

Fundamental Difficulties Prevent the Reconstruction of the Deep Phylogeny of Viruses

Jean-Michel Claverie

▶ **To cite this version:**

Jean-Michel Claverie. Fundamental Difficulties Prevent the Reconstruction of the Deep Phylogeny of Viruses. Viruses, 2020, 12 (10), pp.1130. 10.3390/v12101130 . hal-02965304

HAL Id: hal-02965304
https://amu.hal.science/hal-02965304

Submitted on 13 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

viruses

MDPI

Opinion

Fundamental Difficulties Prevent the Reconstruction of the Deep Phylogeny of Viruses

Jean-Michel Claverie

Structural & Genomic Information Laboratory (IGS, UMR 7256), Mediterranean Institute of Microbiology (FR3479), Aix-Marseille University and CNRS, 13288 Marseille, France; jean-michel.claverie@univ-amu.fr

Received: 17 September 2020; Accepted: 3 October 2020; Published: 6 October 2020



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 isolations from unicellular eukaryotic hosts (heterotrophic and photosynthetic protozoans) revealed numerous viral types previously unexpected in terms of virion structure, gene content, or mode of replication. Complemented by large-scale metagenomic analyses, these discoveries have rekindled interest in the enigma of the origin of viruses, for which a description encompassing all their diversity remains not available. 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 "core" 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 conceptual arguments highlighting several fundamental limitations plaguing the reconstruction of the deep evolutionary history of viruses, and even more the identification of their unique or multiple origin(s). These arguments also underline the risk of establishing premature high level viral taxonomic classifications. Those limitations are direct consequences of the random mechanisms governing the reductive/retrogressive evolution of all obligate intracellular parasites.

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

1. Introduction

Since the serendipitous discovery of the tobacco mosaic virus by Dmitri Ivanovsky in 1892 [1], virology has mostly focused on those viruses responsible for (often dreadful) diseases of human, animals, or plants. Most of the research was then dedicated to the mechanisms of pathogenicity rather than to the physiology of the viruses as biological entities [2]. One noticeable exception was the study of bacteriophages that led to the basic concepts of modern cellular biology and to many of today's molecular biology tools [3]. About fifteen years ago, a second wind of virology occurred with the discovery of the first giant viruses infecting amoeba [4,5], rapidly followed by that of an unexpected diversity of related viruses associated with other protozoan and algal hosts [6–11]. Through isolation studies [12–16] complemented by large-scale environmental metagenomic explorations [17–19], a flurry of new viruses has since been uncovered exhibiting unexpected virion sizes and morphologies, unusual gene contents, or exotic modes of replication [12,20–23]. Beyond their unanticipated diversity, these "unconventional viruses" (not to be confused with the infectious agents now known as prions) [24,25] challenged the established borders between the viral and cellular worlds and revived interest in the study of the deep phylogeny of eukaryotic dsDNA viruses, right down to the question of their origin. However, these works are struggling to converge towards a unanimous scenario, due to the

difficulty of applying the usual molecular phylogeny approaches to the extreme diversity of viral gene contents, and to the low sequence similarity exhibited by the rare "core" proteins that they have in common [12,26–35]. Beyond these practical difficulties, this article will focus on even more fundamental limitations plaguing the reconstruction of the deep evolutionary history of all viruses.

2. All Viruses Are "Unconventional"

All viruses are unconventional, in the sense that this category of microorganisms lacks 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. To clarify the significance of this unique feature of viruses, I will take the counterexample of the cellular world which encompasses an enormous diversity of cell types, from the already highly diverse prokaryotes (Archaea and Bacteria domains), to the eukaryotes. Yet, all the "conventional" cell types 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 *Escherichia 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 [36], 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*) [37], ATP synthesis (e.g., *Rickettsia*) [38], and/or amino-acid or nucleotide synthetic pathways (e.g., *Chlamydiae* or *Tremblaya princeps*) [39,40].

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 for the sake of clarity), or even within the same kingdom. For instance, Mimivirus (with a 1.2 Mb genome) [5] is by no mean a "model" or a prototype for all the dsDNA viruses included in the recently defined "*Bamfordvirae*" kingdom [41] that also includes the adenoviruses (with a 30 kb genome) [42] or a new type of viral parasite, the virophages [20] (Table 1). In other words: 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.

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 [2,43]. Amazingly, the fundamental difficulty of formally defining viruses was already perceived by Lwoff in the early days of virology when the best he could do was to list the missing properties that precluded them to belong to the cellular world: viruses could not divide, could not synthetize ATP, and could not synthetize their proteins [44].

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

Family/Genus Name	DNA Polymerase	RNA Polymerase	Genome Size Range
In 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
In 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 DNA-dependent RNA polymerase is not packaged in the virion.

3. Viruses Display a Huge Gradation in "Absolute" Parasitism

One of the main properties 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 believe 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) [37–40]. 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 (e.g., by the presence of ribosomes).

At this point, it is interesting to notice 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 cases, for instance, supplementing a culture medium with a specific metabolite was found sufficient to turn an absolute intracellular parasitic bacterium into a free-living one (e.g., *Tropheryma*) [45]. In other cases, the parasitic organism is short of achieving free-living in a rich medium by hundreds of missing genes (e.g., *Tremblaya*) [40]. 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) [46,47], to giant viruses encoding, in addition to hundreds of their virion components, the blueprint of the transient intracellular factory used to synthetize them, as well as the regulatory elements required to hijack the systems unique to the host cell (e.g., the ribosomes) [12,20]. 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 [20], many protein translation components [48], and numerous biosynthetic pathways [11,12,19], down to nuclear polyomaviruses with 5 kb genomes encoding 5 proteins [46,47]. Such huge variations in genome sizes and gene contents are difficult to interpret in the context of a

unique one-fit-all evolutionary scenario driven by a fixed set of fitness constraints. A similar variation in genome sizes (14–735 kb) is also seen among dsDNA bacteriophages [49,50]. 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., [29,33,51,52]). In the following sections, I present several fundamental reasons why deep reconstructions of viral phylogenies might not be tractable beyond the level of individual virus families.

