsura, I. (1983), Tail assembly and injection, *in* "Lambda II" (Hendrix, R.W., Roberts, J.W., Stahl, F.W. & Weisberg, R.A.) (pp. 331-346). Cold Spring Harbor Laboratory, New York.

Keese, P., Mackenzie, A. & Gibbs, A. (1989), Nucleotide sequence of the genome of an Australian isolate of turnip yellow mosaic tymovirus. *Virology,* 172, 536-546.

Kingsbury, D.W. (1987), Nomenclature of plant viruses, *in* "Biological Nomenclature Today" (Ride, W.D.L. & Younes, T.) (pp. 49-57). IVBS Monograph Series 2. IRL Press, Arlington, VA.

Kingsman, A.J., Adams, S.E., Burns, N.R. & Kingsman, S.M. (1991), Retroelement particles as purification, presentation and targeting vehicles. *Trends in Biotech.,* 9, 303-309.

Klein, C., Fritsch, C., Briand, J.P., Richards, K.E., Jonard, G. & Hirth, L. (1976), Physical and functional heterogeneity in TYMV RNA: evidence for the existence of an independent messenger coding for the coat protein. *Nucl. Acids Res.,* 3, 3043-3061.

Koonin, E.V. (1991), The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J. Gen. Virol.,* 72, 2197-2206.

Koonin, E.V. (1992), Evolution of double-stranded RNA viruses: a case for polyphyletic origin from different groups of positive-stranded RNA viruses. *Sem. Virol.,* 3, 327-339.

Koonin, E.V., Choi, G.H., Nuss, D.L., Shapira, R. & Carrington, J.C. (1991), Evidence for common ancestry of a chestnut blight hypovirulence-associated double-stranded RNA and a group of positive strand RNA viruses. *Proc. Natl. Acad. Sci.* (Wash.), 88, 10647-10651.

Lawrence, M.C., Suzuki, E., Varghese, J.N., Davis, P.C., Van Donkelaar, A., Tulloch, P.A. & Colman, P.M. (1990), The three-dimensional structure of the seed storage protein phaseolin at 3A resolution. *EMBO J.,* 9, 9-15.

Lommel, S.A., Weston-Fina, M., Xiong, Z. & Lomonossoff, G.P. (1988), The nucleotide sequence and gene organisation of red clover necrotic mosaic virus RNA-2. *Nucl. Acids Res.,* 16, 8587-8602.

Lwoff, A. (1967), Principles of classification and nomenclature. *Nature* (Lond.), 215, 13-14.

Mans, R.M.W., Pleij, C.W.A. & Bosch, L. (1991), tRNA-like structures; structure, function and evolutionary significance. *Eur. J. Biochem.,* 201, 303-324.

Martelli, G.P. (1992), Classification and nomenclature of plant viruses: state of the art. *Plant Dis.,* 76, 436-442.

Martin, R.R., Keese, P.K., Young, M.J., Waterhouse, P.M. & Gerlach, W.L. (1990), Evolution and molecular biology of luteoviruses. *Ann. Rev. Phytopathol.,* 28, 341-363.

Matthews, R.E.F. (1983), The history of viral taxonomy, *in* "Critical appraisal of viral taxonomy" (Matthews, R.E.F.) (pp. 1-35). CRC Press, Boca Raton, FL.

Matthews, R.E.F. (1985a), Viral taxonomy for the non-virologist. *Ann. Rev. Microbiol.*, 39, 451-474.

Matthews, R.E.F. (1985b), Viral taxonomy. *Microbiol. Sci.*, 2, 74-75.

Matthews, R.E.F. (1991), Plant Virology 3rd Edition, Chapter 17 (pp. 635-682). Academic Press, New York.

Mayo, M.A., Reavy, B., Turnbull-Ross, A.D. & Murant, A.F. (1993), Sequence relationships among parsnip yellow fleck virus, rice tungro spherical virus and picorna-like viruses. IXth Intl Congr. Virol. Glasgow, Scotland, Abstract P76-9.

