Opinion

All Viruses are Unconventional

Jean-Michel Claverie [1,*]

1 Aix Marseille Univ, CNRS, IGS, Structural and Genomic Information Laboratory (UMR7256), Mediterranean Institute of Microbiology (FR3479), Marseille, France ; jean-michel.claverie@univ-amu.fr

* Correspondence: jean-michel.claverie@univ-amu.fr

Abstract: The extension of virology beyond its traditional medical, veterinary or agricultural applications, now called environmental virology, has shown that viruses are both the most numerous and diverse biological entities on earth. In particular, virus isolation studies involving unicellular eukaryotic hosts (heterotrophic and photosynthetic protozoans) revealed numerous viral types previously unexpected in terms of virion structure and morphology, genome size and gene content, or mode of replication. Complemented by large-scale metagenomic analyzes, these discoveries have rekindled interest in the enigma of the evolutionary origin of viruses, for which no simple definition encompassing all of their diversity is still unanimous. Several laboratories have repeatedly tackled the deep reconstruction of the evolutionary history of viruses, using various methods of molecular phylogeny applied to the few shared genes detected in certain virus groups (*e.g.* the *Nucleocytoviricota*). Beyond the practical difficulties of establishing reliable homology relationships from extremely divergent sequences, I present here purely conceptual arguments highlighting several fundamental limitations plaguing the reconstruction of the deep evolutionary history of viruses, and even more the identification of their unique of multiple origin (s). Those limitations are direct consequences of the particularly random mechanisms which govern the reductive evolution of obligate intracellular parasites.

Keywords: origin of viruses; phylogenetic reconstruction; reductive evolution; obligate intracellular parasites; *Varidnaviria*; *Bamfordvirae*; *Nucleocytoviricota*

Introduction: all viruses are unconventional

All viruses are unconventional, in the sense that this category of microorganism lack a set of universal homologous components that they all share, and that could be used to define the "norm" that any virus should obey to be recognized as such.

Let's clarify this unique feature of viruses by taking the counterexample of the cellular world. The diversity of cell types is enormous, from the already highly diverse prokaryotes (Archaea and Bacteria domains), to the eukaryotes. Yet, all the "conventional" type of cells will share homologs of the enzymes and structural proteins required to replicate their DNA genome, to express their genes, synthetize their proteins, together with a number of metabolic pathways to synthetize amino-acids, nucleotides, and generate ATP.

There are two fundamental consequences to the existence of such common set of components and subsystems. First, except for the smallest details, what is learned on a given cell type can usually be transposed to many others. Ribosomes, for instance, work the same way in all cells. The notion of "model" systems (such as *E. coli* or yeast) has thus become a founding stone of modern Biology. This is well summarized by the aphorism: "what is true for *E. coli* is true for the elephant", attributed to Jacques Monod and Francois Jacob [1], two of the most prominent founders of molecular biology. Second, unconventional "cells" are then easy to distinguish from the "normal" crowd, because they lack one of these (almost) common capability: a functional cell division apparatus (e.g. *Babela massiliensis*) [2], ATP synthesis (*e.g. Rickettsia*) [3], and/or amino-acid or nucleotide synthetic pathways (*e.g. Chlamydiae* or *Tremblaya princeps*) [4, 5].

If we now try to do the same exercise for viruses, *i.e.* identify a "model" species and then define "unconventional viruses" versus "regular ones", one soon realizes that it is impossible. Amazingly,

this is not even possible among members belonging to the same Baltimore's classification, such as the dsDNA viruses (on which I will focus in this article for the sake of clarity, or even within the same kingdom. For instance, Mimivirus (with a 1.2 Mb genome) [6] is by no mean a "model" or a prototype for all the dsDNA viruses included in the recently defined "*Bamfordvirae*" kingdom [7] that also include the adenoviruses (with a 30 kb genome)[8] or its own viral parasite, the virophage [9] (Table 1). In other word: would a complete knowledge of the physiology/replication cycle of one of these viruses be of any help to elucidate that of the others? The answer is clearly no.

Table 1. Virus-encoded DNA/RNA polymerases in various eukaryotic dsDNA viruses.

Family/genus name	DNA polymerase	RNA polymerase	Genome Size range
kingdom *Bamfordvirae*			
Mimiviridae	+	+	0.4-1.6 Mb
Poxviridae	+	+	185-360 kb
Iridoviridae	+	+	100-212 kb
Asfarviridae	+	+	171-190 kb
Ascoviridae	+	+	120-200 kb
Coccolithovirus[1]	+	+/-	407 kb
Marseilleviridae[1]	+	+/-	350-376 kb
Chlorovirus	+	-	280-300 kb
Prasinovirus	+	-	173-199 kb
Adenoviridae	+	-	25-45 kb
Lavidaviridae	-	-	17-30 kb
Other kingdoms or unclassified			
Pithoviridae	+	+	610 kb
Pandoraviridae	+	-	1.8-2.5 Mb
Nimaviridae	+	-	309 kb
Herpesviridae	+	-	108-236 kb
Nudiviridae	+	-	97-232 kb
Baculoviridae	+	-	80-160 kb
Polydnaviridae	-	-	up to 800 kb
Papillomaviridae	-	-	7 kb
Polyomaviridae	-	-	4-5 kb

[1] The virus-encoded RNA polymerase is not packaged in the virion.

The fundamental reason a virus prototype cannot exist is because the term "virus" does not designate an "object" (alive or not, this question is still debated) of which a model can be built, but a conceptual process. What makes viruses alike is not what they are made of, but the cyclic scenario they use to reproduce themselves. In the most general terms, this scenario is as follows: transported in a molecular box, a genome of some kind (RNA or DNA) gain access to a cellular system that is used to produce more copies of itself, and package them into neo-synthetized boxes that are then released in the environment. This abstract scenario can be materially implemented in many different ways, many of which may not have been discovered yet, making environmental virology one the frontiers of the Terra Incognita of basic Biology [10, 11]. Amazingly the fundamental difficulty of defining viruses was already perceived by Lwoff in the early days of virology when the best he could do to propose the first formal definition of viruses, was to list the ***missing*** properties making them ***NOT*** belong to the cellular world: they could not divide, could not synthetize ATP, and could not synthetize their proteins (12].