4. First Argument in Favor of a Retrogressive Evolutionary Scenario

Even if the gene contents of viruses (including those infecting the same host) appear largely uncorrelated, there is some order in this apparent chaos. For instance, there seems to be a strict hierarchy governing the presence of encoded DNA-dependent DNA polymerases and of DNA-dependent RNA polymerases in viral genomes (Table 1). As of today, all eukaryotic dsDNA viruses encoding their own RNA polymerase, also encode a DNA polymerase. If the RNA polymerase can be absent from viruses encoding a DNA polymerase, the converse is not true. In other words, DNA polymerases are only absent from viruses also lacking a DNA-dependent 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 arguments 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 DNA-dependent RNA polymerase that, most surprisingly, is not packaged in their particles, forcing them to initiate their cytoplasmic replication cycle by first recruiting nuclear functions [23,53].

However, the reason why the loss of the 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 virus 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 "unconventional" dsDNA virus. Interestingly, the presence of a DNA-dependent RNA polymerase strictly conditioned to that of a DNA polymerase is also observed in all known dsDNA bacteriophages, which suggests that it is not linked to the presence of a nucleus.

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 not contain any of the genes required for its replication or the production of virions (reviewed in [54]). 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 a 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 formal criteria) does not help in designing a rigorous definition that will include them all.

5. The Main Conceptual Difficulty Plaguing the Deep Phylogenetic Reconstruction of Viruses' 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 [35]).

I never personally 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 parts of a consortium). The first viruses— in the sense that we 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 Figures 1 and 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, top). A different representation of this tree then suggests that these 3 families originated from a common "parasitic" ancestor (Figure 1, bottom), a convergence obviously imposed by all tree-building algorithms.

These conclusions are of course totally erroneous, as shown in Figure 2, where I incorporated free-living relatives to the analysis of the same 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. Numerous whole genome comparisons have demonstrated that obligate intracellular parasitic bacteria 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 [37–40,55–57]. Interestingly, the 7 parasitic bacteria compared above (some of which encode close to 1000 proteins) 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 [58]. Without knowing the actual evolutionary history of these parasites (Figure 2), one might interpret these 100 core genes as those characteristic of their common parasitic lifestyle. This is of course wrong, as those genes were inherited from (and essential to) their free-living ancestors. On the other hand, these 100 core genes do not either constitute the entire genome of a hypothetical common ancestor (that had to possess a whole set of free-living functions), as it is most often concluded when similar phylogenetic reconstructions are applied to viruses. This then leads to the erroneous conclusion that obligate intracellular parasites (and thus viruses) evolved from simpler ancestors by acquiring genes instead of losing them [28,59].

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 be limited to the family level, i.e., to group of viruses the ancestry of which can be traced back to a common quasi-extant virus by multiple genes and physiological criteria (a situation similar to the *Alphaparasites*/*Rickettsiales* in Figures 1 and 2). For the same reasons explained above, the small number of core genes strictly shared by different eukaryotic dsDNA virus

families (3 within phylum *Nucleocytoviricota*, zero within kingdom *Bamfordvirae*) [33] (Table 1) should not be interpreted as a characteristic of their putative common ancestor, rather as the expected result of a random succession of genes losses starting from an ancestral genome impossible to reconstruct, or as evidence of different origins altogether.

