

Rethinking the Virus Species Concept

Egor Alimpiev

March 15, 2019

Viruses are drastically different from the cellular forms of life. Initial attempts to use the same taxonomic categories and criteria led to viral systematics being less robust, predictive, and generally less reflective of the true evolution of viruses than we want it to be. We discuss the current attempts to rethink and redesign the taxonomic definitions for viruses, the motivations behind them, and critiques by the other members of the scientific community. Later, we propose our solution to the *virus species problem* by integrating several important concepts and methods.

THE ROOTS OF BIOLOGY AS A SCIENCE lie in the area of descriptive approach to the world. Since the beginning of humanity, *naturalists* were exploring the world around them in attempts to describe and classify the phenomena of life. Karl Linnaeus^{1,2} then introduced a systematic approach to describing lifeforms by the means of binominal nomenclature of organisms and sorting them into higher-order groups. Up to this day, we use his approach in all areas of modern biology, and the descriptive approach is still very much relevant in science³. This paper is the account of the current attempts to describe and organize viral diversity, of the challenges one encounters in this process and their possible solutions. To do this properly, it is necessary to start with a broad overview of biological systematics.

What is systematics?

In this paper, the words *taxonomy* and *systematics* are used interchangeably and are used to mean the science of naming and categorizing the biological organisms. These groups of organisms are called *taxa*, and the taxa can be organized on various levels that we call *taxonomic ranks*⁴.

The words we use to describe taxonomy have stayed the same for centuries. However, the principles that govern the way we group the organisms have changed dramatically. The biologists went from the visual similarity to the morphological, paleontological, and developmental criteria, to using the genetic and molecular traits – what is now called *phylogenetics*⁵.

Our ultimate goal is to reflect the actual evolutionary relationships of ancestry and descent. We would call such system *natural*, and encode these relationships in the form of *phylogenetic tree*. (an example can be seen in Figure 1 from the paper by Telford et al., 2015⁶). This

¹ Calisher, C. H. (2007). Taxonomy: What's in a name? Doesn't a rose by any other name smell as sweet? *Croatian medical journal*, 48(2):268

² Karl Linnaeus (1707 – 1778) – Swedish botanist and zoologist, "the father of modern taxonomy".

³ Grimaldi, D. A. and Engel, M. S. (2007). Why Descriptive Science Still Matters. *BioScience*, 57(8):646–647

⁴ These are the usual terms like *species*, *genus*, *family*, etc.

⁵ The study of evolutionary relationships of organisms through using the heritable traits, such as DNA. The term *phylogenetics* is sometimes used interchangeably with systematics, however it has an emphasis on relations, not on classification.

⁶ Telford, M. J., Budd, G. E., and Philippe, H. (2015). Phylogenomic insights into animal evolution. *Current Biology*, 25(19):R876–R887

approach can be contrasted with *phenetics* – the method of classification based on the overall morphological and phenotypic⁷ similarity that was widespread in the 1900s, but became rare in the 21st century⁸.

The reason for our desire to construct such a *natural system* is the fact that such taxonomy would be potentially robust to new data – meaning that it would be very improbable for a newly discovered species to disrupt and change higher-order taxa. Of course, there would still be a possibility of a new taxon of any rank being introduced. Another important feature would be the certain amount of *predictive and explanatory power* that such a natural system would give us: knowing the phylogenetic relationships of an organism allows us to reason about the physiological, morphological, and other phenotypic traits.

THE DEBATES ABOUT SYSTEMATICS would probably never be settled, even with the advent of DNA sequencing⁹, and revisions of classification happen all the time^{10,11}, but the topic that was the main point of arguments in the evolutionary biology community is the *species concept*.

The Elusive Definition of Species

The *species problem* is the long-standing failure of evolution researchers and thinkers to agree on how we define the concept of species and how to then delineate and identify species in the real world¹². The debates on this topic can be traced back at least to the works of E. Mayr and T. Dobzhansky in the 1940s, and since then more than a dozen different species concepts were suggested. To complicate the topic even more, various concepts of species may have several alternative definitions, and the problem of defining the species concept was often confused with the more practical issue of species delimitation – that is, applying the concept/definition in practice to a particular group of organisms. Moreover, some of these concepts are incompatible, giving inconsistent results and leading to different numbers and boundaries of species¹³.