1. Viruses display a huge gradation in "absolute" parasitism

One of the main property that is common to all viruses is that the "active" part of their replication cycle can only happen inside a cell. They are "obligate" intracellular parasites, a property that most non-specialists think is unique to viruses. However, we now know that this property alone is not sufficient to discriminate viruses from the cellular world, as modern microbiology revealed a fascinating underworld of "unconventional" parasitic cells that can only live within other cells (such as those already listed above)[2-5]. Yet, these obligate intracellular parasites (defective for different subsets of essential genes and metabolic pathways) manage to retain enough common macromolecular components so their classification as members of (or derived from) the cellular world (*e.g.* the bacterial domain) remains straightforward (for instance by the presence of ribosomes).

At this point, it is interesting to point out that although "obligate" parasitism sounds like a qualitative character (*i.e.* either you are or are not an absolute parasite), it actually covers a whole gradation of dependency toward the host cell. In some case, for instance, supplementing a culture medium with a specific metabolite was found sufficient to turn an absolute intracellular parasitic bacteria into a free-living one (*e.g. Tropheryma*) [13]. In other cases, the parasitic organism is short of achieving free-living by hundreds of missing genes (*e.g. Tremblaya*) [5]. Absolute parasitism could thus be quantified by the number of essential genes that a thought experiment would need to reintroduce into an absolute parasite to restore its free-living capacity.

Viruses can actually be ranked relative to each other in a similar way, from minimal viral genomes merely encoding the blueprints of their particle (*i.e.* less than a handful of structural proteins) [14, 15], to giant viruses encoding, in addition to hundreds of particle components, the blueprint of the transient intracellular factory used to synthetize them, and the regulatory elements required to hijack the systems unique to the host cell (*e.g.* the ribosomes) [9, 16]. In that respect, the range of "absolute" parasitism covered by the eukaryotic dsDNA viruses is particularly baffling with cytoplasmic giant viruses encoding largely more than thousand proteins including complete DNA replication and transcription machineries (9], many protein translation components [17], and numerous biosynthetic pathways [16-19], down to nuclear polyomaviruses with 5 kb genomes encoding 5 or less proteins [14,15]. Such a factor of 500 in genome sizes, and the apparent randomness of the various gene contents, are difficult to interpret in the context of a unique fit-all evolutionary scenario driven by a fixed set of fitness constraints. A huge variation in genome size (14-735kb) is also seen among dsDNA bacteriophages [20, 21]. Given that huge range of genomic complexity together with the lack of a sizable common set of conserved genes, it may seem quite unrealistic and artificial to postulate a common origin for all dsDNA viruses, even limiting ourselves to those infecting eukaryotes. Yet such a feat is periodically attempted [*e.g.* 22-25]. In the sections below, I present several fundamental reasons why deep reconstructions of viral phylogenies might not be tractable beyond the level of individual virus families.

2. First arguments in favor of a genome reduction evolutionary scenario

Even if the gene content of viruses (even among those infecting the same host) appears to be random, there is some order in this apparent chaos. For instance, there seems to be a strict hierarchy governing the presence of encoded DNA and/or DNA-dependent RNA polymerases in viral genomes (Table 1). As of today, all viruses encoding their own RNA polymerase, also encode a DNA polymerase. If the RNA polymerase can be absent from virus encoding their DNA polymerase, the converse is not true. In other words, DNA polymerases are only absent from viruses also lacking RNA polymerase. Conditioning the presence of a virus-encoded transcription apparatus to that of a replication apparatus strongly suggests an irreversible reductive evolutionary process with a progressive loss of functions from ancestors equipped with both machineries. This is one of the argument in favor of a cell-like origin of dsDNA viruses, including those from the newly defined

Bamfordvirae kingdom (that include a whole spectrum of virus families with and without virus-encoded DNA/RNA polymerases) (Table 1). Further supporting such a progressive loss-of-function scenario, two intermediate virus groups (*Coccolithovirus* and the *Marseilleviridae*) encode a RNA polymerase that, most surprisingly, is not packaged in their particles, forcing them to initiate their cytoplasmic replication cycle by first recruiting nuclear functions [26, 27].

However, the reason why the loss of RNA polymerase should always precede that of the DNA polymerase is not clear, as the absence of any of these genes will force a previously cytoplasmic virus to become dependent of cellular functions located in the nucleus. Once evolved to gain access to the nucleus, a viruses devoid of its own DNA polymerase could use that of the cellular host, independently of the presence/absence of its own RNA polymerase. Thus, no basic biological rule would be violated by the eventual discovery of such type of "unconventional" dsDNA virus.

Finally, another "unconventional" type of large dsDNA virus is represented by the polydnaviruses of parasitic wasps. Amazingly, if their particles could package up to 800 kb of DNA, it does contain not the genes required for viral DNA replication and virion production! [Reviewed in 28]. In this extreme case, even the minimal blueprint of the viral particle has been subcontracted to the host cell. It is difficult to interpret the emergence of such virus other than as the end-point of a reductive evolution. Clearly, the fact that some viruses do not even encode the constituents of their own particles (yet a feature that does not contradict Lwoff's criteria) does not help in designing a rigorous definition that will fit them all.

3. The main conceptual difficulty with the deep phylogenetic reconstruction of virus evolution

Three main scenarios have been proposed to explain the origin of viruses. The "virus-first" theory states that viruses predated the emergence of cells. At the opposite, the "reduction hypothesis" states that viruses evolved as reduced parasitic forms of early cellular organisms. The third one, "the escape hypothesis", is a variation of the later stating that ancestral viral genomes were constituted of subsets of cellular genes that escaped cell control [reviewed in 29].