Memelink, J., van der Vlugt, C.I.M., Linthorst, H.J.M., Derks, A.F.L.M., Asjes, C.J. & Bol, J.F. (1990), Homologies between the genomes of a carlavirus (lily symptomless virus) and a potexvirus (lily virus X) from lily plants. *J. Gen. Virol.*, 71, 917-924.

Milinkovitch, M.C., Orti, G. & Meyer, A. (1993), Revised phylogeny of whales by mitochondrial ribosomal DNA sequences. *Nature* (Lond.), 361, 346-348.

Morgan, M.M., Macreadie, I.G., Harley, V.R., Hudson, P.J. & Azad, A.A. (1988), Sequence of the small double-stranded RNA genomic segment of infectious bursal disease virus and its deduced 90-kDa product. *Virology*, 163, 240-242.

Morse, M.A., Marriot, A.C. & Nuttall, P.A. (1992), The glycoprotein of thogoto virus (a tick-borne orthomyxo-like virus) is related to the baculovirus glycoprotein GP64. *Virology*, 186, 640-646.

Murphy, F.A. & Kingsbury, D.W. (1990), Virus Taxonomy, *in* "Virology" (Fields, B.N., Knipe, D.M., Chanock, R.M., Hirsch, M.S., Melnick, J.L., Monath, T.P. & Roizman, B.) 2nd Edition (pp. 9-35). Raven Press, Ltd., New York.

Nutter, R.C., Scheets, K., Panganiban, L.C. & Lommel, S.A. (1989), The complete nucleotide sequence of the maize chlorotic mottle virus genome. *Nucl. Acids Res.*, 17, 3163-3177.

Parge, H.E., McRee, D.E., Capozza, M.A., Bernstein, S.L., Getzoff, E.D. & Tainer, J.A. (1987), Three dimensional structure of bacterial pili. *Antonie van Leeuwenhoek*, 53, 447-453.

Prasad, B.V.V., Yamaghuchi, S. & Roy, P. (1992), Three-dimensional structure of single-shelled bluetongue virus. *J. Virol.*, 66, 2135-2142.

Pringle, C.R. (1990), The order mononegavirales — a proposal for introduction of the taxonomic category of the order into virus classification, *in* "Abstracts". VIIIth International Congress of Virology, Berlin. Abstract ICTV-10, p. 500.

Rabussay, D. & Geiduschek, E.P. (1977), Regulation of gene action in the development of lytic bacteriophages, *in* "Regulation and Genetics: Bacterial DNA Viruses" (Fraenkel-Conrat, H. & Wagner, R.R.) (pp. 1-196). Plenum Press, New York.

Rupasov, V.V., Morozov, S.Y., Kanyuka, K.V. & Zavriev, S.K. (1989), Partial nucleotide sequence of potato virus M RNA shows similarities to potexviruses in gene arrangement and the encoded amino acid sequences. *J. Gen. Virol.*, 70, 1861-1869.

Rybicki, E. (1990), The classification of organisms at the edge of life or problems with virus systematics. *South African J. Sci.*, 86, 182-186.

Shukla, D.D. & Ward, C.W. (1989), Structure of potyvirus coat proteins and its application in the taxonomy of the potyvirus group. *Adv. Virus Res.*, 36, 273-314.

Stackebrandt, E. (1994), Origin and evolution of prokaryotes, *in* "Molecular basis of virus evolution" (Gibbs, A.J., Calisher, C.H. & Garcia-Arenal, F.). Cambridge University Press, Cambridge (in press).

Stanley, J., Markham, P.G., Callis, R.J. & Pinner, M.S. (1986), The nucleotide sequence of an infectious clone of the geminivirus beet curley top virus. *EMBO J.*, 5, 1761-1767.

Strauss, J.H. (1991), Togaviridae, *in* "Classification and Nomenclature of Viruses: Fifth Report of the International Committee on Taxonomy of Viruses" (Francki, R.I.B., Fauquet, C.M., Knudson, D.L. & Brown, F.) (pp. 216-222). Springer-Verlag, Heidelberg, Berlin.

