

Viruses in Biology

Purificación López-García · David Moreira

Published online: 1 September 2012
© Springer Science+Business Media, LLC 2012

Abstract During the first half of the twentieth century, many scientists considered viruses the smallest living entities and primitive life forms somehow placed between the inert world and highly evolved cells. The development of molecular biology in the second half of the century showed that viruses are strict molecular parasites of cells, putting an end to previous virocentric debates that gave viruses a primeval role in the origin of life. Recent advances in comparative genomics and metagenomics have uncovered a vast viral diversity and have shown that viruses are active regulators of cell populations and that they can influence cell evolution by acting as vectors for gene transfer among cells. They have also fostered a revival of old virocentric ideas. These ideas are heterogeneous, extending from proposals that consider viruses functionally as living beings and/or as descendants of viral lineages that preceded cell evolution to other claims that consider viruses and/or some viral families a fourth domain of life. In this article, we revisit these virocentric ideas and analyze the place of viruses in biology in light of the long-standing dichotomic debate between metabolist and geneticist views which hold, respectively, that self-maintenance (metabolism) or self-replication and evolution are the primeval features of life. We argue that whereas the epistemological discussion about whether viruses are alive or not and whether some virus-like replicators precede the first cells is a matter of debate that can be understood within the metabolism-versus-genes dialectic; the claim that viruses form a fourth domain in the tree of life can be solidly refuted by proper molecular phylogenetic analyses and needs to be removed from this debate.

Keywords Tree of life · Metabolism · Self-replication · Origin of life · Fourth domain of life · Giant viruses · Horizontal gene transfer · Definition of life · Evolution

Introduction

Viruses were discovered at the end of the nineteenth century following several observations. In 1892, the Russian botanist Dmitri Ivanovsky noticed that suspensions of plant tissues afflicted with mosaic tobacco disease were still infectious after passage through ceramic filters that retained bacteria. He thought that his filter had most probably leaked and that the causative agent was a bacterium. In 1898, Martinus Beijerinck made a similar observation and thought that the infectious agent, the “virus,” was the liquid. The same year Friedrich Loeffler and Paul Frosch, while studying the cause of the foot-and-mouth disease, determined that the causative agent was also constituted by “filterable” particles that were nonetheless retained by filters of finer grains than those used for bacteria (Murphy 2011). Viruses had been identified as extremely small infectious particles.

After one century of research, notably thanks to the development of molecular biology and electron microscopy, our knowledge about viruses, their nature, their diversity, their infectious cycles, and their role in biology has enormously increased. Viruses are strict molecular parasites which depend on a cell to develop their reproductive cycle. Viral infectious particles, or virions, are composed of a nucleic acid (DNA or RNA) surrounded by a protein shell (the capsid) and sometimes by an additional lipid envelope. The infective cycle begins when a virion injects or releases its nucleic acid in a cell, leaving its capsid and envelope outside. In eukaryotic cells, virions can be incorporated in endocytotic vesicles where their capsid is degraded before the nucleic acid gets released into the cytoplasm. Once in

P. López-García (✉) · D. Moreira
Unité d'Ecologie, Systématique et Evolution,
CNRS UMR8079, Université Paris-Sud,
CNRS - Centre National de la Recherche Scientifique,
91405 Orsay, France
e-mail: puri.lopez@u-psud.fr

the cell, the viral genome is transcribed (although some viral RNA genomes may act directly as mRNAs) and translated by the cell using the cellular machinery and energetic resources. In this way, the proteins required for the viral genome replication and for the capsid encoded by the virus are synthesized. At the same time, the viral genome gets replicated (directly or via nucleic acid intermediates), generally by specific viral polymerases and using cell resources. Capsid proteins self-assemble spontaneously encapsulating viral genomes to produce novel infective particles that will be released from the cell. If the infection results in a high number of virions rapidly synthesized, the cell lyses, liberating all the virions. Such cycles are called lytic. Sometimes, the viral genomes can be incorporated in the host genome and remain there silently, being reproduced with the host cell for generations until an external signal, usually some kind of stress, activates a lytic cycle. Such silent reproductive cycles are called lysogenic. Some viruses, called satellite viruses, require the co-infection with another (helper) virus to complete their reproductive cycle.