I never understood how the first hypothesis could even be proposed, since it is properly absurd if we respect the precise meaning, accepted by all, of the word "virus": an obligate intracellular parasite. This mere definition immediately implies that the first virus (es) had to emerge in the context of preexisting cell-like organisms (free-living either as individualities, or as a consortium). The first viruses, - in the sense that with give it today - could not precede the emergence of their hosts. Furthermore, the ancestor of the first virus (es) could not be one itself, but had to be an (or several) unknown free-living cell-like organism(s). From this point on, only some sort of reduction hypotheses should constitute the theoretical context on which to base the comparison of extant viral genomes and the reconstruction of their phylogeny.

There is, however, a fundamental difficulty in reconstructing the evolutionary history of obligate intracellular parasites by comparing them without reference to the free-living organisms from which they originated. For tree-based phylogenetic approaches to deliver a sensible scenario, all protagonists of the evolutionary game must be included in the analysis. This difficulty is illustrated in Figure1 and Figure 2 where I attempted the phylogenetic reconstruction of 7 bacterial obligate intracellular parasites. Like viruses, those microorganisms cannot survive and multiply outside of eukaryotic cells which provide them with essential metabolites and enzymatic functions they no longer have. For the sake of my demonstration, I first pretended to reconstruct the phylogeny of these parasites, as if in search of the ancestral obligate intracellular parasite from which they might all have derived. Interestingly, these false premises resulted in a normal-looking tree, suggesting the existence of 3 different parasite "families" with strong statistical support (Figure 1). A different representation of this tree then suggests that these 3 families originated from a common "parasitic" ancestor (Figure 1), a convergence obviously imposed by all tree-building algorithms.

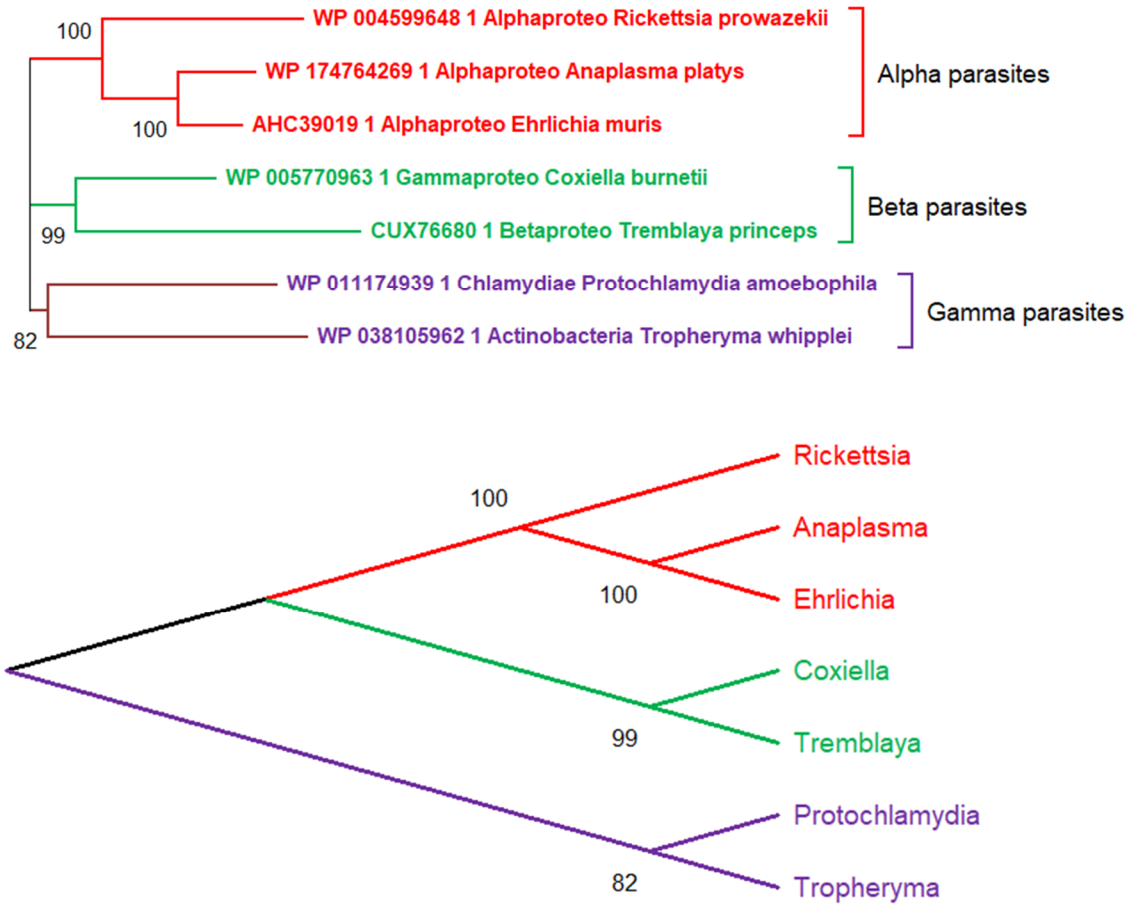

Figure 1. **Erroneous phylogenetic relationship between seven obligate intracellular parasitic bacteria.** Top): The neighbour-joining tree was generated from 1014 conserved sites in the multiple alignment of their DNA polymerase alpha subunits using the JTT substitution matrix. The protein NCBI identifiers are indicated. In absence of free-living bacterial relatives, the tree erroneously suggests (with a strong statistical support) the existence of 3 separate "parasite families" emerging from 3 distinct evolutionary branches (Alpha, Beta, Gamma). Bottom): Using a different representation, the tree topology (inherent to the tree-building algorithm) can be interpreted as supporting the existence of an ancestral obligate parasite from which all three (virus-like) families of extant parasites derived. The true evolutionary history of these parasitic bacteria is shown in Figure 2.