Here are some examples, taken from the paper by De Queiroz, 2007. Some of them may be familiar to the reader, and some may be more exotic:

- (a) *Biological* species concepts are concerned with the reproductive isolation that often exists in different species, and with the mate recognition systems that are involved in mating and fertilization,
- (b) *Ecological* concepts are concerned with the niches and habitats,

⁷ *Phenotype* is the set of all observable traits of an organism.

⁸ Yablokov, A. (1986). *Phenetics – Evolution, Population, Trait*. Columbia University Press

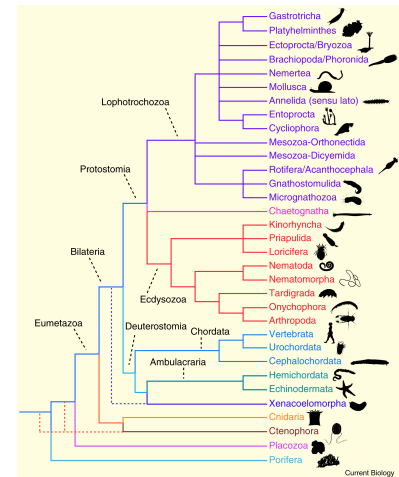


Figure 1: Phylogenetic tree of Metazoa

⁹ There are lots of subtle reasons for this, such as insufficient precision, adaptive radiations, or just the fact that sometimes for a given group of organisms a phenetic approach is more useful in practical applications.

¹⁰ Adl, S. M., Simpson, A. G., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S., Brown, M. W., Burki, F., Dunthorn, M., Hampl, V., et al. (2012). The revised classification of eukaryotes. *Journal of eukaryotic microbiology*, 59(5):429–514

¹¹ Ruggiero, M. A., Gordon, D. P., Orrell, T. M., Bailly, N., Bourgoin, T., Brusca, R. C., Cavalier-Smith, T., Guiry, M. D., and Kirk, P. M. (2015). A higher level classification of all living organisms. *PloS one*, 10(4):e0119248

¹² Hey, J. (2001). The mind of the species problem. *Trends in Ecology and Evolution*, 16:326–329

¹³ De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic biology*, 56(6):879–886

- (c) *Evolutionary* species concepts use the evolutionary roles, tendencies, and history to delineate organisms,
- (d) *Phylogenetic* concepts use the ancestor-descendant relations by looking either at the coalescent events or at the monophylies.
- (e) There are other less frequently mentioned concepts, such as *phenetic*, *genotypic cluster*, *cohesion* criteria.

A lot of the problems arise when we witness a group of species during the process of speciation. One can then argue indefinitely whether they should already be called different species, or just a separate subspecies.

For a more and familiar concrete example, let's look at the definition of species we have all encountered in high school: it can be roughly summarised as "*one of the basic units of biological classification, that most commonly refers to a group of physically similar organisms that can exchange genetic information and produce fertile offspring*". Objectively, this is a very bad definition. First of all, there is no definite measure of similarity that will allow us to consistently delineate similar species. Second, there are lots of examples of the organisms that we traditionally consider different species giving fertile offspring, for example, we can look at the case of grizzly bear and polar bear hybrid in Figure 2.



Figure 2: A grolar bear, hybrid of grizzly and polar bears.

HOPEFULLY, all of the above motivates to the reader the search for the unifying features of these concepts, some kind of a common element that is the most general definition. For us, the best attempt to do this so far is the paper by Kevin De Queiroz¹⁴, that equates species with "*separately evolving metapopulation lineages*¹⁵", solving the philosophical aspect of the problem, and leaving out all of the technicalities of species delineation. This is a very good point of view, because usually the actual methods of species delineation vary from case to case, and these practical issues are usually less problematic to settle¹⁶.

