Current Topics in
Virology
Vol. 13, 2016

Mini-Review

Only viruses, but not their genome sequences, can be classified into hierarchical species and genus classes

M. H. V. van Regenmortel*

UMR 7242, CNRS, University of Strasbourg, ESBS, Pôle API, Parc d'Innovation d'Illkirch,
300, Boulevard Sébastien Brant, CS 10413, 67412 ILLKIRCH Cedex, France.

ABSTRACT

Viruses are obligate intracellular parasites found in all living organisms and they were initially differentiated by the type of host they infected. As increasing knowledge of the structure of virus particles and of the biochemical and biological properties of viruses became available, viruses were classified in taxa known as orders, on the basis of the type and strandedness of their nucleic acids and presence of a reverse transcription process, and in family taxa on the basis of their particle structure, genome organization and virus replication strategy. Lower taxa such as genera and species were more difficult to establish because many additional criteria are needed for demarcating them, for instance natural host range, cell and tissue tropism, pathogenicity, vector transmission, small genome differences, antigenic properties etc. In spite of this, in 2016, a total of 3704 different virus species had already been established. With the recent advent of high-throughput sequencing and metagenomic approaches, tens of thousands of previously unrecorded viral sequences unrelated to any known viral genome were found to be present in a variety of marine and biological environments. Only a tiny fraction of these new sequences correspond to genomes of known virus species and no information is available regarding the host and biological properties of the unknown viruses harbouring these genome sequences. Recently, a group of 26 viral taxonomists convened a workshop to examine the possibility of incorporating these viral metagenomic sequences in the existing official virus classification system. Although viruses are currently classified using a combination of their chemical, structural, biological and genetic properties, virologists increasingly tend to rely mostly or exclusively on nucleotide genome sequences for assigning viruses to different taxa. The present review will discuss some of the problems that will be encountered when attempts are made to create viral species and genus taxa solely on the basis of genome nucleotide sequences.

KEYWORDS: monophyly, nucleotide sequences, viral metagenomics, viral taxonomy, virus species definitions

1. Introduction

A group of 26 viral taxonomists recently published a Consensus Statement article [1] in which they suggested that it should be possible to incorporate large sets of viral metagenomic data in the current, official classification of viruses established on the basis of the chemical and biological properties of viruses. Metagenomic sequencing identifies a metapopulation consensus sequence instead of a single viral haplotype and some of the difficulties encountered when trying to use viral metagenomics for identifying and studying viruses have been described in recent publications [2, 3]. Viral metagenomic analysis has explored less than 1% of the extant viral diversity and most of the viral sequences that have been discovered are not homologous to known

---

*Email id: vanregen@unistra.fr

viruses [4]. If some sequences correspond to known viruses, it is reasonable to expect that the phenotypic properties of the presumed viruses that harbour these sequences are probably very similar to those of the known viruses. On the other hand, it is not possible to predict most of the viral properties if the sequences have never been observed before. The present review will argue that it is in fact not possible to establish new virus species and genera solely on the basis of genome nucleotide sequences in order to incorporate these taxa in the current system of virus classification.

Since finding a virus in a patient with an illness of unknown etiology does not prove that the virus causes the disease, the classical Koch's postulates used for attributing causal efficacy to an infectious agent have been modified in order to establish if metagenomic sequence reads are associated with disease individuals and absent in healthy controls [4]. This could possibly lead to the development of new diagnostic assays for diseases of unknown origin (see section 7).

2. Why do virologists want to classify viruses?

For many years the main purpose of virus taxonomy was to facilitate and organize the study of viruses that were either human pathogens or pathogens that infected animal and plants of particular interest to humans. The properties that seemed most relevant for virus classification were the biological and pathogenic properties of viruses as well as the biochemical, structural and antigenic properties of virus particles that are useful for diagnosing and controlling virus infections. To fully describe a virus, it is necessary to include the different stages of its replication cycle, namely the extracellular infectious state corresponding to virions, the replicating or lytic state in an infected cell and the latent state that occurs when the viral genome has been integrated in the host genome and no virions are produced. In an analogous way, the full description of an insect as an organism includes the different stages of its life cycle namely the egg, larva, pupa and adult. For this reason, both living organisms and viruses have been described as living or non-living processes corresponding to a life cycle or a replication cycle [5] which makes any biological classification more arduous than a classification of simple, easily recognizable individual objects. It is sometimes argued that the genome sequence constitutes the stable, continuing property of an individual viral process, although the genome is only a part of the virion and it is also not possible for a process to have a constant property. As will be discussed in section 6, the persistence of a viral nucleotide sequence during the replication cycle arises from a variety of interactions between individual viral genes and certain gene products of hosts and vectors and this persistence is thus not the cause but the consequence of various dynamic processes involving the virus and its environment.

