Arch Virol 142/7 (1997)

VDN Virology Division News

Guidelines to the demarcation of virus species

M. H. V. Van Regenmortel[1], D. H. L. Bishop[2], C. M. Fauquet[3], M. A. Mayo[4], J. Maniloff[5], and C. H. Calisher[6]

[1] UPR 9021-Immunochimie des Peptides et des Virus, Institut de Biologie Moléculaire et Cellulaire du CNRS, Strasbourg, France
[2] Oxford University, St. Cross College, St. Giles, Oxford, U.K.
[3] ORSTOM/ILTAB, The Scripps Research Institute, Division of Plant Biology, La Jolla, California, U.S.A.
[4] Scottish Crop Research Institute, Invergowrie, Dundee, U.K.
[5] Department of Microbiology and Immunology, University of Rochester, Rochester, New York, U.S.A.
[6] Arthropod Borne & Infectious Disease Laboratory, Department of Microbiology, Colorado State University, Ft. Collins, Colorado, U.S.A.

1. Introduction

The practical need to partition the world of the viruses into distinguishable, universally agreed upon, entities is the ultimate justification for developing a classification system. These entities correspond to the individual viruses that are given the status of virus species. It is the purpose of this paper to review both the theoretical concepts and the practical options for defining virus species.

Species is the universally accepted term for the lowest taxonomic clustering of living organisms. Species of animals and plants are usually defined in terms of the biological species concept; this concept is based upon gene pools and reproductive isolation, which of course are features relevant only to sexually reproducing organisms [18]. For many years, virologists were reluctant to apply the species concept, arguing that entities such as viruses which reproduce by clonal means could not be accommodated within the classical definition of biological species [13, 20]. However, other species concepts have been developed which are applicable to asexual organisms [2, 15, 16, 18] and it is now generally accepted that the species concept is applicable in virology – after all, the viruses have genomes, replicate, evolve and occupy particular ecological niches [30, 31]. The International Committee on Taxonomy of Viruses (ICTV) agreed in 1991 that the hierarchical level of species would be defined and added to the categories of genus, subfamily, family and order which were already in use in the universal virus classification system [24, 32].

2. Definition of virus species

The following definition of virus species has been endorsed by the ICTV:

"A virus species is a polythetic class of viruses that constitutes a replicating lineage and occupies a particular ecological niche" [31].

Species as a polythetic class

Inherent in the definition of virus species as a polythetic class is the point that there can be no single attribute required for a virus to qualify as a member of a particular species. A polythetic class consists of members that have a number of properties in common, but in general, there is no single property which is necessarily shared by all the members and which, because it is absent in other species, can be used as a defining property of a particular species. A single discriminating character, for instance a particular host reaction or a certain percentage of genome sequence identity, is usually unreliable by itself. For example, plant viruses that are transmitted by mechanical inoculation in the greenhouse for many years tend to lose their ability to be transmitted by their usual insect vectors, but this kind of adaptation involving a single property is devoid of taxonomic significance.

The use of a single discriminating character for distinguishing species contradicts the inherent variability of the members of a species and the fact that species are not so-called universal classes definable by a single property. The situation is different with higher taxa such as virus families or genera which are universal classes and consist of members which all share one or more defining properties that are both necessary and sufficient for class membership [32]. In the case of virus species, it is always a combination of properties that provides the rationale for deciding whether a particular virus should be considered a member of the species. ICTV Study Groups, with their in depth knowledge of particular virus genera and families, serve as judges of such issues for the international virology community [21]. Only specialists are aware of the facts, issues and nuances about viruses and their biotic interactions and of the importance of making certain distinctions, that may not be the same in all virus genera, for achieving a convenient and practical classification.

Species as a replicating lineage

The second element in the definition acknowledges that a virus species represents a replicating lineage and that taxonomic distinctions should reflect genealogy and evolution. All members of a virus species share descent from a common ancestor and any classification scheme should be consistent with this genealogical principle. In spite of being a necessary prerequisite, phylogeny is obviously not a sufficient criterion for demarcation between species since shared descent is a common feature linking all members of groups of biological entities, including groups of different species or genera.

Species undergo continuous variation in time and transition from one species to another during evolution occurs within the continuity of replication. As variations accumulate, a point will be reached where the importance of genotypic and phenotypic differences would lead an observer to conclude that he is dealing with a separate entity. In general while no particular value of genome sequence dissimilarity can be used as a cut-off point to differentiate between two species within a single genus, sequence data do provide a good starting point, as they indicate the extent to which the viruses in question have diverged over time. The genotypic variations inherent in any replicating lineage are expressed in continual phenotypic variation such that no property can be used as a single criterion for a practical and meaningful demarcation between viral species. Use of a single criterion, such as the potential for genetic reassortment or the degree of sequence difference, takes no regard, for example, of biotic differences. The need to record phylogeny should not overshadow the importance of phenotypic and biotic distinctions which, along with genotypic distinctions,

are the ultimate justification for species demarcation. Further, classifying viral genomes should not be confused with classifying viruses. Although correlation between the results of the two activities may be high in some cases, (e.g. picornaviruses, geminiviruses and potyviruses), genome comparisons expressed as the degree of similarity (homology analysis) cannot by themselves amount to formal taxonomic species demarcation.