THE UNIQUE FEATURES OF VIRAL BIOLOGY makes it hard to settle on a single definition specific to viruses. In the rest of this paper, we argue that this unifying definition of species is the only one that really fits the purposes of viral taxonomy.

What is a virus species? History of the concept

Viruses present a huge challenge to taxonomists. Most of the usual approaches fail to work for establishing the species-level taxonomy

¹⁴ De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic biology*, 56(6):879–886

¹⁵ Here the term *lineage* refers to an ancestor-descendant series. The beauty of this definition is that we can think of every unit of heredity as a lineage, be it the eukaryotic genome, or a single strand of viral RNA.

¹⁶ Later we discuss the possible practical interpretation of this definition.

of viruses. First, viruses do not mate, and neither do they need to recognize each other during their lifecycle¹⁷. The habitat of viruses is directly linked to the habitat and geographic distribution of host organisms, so we can not explicitly use any ecological criteria. The only phenotypic traits of a viral particle are the structures of the capsid, which also may not be enough for our purposes. The fast rate of evolution and mutation¹⁸ (about 0.76 mutations per round of replication)¹⁹ and simply the small size of viral genomes makes it quite hard to get consistent and unambiguous results on the level of species. However, it should be noted that viruses are now the group of organisms for which we have the most complete genetic information in terms of the number of genomes sequenced²⁰.

The viral species discourse dates back to the 1980s. The idea that it is possible to define the notion of species for viruses was rejected by most of the scientific community, primarily because the main definition at that time was Mayr's reproductive criterion. Another reason was the absence of the official definition of the virus species by the International Committee on Taxonomy of Viruses (ICTV). The simultaneous presence of concepts like *species*, *strains*, and *viral isolates* was also causing much confusion.

Then, in 1982, the Fourth ICTV Report²¹ finally introduced an official consensus definition of *virus species as a concept that will normally be represented by a cluster of strains from a variety of sources, or a population of strains from a particular source, which have in common a set of stable properties that separate the cluster from other clusters of strains*. One of the disadvantages here is that this definition often fails to work in distinguishing closely related or recently diverged species, and some "clusters of strains" that we want to call one species do not possess any properties common to each and every strain. However, it finally settled the hierarchy of terms: a group viral isolates from one specific place or group of hosts is called a strain, and species is a cluster of strains.

The state of affairs became more stable with the definition that was suggested by Van Regenmortel²² in 1989 and later endorsed by the ICTV²³ in 1991. It introduced the notion of polythetic and monothetic classes into virology – a notion at that time already well established in the taxonomy of cellular organisms. Van Regenmortel proposed to call a virus species "*a polythetic class of viruses that constitute a replicating lineage and occupy a particular ecological niche*". The word *virus* is used by the author as a synonym of a strain. The concepts used here played a very important role in further development of virology, so it is necessary to talk more about the notions used in the definition. First, a *monothetic* class requires each and every one of its members to share several so-called "universal" traits – reminiscent of the old def-

¹⁷ Note however that the recognition of the host cell by the virus is a perfectly valid trait used for classification.

¹⁸ Drake, J. W. and Holland, J. J. (1999). Mutation rates among rna viruses. *Proceedings of the National Academy of Sciences*, 96(24):13910–13913

¹⁹ Compare this for example with 0.01 mutations per genome replication in eukaryotes.

²⁰ Even the most modest estimates taken from NCBI genome database at ncbi.nlm.nih.gov/genome, tell us that we have almost 30000 viral genomes, but only around 8000 eukaryotic ones.