In order to communicate among themselves about their work, virologists need a practical and workable scheme of recognizable viral entities and the International Committee on Taxonomy of Viruses (ICTV), which is the body empowered by the International Union of Microbiological Societies to make decisions on matters of virus classification and nomenclature, has for many years developed a classification system based mainly on biological and phenotypic characteristics of viruses. This produced a practical classification based on the type of nucleic acid (DNA or RNA), the single or double strandedness of the nucleic acid, the use of a reverse transcription process and different replication strategies, the morphology of virus particles, the presence or absence of an envelope as well as the hosts and vectors of viruses [6]. This classification was extremely useful to laboratory scientists who investigate the biology, pathogenic properties, immune responses and epidemiology of viruses and the diseases they cause.

In recent years, the advent of high-throughput sequencing technologies has enormously increased our ability to rapidly determine the nucleotide sequences of viral genomes and this made it possible to study virus diversity and evolution and to reconstruct comprehensive virus phylogenies with the aim to have taxonomy reflect phylogeny [7]. However, it should be recognized that the most conspicuous and overwhelming purpose of a phylogenetic virus classification is mainly to satisfy an intellectual curiosity about the evolutionary history of viruses on our planet and such knowledge may not necessarily be the major concern of laboratory virologists involved in the study of present day viruses.

3. What is involved in defining a virus species?

Since viruses are non-living, biological entities possessing genomes [8, 9] they have been classified using the hierarchical classes of species, genera and families used in all biological classifications. Classifications deal with classes which are conceptual constructions of the human mind and should not be confused with the concrete viral objects that are the members of each class.

In the case of the higher categories of genus and family, virologists readily accept that the corresponding classes are abstract concepts and they do not confuse them with the real objects that are the members of such classes. It is indeed obvious, for instance, that a virus family cannot be purified by centrifugation or visualized by electron microscopy. The situation is different in the case of virus species since many virologists consider them to be more "real" than genera or families because they often think of a virus species as an individual kind of virus that is able to infect a particular host. The confusion between species as abstract classes and species as concrete objects has been a persistent problem in the whole of biology [10].

The ICTV accepted that viruses could be classified into species only in 1991 when the following definition was endorsed: *A virus species is a polythetic class of viruses that constitute a replicating lineage and occupy a particular ecological niche* [11]. Whereas monophyletic classes are defined by one or a few properties that are both necessary and sufficient for membership in a class, polythetic classes (also known as cluster classes) are defined by a variable combination of statistically covariant properties, none of which is a defining property necessarily present in every member of the species class [12]. The ICTV Master Species List currently contains 3704 different virus species taxa [13] and the class of all these species taxa itself constitutes a class known as the species category. This category is the lowest one in a hierarchical biological classification, i.e. below the genus, family and order categories. The 1991 ICTV definition of virus species is a definition of the virus species category and not of any particular species taxon. This explains why this definition was of little use for establishing new species taxa, a task that is carried out by specialized ICTV Study Groups composed of virologists who have an in-depth knowledge of specific areas of virology [14].

It is important to realize that abstract classes, like any abstraction, can only be defined whereas concrete viruses or any other concrete object cannot be defined but can only be described [15]. It must also be stressed that the variable distribution of properties that characterizes a polythetic class is not itself a single common property of all the virus members of the class, since if this were the case, it would lead to the paradox that every polythetic class is also a monothetic one [12]. Viruses themselves do not have polythetic or monothetic properties.

A conceptual species class also cannot itself possess any intrinsic physical property (such as a chemical composition) nor any relational properties (arising for instance from a relation with a host or a vector) because the class is only defined by listing certain properties of its concrete members. The viruses that the species class refers to are known as its extension or referents which are the concrete instances that satisfy the membership conditions of the class. The extension of the class must be distinguished from its so-called intension which is its meaning or definition. Since the extension of the class can only be determined if one can distinguish members from non-members, the definition or intension of the class must be established by taxonomists before it becomes feasible for virologists to determine its actual extension ([16], p. 227).