These conclusions are of course totally erroneous, as shown in Figure 2, where I incorporated free-living relatives to the analysis of the seven parasitic bacteria. The resulting tree suggests an evolutionary scenario totally different from the previous one. The seven parasitic bacteria are now seen to relate to 5 different bacterial domains, 4 of which include a majority of free-living representatives (*Chlamydiae* being a noticeable exception). In contrast to Figure 1, this more realistic tree (further supported by a large body of genomic data) does invalidate the existence of an "ancestral" bacterial parasite from which all extant parasites would have derived. Through extensive whole genome comparisons it has been well-demonstrated that obligate intracellular parasitic members originated from their free-living relatives by the loss of essential genes and functions, an irreversible process of genome reduction through which they become increasingly dependent toward their hosts [2-5, 30-32]. Interestingly, the 7 parasitic bacteria I compared above (some of which encode close to 1000 genes) share less than 100 "core genes" (involved in translation, DNA replication and transcription), thus much less than the 400 or so genes considered to constitute a minimal free-living

bacterial genome [33]. Without knowing the actual evolutionary history of these parasites (Figure 2) one might see these 100 core genes as characteristic of their common parasitic life-style. This is of course wrong, as these genes were inherited from (and essential to) their free-living ancestor. On the other hand, these 100 common genes also do not match the entire genome of the hypothetical common ancestor of these parasites, as it is most often concluded from similar phylogenetic reconstruction involving viruses. This then lead to the erroneous conclusion that obligate intracellular parasites (and thus viruses) evolved from simpler ancestors by acquiring genes instead of losing them [34-35].

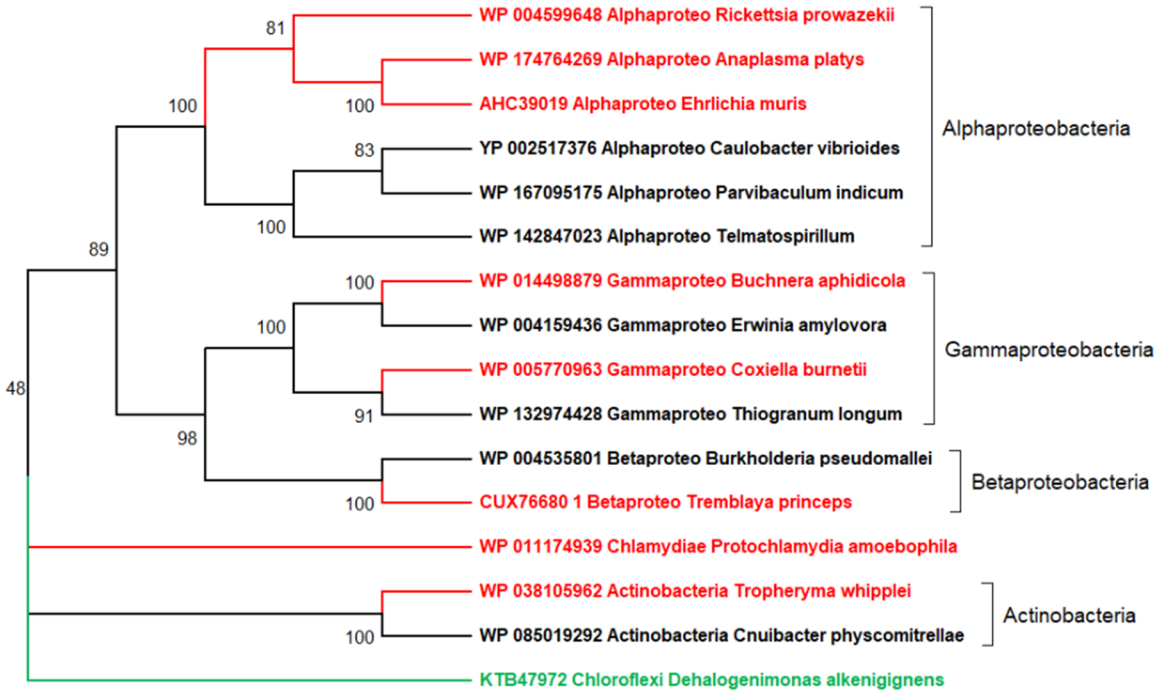

Figure 2. A more realistic representation of the origin and evolution of the obligate intracellular parasitic bacteria depicted in Figure 1. The neighbor-joining tree was generated from 974 conserved sites in the multiple alignment of 16 DNA polymerase alpha subunits using the JTT substitution matrix. The protein NCBI identifiers are indicated. The red branches correspond to the 7 obligate intracellular parasites while free-living relatives are in black. The green branch correspond to a distant bacteria from the *Chloroflexi* phylum, used as outgroup. This tree suggests (with strong statistical support) that parasitic bacteria independently originated at least 5 times from within 5 lineages also containing free living members: once from within *Actinobacteria* and *Betaproteobacteria*, twice from *Gammaproteobacteria*, and once early in the *Alphaproteobacteria* class from which three member of the order *Rickettsiales* emerged. In each case, the switch to a parasitic lifestyle was associated to the loss of essential genes (reductive evolution) nowadays documented by direct comparative genomics. One exception, visible in the tree, is *P. amoebophila* for which no free living relative could be convincingly suggested. *P. amoebophila* belongs to *Chlamydiae*, a phylum of highly diverse members all of which have –like viruses- an obligate intracellular life style. In absence of known free-living relatives, the origin of this bacterial phylum remains mysterious. Compared to Figure 1, this figure illustrates how the lack of known free-living relatives might suggest totally erroneous evolutionary scenarios. The DNA polymerase was used as a conserved protein present in all bacteria (parasitic of not). Its viral version is frequently used in global phylogenetic reconstructions of eukaryotic dsDNA viruses.