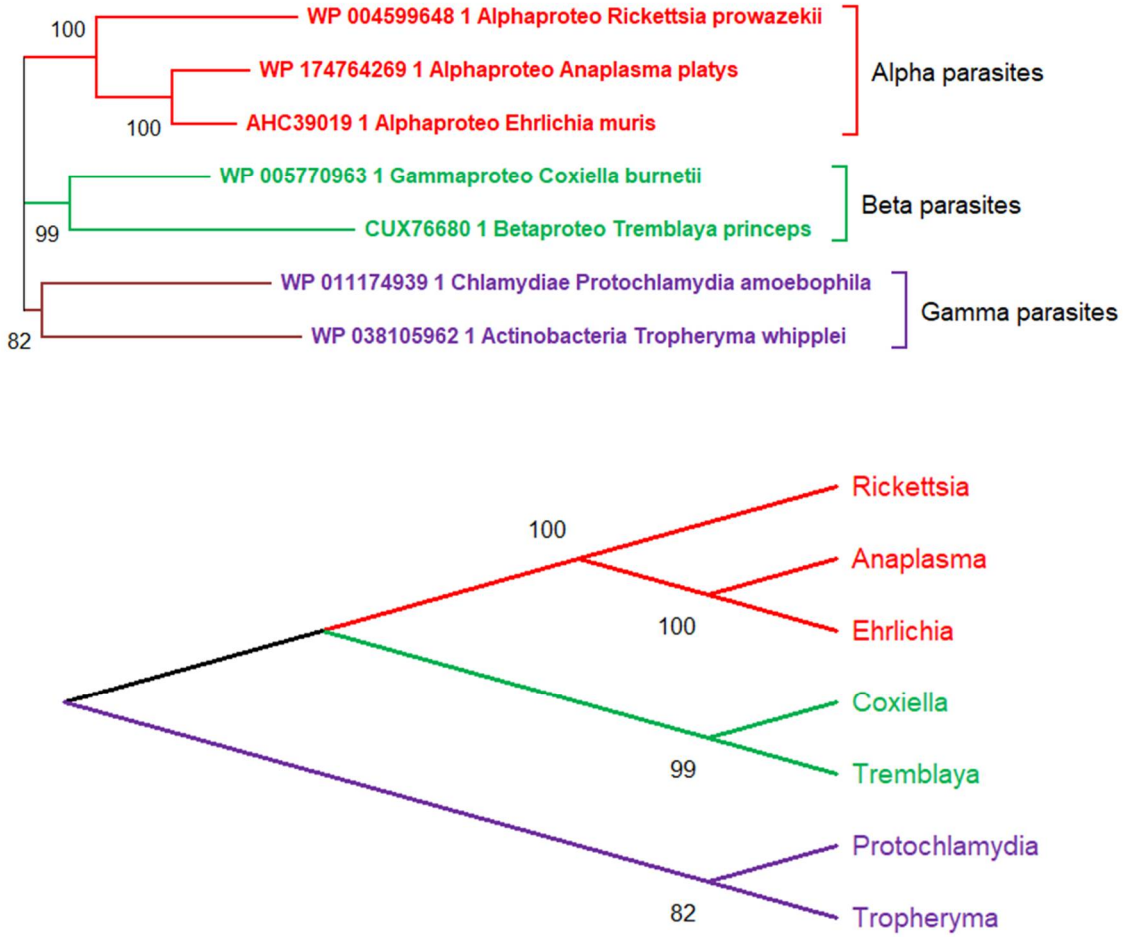


Figure 1. Erroneous phylogenetic relationships between seven obligate intracellular parasitic bacteria. **Top**: The neighbor-joining tree was generated from 1014 conserved sites in the multiple alignment of their DNA polymerase alpha subunits using the JTT substitution matrix. The protein NCBI identifiers 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 families of extant parasites derived. The true evolutionary history of these parasitic bacteria is shown in Figure 2.

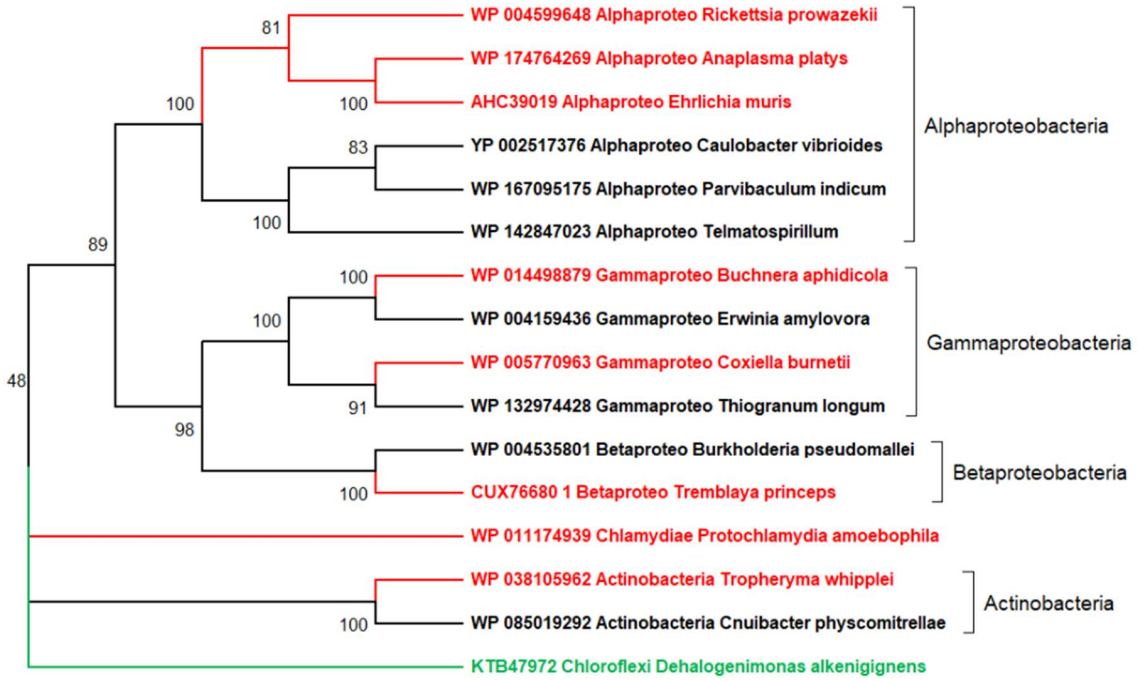


Figure 2. A more realistic representation of the origin and evolution of the obligate intracellular parasitic bacteria depicted in Figure 1. The neighbor-joining tree was generated from 974 conserved sites in the multiple alignment of 16 DNA polymerase alpha subunits using the JTT substitution matrix. 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 corresponds to a distant bacterium from the *Chloroflexi* phylum, used as outgroup. This tree suggests (with strong statistical support) that the 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 members 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 *Protochlamydia amoebophila* for which no free living relative could be found. *P. amoebophila* belongs to *Chlamydiae*, a phylum of highly diverse members all of which have—like viruses—an obligate intracellular lifestyle. In the 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.