4. Class membership and class inclusion

Class membership is the logical relation that allows a bridge to be established between two different logical categories namely, a mental, abstract construct or taxon, and the concrete members of the taxon that are objects located in space and time. The membership relation is different from a part-whole relationship that can exist only between two concrete objects, one being a part of the other. It is, indeed, not possible for a viral object to be a physical part of an abstract viral taxon.

Virus taxonomy uses a hierarchy of taxa, the lowest taxon class being a virus species. The viruses that are members of a species taxon are also members of a genus taxon immediately above it as well as members of higher taxa such as families and orders. The relation between a lower taxon and a higher taxon immediately above it is called *class inclusion*. Class inclusion in an hierarchical classification obviates the need to repeat the properties used for defining higher taxa in the definition of the lower taxa. The defining properties of a genus are also present in the members of the lower species taxa that are included in the genus. Higher taxa such as genera and families are usually defined by a small number of stable properties such as virion structure and virus replication strategy that are both necessary and sufficient for membership in the class, and this is the reason why these classes are often considered to be universal classes. Since higher taxa have more virus members than species taxa, they require fewer defining properties than species taxa that need more properties to meet the qualifications for membership. The logical principle that increasing the number of required qualifications decreases membership [17] invalidates the currently fashionable claim that a single property could be sufficient for defining a virus species [12].

It must be emphasized that the relation of class inclusion does not imply that the stable defining properties of a genus which can be used to discriminate between different genera are themselves defining properties of a species. Properties useful for distinguishing individual species within a genus obviously cannot be properties that are the same in all the members of the genus, but are properties that vary considerably in different members of the same species. Species-defining properties such as natural host range, cell and tissue tropism, pathogenicity, disease symptoms, mode of transmission and small genome differences are virus properties that can be altered by a few mutations. This is the main reason why species are defined polythetically by a variable combination of changeable properties and not by a single, stable property necessarily present in every member of the species and always absent in the members of other species. All the species-defining properties are not necessarily present in every member of the species and a single property cannot be used for establishing membership in a species class [12].

It is ironic that RNA viruses are said to form so-called quasi-species, since the term species in this case is borrowed from chemistry where chemical species always consist of identical molecules. It should be evident that no virus species can ever correspond to a chemical species and referring to virus species as imperfect, quasi-species simply reflects a confusion between chemical and biological categories [14].

The notion of what is a trait, a character, an attribute or a property of a concrete object that could be used for classifying it, has been called the central problem of taxonomy [18]. Although a part of an object is often confused with a property, a part of a thing is a thing and not a property ([16], p.11). The possession of a certain part may perhaps be interpreted as a property of an object but certainly not the part itself. It is often overlooked that the genomic DNA or RNA sequence present in a virion is part of the virus phenotype since it is part of the virion chemical composition and structure. A virus classification based on nucleotide sequences found in virions can thus be viewed as a phenotypic classification that relies on the properties of molecular sequences instead of on the biological and functional properties of viruses. The nucleotide sequence of a virus is clearly an essential constituent part of a virus in the same way that an engine is an essential constituent part of a motor car. If one classifies car engines, the result will be a classification of engines and not of motor cars. Similarly, it is possible to classify the nucleotide sequences found in viruses on the basis of characteristics such as overall genome organization, genome compositional features, gene content and order, particular nucleotide motifs and inferred replication strategies, but this produces a classification of nucleotide sequences or of viral genomes and not a classification of viruses.

5. The definition of virus species adopted by the ICTV in 2013 suggests that the essence of virus species is descent and monophyly

Species have been viewed as the fundamental units of biological classification and taxonomy long before the theory of evolution had been formulated. In his article entitled: "*In Defense of Classification*", John Dupré [19] argued that the units of classification should not be confused with the units of evolution. A unit of classification such as a species can be viewed as a natural kind that exists independently of any human conceptualization, in the same way that chemical elements exist independently of their discovery and naming by humans [20]. Whereas chemical elements are kinds readily distinguished by their atomic number, in the case of virus species it is more difficult to decide whether differences observed between two virus isolates are of sufficient magnitude to justify the conclusion that they correspond to members of two different kinds of viruses (i.e. to two separate species) and not to members of the same species. It used to be believed that the members of a species always shared certain essential properties and that species, therefore, possessed a metaphysical Platonic essence that made each species unique [21]. However, after Darwin, such a view is no longer tenable.