In the context of the reduction hypothesis, the simulated analysis in Figure 1 parallels the protocol by which viruses are compared to investigate their evolutionary history. By definition, viruses do not have "free-living" relatives alongside which they could be compared. Without such reference, the deep reconstruction of the evolution of viruses seems unattainable and even conceptually flawed. Phylogenetic reconstruction must thus limited to the family level, that is to group of viruses the ancestry of which can be traced back to a common quasi-extant virus (a situation similar to the *Alphaparasites*/*Rickettsiales* in Figure 1 and Figure 2). Similarly, the small number of core genes strictly shared by different eukaryotic dsDNA virus families (now equal to zero, when including all of them) should not be interpreted as a characteristic of their putative common ancestor, but as the expected result of a non-coordinated succession of genes losses starting from ancestors too old to be scientifically reconstructed, as explained in the next section. It could also sign different origins altogether.

4. The evolutionary random walk of gene losses: another main hurdle in the reconstruction of virus evolution

As soon as an organism switches to the lifestyle of absolute parasitism, the usual laws of neo-Darwinian selection which apply to the conservative evolution of its genes change radically. The previously careful preservation of essential genes is replaced by the possibility of losing functions which can be subcontracted to the host. This trend is irreversible, and no obligate intracellular parasite has ever been documented to revert to a free-living life-style. "Once a parasite, always a parasite" appears to be one of the few absolutely respected motto of microbial evolution [36]. As viruses are archetypes of obligate intracellular parasites, I do believe that the irreversible succession of gene losses constitutes the dominant force in their evolution [10].

Given the central place held by the absence of the protein translation function in the definition of viruses [10,12], it is natural to postulate that the cascade of gene losses that led to the diversity of viruses we know today was initiated by that of an essential ribosomal protein (or rRNA). This would immediately make the defective microbe an "obligate intracellular parasite", moreover confined in the cytoplasm of its host. This new environment, rich in metabolites of all kind as well as ATP could then be used as a rich culture medium for the emerging parasite. This would then open the door to further genome reduction by the losses of the redundant biosynthetic and bioenergetic pathways, until reaching the bare bones of the virus-encoded DNA transcription and replication apparatus.

During this phase, the neo-darwinian selection process will continuously select the viruses for an optimized parasitic life-style generating more progeny at each infection. However, this goal can be achieved in many different ways, sometimes contradictory, depending on the host and ecological situations. It could be via further genomic reduction (thus alleviating the energetic burden of DNA replication on the host) [37], by improving the efficiency of host infection (by innovating on virion structures and infection strategies) [16], by helping the host viability (thus increasing burst sizes) [38, 39], or by using molecular defense against viruses competing for the same host [40]. In the general context of genome reduction, such complex web of evolutionary constraints is expected to generate a huge diversity of "optimal" solutions (sometimes involving moderate gene gains) realized through a variety of viral gene contents without much apparent rationale, as observed.

If the loss of viral genes duplicating cytoplasmic functions (amino acids and nucleotide synthesis, energy metabolism, protein translation) probably can happen quickly, in an almost random manner (the virus benefiting from a free lunch within the cell), the loss of a functional virus-encoded DNA replication or transcription machinery must be concomitantly compensated by an access to the cellular ones, in the host nucleus. The viruses must thus evolved a strategy to either transport its genome to the intact nucleus [41, 42], or make it functionally "leaky" [27, 43], or even dissolve it [44] altogether. The passage from a purely cytoplasmic replication cycle to a nuclear one is thus a major step in the continuous reductive evolution of viruses. I previously noticed that such transition appears to obey a strict order (loss of transcription first, then of replication). However, it

seems to happen randomly at various stages of the reductive evolution process, concerning viruses with vastly different genome sizes (Table 1).

A general model of virus evolution through genome reduction from a free-living (non-virus) ancestor is represented in Figure 3. It illustrates two main points.

First, soon after a parasitic lifestyle is initiated, the process of random gene loss of genes generate very different genomes, from which common-to-all genes (the so-called "core" genes) could disappear rapidly. This corresponds to reality, where the discovery of new virus families steadily lead to the reduction of the number of core genes [23-25, 45-47]. However, while other authors attribute this phenomenon to the "high rate of horizontal transfer and fast sequence divergence of virus evolution" [sic 48], it is it intrinsic to the reductive hypothesis. "Core" genes are borne to disappear as our knowledge of the viral diversity increase because the concept of virus simply does not imply the existence/conservation of any specific virus-encoded function. Core genes are not intrinsically "essential", but are just the artefactual (provisional) consequence of finite, incomplete, sampling.

Second, even when similar genotypes (*i.e.* assortment of core genes) are recognized in different viruses they cannot be used as reliable evidence of common ancestry as they could originate from totally unrelated evolutionary pathways. Such cases are illustrated by the genotypes indicated in red in Figure 3. Thus the more ubiquitous a core (or quasi-core gene) appears to be, the more likely are the virus groups exhibiting it to be polyphyletic (see gene D as an example, Figure 3). One then expects phylogenetic trees build from the most shared genes to be highly discordant. This has been a common finding [23-25, 46, 47], forcing authors to abandon tree-based phylogenetic reconstruction methods for network representations [45, 48], to extrapolate homologies from non-significant sequence similarity [49], or abandon the grail of a unique common ancestry [50], or discard whole virus families causing troubles [25].