²¹ Mathews, R. (1982). Fourth report of the international committee on taxonomy of viruses. classification and nomenclature of viruses. *Intervirology*, 17:1–199

²² Van Regenmortel, M. (1989). Applying the species concept to plant viruses. *Archives of virology*, 104(1-2):1–17

²³ Pringle, C. (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

inition from 1982. On the contrary, the *polythetic* class is defined by a combination of properties that need not be universal²⁴. Each member of such polythetic class shares some of the traits with most but not necessarily every other member of the class. Figure 3, taken from Van Regenmortel, 2016, illustrates the distinction between monothetic and polythetic classes. On it, the individuals 1, 2, 3, and 4 form a polythetic class (note that no property, defined by capital letters, is present in all of the members, however still the class is clearly distinguishable), and individuals 5, 6, 7, and 8 for a monothetic class with respect to the two traits *E* and *F*. This definition respects the natural variability of viruses and their phenotypic properties and gives scientists more freedom in organizing the taxonomy. However, it should be noted that it does not provide the researchers with practical instructions and guidelines for establishing and delineating new and currently known species – a fact that is positive for one part of the scientific community, and frustrating for the other.

Modern definition and its critique

The most recent modifications and clarifications to the official definition by the ICTV were introduced in 2012. The new edition of the International Code of Virus Classification and Nomenclature states that species is “*a monophyletic group²⁵ of viruses whose properties can be distinguished from those of other species by multiple criteria*”²⁶. ICTV rejects the concept of polythetic classes, instead requiring monophyly of the group. The new definition is more concerned with questions of evolutionary relations than with the particular traits possessed by members of the species, agreeing with the trends in modern systematics of cellular organisms. This modification received a lot of criticism both from the scientists that agreed with the previous polythetic definition, its original author included²⁷, and those disagreeing with the old versions of the Code²⁸. In what follows, we will explore these criticisms and at the same time will look at what traits, in general, are suitable and reliable for the needs of modern viral taxonomy *at the species level*.

LOOKING MORE CLOSELY at the definition, one can notice that there is nothing special to the species in it, and indeed it can be used to define any *virus taxon*. This fact alone makes it hard to assign definite taxonomic ranks.

The current definition allows us to define virus species based only on the genomic sequence²⁹. This fact is one of the main arguments against the current definition of virus species, as it is unreasonable to classify viruses using only genetic data. Consequently, one can then

²⁴ Van Regenmortel, M. (2016). Classes,

	A	B	C	D	E	F	G	H
1	+	+	+					
2	+	+		+				
3	+		+	+				
4		+	+	+				
5					+	+	+	
6					+	+	+	
7					+	+		+
8					+	+		+

Figure 3: Figure 1 from Van Regenmortel, 2016, illustrating the concept of monothetic and polythetic classes. Individuals are labelled by numbers and the traits possessed by them are labelled by capital letters.

²⁵ A *monophyletic group* is a group that consists of all living descendants of a common ancestor. It is the synonym to the term *clade*.

²⁶ Adams, M., Lefkowitz, E., King, A., and Carstens, E. (2013). Recently agreed changes to the International Code of Virus Classification and Nomenclature. *Archives of virology*, 158(12):2633–2639

²⁷ Van Regenmortel, M. H., 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., et al. (2013). Virus species polemics: 14 senior virologists oppose a proposed change to the ictv definition of virus species. *Archives of virology*, 158(5):1115–1119

²⁸ Peterson, A. T. (2014). Defining viral species: Making taxonomy useful. *Virology journal*, 11(1):131

²⁹ Multiple nucleotides are considered multiple criteria and therefore suit the definition

argue that two virus strains or isolates belong to different species based solely on the percentage of nucleotide identity, regardless of the fact that they are similar in all other biological aspects, and indeed this is being done, for example for anelloviruses or filoviruses³⁰. The cutoff value of the sequence differentiation is usually completely arbitrary, so it can never reflect the actual evolutionary relationships. Moreover, the different cutoff values then give different numbers of species, which is certainly not biological.

Another often proposed genetic method of taxonomy of viruses is associating the species with a certain nucleotide motif in the genome. This is wrong from the most basic philosophical standpoint – to do such association it would be necessary to know beforehand that this motif is present in all of the members of the species and can not be found in others³¹.

The unique biology of viruses makes questionable even the general applicability of the monophyly. Genetic rearrangements and reassortments within and between species are ubiquitous phenomena in the lifecycles of viruses. A lot of viruses are chimeric, having genes of different (polyphyletic) origins.