6. The Random Walk of Gene Losses: Another Main Hurdle in the Reconstruction of Viruses' Evolution

As soon as an organism switches to the absolute parasitism lifestyle, the laws of neo-Darwinian selection, which apply to the usual conservative evolution of 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 lifestyle. "Once a parasite, always a parasite" appears to be one of the few absolutely respected mottos of microbial evolution [60]. As viruses are archetypes of (acellular) obligate intracellular parasites, I do believe that the irreversible succession of gene losses constitutes the dominant force in their evolution [2].

Given the central place held by the absence of the protein translation function in the definition of viruses [2,44], 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. Alternatively, this process of retrogressive evolution may have been initiated within an already established cellular parasite becoming a virus via the loss of protein translation.

During this phase, the neo-Darwinian selection process will continuously select the viruses for an optimized parasitic lifestyle generating more progenies at each round of 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) [61], by improving the efficiency of host infection (by innovating on virion structures and infection strategies) [12], by helping the host viability (thus increasing burst sizes) [62,63], or by using molecular defenses against viruses competing for the same host [64]. Such a complex web of evolutionary constraints is expected to generate a huge diversity of "optimal" solutions (sometimes involving moderate gene gains via horizontal transfers) resulting into the observed variety of viral gene contents without much apparent rationale.

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 the functional virus-encoded DNA replication or transcription machineries must be concomitantly compensated by an access to the cellular ones, in the host nucleus. The viruses must thus evolve a strategy to either transport their genomes to the intact nucleus [42,65], or make it functionally "leaky" [23,66], or even dissolve it altogether [67]. The passage from a purely cytoplasmic replication cycle to a nuclear one is thus a major step in the continuous retrogressive 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 genome reduction process, concerning viruses with vastly different genome sizes (Table 1).

To illustrate two main points, a diagram of virus evolution through genome reduction from a cell-like (non-virus) ancestor is represented in Figure 3. First, soon after a parasitic lifestyle is initiated, random gene losses generate very different genomes, from which common-to-all genes (the so-called "core" genes) have no reason to be maintained and could disappear rapidly. This corresponds to reality, where the discovery of new virus families steadily led to the reduction of the number of core genes [12,29,33,51,66–70]. However, while other authors attribute this phenomenon to the "high rate of horizontal transfer and fast sequence divergence of virus evolution" [71], it is in fact intrinsic to the reduction hypothesis. "Core" genes are borne to disappear as our knowledge of the viral diversity increases because the concept of virus simply does not imply the existence/conservation of any specific virus-encoded function. Core genes are not intrinsically "essential" to viruses, but are just the artefactual (provisional) consequence of our finite and incomplete samplings of the virosphere.

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 from which intermediate viral forms have disappeared or are not yet discovered. In absence of such intermediates, polyphyletic viral clades might erroneously appear monophyletic (eventually promoting irrelevant taxonomic clustering). 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 built from the most shared genes to be highly discordant. This has been a common finding [29,30,33,51,52,69,70], prompting authors to abandon tree-based phylogenetic reconstruction methods for network representations [68,71], to extrapolate homologies from non-significant sequence similarity [72], to invoke unwarranted combinations of

losses and horizontal gene transfers [33,70,73], to abandon the grail of a unique common ancestry [73], or to ignore whole virus families causing troubles [33].