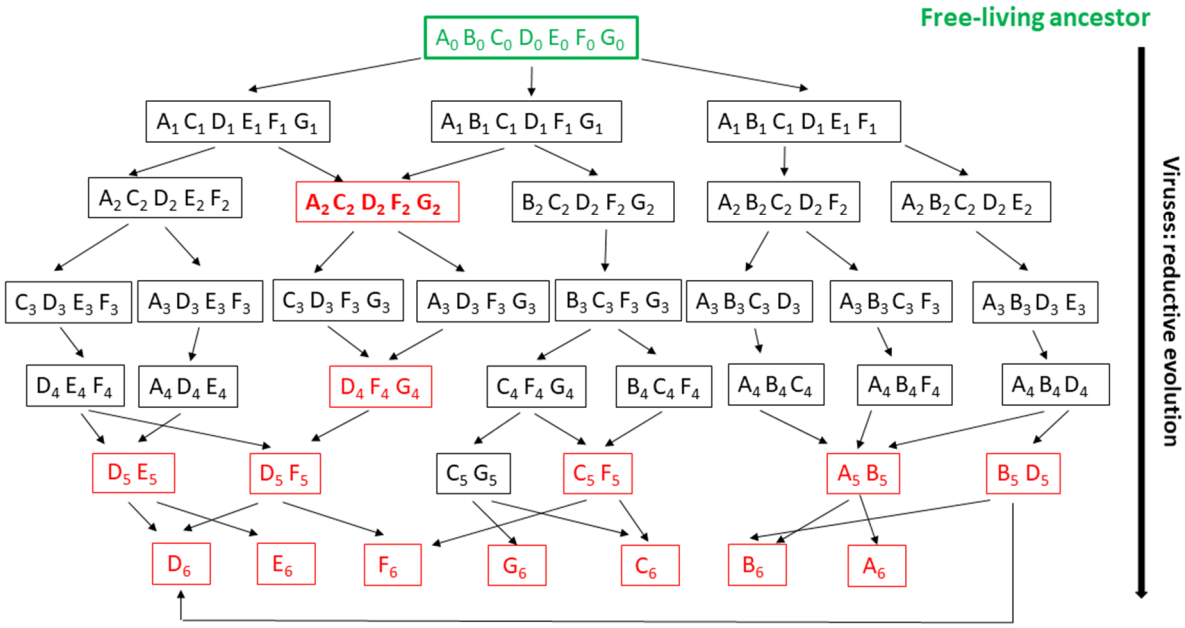

Figure 3. The "virus late" hypothesis: illustration of the intractable evolutionary scenarios of obligate intracellular parasites resulting from random gene/function losses. A toy virus world is represented, starting from a hypothetical ancestral free-living organism (level 0, w/o extant

representative), each box contains the abstract gene content inherited by a given virus (family) from its immediate ancestor. Red "genomes" indicate viruses with similar gene contents albeit resulting from distinct evolutionary scenarios. Random gene losses lead to very diverse overlaps of gene assortment (as in viruses) or to a situation where no single "core gene" is shared by all parasite family (here level 3), as seen between real viral genomes (in particular small ones). Individual gene recurring in multiple combinations (families) or ultimately remaining in the smallest genomes (level 6) are not more characteristic of the parasitic life-style than less ubiquitous ones. In addition, genes shared by more families than other (such as D) may not be better phylogenetic markers than others, as they could have been inherited form different evolutionary pathways. This graph illustrates the intricacy of reconstructing the deep phylogeny of viruses beyond the immediate family level both due to the capacity of random gene losses enjoyed by obligate intracellular parasite and the lack of associated free-living organisms to be used as references.

Conclusion

In this conceptual article, I showed that the phylogenetic reconstruction of the evolution of viruses suffers from several fundamental limitations, in the hypothesis that they were derived from free-living, cell-like, microorganisms, the only logically sound scenario. One limitation is due to the lack of free-living lineage(s) against which to compare the various virus families (Figure 1). The other is due to the almost complete relaxation of functional constraints which characterizes an organism having switched to an intracellular parasitic lifestyle. Paradoxically, the evolutionary trajectory of a virus is much better defined by the way it has lost genes, than by the nature of those it has kept. Unfortunately, one can only compare viral genomes on the basis of the later.

In this paper, I voluntarily neglected two additional confounding evolutionary processes: i) the acquisition of genes by horizontal transfers from cells or other viruses, ii) the de novo creation of genes by the viruses themselves [51]. It is nevertheless clear that these two processes could only make phylogenetic reconstruction even more intractable.

I made no hypothesis on the very nature of the ancestral microorganism(s) at the origin of viruses, but it was in all likelihood equipped with a DNA transcription/replication machinery, and protein synthesis. I proposed that the loss of the later is the evolutionary event that initiated all virus lineages, for which there is no evidence –or logical need- that it only happened once.

Viruses are traditionally classified into families on the basis of common infection and intracellular replication strategies, overall particle structures, and large, specifically shared gene content attesting their descent from a common (quasi-extant) viral ancestor. Cladistics (*i.e.* the presence/absence comparison of entire gene contents) is often a convenient and sufficient method to delineate families that should clearly appear as monophyletic clades [*e.g.* Figure 8 in 51]. It is nevertheless not a fool proof exercise as illustrated by Figure 1 and Figure 2. The recent reclassification of certain members of the Phycodnaviridae family into that of Mimiviridae [52] (which could have been possible prior to the discovery of Mimivirus and its relatives [53]) is a good illustration of the danger of freezing a classification too early.

Having listed the fundamental limitations plaguing the deep phylogenetic reconstruction of viruses beyond the level of the family, I can only wonder about the merits of a spectacular fifteen-rank classification hierarchy recently adopted by the ICTV [7]. Families are now aggregated in a succession of 10 taxonomic levels (suborder, order, subclass, class, subphylum, phylum, etc.), most of which, according to the argument presented here, may remain forever beyond the realm of scientific evidence. In addition of being nearly impossible to pronounce or memorize, many of these abstract clades are dangerously suggesting totally unsupported related ancestries between families as visibly unrelated as *Mimiviridae*, *Adenoviridae*, *Phaeovirus*, and virophages (included in the *Bamfordvirae* kingdom), or the *Herpesviridae* and a large number of phages (now included in the *Heunggongvirae* kingdom). While I fear that this new taxonomical scheme will be taken as a word of the gospel by the incoming generation of virologists, I am also confident that the future discovery

and characterization of many more unconventional viruses will quickly convince them that any attempt to lock viruses into such a deep and rigid classification does not make any biological sense.