Ecological niche occupancy

The third element in the virus species definition, namely ecological niche occupancy, refers to the biotic properties of the species, which include its geographic localization, host range, vector tropism, pathogenesis and other host responses. The ecological niche is an attribute of a virus species related to its biotic habitat. Virus ecology is the study of viruses in relation to their biotic niche and how they interact with their hosts and surroundings. The concept of ecological niche encompasses environmental, biotic and functional aspects. The biotic component of the niche concept includes the host and tissues where a virus is found while the functional component refers to the replication process and to the multiple interactions with host and vector.

The concept of niche in taxonomy has sometimes been denigrated because it has been argued that the notion of a vacant niche is meaningless. However, since a biotic niche does not simply refer to a location in three dimensional space but is an attribute of a species, vacancy is not a problem [6].

Taxonomic species are not molecular quasi-species

In discussions of virus phylogeny, it is common to refer to RNA viruses as quasi-species populations. Since RNA viruses have genomes that replicate in the absence of repair mechanisms, they evolve very rapidly with a mutation frequency per nucleotide site in the genome of 10^{-3} to 10^{-5}. A clone of an RNA virus will therefore always generate many thousands of different genomes all of which compete during replication of the clone [14]. Such a population which consists of a master sequence corresponding to the most fit genome sequence with respect to a given environment together with countless competing virus mutants, is labelled a quasi-species population or swarm. The term quasi-species was introduced by Eigen to describe the self-replicating RNAs believed to be the first genes on earth (for a review, see [9]). It should be stressed that the term quasi-species is used to convey the notion that the viral genome is not a unique molecular species and that a virus cannot be defined by a single genome sequence. In this context the terms species and quasi-species refer respectively to homogeneous and heterogeneous molecular species i.e. purely chemical entities and not to the taxonomic concept of virus species as a biological entity which is the subject of the present paper. When a virus population, because of chemical heterogeneity, is referred to as a quasi-species, i.e. implying some sort of imperfect species, this does not mean that it is possible to find "true" virus species that would possess a single, invariant genome sequence. Whereas all the members of a chemical species are identical molecules, the members of a virus species are not. It has been proposed that virus species could be considered as "an ensemble that occupies a coherent part of the sequence space which is continuously populated for prolonged periods of time and under a wide variety of environmental conditions" [7]. However, such a definition reduces viruses to genome sequences and ignores the phenotypic characteristics which are the reason why viruses are

being classified in the first instance. Furthermore, focusing on sequences that correspond to heterogeneous chemical populations does not help to resolve the inherent fuzziness of all species concepts [19, 33]. Virus species are fuzzy sets with hazy boundaries and it is counter-productive to try to make absolutely clear distinctions where none exist.

3. Species demarcation involves a process of identification based on diagnostic properties

Although the acceptance of a definition of viral species by the ICTV was an important step in establishing a unified virus classification system, it should be emphasized that this definition cannot be used for deciding if a particular isolate is a member of a certain virus species or not. Definitions apply only to abstract concepts such as the notion of species viewed as a class. Individual viruses can be identified and named [17] but cannot be "defined". In an analogous manner, the concept of a human family can be defined in terms of a lineage comprising parents, grandparents, children, siblings, etc. but such a definition would be of little use for identifying the members of an individual family who have gathered for the annual school concert. Identification of real entities relies on the use of so-called diagnostic properties, i.e. phenotypic and genotypic characters that make it possible for instance to discriminate between members of different species or different genera. The theoretical, defining properties of abstract classes are not helpful for recognizing individual species [12]. In order to divide the world of viruses into separate species that fulfill the requirement for a practical classification, it is necessary to reach an agreement about which diagnostic properties will be most useful for identifying the individual members of a species. For different virus species, different properties have to be used to take into account the variations that exist.

One of the principal aims of a classification is to provide a scheme whereby new virus isolates can be identified. Such identification is a comparative process by which new isolates are examined and compared with established virus species. Since species are polythetic, comparison should involve a number of characters to assess the relationship of a new isolate with established species rather than the presence or absence of a single key feature. Whereas the construction of a classification scheme necessarily entails the use of several characters for demarcating individual species, the identification of a virus isolate as a member of an established species may often be achieved by considering only a few characters. Except where a single virus species exists in a taxon (e.g. African swine fever virus), it is essential not to use for species diagnosis characters that are present in all the members of a genus or family, since these obviously will not permit species demarcation within the group. In general, characters such as morphology, genome organization, method of replication and the number and size of structural and non-structural proteins are likely to be family- or genus-defining properties and therefore of little value for demarcating individual species.

The characters that may be useful in discriminating between virus species allocated to a particular genus are:

- genome sequence relatedness
- natural host range
- cell and tissue tropism

- pathogenicity and cytopathology
- mode of transmission
- physicochemical properties
- antigenic properties

Other characters have been listed by Murphy et al. [21] and Murphy [22]. The relative importance of these characters for species demarcation may vary in different genera and families.