When species are considered to be abstract classes, another problem arises because it may seem that man-made concepts cannot “exist” in the same way that viral objects “exist”. This difficulty can be resolved if one accepts that several modes of existence can be distinguished such as real, phenomenal, conceptual and semiotic existence [22] since it is then possible to claim that virus species do have a conceptual existence.

When species are viewed as conceptual constructs and as units of classification, they cannot be said to evolve because only units of evolution evolve in the form of a lineage or a clade of some kind. Lineages are temporal sequences of entities defined by ancestry and the changes that accumulate within lineages will be passed from ancestors to descendants. Lineages are distinct from clades that include all the descendants of a specific ancestral population. A clade may thus contain a number of distinct branches that are no longer part of the same unit of evolution, whereas a lineage does not cease to exist because a branch achieves the status of a distinct lineage [19].

Evolution occurs when the properties characteristic of the individuals in a phylogenetic lineage change over time. However, lineages are not well suited for storing information about biological entities endowed with intrinsic variability although this is one of the main purposes of a classification. This explains why evolutionary considerations are of limited use for classificatory purposes, even if a taxonomy sometimes may coincide with a plausible phylogeny. Since the biological world cannot be divided up into fully distinct and internally coherent lineages, it is counterproductive to assimilate units of classification with units of evolution [19].

Although the ICTV definition of a virus species as a polythetic class has been used successfully for several years to establish many new virus species [23], the availability of increasing numbers of viral genome sequences has convinced many virologists that it should be possible to define a virus species only on the basis of certain genome properties, for instance the presence of a particular nucleotide motif in all the members of a species. This led to a 10-year long, sometimes acrimonious debate between adepts of polythetic and monothetic definitions of virus species, a debate that has been recorded in great detail [12]. Although many objections were raised against the removal of the term polythetic in the definition of virus species [24, 25], the ICTV in 2013 endorsed the following new definition: *A species is a monophyletic group of viruses whose properties can be distinguished from those of other species by multiple criteria* [26]. Although it was acknowledged that these criteria could be any viral property, this definition made it possible for ICTV study groups to establish species as monophyletic groups of viruses only on the basis of a single percentage value of genome divergence or by the presence of a particular nucleotide motif in the viral genome, both criteria being assumed to be species-defining properties. By doing away with the polythetic requirement that a species should not be defined on the basis of a single property, it became feasible to establish a virus species on the basis of a single genome characteristic.

Another defect of the 2013 definition of virus species is that it discards the concept of class universally used in all biological classifications and defines a species as a group of viruses, i.e., a collection of viral objects that together constitute the group. This contradicts the logic of classes based on the relations of class membership and class inclusion that are used in all biological classifications [12].

Since genera are also monophyletic taxa [27] the new ICTV species definition [26] reintroduced the discarded concept of a Platonic essence, in the form of descent and monophyly [21] and made it applicable to both species and genus taxa.[21]. Such a definition will certainly not make it easier for virologists to establish and distinguish new species and new genera simply by examining viral nucleotide sequences.

A classic example of the pernicious consequences of establishing new virus species only on the basis of a single percentage of genome divergence (91% in this case) are the 307 species of begomoviruses that have recently been recognized [28]. This included 42 species of *Tomato leaf curl virus* that were all very similar biologically and could be differentiated only by introducing the geographical location of the first isolate of each species. Although it was claimed that these 42 species all differed in their biological properties, no evidence was presented that this was the case. In contrast, the two species *Bean golden mosaic virus* and *Bean golden yellow mosaic virus*, for instance, differed in the tissue tropism of their members [28]. It may, indeed, have been preferable to consider the 42 species of *Tomato leaf curl virus* as strains of a single species [12]. This new classification was based on Pairwise Sequence Comparisons (PASC) of all available begomovirus genome sequences which produce plots of the frequency distribution of pairwise identity percentages that show multimodal distributions of several peaks [29]. Some of these peaks can be attributed to genera, species, subspecies, strains, or serotypes [14] if phenotypic traits of the viruses are taken into account for producing biologically meaningful categories. Instead of opting for a species demarcation criterion of 91% [28], it would have been possible, for instance, to use the small valley observed at 71% pairwise identity in the twin peaks in the region of 62-82% pairwise sequence identity [29, 30] as a cut off point for differentiating between strains and species. On the other hand, when recombination and reassortment are common in certain virus families, as in the *Geminiviridae*, this always leads to polyphyletic genomes that cannot be represented adequately by using a monophyletic scheme [23].