Funding: This research received no specific funding. The IGS laboratory is supported by recurrent funding from Aix-Marseille University and CNRS.

Acknowledgments: I thank all my colleagues from the IGS laboratory, in particular D. Chantal Abergel, for the numerous and invaluable discussions that allowed the emergence and refinement of the ideas presented here.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Monod J, Jacob F. Teleonomic mechanisms in cellular metabolism, growth, and differentiation. Cold Spring Harb Symp Quant Biol. 1961;26:389-401. doi:10.1101/sqb.1961.026.01.048)
2. Pagnier I, Yutin N, Croce O, et al. Babela massiliensis, a representative of a widespread bacterial phylum with unusual adaptations to parasitism in amoebae. Biol Direct. 2015;10:13. doi:10.1186/s13062-015-0043-z
3. Driscoll TP, Verhoeve VI, Guillotte ML, et al. Wholly Rickettsia! Reconstructed Metabolic Profile of the Quintessential Bacterial Parasite of Eukaryotic Cells. mBio. 2017;8(5):e00859-17. doi:10.1128/mBio.00859-17
4. Omsland A, Sixt BS, Horn M, Hackstadt T. Chlamydial metabolism revisited: interspecies metabolic variability and developmental stage-specific physiologic activities. FEMS Microbiol Rev. 2014;38(4):779-801. doi:10.1111/1574-6976.12059
5. López-Madrigal S, Latorre A, Porcar M, Moya A, Gil R. Mealybugs nested endosymbiosis: going into the 'matryoshka' system in Planococcus citri in depth. BMC Microbiol. 2013;13:74. doi:10.1186/1471-2180-13-74
6. Raoult D, Audic S, Robert C, et al. The 1.2-megabase genome sequence of Mimivirus. Science. 2004;306(5700):1344-1350. doi:10.1126/science.1101485
7. International Committee on Taxonomy of Viruses Executive Committee. The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. Nat Microbiol. 2020;5(5):668-674. doi:10.1038/s41564-020-0709-x
8. Hidalgo P, Anzures L, Hernández-Mendoza A, et al. Morphological, Biochemical, and Functional Study of Viral Replication Compartments Isolated from Adenovirus-Infected Cells. J Virol. 2016;90(7):3411-3427. doi:10.1128/JVI.00033-16
9. Claverie JM, Abergel C. Mimivirus and its virophage. Annu Rev Genet. 2009;43:49-66. doi:10.1146/annurev-genet-102108-134255
10. Claverie JM, Abergel C. Giant viruses: The difficult breaking of multiple epistemological barriers. Stud Hist Philos Biol Biomed Sci. 2016;59:89-99. doi:10.1016/j.shpsc.2016.02.015
11. Dupré J, Guttinger S. Viruses as living processes. Stud Hist Philos Biol Biomed Sci. 2016;59:109-116. doi:10.1016/j.shpsc.2016.02.010
12. Lwoff A. The concept of virus. J Gen Microbiol. 1957;17(2):239-253. doi:10.1099/00221287-17-2-239
13. Renesto P, Crapoulet N, Ogata H, et al. Genome-based design of a cell-free culture medium for Tropheryma whipplei. Lancet. 2003;362(9382):447-449. doi:10.1016/S0140-6736(03)14071-8
14. Rector A, Van Ranst M. Animal papillomaviruses. Virology. 2013;445(1-2):213-223. doi:10.1016/j.virol.2013.05.007
15. Polyomaviridae Study Group of the International Committee on Taxonomy of Viruses, Calvignac-Spencer S, Feltkamp MC, et al. A taxonomy update for the family Polyomaviridae. Arch Virol. 2016;161(6):1739-1750. doi:10.1007/s00705-016-2794-y
16. Abergel C, Legendre M, Claverie JM. The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. FEMS Microbiol Rev. 2015;39(6):779-796. doi:10.1093/femsre/fuv037
17. Abrahão J, Silva L, Silva LS, et al. Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. Nat Commun. 2018;9(1):749. doi:10.1038/s41467-018-03168-1
18. Schvarcz CR, Steward GF. A giant virus infecting green algae encodes key fermentation genes. Virology. 2018;518:423-433. doi:10.1016/j.virol.2018.03.010