Altogether, we can see that a lot of objections to the new Code are based on concerns about it allowing using purely genetic methods in viral taxonomy. Classifying viral genomes is not the same as classifying the viruses is most of the cases.

AS FOR THE OTHER TRAITS, the situation is less controversial but is it important to point out a few caveats. For example, geographic data cannot be reliably used for the purposes of taxonomy. The specification of *Sudan ebolavirus* is *being endemic in the Republic of Sudan and the Republic of Uganda*. This is, of course, a property currently shared by every member of the species. But we would certainly refuse to introduce a new species if some traveler was to bring this virus outside of Africa. Such potential range expansions should prevent us from using geographical traits. Medical and pathological properties, such as symptoms and pathogenicity, are very useful, but mostly specific to human viruses. Molecular properties like cell and tissue specificity, capsid properties are also suitable for taxonomy, however, they can fail to work at the species level³².

The emergence of cheap sequencing technology has allowed us to do large metagenomic assays of soil, ocean water, and other media. Thus, we currently have lots of viral DNA that can not be attributed to any of the known viruses. The problem of classifying viruses from metagenomic samples is extremely hard: we argued that we should not classify viruses based solely on the genetic material, but we have no phenotypic data at all³³. Also, for viruses, we are usually unable

³⁰ As a concrete example, currently the members of genus *Ebolavirus* are defined to belong to the same species if their sequence divergence is less than 30%.

³¹ This simply means that we would need to already have the species labels assigned to assign species. This should not be confused with the problem of assigning a sample to one of the already established samples, in which this method is perfectly useable.

³² The molecular biology of viruses is also of little use at such low taxonomical level because they are usually shared by at least all members of the family.

³³ Simmonds, P., Adams, M. J., Benkő, M., Breitbart, M., Brister, J. R., Carstens, E. B., Davison, A. J., Delwart, E., Gorbalenya, A. E., Harrach, B., et al. (2017). Consensus statement: virus taxonomy in the age of metagenomics. *Nature Reviews Microbiology*, 15(3):161

to predict the phenotype based on the genes because of the small genome and the general lack of phenotype thereof.

THE BEST SUMMARY of the current situation was given by Van Regenmortel et al., 2013, that said that "*the new species definition is in no way superior, and in many ways inferior to the earlier ICTV definition*"³⁴, and it does not make the problem of establishing and delineating virus species easier in practice. Peterson, 2014, points out³⁵ that the democratic way, in which the decisions are made in the ICTV may be the root of the problem. He argues that the decisions on such fundamental and intricate problems as defining the virus species should not be made through the popularity vote, but rather "*based on explicit biological criteria*".

Possible solutions to the debate

What are the ways to settle the debate around this topic? We argue that the answer is to use the universal definition, proposed by De Queiroz³⁶. It solves most of the conceptual and philosophical problems we highlighted above, leaving only the requirement of separate evolution of metapopulation lineages. This approach is supported by Peterson, 2014³⁷, that says only one criterion alone should be dominant: "*evolutionary independence of evolving lineages*", however he was probably unaware of the fact that De Queiroz had already made this into a rigorous definition.

With this approach we no longer care about monophyly, resolving the issue of polyphilic genomes. Instead, we propose using the quantitative statistical criteria of *lineage independence*^{38,39}, supplemented by phenotypic traits ranked in importance by their reliability, to establish species in practice. Bayesian Statistical methods are much more reliable than distinguishing just by looking at the sequence divergence. This may require collecting more data and writing specialized software, but we think this method can be used on a large scale. For chimeric viruses, the approach of consensus species trees can be incorporated, in which the phylogenetic trees are built separately for each of the genes, and then they are combined, by using "averaging" algorithms, in one species phylogenetic tree that reflects the true (without the effects of recombination) origins of the virus with the highest probability.

The metagenomic data can also be classified by measuring lineage independence, but probably not down to the species level.

Another helpful idea may be to make a clearer distinction between the problem discussed in this paper, and the problem of assigning and identifying the sample with one of the already estab-

³⁴ It should be noted however that one of the authors of this quotation is also the author of the earlier ICTV definition.