In each genus recognized by the ICTV, one species is designated as the type species. This type species is usually one for which considerable knowledge is available and it often corresponds to a nominal species in the sense that it is the name-bearing type of a nominal genus (for instance, simian *rotavirus* SA11 is the type species of the genus *Rotavirus*, family *Reoviridae*). It should be stressed that the type species is not, and never could be one which is most typical and representative of the properties of all species in a genus.

4. Examples of species discrimination in some virus families which illustrate the diversity of the problems and some solutions

Examples of criteria that have proved useful for species demarcation are afforded by consideration of viruses assigned to the families *Bunyaviridae*, *Herpesviridae*, *Potyviridae* and *Geminiviridae*.

The family Bunyaviridae

All the viruses assigned to this family have the characteristics of a segmented, single-stranded RNA genome with, depending on the genus, a negative or ambisense coding arrangement and common transcription, replication and morphogenetic strategies – albeit distinguishable between the genera. In virus particles the RNA species are each associated with nucleocapsid protein and an RNA-directed RNA polymerase. The nucleocapsid ensemblage is bounded by a lipid membrane that incorporates the viral surface glycoproteins. Packaging of the viral nucleocapsids is imprecise and frequently virions are formed that are diploid with respect to 1 or more RNA species (usually the smallest species). Apart from the 3 primary gene products that form the structural proteins, there may be 1 or more non-structural proteins, depending on the genus.

Since the genome is segmented, RNA segment reassortment during virus co-infections provides an opportunity for the formation of recombinant viruses (depending on the virus) and hence the designation of the parent viruses as potential contributors to a virus gene pool. Recombinant (reassortant) viruses have been identified among field isolates and in laboratory experiments. The evidence to-date supports the view that there are a number of distinct Bunyavirus gene pools within a genus.

As in other systems, for members of the *Bunyaviridae* the biological property of genetic recombination provides a means of species demarcation. Correspondingly, a number of correlates may be predicted. These include the compatibility of virus proteins with respect to virus replication, transcription and morphogenetic processes. Such compatibilities will be reflected in certain antigenic and sequence homologies between comparable proteins, although the extent that particular sequences, protein motifs and antigenic properties (e.g., those that elicit cell-mediated, humoral, or secretory immune responses) are

conserved will vary depending on the antigen involved and its functions, and for different members of a species (e.g., in relation to their evolution and divergence). Depending on the protein, less sequence conservation may be expected between members representing different virus species.

The data available for viruses that are known to contribute to gene pools among the *Bunyaviridae* bear out these predictions. Such data indicate that overall a number of epitopes and protein homologies reflect the identified genetic compatibilities and species assignments. Cross-neutralisation, cross-haemagglutinin-inhibition, and N protein sequence relationships are specific examples.

Five genera are recognised in the family *Bunyaviridae* (*Bunyavirus*, *Hantavirus*, *Nairovirus*, *Phlebovirus* and *Tospovirus*). As noted above, details of the morphological properties and coding strategies vary among the genera. Generally, virions are variably spherical, some 80–120 nm in diameter with 5–10 nm long surface peplomers. Viral nucleocapsids display helical symmetry and have been observed to exist in circular configurations. The viral RNA species can be recovered from nucleocapsids/virions in the form of non-covalently closed circles. Genome sizes are 11–20 Kb, depending on the genus. Bunyaviruses, phleboviruses and nairoviruses replicate in both vertebrate and arthropod vectors. Tospoviruses replicate in plants and are transmitted between plants by certain plant-feeding insects. Hantaviruses have no known arthropod vector, replicating only in certain vertebrate species.

As of 1996, some 172 serologically distinct viruses (serotypes) were assigned in the genus *Bunyavirus*. Using a variety of serological tests, 168 of these were assigned to 18 different serogroups (Anopheles A, Anopheles B, Bakau, Bunyamwera, Group C, California, Capim, Gamboa, Guama, Koongol, Mintillan, Nyando, Olifantsvlei, Patois, Simbu, Tete and Turlock). Four viruses remain unassigned. For new virus isolates and where little or no serological difference is demonstrable with a known virus serotype, such viruses are considered alternative isolates, or variants, or varieties of the known virus serotype. Apart from the conserved coding, morphogenetic and replicative strategies, the viruses classified to different *Bunyavirus* serogroups have been shown to be related to members of other serogroups of the genus by complement fixation tests, and distinguished, inter se, by cross-neutralisation and haemagglutination-inhibition assays.

Ten other *Bunyaviridae* viruses have been placed in the genus *Hantavirus*, 34 in the genus *Nairovirus*, 51 in the genus *Phlebovirus*, and 4 in the genus *Tospovirus*. A further 57 viruses are considered members of the family, but have not been assigned to a recognised genus. Of these, 34 have been placed into 7 serogroups with the remaining 23 viruses ungrouped [10].