Strauss, J.H. & Strauss, E.G. (1988), Evolution of RNA viruses. *Ann. Rev. Microbiol.*, 42, 657-683.

Strauss, E.G., Strauss, J.H. & Levine, A.J. (1990), Virus evolution, *in* "Virology" (Fields, B.N., Knipe, D.M., Chanock, R.M., Hirsch, M.S., Melnick, J.L., Monath, T.P. & Roizman, B.) 2nd Edition, Chapter 9 (pp. 167-190). Raven Press, Ltd., New York.

Takahashi, M., Toriyama, S., Hamamatsu, C. & Ishihama, A. (1993), Nucleotide sequence and possible ambisense coding strategy of rice stripe virus RNA segment 2. *J. Gen. Virol.*, 74, 769-773.

Tordo, N., De Haan, P., Goldbach, R. & Poch, O. (1992), Evolution of negative-stranded RNA genomes. *Sem. Virol.*, 3, 341-357.

Uhlin, B.E., Baga, M., Goransson, M., Lindbereg, F.P., Lund, B., Norgren, M. & Normark, S. (1985), Genes determining adhesin formation in uropathogeneic *Escherichia coli*. *Curr. Topics Microbiol. Immunol.*, 118, 163-178.

Van Regenmortel, M.H.V. (1989), Applying the species concept to plant viruses. *Arch. Virol.*, 104, 1-17.

Van Vloten-Doting, L., Francki, R.I.B., Fulton, R.W., Kaper, J.M. & Lane, L.C. (1981), *Tricornaviridae:* a proposed family of plant viruses with tripartite, single-stranded RNA genomes. *Intervirology*, 15, 198-203.

Ward, C.W. (1981), Structure of the influenza virus hemagglutinin. *Curr. Topics Microbiol. Immunol.*, 94/95, 1-74.

Ward, C.W. & Shukla, D.D. (1991), Taxonomy of potyviruses: current problems and some solutions. *Intervirology*, 32, 269-296.

Ward, C.W., McKern, N.M., Frenkel, M.J. & Shukla, D.D. (1992), Sequence data as the major criterion for potyvirus classification, *in* "Potyvirus Taxonomy" (Barnett, O.W.) (pp. 283-297). Springer-Verlag, Heidelberg, Berlin.

Ward, C.W., Weiller, G., Shukla, D.D. & Gibbs, A.J. (1994), Molecular evolution of potyviruses, the largest plant virus family, *in* "Molecular basis of viral evolution" (Gibbs, A.J., Calisher, C.H. & Garcia-Arenal, F.). Cambridge University Press, Cambridge (in press).

Weiner, A.M. & Maizels, N. (1987), tRNA-like structures tag the 3' ends of genomic RNA molecules for replication: implications for the origin of protein synthesis. *Proc. Natl. Acad. Sci.* (Wash.), 84, 7383-7387.

White, M. & Gribbin, J. (1992), "Stephen Hawking: A

Life In Science", Penguin Books, London, pp. 102-103.

Wu, S., Rinehart, C.A. & Kaesberg, P. (1987), Sequence and organisation of southern bean mosaic virus genomic RNA. *Virology,* 161, 73-80.

Xiong, Z. & Lommel, S.A. (1989), The complete nucleotide sequence and genome organisation of red clover necrotic mosaic virus RNA-1. *Virology,* 171, 543-554.

Zavriev, S.K., Kanyuka, K.V. & Levay, K.E. (1991), The genome organisation of potato virus M RNA. *J. Gen. Virol.,* 72, 9-14.

Zhu, Y., Hayakawa, T., Toriyama, S. & Takahashi, M. (1991), Complete nucleotide sequence of RNA 3 of rice stripe virus: an ambisense coding strategy. *J. Gen. Virol.,* 72, 763-767.

Zimmern, D. (1987), Evolution of RNA viruses, *in* "RNA genetics" (Holland, J., Domingo, E. & Ahlquist, P.) (pp. 211-240). CRC Press, Boca Raton, FL.