19. Needham DM, Yoshizawa S, Hosaka T, et al. A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. Proc Natl Acad Sci U S A. 2019;116(41):20574-20583. doi:10.1073/pnas.1907517116
20. Petrovski S, Dyson ZA, Seviour RJ, Tillett D. Small but sufficient: the Rhodococcus phage RRH1 has the smallest known Siphoviridae genome at 14.2 kilobases. J Virol. 2012;86(1):358-363. doi:10.1128/JVI.05460-11
21. Al-Shayeb B, Sachdeva R, Chen LX, et al. Clades of huge phages from across Earth's ecosystems. Nature. 2020;578(7795):425-431. doi:10.1038/s41586-020-2007-4
22. Yutin N, Wolf YI, Raoult D, Koonin EV. Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. Virol J. 2009;6:223. doi:10.1186/1743-422X-6-223
23. Koonin EV, Yutin N. Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses. Intervirology. 2010;53(5):284-292. doi:10.1159/000312913
24. Forterre P, Gaïa M. Giant viruses and the origin of modern eukaryotes. Curr Opin Microbiol. 2016;31:44-49. doi:10.1016/j.mib.2016.02.001
25. Guglielmini J, Woo AC, Krupovic M, Forterre P, Gaia M. Diversification of giant and large eukaryotic dsDNA viruses predated the origin of modern eukaryotes. Proc Natl Acad Sci U S A. 2019;116(39):19585-19592. doi:10.1073/pnas.1912006116
26. Allen MJ, Howard JA, Lilley KS, Wilson WH. Proteomic analysis of the EhV-86 virion. Proteome Sci. 2008;6:11. doi:10.1186/1477-5956-6-11
27. Fabre E, Jeudy S, Santini S, et al. Noumeavirus replication relies on a transient remote control of the host nucleus. Nat Commun. 2017;8:15087. doi:10.1038/ncomms15087
28. Drezen JM, Leobold M, Bézier A, Huguet E, Volkoff AN, Herniou EA. Endogenous viruses of parasitic wasps: variations on a common theme. Curr Opin Virol. 2017;25:41-48. doi:10.1016/j.coviro.2017.07.002
29. Nasir A, Kim KM, Caetano-Anollés G. Viral evolution: Primordial cellular origins and late adaptation to parasitism. Mob Genet Elements. 2012;2(5):247-252. doi:10.4161/mge.22797
30. Moran NA, McLaughlin HJ, Sorek R. The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. Science. 2009;323(5912):379-382. doi:10.1126/science.1167140
31. Blanc G, Ogata H, Robert C, et al. Reductive genome evolution from the mother of Rickettsia. PLoS Genet. 2007;3(1):e14. doi:10.1371/journal.pgen.0030014
32. Lescot M, Audic S, Robert C, et al. The genome of Borrelia recurrentis, the agent of deadly louse-borne relapsing fever, is a degraded subset of tick-borne Borrelia duttonii. PLoS Genet. 2008;4(9):e1000185. doi:10.1371/journal.pgen.1000185
33. Gibson DG, Glass JI, Lartigue C, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. Science. 2010;329(5987):52-56. doi:10.1126/science.1190719
34. Yutin N, Wolf YI, Koonin EV. Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. Virology. 2014;466-467:38-52. doi:10.1016/j.virol.2014.06.032
35. Koonin EV, Krupovic M, Yutin N. Evolution of double-stranded DNA viruses of eukaryotes: from bacteriophages to transposons to giant viruses. Ann N Y Acad Sci. 2015;1341(1):10-24. doi:10.1111/nyas.12728
36. Wolf YI, Koonin EV. Genome reduction as the dominant mode of evolution. Bioessays. 2013;35(9):829-837. doi:10.1002/bies.201300037
37. Mahmoudabadi G, Milo R, Phillips R. Energetic cost of building a virus. Proc Natl Acad Sci U S A. 2017;114(22):E4324-E4333. doi:10.1073/pnas.1701670114
38. Rosenwasser S, Ziv C, Creveld SGV, Vardi A. Virocell Metabolism: Metabolic Innovations During Host-Virus Interactions in the Ocean. Trends Microbiol. 2016;24(10):821-832. doi:10.1016/j.tim.2016.06.006
39. Howard-Varona C, Lindback MM, Bastien GE, et al. Phage-specific metabolic reprogramming of virocells. ISME J. 2020;14(4):881-895. doi:10.1038/s41396-019-0580-z
40. Jeudy S, Rigou S, Alempic JM, Claverie JM, Abergel C, Legendre M. The DNA methylation landscape of giant viruses. Nat Commun. 2020;11(1):2657. doi:10.1038/s41467-020-16414-2
41. Van Etten JL, Agarkova IV, Dunigan DD. Chloroviruses. Viruses. 2019;12(1):20. doi:10.3390/v12010020

42. Hidalgo P, Anzures L, Hernández-Mendoza A, et al. Morphological, Biochemical, and Functional Study of Viral Replication Compartments Isolated from Adenovirus-Infected Cells. J Virol. 2016;90(7):3411-3427. doi:10.1128/JVI.00033-16
43. Legendre M, Lartigue A, Bertaux L, et al. In-depth study of Mollivirus sibericum, a new 30,000-y-old giant virus infecting Acanthamoeba. Proc Natl Acad Sci U S A. 2015;112(38):E5327-E5335. doi:10.1073/pnas.1510795112
44. Philippe N, Legendre M, Doutre G, et al. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science. 2013;341(6143):281-286. doi:10.1126/science.1239181
45. Iranzo J, Krupovic M, Koonin EV. The Double-Stranded DNA Virosphere as a Modular Hierarchical Network of Gene Sharing. mBio. 2016;7(4):e00978-16. doi:10.1128/mBio.00978-16
46. Yutin N, Koonin EV. Pandoraviruses are highly derived phycodnaviruses. Biol Direct. 2013;8:25. doi:10.1186/1745-6150-8-25
47. Yutin N, Koonin EV. Hidden evolutionary complexity of Nucleo-Cytoplasmic Large DNA viruses of eukaryotes. Virol J. 2012;9:161. doi:10.1186/1743-422X-9-161
48. Iranzo J, Krupovic M, Koonin EV. A network perspective on the virus world. Commun Integr Biol. 2017;10(2):e1296614. doi:10.1080/19420889.2017.1296614
49. Krupovic M, Yutin N, Koonin E. Evolution of a major virion protein of the giant pandoraviruses from an inactivated bacterial glycoside hydrolase. Virus Evol. 2020; veaa059. doi:10.1093/ve/veaa059
50. Krupovic M, Koonin EV. Multiple origins of viral capsid proteins from cellular ancestors. Proc Natl Acad Sci U S A. 2017;114(12):E2401-E2410. doi:10.1073/pnas.1621061114
51. Legendre M, Fabre E, Poirot O, et al. Diversity and evolution of the emerging Pandoraviridae family. Nat Commun. 2018;9(1):2285. doi:10.1038/s41467-018-04698-4
52. Gallot-Lavallée L, Blanc G, Claverie JM. Comparative Genomics of Chrysochromulina Ericina Virus and Other Microalga-Infecting Large DNA Viruses Highlights Their Intricate Evolutionary Relationship with the Established Mimiviridae Family. J Virol. 2017;91(14):e00230-17. doi:10.1128/JVI.00230-17
53. Claverie JM, Abergel C. Mimiviridae: An Expanding Family of Highly Diverse Large dsDNA Viruses Infecting a Wide Phylogenetic Range of Aquatic Eukaryotes. Viruses. 2018;10(9):506. doi:10.3390/v10090506