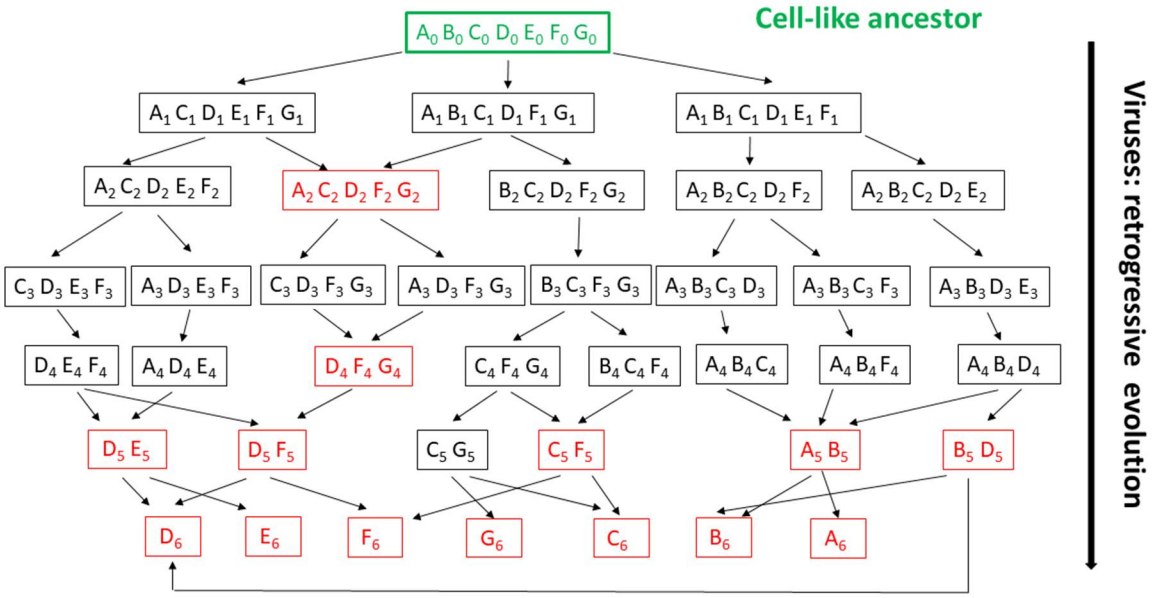


Figure 3. The "virus late" hypothesis: illustration of the intractable evolutionary scenarios resulting from random gene/function losses. A toy virus world is represented, starting from a hypothetical ancestral cell-like organism (level 0, w/o extant representative). Each box contains the abstract gene content inherited by a given virus (family) from its immediate ancestor. Red "genomes" indicate viruses with shared gene contents albeit possibly resulting from distinct evolutionary pathways. Random gene losses lead to very diverse overlaps of gene assortments (as in viruses) or to situations where no single "core gene" is shared by all virus family (here starting at level 3), as observed in actual viral genomes (in particular small ones). Individual genes recurring in multiple combinations (families) or ultimately remaining in the smallest genomes (level 6) are not more characteristic of the parasitic lifestyle 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 from different ancestors (DE, DF, BD). Their polyphyly will not be detected if some of their above ancestors are extinct or unknown. This graph illustrates the difficulty 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 parasites and the lack of associated free-living organisms to be used as references.

7. Conclusions

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 cell-like microorganisms, the only logically sound scenario if we respect the definition of viruses as "obligate intracellular parasites". 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 a microorganism having switched to an obligate intracellular parasitic lifestyle. Paradoxically, the evolutionary trajectory of a virus is much better defined by the way it has lost genes, rather 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 [74]. 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 dsDNA 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 contents 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., in [74]). It is nevertheless not a fool proof exercise as illustrated by Figures 1 and 2. The recent reclassification of certain members of the *Phycodnaviridae* family into that of *Mimiviridae* [10] (which could not have been possible prior to the discovery of Mimivirus and its relatives [6]) is a good illustration of the risk of freezing a taxonomic classification too early.

Having listed the fundamental limitations plaguing the deep phylogenetic reconstruction of viruses beyond the family level, I can only wonder about the merits of a spectacular fifteen-rank classification hierarchy recently adopted by the ICTV [41]. Families are now aggregated in a succession of 10 taxonomic levels (suborder, order, subclass, class, subphylum, phylum, etc.), the monophyly of which, according to the argument presented here, may remain forever beyond the realm of scientific evidence. In addition to being nearly impossible to pronounce or memorize, many of these abstract clades are dangerously suggesting totally unsupported related ancestries between families as different as *Mimiviridae*, *Adenoviridae*, *Phaeovirus*, and virophages (all 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 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. Ivanovski, D. *Über die Mosaikkrankheit der Tabakspflanze*; Izv. Imp. Akad. Nauk. 1892; 35, 67; Johnson, J., Translator; Phytopathological classics; American Phytopathological Society: St. Paul, MN, USA, 1892; pp. 27–30.
2. Claverie, J.-M.; Abergel, C. Giant viruses: The difficult breaking of multiple epistemological barriers. *Stud. Hist. Philos. Biol. Biomed. Sci.* **2016**, *59*, 89–99. [CrossRef] [PubMed]
3. Salmond, G.P.C.; Fineran, P.C. A century of the phage: Past, present and future. *Nat. Rev. Genet.* **2015**, *13*, 777–786. [CrossRef] [PubMed]
4. Scola, B.L. A Giant Virus in Amoebae. *Science* **2003**, *299*, 2033. [CrossRef]
5. Raoult, D.; Audic, S.; Robert, C.; Abergel, C.; Renesto, P.; Ogata, H.; La Scola, B.; Suzan, M.; Claverie, J.-M. The 1.2-Megabase Genome Sequence of Mimivirus. *Science* **2004**, *306*, 1344–1350. [CrossRef] [PubMed]
6. Claverie, J.-M.; Abergel, C. Mimiviridae: An Expanding Family of Highly Diverse Large dsDNA Viruses Infecting a Wide Phylogenetic Range of Aquatic Eukaryotes. *Viruses* **2018**, *10*, 506. [CrossRef] [PubMed]
7. Fischer, M.G.; Allen, M.J.; Wilson, W.H.; Suttle, C.A. Giant virus with a remarkable complement of genes infects marine zooplankton. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 19508–19513. [CrossRef] [PubMed]
8. Moniruzzaman, M.; LeCleir, G.R.; Brown, C.M.; Gobler, C.J.; Bidle, K.D.; Wilson, W.H.; Wilhelm, S.W. Genome of brown tide virus (AaV), the little giant of the Megaviridae, elucidates NCLDV genome expansion and host–virus coevolution. *Virology* **2014**, *467*, 60–70. [CrossRef] [PubMed]