³⁵ Peterson, A. T. (2014). Defining viral species: Making taxonomy useful. *Virology journal*, 11(1):131

³⁶ De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic biology*, 56(6):879–886

³⁷ Peterson, A. T. (2014). Defining viral species: Making taxonomy useful. *Virology journal*, 11(1):131

³⁸ Yang, Z. and Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, 107(20):9264–9269

³⁹ Fujita, M. K., Leaché, A. D., Burbrink, F. T., McGuire, J. A., and Moritz, C. (2012). Coalescent-based species delimitation in an integrative taxonomy. *Trends in ecology & evolution*, 27(9):480–488

lished species. The second problem has more direct solutions, such as searching for common DNA motifs, as discussed above, and antibody assays – but these are not the species-defining properties.

References

- Adams, M., Lefkowitz, E., King, A., and Carstens, E. (2013). Recently agreed changes to the International Code of Virus Classification and Nomenclature. *Archives of virology*, 158(12):2633–2639.
- Adl, S. M., Simpson, A. G., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S., Brown, M. W., Burki, F., Dunthorn, M., Hampl, V., et al. (2012). The revised classification of eukaryotes. *Journal of eukaryotic microbiology*, 59(5):429–514.
- Calisher, C. H. (2007). Taxonomy: What’s in a name? Doesn’t a rose by any other name smell as sweet? *Croatian medical journal*, 48(2):268.
- De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic biology*, 56(6):879–886.
- Drake, J. W. and Holland, J. J. (1999). Mutation rates among rna viruses. *Proceedings of the National Academy of Sciences*, 96(24):13910–13913.
- Fujita, M. K., Leaché, A. D., Burbrink, F. T., McGuire, J. A., and Moritz, C. (2012). Coalescent-based species delimitation in an integrative taxonomy. *Trends in ecology & evolution*, 27(9):480–488.
- Grimaldi, D. A. and Engel, M. S. (2007). Why Descriptive Science Still Matters. *BioScience*, 57(8):646–647.
- Hey, J. (2001). The mind of the species problem. *Trends in Ecology and Evolution*, 16:326–329.
- Mathews, R. (1982). Fourth report of the international committee on taxonomy of viruses. classification and nomenclature of viruses. *Intervirology*, 17:1–199.
- Peterson, A. T. (2014). Defining viral species: Making taxonomy useful. *Virology journal*, 11(1):131.
- Pringle, C. (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.

- Ruggiero, M. A., Gordon, D. P., Orrell, T. M., Bailly, N., Bourgoin, T., Brusca, R. C., Cavalier-Smith, T., Guiry, M. D., and Kirk, P. M. (2015). A higher level classification of all living organisms. *PloS one*, 10(4):e0119248.
- Simmonds, P., Adams, M. J., Benkő, M., Breitbart, M., Brister, J. R., Carstens, E. B., Davison, A. J., Delwart, E., Gorbalenya, A. E., Har-rach, B., et al. (2017). Consensus statement: virus taxonomy in the age of metagenomics. *Nature Reviews Microbiology*, 15(3):161.
- Telford, M. J., Budd, G. E., and Philippe, H. (2015). Phylogenomic insights into animal evolution. *Current Biology*, 25(19):R876–R887.
- Van Regenmortel, M. (1989). Applying the species concept to plant viruses. *Archives of virology*, 104(1-2):1–17.
- Van Regenmortel, M. (2016). Classes, taxa and categories in hierarchical virus classification: a review of current debates on definitions and names of virus species. *Bionomina*, 10(1):1–21.
- Van Regenmortel, M. H., 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., et al. (2013). Virus species polemics: 14 senior virologists oppose a proposed change to the ictv definition of virus species. *Archives of virology*, 158(5):1115–1119.
- Yablokov, A. (1986). *Phenetics – Evolution, Population, Trait*. Columbia University Press.
- Yang, Z. and Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, 107(20):9264–9269.