6. It is not possible to infer the phenotypic properties of a virus simply from its genome sequence

Although it is often claimed that the properties of a virus are largely or entirely encoded by its genome, it is actually not possible to infer all the phenotypic and biological characteristics of a virus simply from its nucleotide sequence and there are many reasons for this. The first one is that genome sequences are not causally linked to particular phenotypic traits since alternative splicing and the discarding of introns produces unpredictable RNA transcripts and proteins. These proteins interact with vector and host gene products through numerous mechanisms that have not been elucidated. Causal connections between viral genes and the development of complex phenotypic traits are always indirect and contingent upon many other unknown causal factors that eventually result in the development of a viral phenotype. Treating viral genes as representatives of phenotypic traits or as instructions for producing them reduces the role of these other causal factors to a simple nonspecific support for reading genomic information imbedded in undefined regions of the viral nucleotide sequence [31]. It is well-known that small DNA and RNA viruses are composites of replication and structural gene modules that are reorganized during their evolutionary histories and can lead to changes in host range and other properties [3]. The presence of structurally homologous capsid proteins in many DNA and RNA viruses suggests that the evolutionary history of structural genes is distinct from that of non-structural genes [32].

The multiple causal mechanisms that arise from complex interactions between viral genome products and the biological environment of viral vectors and hosts, leading to phenotypic traits, remain totally unknown. A major difficulty for elucidating the underlying causal mechanisms responsible for the appearance of phenotypic traits is that both a reductive decomposition and an integrative recomposition are required in order to understand and reproduce the development of a phenotype.

The argument that certain viral phenotypic properties are not deducible from the viral genome may seem to be contradicted by the considerable success of metagenomic approaches in reliably predicting protein structures from nucleotide sequences using for instance the Rosetta server [33]. Such methods that rely on the accurate prediction of individual residue-residue contacts using millions of metagenomic protein sequences as well as on coevolutionary analyses have indeed made it possible to identify many novel protein families. Successful prediction of a protein structure from a viral genome sequence will certainly make it easier in the future to allocate a newly discovered virus to a viral order or family since membership in such categories can be ascertained from the presence of very few conserved structural features. However, this is not the case when trying to allocate a virus to a species which is always defined by a combination of changeable and genetically variable properties rather than by the presence of a single and stable structural property that is predictable from the genome sequence (see section 4).

A second reason for our ignorance of the links between genotype and phenotype is that epigenetic factors have been shown to play a major role in controlling the expression of viral genes. Epigenetic modifications are potentially reversible changes in gene expression that do not alter the nucleic acid sequence but regulate the interactions between the virus and its hosts and vectors [34]. The role of DNA methylation, histone modifications such as methylation, acetylation and phosphorylation and of DNA silencing mediated by noncoding microRNAs in regulating viral gene expression has been studied extensively [35]. Direct or indirect epigenetic regulation plays a major role in the replication cycle of DNA tumor viruses as well as in the transformation of a normal cell to a cancer cell. It has also been shown that nucleosomes within viral genomes are able to control the expression of genes presenting histone modifications and that DNA methylation can induce gene silencing during latent infection by herpesviruses [36]. The major role played by epigenetic factors in determining viral phenotypes is unlikely to make it possible to unravel the complex contribution of linear genetic information present in portions of viral nucleotide sequences to the biological and functional characteristics of the members of a particular virus species.

A third reason that prevents viral phenotypes from being deduced from genotypes is that a single reference genome is no longer considered to be adequate for representing the total amount of genetic variation present in a species. It is now known that all the members of a biological species do not always share the same genes and this has given rise to the concept of the “pangenome” which is defined by the entire set of genes possessed by all the members of a particular species. In the case of the marine cyanobacterium *Prochlorococcus* which contains only 2000 genes, it was found after sequencing many strains that only about 1000 genes represented the core genome of the organism and that its pangenome comprised as many as 6000 genes [37].

In the case of the alphabaculovirus *Anticarsia gemmatalis* multiplenucleopolyhedrovirus (AgMNPV), the pangenome was found to contain at least 167 genes of which 16 were not shared by all the genomes that had been sequenced [38]. The evolution of AgMNPV is driven by substitutions, gain and loss of nucleotide stretches or entire coding sequences although it is not known which variations cause phenotypic alterations.