9. Deeg, C.M.; Chow, C.-E.T.; Suttle, C.A. The kinetoplastid-infecting Bodo saltans virus (BsV), a window into the most abundant giant viruses in the sea. *eLife* **2018**, *7*, e33014. [CrossRef] [PubMed]
10. Gallot-Lavallée, L.; Blanc, G.; Claverie, J.-M. 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*, e00230-17. [CrossRef]
11. Schvarcz, C.R.; Steward, G.F. A giant virus infecting green algae encodes key fermentation genes. *Virology* **2018**, *518*, 423–433. [CrossRef] [PubMed]
12. Abergel, C.; Legendre, M.; Claverie, J.-M. The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. *FEMS Microbiol. Rev.* **2015**, *39*, 779–796. [CrossRef] [PubMed]
13. Aherfi, S.; La Scola, B.; Pagnier, I.; Raoult, D.; Colson, P. The expanding family Marseilleviridae. *Virology* **2014**, *466*, 27–37. [CrossRef] [PubMed]
14. Andreani, J.; Aherfi, S.; Khalil, J.Y.B.; Di Pinto, F.; Bitam, I.; Raoult, D.; Colson, P.; La Scola, B. Cedratvirus, a Double-Cork Structured Giant Virus, is a Distant Relative of Pithoviruses. *Viruses* **2016**, *8*, 300. [CrossRef] [PubMed]
15. Reteno, D.G.; Benamar, S.; Khalil, J.B.; Andreani, J.; Armstrong, N.; Klose, T.; Rossmann, M.; Colson, P.; Raoult, D.; La Scola, B. Faustovirus, an Asfarvirus-Related New Lineage of Giant Viruses Infecting Amoebae. *J. Virol.* **2015**, *89*, 6585–6594. [CrossRef] [PubMed]
16. Yoshikawa, G.; Blanc-Mathieu, R.; Song, C.; Kayama, Y.; Mochizuki, T.; Murata, K.; Ogata, H.; Takemura, M. Medusavirus, a Novel Large DNA Virus Discovered from Hot Spring Water. *J. Virol.* **2019**, *93*. [CrossRef]
17. Hingamp, P.; Grimsley, N.; Acinas, S.G.; Clerissi, C.; Subirana, L.; Poulain, J.; Ferrera, I.; Sarmento, H.; Villar, E.; Lima-Mendez, G.; et al. Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. *ISME J.* **2013**, *7*, 1678–1695. [CrossRef]
18. Schulz, F.; Alteio, L.; Goudeau, D.; Ryan, E.M.; Yu, F.B.; Malmstrom, R.R.; Blanchard, J.; Woyke, T. Hidden diversity of soil giant viruses. *Nat. Commun.* **2018**, *9*, 4881. [CrossRef]
19. Needham, D.M.; Yoshizawa, S.; Hosaka, T.; Poirier, C.; Choi, C.J.; Hehenberger, E.; Irwin, N.A.T.; Wilken, S.; Yung, C.-M.; Bachy, C.; et al. A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 20574–20583. [CrossRef]
20. Claverie, J.-M.; Abergel, C. Mimivirus and its Virophage. *Annu. Rev. Genet.* **2009**, *43*, 49–66. [CrossRef] [PubMed]
21. Fischer, M.G.; Hackl, T. Host genome integration and giant virus-induced reactivation of the virophage mavirus. *Nat. Cell Biol.* **2016**, *540*, 288–291. [CrossRef]
22. Schrad, J.R.; Abrahão, J.S.; Cortines, J.R.; Parent, K.N. Structural and Proteomic Characterization of the Initiation of Giant Virus Infection. *Cell* **2020**, *181*, 1046–1061. [CrossRef] [PubMed]
23. Fabre, E.; Jeudy, S.; Santini, S.; Legendre, M.; Trauchessec, M.; Couté, Y.; Claverie, J.-M.; Abergel, C. Noumeavirus replication relies on a transient remote control of the host nucleus. *Nat. Commun.* **2017**, *8*, 15087. [CrossRef] [PubMed]
24. Gajdusek, D.C. Unconventional viruses and the origin and disappearance of kuru. *Science* **1977**, *197*, 943–960. [CrossRef] [PubMed]
25. Igel-Egalon, A.; Bohl, J.; Moudjou, M.; Herzog, L.; Reine, F.; Rezaei, H.; Béringue, V. Heterogeneity and Architecture of Pathological Prion Protein Assemblies: Time to Revisit the Molecular Basis of the Prion Replication Process? *Viruses* **2019**, *11*, 429. [CrossRef] [PubMed]
26. Claverie, J.-M.; Abergel, C. Open Questions About Giant Viruses. *Adv. Virus Res.* **2013**, *85*, 25–56. [CrossRef] [PubMed]
27. Nasir, A.; Sun, F.-J.; Kim, K.M.; Caetano-Anollés, G. Untangling the origin of viruses and their impact on cellular evolution. *Ann. N. Y. Acad. Sci.* **2015**, *1341*, 61–74. [CrossRef] [PubMed]
28. Koonin, E.V.; 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*, 10–24. [CrossRef]
29. Forterre, P.; Gaïa, M. Giant viruses and the origin of modern eukaryotes. *Curr. Opin. Microbiol.* **2016**, *31*, 44–49. [CrossRef] [PubMed]
30. Koonin, E.V.; Yutin, N. Multiple evolutionary origins of giant viruses. *F1000Research* **2018**, *7*, 1840. [CrossRef]
31. Colson, P.; Levasseur, A.; La Scola, B.; Sharma, V.; Nasir, A.; Pontarotti, P.; Caetano-Anollés, G.; Raoult, D. Ancestrality and Mosaicism of Giant Viruses Supporting the Definition of the Fourth TRUC of Microbes. *Front. Microbiol.* **2018**, *9*, 2668. [CrossRef]

