



Virus taxonomy and classification: naming of virus species

Roger Hull^{1,2} · Bert Rima³

© Springer-Verlag GmbH Austria, part of Springer Nature 2020

Abstract

Virus classification arranges viruses showing similar properties into groups and, even though this depends on choices of which specific properties have a value for classification, it does have certain important features. It aims to give a structured arrangement of viruses so that the human mind can comprehend them more easily. It helps with communication between virologists, and between virologists and non-virologists (e.g. regulators, advisers, other stakeholders etc.). It enables properties of new viruses to be predicted, and it could reveal possible evolutionary relationships. We need appropriate unambiguous names for virus species, which is the keystone taxon, howsoever these are defined. We react to the recent consultation paper [1] and suggest that, before deciding on a binomial (Latinized or non-Latinized) system for virus species names, the International Committee on Taxonomy of Viruses develops a 21st century virus classification system that handles the large numbers of new virus species expected from metagenomic studies. This system should be user-friendly for easy communication, especially between virologists and non-virologist stakeholders.

Introduction

Humans have an innate desire to name and classify, and the classification of any object is a totally artificial and human-driven activity without a natural base. But why do we classify anything, natural objects, books etc.? It gives an accepted name of each object and a structured arrangement of the objects which enables the human mind to comprehend them more easily. Although classification systems can be traced back to Plato and Aristotle, most of the classification of natural organisms is based on that of Linnaeus published in 1735 in *Systema Naturae*. This proposed a taxonomical structure of natural objects, giving them binomial names and hierarchical arrangement.

Virologists are no exception to this innate desire to classify. The International Committee on Taxonomy of Viruses (ICTV) was charged with organising the taxonomy of viruses. Its Statutes include statements that the objects of the Committee shall be for the public benefit and, in particular, to advance education in the taxonomy of viruses, and to achieve the objectives that are described by Siddell et al. [1]. In the last rounds of discussion at the ICTV, the potential to classify a virus has been expanded to include those “viruses” for which the only known attribute is its RNA or DNA sequence [2]. Thus, computer-generated sequences which, hopefully but not always certainly, represent a contiguous viral genome, must now be assigned to a virus species with the further requirement to also generate an unambiguous genus name to complete the classification of the sequence.

The Executive Committee (EC) of the ICTV initiated a “thought experiment” on the possibility of generating Latin Linnaean binomial names (LLBNs) for virus species and genera in line with those used in the taxonomy of plants, animals and other organisms. This experiment resulted in a proposal [3] that showed that it was possible to name virus species with LLBNs, such as *Morbillivirus hominis* aka the species currently known as *Measles morbillivirus* (a non-Latin binomial name) or just measles virus. Whilst English is not the only language of science, Latin is rarely used, except by tradition in biological taxonomy. Furthermore, the application of LLBNs raises several serious problems [1, 3].

Handling Editor: Sead Sabanadzovic.

✉ Bert Rima
b.rima@qub.ac.uk

Roger Hull
rogerhull@btinternet.com

¹ John Innes Centre, Norwich Research Park, Norwich, England, UK

² Present Address: Portman Drive, Child Okeford, Blandford Forum, Dorset, England, UK

³ Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, Belfast, Northern Ireland

This paper presents some thoughts on the discussion on binomial naming of virus species raised by Siddell et al. [1] and also requested by the ICTV through the various study groups. It goes on to make suggestions that should be considered in updating the taxonomic system to take into account the large changes in available data and in the usage of the system.

Why classify and name virus species?

A classification system on its own is a sterile entity, and one has to consider who want to use it (the customers) and why they want to use it? Hull [4] made some suggestions on this question:

1. It helps communication between virologists.
2. It enables properties of new viruses to be predicted
3. It could reveal possible evolutionary relationships
4. It helps communication between virologists and non-virologists (e.g., regulators, advisers, other stakeholders, lay people, etc.).

The keystone to a biological classification system is the species taxon upon which all the other taxa are based. For most living organisms the standard definition of a species used in Linnaean classification has been “a group of closely related organisms that are very similar to each other and are usually capable of interbreeding and producing fertile offspring.” However, this species concept becomes more arguable when applied to those organisms such as viruses (and retrotransposons and bacteria) in which sex does not play a role in reproduction, and especially when horizontal gene transfer takes place frequently. Consequently, in bacteriology, which does use a Latin binomial naming system for its species classifications, the species concept is still debated [5–7].