It is sometimes claimed that the hosts of a virus can be inferred from its genome sequence, and it has indeed been found that the abundance of certain dinucleotides (such as CpG, UpA and CpA) in viral genomes of ssRNA viruses that infect vertebrates, plants or arthropods tend to differ significantly. Although the origin of this dinucleotide bias is unknown, it is likely to be due to the fact that vertebrate and arthropod lineages diverged about 600 million years ago and that viruses that infect these groups were subjected to distinct, host-specific evolutionary pressures [39]. This made it possible, using nucleotide composition analysis, to determine that three novel RNA viruses isolated from the stool of an Afghan child were actually insect viruses acquired by contamination of the child's food by insects [40]. Identifying the kingdom or phylum of a viral host on the basis of the nucleotide composition of the viral genome is certainly a remarkable achievement, although it does not amount to identifying the particular host of a virus which is the information that clinical virologists usually would like to have. It cannot be denied, however, that studies of the size and structure of viral genomes is key to unravelling the evolutionary history of certain viral families. For instance, the recent discovery of flavi-like viruses that possess long, segmented genomes and infect fish and insects has completely transformed our knowledge of the evolutionary history of the *Flaviviridae* family by demonstrating that these viruses found in invertebrates were the ancestral forms from which the vertebrate-infecting flaviviruses evolved [27].

7. Diagnostic markers used for identifying viruses are not species-defining properties

As discussed in sections 3 and 4, when taxonomists establish a new virus species, they make use of species-defining properties of the concrete members of the newly created species class that must be properties that make it possible to discriminate between different species in a genus. The initial definition of a virus species as a polythetic or cluster class [11] required that a variable combination of about half a dozen properties of viruses had to be used for distinguishing individual species within a genus [14, 29]. An important feature of a polythetic species class is that a single, conserved property is not sufficient for establishing membership in such a class. It is the presence of a covariant set of shared properties in most members of a polythetic species that makes it useful to identify a virus as a member of a particular species since this makes it possible to predict many of the properties of any newly discovered virus isolate.

In contrast, the new species definition endorsed by the ICTV in 2013 [26] suggests that it is possible to establish a new monophyletic virus species (as well as a new genus) only on the basis of a single characteristic property of a viral genome which would considerably reduce the usefulness of identifying a new virus isolate as a member of such a species.

It is important not to confound the process whereby a taxonomist defines and establishes a new virus species with the procedure used by a virologist for identifying a virus isolate as a member of a known species. For identifying a virus, it may indeed be sufficient to rely on a single diagnostic marker [15] such as a particular nucleotide motif in a genome sequence or the reactivity of a specific monoclonal antibody with virus particles, although neither of these diagnostic markers could have been used beforehand for defining and establishing a new virus species [12, 14]. As explained in section 3, the intension or definition of a species class must necessarily precede the unrealistic attempt to determine the full extension of the class since it would require identifying all potential species members and all the virus isolates that may be assigned to the species in the future! [16, p. 227].

The approach known as DNA barcoding [41] is a tool claimed to be useful for identifying members of previously established species on the basis of short, conserved nucleotide motifs that are believed to be correlated with the full genome sequence. Although it has been advocated as a method for defining and recognizing the thousands of virus species that remain hidden in viral metagenomic databases, correlation is not causation and since such motifs are not causally linked to the phenotypic properties of viruses, they cannot be used as species-defining parts or properties that could identify the members of an existing virus species. For the same reason, nucleotide motifs cannot be used for establishing new virus species or genera on the basis of a correlation with full genome sequences. Although the seasonal appearance of storks is sometimes correlated with increased numbers of human births in a certain geographical location, this does not justify a belief in fairy tales. In the absence of any phenotypic information on the viruses themselves, it would be unwise to incorporate virus species derived solely from nucleotide sequences in the current scheme of species recognized in the official ICTV virus classification [12, 42, 43].

8. Conclusion

Virus species and genus taxa must first be established and defined by taxonomists before it becomes feasible to ascertain if a sufficient number of the species and genus-defining properties are present in an individual virus to make it a member of a certain taxon.

The proposal that a monophyletic species (or genus) can be established by relying on a single part or section of a viral nucleotide sequence determined by metagenomic analysis overlooks the fact that it is not possible to know beforehand that such a nucleotide motif is present in all the members of the class and absent in the members of other classes.