32. Koonin, E.V.; Yutin, N. Evolution of the Large Nucleocytoplasmic DNA Viruses of Eukaryotes and Convergent Origins of Viral Gigantism. *Adv. Virus Res.* **2018**, *103*, 167–202. [CrossRef] [PubMed]
33. Guglielmini, J.; Woo, A.C.; Krupovic, M.; Forterre, P.; Gaia, M. Diversification of giant and large eukaryotic dsDNA viruses predated the origin of modern eukaryotes. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 19585–19592. [CrossRef] [PubMed]
34. Mughal, F.; Nasir, A.; Caetano-Anollés, G. The origin and evolution of viruses inferred from fold family structure. *Arch. Virol.* **2020**, *165*, 2177–2191. [CrossRef] [PubMed]
35. Nasir, A.; Romero-Severson, E.; Claverie, J.M. Investigating the concept and origin of viruses. *Trends Microbiol.* **2020**, in press.
36. Monod, J.; Jacob, F. General Conclusions: Teleonomic Mechanisms in Cellular Metabolism, Growth, and Differentiation. *Cold Spring Harb. Symp. Quant. Biol.* **1961**, *26*, 389–401. [CrossRef]
37. Pagnier, I.; Yutin, N.; Croce, O.; Makarova, K.S.; Wolf, Y.I.; Benamar, S.; Raoult, D.; Koonin, E.V.; La Scola, B. Babela massiliensis, a representative of a widespread bacterial phylum with unusual adaptations to parasitism in amoebae. *Boil. Direct* **2015**, *10*, 13. [CrossRef] [PubMed]
38. Driscoll, T.; Verhoeve, V.I.; Guillotte, M.L.; Lehman, S.S.; Rennoll, S.A.; Beier-Sexton, M.; Rahman, M.S.; Azad, A.F.; Gillespie, J.J. Wholly Rickettsia! Reconstructed Metabolic Profile of the Quintessential Bacterial Parasite of Eukaryotic Cells. *mBio* **2017**, *8*, e00859-17. [CrossRef] [PubMed]
39. Omsland, A.; Sixt, B.S.; Horn, M.; Hackstadt, T. Chlamydial metabolism revisited: Interspecies metabolic variability and developmental stage-specific physiologic activities. *FEMS Microbiol. Rev.* **2014**, *38*, 779–801. [CrossRef] [PubMed]
40. López-Madrigal, S.; Latorre, A.; Porcar, M.; Moya, A.; Gil Benso, R. Mealybugs nested endosymbiosis: Going into the 'matryoshka' system in *Planococcus citri* in depth. *BMC Microbiol.* **2013**, *13*, 74. [CrossRef]
41. International Committee on Taxonomy of Viruses Executive Committee; Gorbalenya, A.E. The new scope of virus taxonomy: Partitioning the virosphere into 15 hierarchical ranks. *Nat. Microbiol.* **2020**, *5*, 668–674. [CrossRef]
42. Hidalgo, P.; Anzures, L.; Hernandez-Mendoza, A.; Guerrero, A.; Wood, C.; Valdés, M.; Dobner, T.; Gonzalez, R.A. Morphological, Biochemical, and Functional Study of Viral Replication Compartments Isolated from Adenovirus-Infected Cells. *J. Virol.* **2016**, *90*, 3411–3427. [CrossRef] [PubMed]
43. Dupré, J.; Guttinger, S. Viruses as living processes. *Stud. Hist. Philos. Biol. Biomed. Sci.* **2016**, *59*, 109–116. [CrossRef]
44. Lwoff, A.; Nicholas, D.J.D. The Concept of Virus. *Microbiology* **1957**, *17*, 239–253. [CrossRef] [PubMed]
45. Renesto, P.; Crapoulet, N.; Ogata, H.; La Scola, B.; Vestris, G.; Claverie, J.-M.; Raoult, D. Genome-based design of a cell-free culture medium for *Tropheryma whipplei*. *Lancet* **2003**, *362*, 447–449. [CrossRef]
46. Rector, A.; Van Ranst, M. Animal papillomaviruses. *Virology* **2013**, *445*, 213–223. [CrossRef] [PubMed]
47. Calvignac-Spencer, S.; Polyomaviridae Study Group of the International Committee on Taxonomy of Viruses; Feltkamp, M.C.; Daugherty, M.D.; Moens, U.; Ramqvist, T.; Johne, R.; Ehlers, B. A taxonomy update for the family Polyomaviridae. *Arch. Virol.* **2016**, *161*, 1739–1750. [CrossRef]
48. Abrahao, J.; Silva, L.; Silva, L.S.; Khalil, J.Y.B.; Rodrigues, R.A.L.; Arantes, T.; Assis, F.; Boratto, P.; Andrade, M.; Kroon, E.G.; et al. Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. *Nat. Commun.* **2018**, *9*, 1–12. [CrossRef]
49. Petrovski, S.; Dyson, Z.A.; Seviour, R.J.; Tillett, D. Small but Sufficient: The Rhodococcus Phage RRH1 Has the Smallest Known Siphoviridae Genome at 14.2 Kilobases. *J. Virol.* **2011**, *86*, 358–363. [CrossRef]
50. Al-Shayeb, B.; Sachdeva, R.; Chen, L.-X.; Ward, F.; Munk, P.; Devoto, A.; Castelle, C.J.; Olm, M.R.; Bouma-Gregson, K.; Amano, Y.; et al. Clades of huge phages from across Earth's ecosystems. *Nat. Cell Biol.* **2020**, *578*, 425–431. [CrossRef] [PubMed]
51. Yutin, N.; Wolf, Y.I.; Raoult, D.; Koonin, E.V. Eukaryotic large nucleo-cytoplasmic DNA viruses: Clusters of orthologous genes and reconstruction of viral genome evolution. *Virol. J.* **2009**, *6*, 223. [CrossRef] [PubMed]
52. Koonin, E.V.; Yutin, N. Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses. *Intervirology* **2010**, *53*, 284–292. [CrossRef]
53. Allen, M.J.; A Howard, J.; Lilley, K.S.; Wilson, W.H. Proteomic analysis of the EhV-86 virion. *Proteome Sci.* **2008**, *6*, 11. [CrossRef] [PubMed]
54. Drezen, J.-M.; Leobold, M.; Bézier, A.; Huguet, E.; Volkoff, A.-N.; A Herniou, E. Endogenous viruses of parasitic wasps: Variations on a common theme. *Curr. Opin. Virol.* **2017**, *25*, 41–48. [CrossRef] [PubMed]