Historically, virus assignment to the *Bunyaviridae* was the result of comprehensive serological analyses using polyclonal sera to differentiate viruses (tospoviruses are an exception). When shown by such tests to be significantly different from other viruses (serotypes), the viruses were named according to the location where an isolate was recovered (e.g., Tahyna virus), or a known host (snowshoe hare virus), or a common vector (Trivittatus virus), or a particular disease (Rift valley fever virus). From multiple isolations of a virus (in particular for viruses associated with human/animal disease, and where intensive studies of insect or vertebrate fauna or diseased plants were undertaken) it has been observed that similar viruses can be recovered from other places, other hosts, other vectors, sometimes causing different diseases. An example of the latter is Oriboca virus

(Group C) for which isolates are known that cause virulent hepatitis in rodents while other isolates cause only a delayed onset neurological disease. In most cases, however, there are insufficent data to know the extent of virus distribution, or the biological variation that may exist between isolates. Thus defining distribution, hosts, or vector properties, is generally unreliable in terms of species demarcation for most bunyaviruses.

As noted above, certain serological tests (e.g., for members of *Bunyavirus* genus cross-neutralisation, cross-haemagglutination-inhibition) have allowed different viruses (serotypes) to be grouped into so-called serogroups, and distinguished from other serogroups by the lack of such cross-reactivities. Thus Tahyna, snowshoe hare and Trivittatus viruses have been grouped with some 12 other viruses into the California serogroup with California encephalitis virus (the original virus isolate) considered as the type member. Confirmation that many members of the California serogroup constitute a single virus species has come from reassortment analyses. Reassortment has been demonstrated in laboratory tests between California encephalitis, La Crosse, snowshoe hare, Tahyna, Lumbo and Trivittatus viruses (data for other members of the California serogroup are not available and strictly, therefore, these viruses are unassigned with respect to California encephalitis virus species and its gene pool).

To-date, no reassortment has been demonstrated between California group members and Bunyamwera, Group C and Simbu serogroup viruses (see [10] and D. H. L. Bishop, unpublished data). However, reassortment studies undertaken with Bunyamwera virus and certain other members of the Bunyamwera serogroup indicate that they constitute a gene pool, distinct from other members (gene pools) of the same serogroup, or the California encephalitis virus species referred to above.

In relation to protein sequence relationships, among the California group viruses (for which most data exist), the nucleocapsid proteins exhibit > 60% identity. By comparison, there is < 40% sequence identity between California group viruses and members of the Bunyamwera and Simbu groups, and none identifiable with members of the other genera of the family. For the viral glycoproteins, and again depending on the virus, very little identity or homology is recognised even among viruses that contribute to the same gene pool although certain features are conserved (e.g., the relative positions of cysteines and other residues and motifs).

In summary, among the *Bunyavirus* genus of the *Bunyaviridae*, host and vector specificities, geographic distribution and pathogenic phenotypes may not always be useful in terms of species demarcation either due to the lack of sufficient data or the lack of uniqueness (the same host may vector different virus species). However, certain serological properties (cross-neutralisation, cross-haemagglutination inhibition) and sequence relationships appear to correlate well with genetic data (reassortment and virus interference studies) and allow virus isolates and serotypes to be classified together as a species.

There is an obvious value and need to recognise viruses as species not only as taxonomic entities but also in biological terms, for example in relation to recombination potential (virus interference), and to differentiate viruses from other virus species. Apart from taxonomic and evolutionary issues, species designation can provide much more information, for example, in relation to the biological properties of the component viruses. The problem that may be voiced is that in many cases insufficient data are available to make species assignments. In view of the added value that species designation provides this is not a reason to resist making species assignments and tentative classifications when the data are

incomplete. For example, for those bunyaviruses for which only serological data exist and using the California encephalitis and Bunyamwera virus species as examples, they can be classified as tentative species (or possible members of a species) depending on the information available and until further data are obtained.

The family Herpesviridae

The family *Herpesviridae* comprises hundreds of viruses infecting virtually all vertebrates. At present the family contains three subfamilies which in turn contain numerous genera. Virions range from 150 to 200 nm in diameter; they are enveloped, are covered by surface projections, contain a tegument beneath their envelope, and contain a characteristic 100-110 nm icosahedral capsid. Their genome is composed of linear, double stranded DNA, ranging from 124 to 235 Kbp in size. Many herpesvirus genomes contain internal repeats of one or both terminal sequences which cause the sequences flanked by the repeats to invert relative to the remainder of the genome and therefore result in the formation of 2 or 4 isomeric genome forms (presently designated by letters A–F).

The division of the family *Herpesviridae* into three subfamilies was done on the basis of biological properties. The subfamily *Alphaherpesvirinae* comprises the viruses that resemble human herpesvirus 1 (herpes simplex virus 1); the subfamily *Betaherpesvirinae* comprises the cytomegaloviruses and the subfamily *Gammaherpesvirinae* comprises the viruses that resemble human Epstein-Barr virus.