As described by Siddell et al. [1], there have been decades of debate on a standardised naming system for virus species which is contrasted to, but closely linked to, naming viruses themselves. The subject of this current paper is not primarily the definition of a virus species, although the actual definition is of paramount importance.

In considering how to deal with virus species in a taxonomy system, a further element that should be taken into account is the vast expansion of the current number of recognized virus species, mainly from metagenomic studies [2, 8]. A recent guestimate by Edward Holmes [9] suggested that there might be 10^8 human and animal virus species when sampling and metagenomics are extended to all hitherto neglected animal species. A recent (2020) virus species count from the ICTV Virus Metadata Resources indicated that the overall total of virus species was 5560 (51 from

algae, 49 archaea, 1396 bacteria, 131 fungi, 534 invertebrates, 238 invertebrates/vertebrates, 13 marine (S), 5 phytobiome (S), 6 plants/invertebrates, 1632 plants, 1 plants/fungi, 62 plants/invertebrates, 20 protozoa, 9 sewage (S), 1413 vertebrates). When viewed in the light of the estimate of 405 for characterised human-infecting viruses [10], the overall numbers of viruses (more than 10^9) that might have to be named and classified become a significant challenge. Even for virus names, we would need more than 6 or 7 letters for the abbreviated version.

How the current and proposed systems fit the requirement of users

As noted above, a classification system is artificial and should be of use to a range of users (customers):

1. *It helps communication between virologists;* 2. *It enables properties of new viruses to be predicted;* 3. *It could reveal possible evolutionary relationships.* Ideally, the classification of viruses should allow a relatively expert reader (or listener) to immediately recognize a set of likely characteristics of the virus. This provides meaning to a statement that a specific virus is classified in a given taxon. There is a general and immediate perception of what the properties of viruses in the order *Mononegavirales* are (ignoring the rhabdoviruses assigned to the genus *Dichornavirus*). The classification of a virus as a member of the *Paramyxoviridae*, which was recently revised [11] immediately has implications in terms of strategy of gene expression, replication, numbers and likely types of genes and even the likely gene order. In that sense, classification is useful for the virologists.

The current system of virus species nomenclature has evolved from everyday usage by virologists, especially if it includes the genus name as well as the species name in the first formal mention (e.g., in a publication). The discussion about renaming the viral species in the classification system with either Linnaean binomials in English or Latin [1] has implications for the immediacy of the recognition of these implied characteristics and for a potential breakage of the link to the past literature, which could be disastrous.

4. *It helps communication between virologists and non-virologists.* The arguments about naming virus species are not without a consequence for other stakeholders. While it is recognized that there is a clear distinction between the taxonomic naming of virus species and the names of the viruses themselves, it is important for the wide range of users of the taxonomy that they do not diverge too much. Virus names and virus taxonomy are used in communication between virologists and non-virologists such as regulators, quarantine officers and advisors (and also between non-virologists themselves). For example, to move plants or samples around

the world, the requirements of regulatory and quarantine agencies in each country, such as the United States Animal and Plant Health Service (APHIS), must be satisfied. APHIS follows the current ICTV nomenclature but does not offer guidance to plant protection and quarantine (PPQ) officers who are unlikely to be virologists. An example of a recent notification (12 December 2019, personal communication) by APHIS to PPQ officers is “First detection of the tobamovirus *Tomato brown rugose fruit virus* (ToBRFV) in *****”. Creating new binomial species names (especially Latin ones) for viral pathogens is really going to complicate this basic reporting system to front-line quarantine and advisory officers in general.

A virus classification system fit for the 21st century

We are going through a period of rapid changes in the potential numbers of viruses that need to be classified, coupled with increases in the amount of information that is available on viruses, new approaches to analysing this information, and new needs for structures that accommodate all this information. The information should be easily available for all interested parties, whether they be research virologists, industry, regulators, advisors, or other stakeholders. We feel that the ICTV should consider developing a framework that can accommodate these needs built upon the current taxonomic structure. For this framework there are two categories of virus species – **characterised species** (with sequence, symptomatology, host range, transmission, etc. known) and **sequence-species**. The latter are often derived from ecological samples and the only information available is the nucleic acid sequence and possibly a host. As there are as often several apparently new viruses in a single sample without any relational properties, these would simply have to be numbered.