55. Moran, N.A.; McLaughlin, H.J.; Sorek, R. The Dynamics and Time Scale of Ongoing Genomic Erosion in Symbiotic Bacteria. *Science* **2009**, *323*, 379–382. [CrossRef] [PubMed]
56. Blanc, G.; Ogata, H.; Robert, C.; Audic, S.; Suhre, K.; Vestris, G.; Claverie, J.-M.; Raoult, D. Reductive Genome Evolution from the Mother of Rickettsia. *PLoS Genet.* **2007**, *3*, e14. [CrossRef]
57. Lescot, M.; Audic, S.; Robert, C.; Nguyen, T.T.; Blanc, G.; Cutler, S.J.; Wincker, P.; Couloux, A.; Claverie, J.-M.; Raoult, D.; 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*, e1000185. [CrossRef]
58. Gibson, D.G.; Glass, J.I.; Lartigue, C.; Noskov, V.N.; Chuang, R.-Y.; Algire, M.A.; Benders, G.A.; Montague, M.G.; Ma, L.; Moodie, M.M.; et al. Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science* **2010**, *329*, 52–56. [CrossRef]
59. Yutin, N.; Wolf, Y.I.; Koonin, E.V. Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. *Virology* **2014**, *467*, 38–52. [CrossRef]
60. Wolf, Y.I.; Koonin, E.V. Genome reduction as the dominant mode of evolution. *BioEssays* **2013**, *35*, 829–837. [CrossRef]
61. Mahmoudabadi, G.; Milo, R.; Phillips, R. Energetic cost of building a virus. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E4324–E4333. [CrossRef]
62. Rosenwasser, S.; Ziv, C.; Van Creveld, S.G.; Vardi, A. Virocell Metabolism: Metabolic Innovations During Host–Virus Interactions in the Ocean. *Trends Microbiol.* **2016**, *24*, 821–832. [CrossRef]
63. Howard-Varona, C.; Lindback, M.M.; Bastien, G.E.; Solonenko, N.; Zayed, A.A.; Jang, H.; Andreopoulos, B.; Brewer, H.M.; Del Rio, T.G.; Adkins, J.N.; et al. Phage-specific metabolic reprogramming of virocells. *ISME J.* **2020**, *14*, 881–895. [CrossRef]
64. Jeudy, S.; Rigou, S.; Alempic, J.-M.; Claverie, J.-M.; Abergel, C.; Legendre, M. The DNA methylation landscape of giant viruses. *Nat. Commun.* **2020**, *11*, 2657. [CrossRef]
65. Van Etten, J.L.; Agarkova, I.; Dunigan, D.D. Chloroviruses. *Viruses* **2019**, *12*, 20. [CrossRef]
66. Legendre, M.; Lartigue, A.; Bertaux, L.; Jeudy, S.; Bartoli, J.; Lescot, M.; Alempic, J.-M.; Ramus, C.; Bruley, C.; Labadie, K.; et al. In-depth study of Mollivirus sibericum, a new 30,000-y-old giant virus infecting Acanthamoeba. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E5327–E5335. [CrossRef]
67. Philippe, N.; Legendre, M.; Doutre, G.; Couté, Y.; Poirot, O.; Lescot, M.; Arslan, D.; Seltzer, V.; Bertaux, L.; Bruley, C.; et al. Pandoraviruses: Amoeba Viruses with Genomes Up to 2.5 Mb Reaching That of Parasitic Eukaryotes. *Science* **2013**, *341*, 281–286. [CrossRef]
68. Iranzo, J.; Krupovic, M.; Koonin, E.V. The Double-Stranded DNA Virosphere as a Modular Hierarchical Network of Gene Sharing. *mBio* **2016**, *7*. [CrossRef]
69. Yutin, N.; Koonin, E.V. Pandoraviruses are highly derived phycodnaviruses. *Biol. Direct* **2013**, *8*, 25. [CrossRef]
70. Yutin, N.; Koonin, E.V. Hidden evolutionary complexity of Nucleo-Cytoplasmic Large DNA viruses of eukaryotes. *Virol. J.* **2012**, *9*, 161. [CrossRef]
71. Iranzo, J.; Krupovic, M.; Koonin, E.V. A network perspective on the virus world. *Commun. Integr. Biol.* **2017**, *10*, e1296614. [CrossRef] [PubMed]
72. Krupovic, M.; Yutin, N.; Koonin, E.V. Evolution of a major virion protein of the giant pandoraviruses from an inactivated bacterial glycoside hydrolase. *Virus Evol.* **2020**, veaa059. [CrossRef]
73. Krupovic, M.; Koonin, E.V. Multiple origins of viral capsid proteins from cellular ancestors. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E2401–E2410. [CrossRef]
74. Legendre, M.; Fabre, E.; Poirot, O.; Jeudy, S.; Lartigue, A.; Alempic, J.-M.; Beucher, L.; Philippe, N.; Bertaux, L.; Christo-Foroux, E.; et al. Diversity and evolution of the emerging Pandoraviridae family. *Nat. Commun.* **2018**, *9*, 2285. [CrossRef]



© 2020 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).