A relatively small number of herpesviruses have been placed into genera based on DNA sequence homology, similarities in genome structure (e.g., presence, nature and placement of terminal repeats), and relatedness of important viral proteins (demonstrable by immunologic methods). While a few genes (e.g., homologs of glycoprotein B or H of HHV-1) are conserved among some members of different subfamilies, nucleic acid and protein sequence homologies do serve to distinguish closely related viruses and are therefore useful in placing viruses in genera.

Species status has been assigned to each of the well known herpesviruses of humans and domestic animals, and to a few well known viruses of other hosts. The criteria used for such preliminary speciation have started with the host and have been influenced by the characters noted above as defining the subfamilies. Next, traditionally, identification of isolates has depended upon serological assays. Presently, more and more partial sequencing is being done and is proving to be a powerful complement to serological methods in discriminating species. According to the present ICTV nomenclature, all herpesviruses are designated by the family name (in most cases) or subfamily name (for primates and domestic animals) of the natural host of the virus followed by a serial arabic number (e.g., human herpesvirus 6, circopithecine herpesvirus 1). Clearly, speciation is the dominant taxonomic issue facing herpesvirologists at this point in time. The problem is particularly acute in the case of viruses that share considerable DNA homology and some antigenic sites but that are readily differentiated by unambiguous tests and differ in biologic properties [25].

The families Potyviridae and Geminiviridae

For a number of years, and for many plant viruses, the main characters used to distinguish between different viruses (sensu species) have been combinations of biological features such as transmission characteristics and host range, physico-chemical features of virus

particles, such as shape, and the serological cross-reactivity of antibody preparations made against purified virus particles. As the concept of clustering viruses in groups, and then higher taxa, has been developed and applied to plant viruses, discrimination of viruses using some criteria (e.g. particle morphology) has been thought to be indicative of a separation into distinct higher taxa (genera or families). Other criteria, for example serological relatedness, have been used to discriminate between individual viruses; normally a serological differentiation index of about 2 or less was deemed strongly suggestive that two virus isolates were related as strains of a single virus [29]. However, not all groups of viruses have been amenable to non-controversial clustering into species or strains of particular species.

Two examples are discussed below which illustrate the difficulties and their resolution by applying a polythetic approach. The best discussed example is that of viruses which are now classified in the family *Potyviridae*. There are several hundreds of potyvirus diseases recorded so far. The viruses are characterized by having filamentous particles containing genomic positive-sense linear ssRNA which is expressed as a polyprotein. The family contains three genera which are distinguished on the basis of the genome organization and the type of vector responsible for transmission; genus *Bymovirus* contains viruses with bipartite genomes which are transmitted by soil-inhabiting plasmodiophoromycete fungi, genus *Rymovirus* contains viruses with monopartite genomes which are transmitted by mites; genus *Potyvirus* contains viruses also with monopartite genomes but which are transmitted by aphids. The family also contains a number of viruses which are transmitted by whiteflies or aphids and which are currently unassigned [21] but which might form new genera. The genera *Bymovirus* and *Rymovirus* each contain 5 definitive species and there are few problems in discriminating among species. In contrast, the genus *Potyvirus* contains many viruses; there are already more than 75 recognized species and another 93 tentative species are listed in the most recent ICTV Report [21]. Species demarcation among these viruses has always been problematic.

Traditionally, biological criteria like seed transmission, cross-protection, aphid vector specificity, host range and symptomatology were used to classify potyviruses [4, 27]. Particular criteria, such as the morphology of the cytoplasmic inclusion bodies formed in infected cells, can be applied to species clustering within a genus and to particular discrimination problems between or even within a species [8]. The most frequently used criterion for discrimination among similar plant viruses has been serological relatedness, but this proved also to be of only limited use [3, 26]. Sequence analysis has shed considerable light on this problem. *Potyvirus* coat proteins were found to consist of a conserved core sequence, which elicits highly cross-reactive antibodies, and a relatively immunogenic N-terminal region that varies among (most) viruses and which elicits potentially discriminatory antibodies. A complication is that the N-terminal sequence is readily lost when sap extracts are prepared but the virus particles remain intact. Nevertheless, when antibodies are appropriately raised, and when serological tests are well conducted, the results do help to differentiate between species of potyviruses.

It was only with the application of molecular analysis methods to reveal genome organization structures, and above all sequences, that demarcation of potyvirus species and the definition of the family structure could be addressed with a high degree of confidence [35, 36]. There are now complete sequences of the genomes of 34 potyviruses and many more partial sequences; for example there are more than 219 potyvirus coat protein sequences available in the database. Extensive sequence comparisons among these se-

quences have shown that the application of such quantitative taxonomy to all members of the family *Potyviridae* results in strong evidence for distinctions between each of the different taxonomic levels; strains, species, and genera [26]. Figure 1 shows the discontinuous distribution of pairwise similarities between coat protein sequences of different potyviruses. The "minima", or gaps between the peaks, provide the necessary "cut-off" values that allow the criterion to be used. Further analysis (Aleman and Fauquet, unpublished observations) has shown that this discriminatory structure is conserved all along the potyvirus genome, in the highly conserved and variable regions, as well as in the coding and non-coding regions. The addition of this criterion to the list of discriminatory characters (Table 1) has recently led to proposals to upgrade some strains to the species level and to downgrade some species to the strain level as well as to reorganize the family with the creation of two new genera [1, 5].