In developing such a framework there are two factors to be considered: a) how a species is defined; this has been discussed almost as long as ICTV has existed [see reference 12]. b) how to name them – it would be impossible to give them word names and even more impossible to give the binomials (either Latinized or non-Latinized) for sequence-species because, often in metagenomic studies on environmental samples, neither the host nor the location is well defined.

In light of the large numbers of virus species, a numbering system should be developed. This could distinguish between characterised species (linking to accepted virus names) and viral sequence-species (useful information for quarantine officers for potential new infections) and comprise numerical indicators for other characters such as taxa, major host and/or ecological source of the species.

This designation of a virus species fitting into the current taxonomic framework could be used for communication of information: a) For formal use, such as in publications or oral presentations on characterised species, use the current formal name or common name; b) for publication or presentation of sequence-species, use the species number for formal use; c) for communication with quarantine officers, regulators, other stakeholders, etc., use the common name for characterised species and the species number for sequence-species; in the latter case the person concerned can glean data from the numerical designations in the species number.

Postscript

1. The above discussion removes the need for the ICTV to have long discussions on binomial names (Latinized or not) for virus species. It (and its study groups) can continue the current good work it does on classifying new viruses, whether they be characterized or sequence-species.
2. Organizations dealing with the classification of other biological organisms (e.g., bacteria) are beginning to be presented with problems from increasing amounts of sequence data. An initiative taken by virologists could be a paradigm for how to approach the situation.

References

1. Siddell SG, Walker PJ, Lefkowitz EJ, Mushegian AR, Dutilh BE, Harrach B, Davison AJ (2020) Binomial nomenclature for virus species: a consultation. *Arch Virol* 165:519–525. <https://doi.org/10.1007/s00705-019-04477-6>
2. Simmonds P, Adams MJ, Benk M, Breitbart M, Brister JR, Carstens EB, Zerbini FM (2017) Consensus statement: Virus taxonomy in the age of metagenomics. *Nat Rev Microbiol* 15:161–168. <https://doi.org/10.1038/nrmicro.2016.177>
3. Postler TS, Clawson AN, Amarasinghe GK, Basler CF, Bavari S, Benko M, Kuhn JH (2017) Possibility and challenges of conversion of current virus species names to Linnaean binomials. *Syst Bot* 66:463–473. <https://doi.org/10.1093/sysbio/syw096>
4. Hull R (2014) *Plant virology*, 5th edn. Academic Press, London, Waltham, San Diego
5. Georgiades K, Raoult D (2011) Defining pathogenic bacterial species in the genomic era. *Front Microbiol* 1:Article 151
6. Land M, Hauser L, Jun SR, Nookaew I, Leuze MR, Ahn TH, Ussery DW (2015) Insights from 20 years of bacterial genome sequencing. *Funct Integr Genom* 15:141–161. <https://doi.org/10.1007/s10142-015-0433-4>
7. Konstantinidis K, Kolter R (2017) Do bacterial species really exist and why should we care? <https://schaechter.asmblog.org/schaechter/2017/01/do-bacterial-species-really-exist-and-why-should-we-care.html>. Accessed 16 Jan 2017
8. Simmonds P (2018) A clash of ideas—the varying uses of the ‘species’ term in virology and their utility for classifying viruses

- in metagenomic datasets. *J Gen Virol* 99:277–287. <https://doi.org/10.1099/jgv.0.001010>
9. Skern T (2020) (2019): archives of Virology celebrates its eightieth birthday with a scientific symposium. *Arch Virol* 165:1–3. <https://doi.org/10.1007/s00705-019-04499-0>
 10. Wells K, Morand S, Wardeh M, Daylis M (2020) Distinct spread of DNA and RNA viruses among mammals amid prominent role of domesticated species. *Glob Ecol Biogeogr* 29:470–481
 11. Rima B, Balkema-Buschmann A, Dundon WG, Duprex P, Easton A, Fouchier R, Kurath G, Lamb R, Lee B, Rota P, Wang L, ICTV Report Consortium (2019) ICTV virus taxonomy profile: *Paramyxoviridae*. *J Gen Virol* 2019(100):1593–1594. <https://doi.org/10.1099/jgv.0.001328>
 12. van Regenmortel MHV, Ackermann H-W, Calisher CH, Dietzgen RG, Horzinek MC, Keil GM, Weaver SC (2013) Virus species polemics: 14 senior virologists oppose a proposed change to the ICTV definition of virus species. *Arch Virol*. <https://doi.org/10.1007/s00705-012-1583-5>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.