Virus species and genus taxa cannot be established solely by analyzing viral nucleotide sequences obtained by a metagenomic approach. The phenotypic and biological properties of members of new species taxa proposed on the basis of metagenomic data must be known before it becomes feasible to try to incorporate such hypothetical species in the current official system of virus classification.

ACKNOWLEDGMENT

The author is grateful to Florence Diemer for her usual outstanding secretarial assistance.

CONFLICT OF INTEREST STATEMENT

The author has no conflicts of interests.

REFERENCES

1. Simmonds, P., Adams, M. J., Benkő, M., Breitbart, M., Rodney Brister, J., Carstens, E. B., Davison, A. J., Delwart, E., Gorbalenya, A. E., Harrach, B., Hull, R., King, A. M. Q., Koonin, E. V., Krupovic, M., Kuhn, J. H., Lefkowitz, E. J., Nibert, M. L., Orton, R., Roossinck, M. J., Sabanadzovic, S., Sullivan, M. B., Suttle, C. A., Tesh, R. B., van der Vlugt, R. A., Varsani, A. and Zerbini, F. M. 2017, Nat. Rev. Microbiol., doi:10.1038/nrmicro.2016.177.
2. Reyes, A., Semenkovich, N. P., Whiteson, K., Rohwer, F. and Gordon, J. I. 2012, Nature Reviews Microbiology, 10, 607. doi:10.1038/nrmicro2853.
3. Simmonds, P. 2015, J. Gen. Virol., 96, 1193.
4. Mokili, J. L., Rohwer, F. and Dutilh, B. E. 2012, Curr. Opin. Virol., 2, 63.
5. Dupré, J. and Guttinger, S. 2016, Stud. Hist. Philos. Biol. Biomed. Sci., 59, 109.
6. Fauquet, C. M. 2010, Taxonomy, Classification and Nomenclature of Viruses. In: B. W. J. Mahy and M. H. V. van Regenmortel (Eds.), Desk Encyclopedia of General Virology, San Diego (Elsevier), pp. 80-95.
7. Gorbalenya, A. E. 2010, Phylogeny of viruses. In: B. W. J. Mahy and M. H. V. van Regenmortel (Eds), Desk Encyclopedia of General Virology, San Diego (Elsevier), 66-69.
8. van Regenmortel, M. H. V. 2010, Syst. Appl. Microbiol., 33, 1.
9. van Regenmortel, M. H. V. 2016, Stud. Hist. Philos. Biol. Biomed. Sci., 59, 117.
10. Ereshefsky, M. 2011, Cladistics, 27, 67.
11. Pringle, C. R. 1991, Arch. Virol., 119, 303.
12. van Regenmortel, M. H. V. 2016, Bionomina, 10, 1.
13. Adams, M. J., Lefkowitz, E. J., King, A. M., Harrach, B., Harrison, R. L., Knowles, N. J., Kropinski, A. M., Krupovic, M., Kuhn, J. H., Mushegian, A. R., Nibert, M., Sabanadzovic, S., Sanfaçon, H., Siddell, S. G., Simmonds, P., Varsani, A., Zerbini, F. M., Gorbalenya, A. E. and Davison, A. J. 2016, Arch. Virol., 161, 2921.
14. van Regenmortel, M. H. V. 2011, Virus Species. In: M. Tibayrenc (Ed.), Genetics and Evolution of Infectious Diseases, London, Burlington (Elsevier), 3-19.
15. Ghiselin, M. T. 1984, Syst. Zool., 33, 104.
16. Mahner, M. and Bunge, M. 1997, Foundations of Biophilosophy, Berlin (Springer-Verlag).
17. Buck, R. C. and Hull, D. L. 1966, Syst. Zool., 15, 97.
18. Inglis, W. G. 1991, Biol. J. Linn. Soc. Lond., 44, 121.
19. Dupré, J. 2001, Stud. Hist. Philos. Biol. Biomed. Sci., 32, 203.
20. Bird, A. and Tobin, E. 2015, Natural Kinds. In: E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Spring 2015 Ed.).
21. Ruse, M. 1987, Br. J. Philos. Sci., 38, 225.
22. Bunge, M. 2016, Rev. Metaphys., 70, 225.
23. Ball, L. A. 2005, The universal taxonomy of viruses in theory and practice. In: C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger and L. A. Ball (Ed.), Virus taxonomy: eight report of the international committee on taxonomy of viruses. Amsterdam (Elsevier Academic Press), 3-8.
24. King, A. 2012, Comments to proposed modification to code rule 3.21 (defining virus species). ICTV Discussions. <https://talk.ictvonline.org/ictv1/f/general_ictv_discussions-20/3930/comments-to-proposed-modification-to-code-rule-3-21-defining-virus-species> [Accessed on 13 January 2017].