Other characters can be deduced from sequence data. The expression of potyvirus genome RNA results in the synthesis of a polyprotein which is cleaved at specific sites by virus-coded protease(s) to form mature virus proteins. Comparison between the amino acid sequences at the cleavage sites in the polyproteins of two viruses can thus be a sensitive measure of similarity between the respective proteases.

The current discriminatory characters, summarized as those characters which would be taken to indicate that two species are distinct are shown in Table 1. It is clear that no one criterion has an absolute supremacy over others, some are more informative and discriminate better or they are easier to acquire, but it is the sum of the information accumulated that has built up a clear and generally accepted taxonomy for potyviruses [21].

The other example is the family *Geminiviridae* which contains many viruses, a number of which cause economically important diseases in various parts of the world. In particular, whitefly- transmitted geminiviruses are currently spreading into new parts of the world and are causing devastating diseases. Geminiviruses have single stranded circular DNA genomes and capsids which have characteristic shapes, often of twin icosahedra. There are about

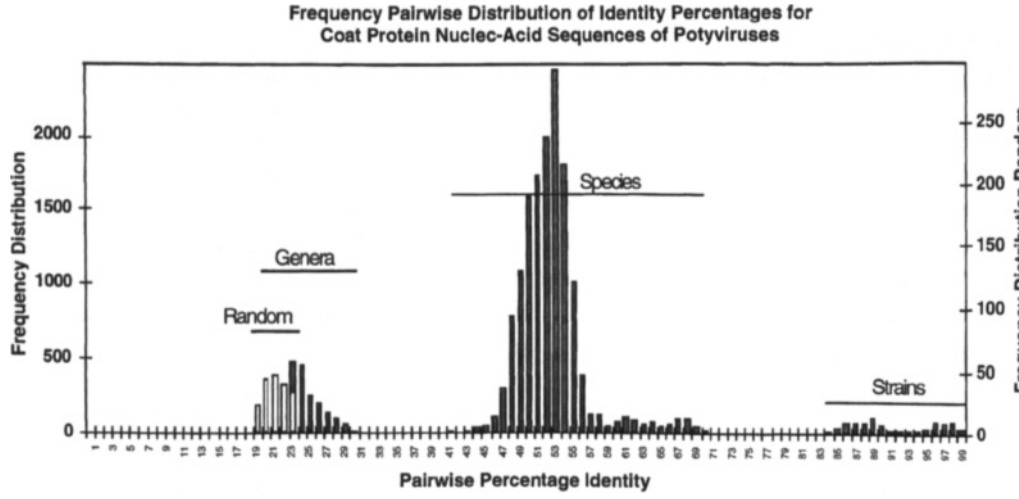

Fig. 1. Frequency distribution of pairwise sequence comparisons of 219 coat protein nucleic acid sequences of members of the family *Potyviridae*

Table 1. Characters which demarcate virus isolates as distinct species in the families *Potyviridae* and *Geminiviridae*

Character	*Potyviridae*	*Geminiviridae*
Genome features	–	Different numbers of genome components
	–	Different organization of genes in the genome
	–	No transcomplementation of gene products
	–	No pseudorecombination between components
Genome sequence	< circa 85% identical over whole sequence	< circa 90% identical over genome component A
	< circa 75% identical in 3' NCR	–
Protein features	Different polyprotein cleavage sites	–
	Virions react differently with key antibodies	Virions react differently with key antibodies
	< circa 90% identical in coat protein sequence	< circa 90% identical in coat protein sequence
Transmission	Different vector species	Different vector species
	Different seed transmissibility	–
Effects in infected tissue	Different inclusion body morphology	–
	No cross protection effects	–
	–	Different tissue tropism
Host range	Different in key species	Different in key species

118 strains and species of geminiviruses adequately described and there are certainly a great many poorly described and as yet undiscovered geminiviruses. The family is currently divided into three genera according to transmission vector and genome organization [21].

- Subgroup I geminivirus
  Members have monopartite genomes, are transmitted by leafhoppers and mostly infect monocotyledonous plants.
- Subgroup II geminivirus
  Members also have monopartite genomes but are transmitted by leafhoppers or treehoppers and infect only dicotyledonous plants.
- Subgroup III geminivirus
  Members are transmitted by whiteflies, infect only dicotyledonous plants and have either monopartite or bipartite genomes.

There are 82 viruses in this last genus and, as some viruses can cause similar symptoms on the same plants, their demarcation into strains and species is problematic. At least 8 different virus species are thought to cause tomato yellow leaf curl or tomato leaf curl diseases in the world.

Few geminiviruses can be transmitted mechanically and all are very difficult to purify. As a result, all the classical biological techniques have been very difficult to implement with these viruses. Serology was not useful until monoclonal antibodies (Mabs) were available, but since these reagents are specific for a single epitope, it is necessary to use a fairly large panel of different Mabs to obtain a reliable picture of the overall degree of antigenic

similarity between two viruses [28]. Depending on the particular epitope that is recognized by an individual antibody, some Mabs will emphasize what is common between two viruses while others will emphasize what is different between them [34].

Geminiviruses induce several different types of symptoms like yellow mosaic, leaf curling, stunting and enations. All are phenotypic effects of tissue-specific expression of virus replication, and although not very well documented in a precise biological manner, such symptomatology can be a useful criterion to distinguish between certain species.

As for potyviruses, it is the application of molecular biology that has improved greatly geminivirus taxonomy and our capacity to distinguish geminivirus species. It is obvious that genome organization, starting with the number of genomic components, is a major criterion for species identification. Members of the same genus may have one or two genomic components but it is a species characteristic to possess a mono or bipartite genome. It has also been shown that gene trans-complementation as well as pseudo-recombination by exchanging genomic components is only possible between individuals of the same species [11]. Full genome sequences are known for more than 60 geminiviruses and comparisons among them suggest that quantitative taxonomy can be applied to geminiviruses to determine if individuals are likely to be strains, or belong to distinct species or genera. It appears that the impact of molecular evolution has been homogenous on the entire genome of geminiviruses and therefore the sequence of a particular gene, such as the coat protein, can be used as a discriminatory criterion [23]. The comparisons also suggest that although recombination between species has occurred, at least for whitefly-transmitted viruses, the extent does not invalidate the use of sequence comparisons as taxonomic criterion.

Table 1 lists characters that can be used when deciding if two virus isolates in either the *Potyviridae* or *Geminiviridae* are different species or not. Some criteria are the same for the two families, others are qualitatively the same but quantitatively different, and some criteria have not been, or cannot be, applied to viruses in both families. This is likely to be so for many families. However, even such apparently "hard" characters as percentage sequence identity should be treated with care. If evolutionary constraints on a population of variants of one species in one family is much greater than those on a species of another family, the "cut-off" values (see Fig. 1) for practical discrimination will differ, although they may be equally valid if a discontinuous distribution of pairwise similarities is also present.

In recent discussions among virologists working with luteoviruses (C. J. D'Arcy, personal communication) a similar list of criteria was compiled which differed in the detail and relative importance given to the criteria, but which was very much the same in principle. It is likely that these lists are representative of the approaches likely to be used when discriminating among species in any of the currently recognized genera of plant viruses. It is the task of Study Groups of the Plant Virus Subcommittee considering individual families or clusters of genera, to draw up similar lists for their particular viruses.

5. Conclusion

The major problem facing taxonomic virology in the next few years is the demarcation of many additional virus species. The present situation recorded in the Sixth Report of the ICTV [21] is somewhat unsatisfactory since in many cases only the type species of many viral genera have so far been included. A major task for the ICTV Study Groups in the

coming years will be to devise operational procedures for classifying at the species level the many entities within established genera that practicing virologists consider to be separate viruses. Species demarcation can be achieved only by considering a number of properties that are not shared by all the members of a genus. This is an extremely challenging task that needs the input of many virologists world-wide. The development of virology as a mature discipline that is also accessible to non-specialists requires that its practitioners organize the bewildering variety of viral entities into a coherent scheme of individual species.