25. van Regenmortel, M. H. V., Ackermann, H. W., Calisher, C. H., Dietzgen, R. G., Horzinek, M. C., Keil, G. M., Mahy, B. W., Martelli, G. P., Murphy, F. A., Pringle, C., Rima, B. K., Skern, T., Vetten, H. J. and Weaver, S. C. 2013, Arch. Virol., 158, 1115.
26. Adams, M. J., Lefkowitz, E. J., King, A. M. Q. and Carstens, E. B. 2013, Arch. Virol., 158, 2633.
27. Shi, M., Lin, X. D., Vasilakis, N., Tian, J. H., Li, C. X., Chen, L. J., Eastwood, G., Diao, X. N., Chen, M. H., Chen, X., Qin, X. C., Widen, S. G., Wood, T. G., Tesh, R. B., Xu, J., Holmes, E. C. and Zhang, Y. Z. 2015, J. Virol., 90, 659.
28. Brown, J. K., Zerbini, F. M., Navas-Castillo, J., Moriones, E., Ramos-Sobrinho, R., Silva, J. C., Fiallo-Olivé, E., Briddon, R. W., Hernández-Zepeda, C., Idris, A., Malathi, V. G., Martin, D. P., Rivera-Bustamante, R., Ueda, S. and Varsani, A. 2015, Arch. Virol., 160, 1593.
29. van Regenmortel, M. H. V. 2007, Infect. Genet. Evol., 7, 133.
30. Bao, Y., Kapustin, Y. and Tatusova, T. 2010, Virus Classification by Pairwise Sequence Comparison (PASC). In: M. H. V. van Regenmortel and B. Mahy (Eds.), Desk Encyclopedia of General Virology. Academic Press, Elsevier, Oxford, pp. 95-100.
31. Griffiths, P. E. and Tabery, J. 2013, Adv. Child Dev. Behav., 44, 65.
32. Koonin, E. V. and Dolja, V. V. 2013, Curr. Opin. Virol., 3, 546.
33. Ovchinnikov, S., Park, H., Varghese, N., Huang, P. S., Pavlopoulos, G. A., Kim, D. E., Kamisetty, H., Kyrpides, N. C. and Baker, D. 2017, Science, 355, 294.
34. Galvan, S. C., García Carrancá, A., Song, J. and Recillas-Targa, F. 2015, Front. Genet., 6, 48.
35. Gómez-Díaz, E., Jordà, M., Peinado, M. A. and Rivero, A. 2012, PLoS Pathog., 8, e1003007.
36. Milavetz, B. I. and Balakrishnan, L. 2015, Methods Mol. Biol., 1238, 569.
37. Kettler, G. C., Martiny, A. C., Huang, K., Zucker, J., Coleman, M. L., Rodrigue, S., Chen, F., Lapidus, A., Ferriera, S., Johnson, J., Steglich, C., Church, G. M., Richardson, P. and Chisholm, S. W. 2007, PLoS Genet., 3, e231.
38. Brito, A. F., Braconi, C. T., Weidmann, M., Dilcher, M., Alves, J. M., Gruber, A. and Zanotto, P. M. 2015, Genome Biol. Evol., 8, 94.
39. Lobo, F. P., Mota, B. E., Pena, S. D., Azevedo, V., Macedo, A. M., Tauch, A., Machado, C. R. and Franco, G. R. 2009, PLoS One, 4, e6282.
40. Kapoor, A., Simmonds, P., Lipkin, W. I., Zaidi, S. and Delwart, E. 2010, J. Virol., 84, 10322.
41. Hebert, P. D. N. and Gregory, T. R. 2005, Syst. Biol., 54, 852.
42. Ebach, M. C. and Holdrege, C. 2005, Bioscience, 55, 822.
43. Wheeler, Q. and Valdecasas, A. G. 2007, An. Jard. Bot. Madr., 64, 237.