References

1. Badge J, Robinson DJ, Brunt AA, Foster GD (1997) 3'-terminal sequences of the RNA genomes of narcissus latent and Maclura mosaic viruses suggests that they represent a new genus of the Potyviridae. J Gen Virol 78: 253–257
2. Bishop DHL (1985) The genetic basis for describing viruses as species. Intervirology 24: 79–93
3. Bos L (1992) Potyviruses, chaos or order. In: Barnett OW (ed) Potyvirus taxonomy. Springer, Wien New York, pp 31–46 (Arch Virol [Suppl] 5)
4. Brunt AA (1992) General properties of potyviruses. In: Barnett OW (ed) Potyvirus taxonomy. Springer, Wien New York, pp 3–16 (Arch Virol [Suppl] 5)
5. Colinet D, Kummert J, Lepoivre P (1996) Molecular evidence that the whitefly-transmitted sweetpotato mild mottle virus belongs to a distinct genus of the Potyviridae. Arch Virol 141: 125–135
6. Colwell RJ (1992) Niche. A bifurcation in the conceptual lineage of the term. In: Keller EF, Lloyd EA (eds) Keywords in evolutionary biology. Harvard University Press, Harvard, pp 240–248
7. Domingo E, Holland JJ, Biebricher C, Eigen M (1995) Quasi-species: the concept and the word. In: Gibbs A, Calisher CH, Garcia-Arenal F (eds) Molecular basis of virus evolution. Cambridge University Press, Cambridge, pp 181–191
8. Edwardson JR (1992) Inclusion bodies. In: Barnett OW (ed) Potyvirus taxonomy. Springer, Wien New York, pp 25–30 (Arch Virol [Suppl] 5)
9. Eigen M (1993) Viral quasi-species. Sci Am 269: 32–39
10. Elliott RM (ed) (1996) The Bunyaviridae. Plenum Press, New York
11. Faria JC, Gilbertson RI, Hanson SFG, Morales FJ, Alhquist P, Loniello AO, Maxwell DP (1994) Bean golden mosaic geminivirus type II isolates from the Dominican Republic and Guatemala: nucleotide sequences, infectious pseudorecombinants, and phylogenetic relationships. Mol Plant Pathol 84: 321–329
12. Ghiselin MT (1984) Definition, character and other equivocal terms. Syst Zool 33: 104–110
13. Harrison BD (1985) Usefulness and limitations of the species concept for plant viruses. Intervirology 24: 71–78
14. Holland JJ, De la Torre JC, Steinhauer DA (1992) RNA virus populations as quasi-species. In: Holland JJ (ed) Genetic diversity of RNA viruses. Springer, Wien New York, pp 1–20
15. Kingsbury DW (1988) Biological concepts in virus classification. Intervirology 29: 242–253
16. Kitcher P (1984) Species. Phil Sci 51: 308–333
17. Kitts DB (1984) The names of species: a reply to Hull. Syst Zool 33: 112–115
18. Mayr E (1982) The growth of biological thought. Diversity, evolution, and inheritance. Harvard University Press, Cambridge
19. Mc Neil D, Freiberger P (1993) Fuzzy logic. Simon and Schuster, New York
20. Milne RG (1985) Alternatives to the species concept for virus taxonomy. Intervirology 24: 94–98
21. Murphy FA, Fauquet CM, Bishop DHL, Ghabrial SA, Jarvis AW, Martelli GP, Mayo MA, Summers MD (eds) (1995) Virus Taxonomy. Classification and Nomenclature of Viruses. Sixth Report of the International Committee on Taxonomy of Viruses. Springer, Wien New York (Arch Virol [Suppl] 10)
22. Murphy FA (1996) Virus taxonomy. In: Fields BN, Knipe D (eds) Virology, 3rd ed. Lippincott Raven, Philadelphia, pp 15–58
23. Padidam M, Beachy RN, Fauquet CM (1995) Classification and identification of geminiviruses using sequence comparisons. J Gen Virol 76: 249–263
24. Pringle CR (1991) The 20th Meeting of the Executive Committee of ICTV. Virus species, higher taxa, a universal database and other matters. Arch Virol 119: 303–304
25. Roizman B (1996) Herpesviridae. In: Fields BN, Knipe D (eds) Virology, 3rd ed. Lippincott Raven, Philadelphia, pp 2228–2229

26. Shukla DD, Lauricella R, Ward CW (1992) Serology of potyviruses: current problems and some solutions. In: Barnett OW (ed) Potyvirus taxonomy. Springer, Wien New York, pp 57–69 (Arch Virol [Suppl] 5)
27. Shukla D, Ward CW, Brunt AA (1994) The potyviridae. CAB International, Wallingford
28. Thomas JE, Massalski PR, Harrison BD (1986) Production of monoclonal antibodies to African cassava mosaic virus and differences in their reactivities with other whitefly transmitted geminiviruses. J Gen Virol 67: 2739–2748
29. Van Regenmortel MHV (1985) Antigenic structure of plant viruses. In: Van Regenmortel MHV, Neurath AR (eds) Immunochemistry of viruses. Elsevier, Amsterdam, pp 471–473
30. Van Regenmortel MHV (1989) Applying the species concept to plant viruses. Arch Virol 104: 1–17
31. Van Regenmortel MHV (1990) Virus species, a much overlooked but essential concept in virus classification. Intervirology 31: 241–254
32. Van Regenmortel MHV, Maniloff J, Calisher C, (1991) The concept of virus species. Arch Virol 120: 313–314
33. Van Regenmortel MHV (1997) Viral species. In: Claridge MF, Dawah HA, Wilson MR (eds) Species: the units of biodiversity. Chapman and Hall, London, pp 17–24
34. Van Regenmortel MHV, Dubs MC (1993) Serological procedures. In: Matthews REF (ed) Diagnosis of plant viruses. CRC Press, BocaRaton, pp 159–214
35. Ward CW, McKern NM, Frenkel MJ, Shukla DD (1992) Sequence data as the major criterion for potyvirus classification. In: Barnett OW (ed) Potyvirus taxonomy. Springer, Wien New York, pp 283–297 (Arch Virol [Suppl] 5)
36. Ward CW, Weiller G, Shukla DD, Gibbs AJ (1995) Molecular systematics of the Potyviridae, the largest plant virus family. In: Gibbs AJ, Calisher C, Garcia-Arenal F (eds) Molecular basis of virus evolution. Cambridge University Press, Cambridge, pp 477–500

Authors' address: Dr. M. H. V. Van Regenmortel, Institut de Biologie Moléculaire et Cellulaire du C.N.R.S., Laboratoire d'Immunochimie, 15, rue René Descartes, F-67084 Strasbourg Cedex, France.