Viruses are extremely diverse. They are classified according to the nature of their genome (DNA or RNA, double- or single-stranded, positive or negative ssRNA, ssRNA requiring a DNA intermediate—retroviruses—or dsDNA requiring an RNA intermediate), their shape, capsid structure, presence or absence of an envelope, presence of additional structures (e.g., in head-and-tail viruses), or the type of organisms they parasitize. These features are used to define viral species, which are named and classified following a Linnaean-type hierarchical system according to the rules of the International Committee on Taxonomy of Viruses (ICTV). For the ICTV, viruses are strict molecular parasites that “possess some of the properties of living systems such as having a genome and being able to adapt to a changing environment,” and a virus species is “a polythetic class (a group that cannot be defined on the basis of any single shared character, but on overlapping combinations of characters shared by some of its members) of viruses that constitute a replicating lineage and occupy a particular ecological niche” (van Regenmortel 2000). In the latest release of its Master Species List, the ICTV recognizes 2,475 viral species, distributed in 395 genera, 22 subfamilies, 94 families, and six orders (<http://www.ictvonline.org>).

However, as in the case of prokaryotic and eukaryotic microorganisms, the number of described species is far less than the real species richness in nature. In recent years, a variety of approaches, including most particularly the exponentially growing genomics and metagenomics (or community genomics, the study of collective genomes from environmental samples), reveals not only a large diversity of viruses in nature, but also their influence in general ecology and the evolution of organismal lineages.

Viruses Are Important Players in Ecology and Evolution

Viruses are extraordinarily diverse and abundant in the environment. In ocean plankton, pioneer studies by DNA-staining of ultra-filtrates supposed to be free of bacteria suggested that viral particles (virions) can be up to an order of magnitude more abundant than cells (Fuhrman 1999). Subsequently, metagenomic analysis of those cell-free fractions by direct DNA or retrotranscribed-DNA sequencing revealed immense viral genetic diversity (Culley et al. 2003; Edwards and Rohwer 2005). Metagenomic analyses of viruses have largely expanded since those initial studies, leading to the discovery of novel groups of viruses and virus-like agents that collectively constitute a huge genetic reservoir (Kristensen et al. 2010; Suttle 2007). Due to their abundance and to the effects that they have on infected populations of cells, they play important roles in nutrient cycling, sinking rates, and phytoplankton bloom control (Danovaro et al. 2011; Fuhrman 1999). Viruses control cell populations by inducing cell lysis, which not only contributes to fostering biogeochemical turnover but also to maintaining biodiversity. Indeed, the strong demographic decrease caused in dominant cell populations by viral lysis (the so-called kill-the-winner mechanism) permits other, less competitive species to coexist at intermediate frequencies, resulting in the persistence of a large variety of species (Rodriguez-Valera et al. 2009; Suttle 2007). Viruses also contribute to controlling populations by affecting their evolutionary ecology through “Red Queen” effects, i.e., generating an arms race involving the continuous evolution of resistance by hosts to novel virus variants. A remarkable example is the viral induction of cell cycle changes in the photosynthetic picoeukaryote *Emiliana huxleyi*. Recently, giant phycodnaviruses were shown to infect and lyse only the algal diploid stage, thereby promoting a change from a diploid non-motile to a haploid motile and virus-resistant phase—a “Cheshire cat” escape strategy (Frada et al. 2008).

The discovery of novel viruses comes not only from high-throughput metagenomic sequencing, but also from classical studies of viruses infecting cellular lineages not previously studied. The described viral diversity is dominated by viruses infecting humans, cattle, or plants of agricultural interest. However, in the past 20 years, important progress was made in the description of viruses infecting the third domain of life, the Archaea, which had received little or no attention before. Studies on viruses infecting hyperthermophilic archaea revealed an unsuspected variety of new viral families, including many novel morphotypes. Some of these viruses can experience morphological changes when exposed to high temperatures—a sort of “developmental cycle” due to conformational protein changes—which make them infective only at temperatures where their host is able to grow (Prangishvili et al. 2006). In

more recent years, a remarkable discovery has been that of giant viruses with very large genomes (over 300 Kbp and up to 1.2 Mbp) infecting amoeba and other microbial eukaryotes (protists). Some of these genomes exceed the size of some parasitic bacterial genomes encoding several hundreds of proteins (Arslan et al. 2011; Boyer et al. 2009; Raoult et al. 2004; Van Etten 2011).

Viruses are not only abundant, diverse, and important for ecology, they play a significant role in the evolution of their hosts. In addition to the selective pressure that they exert on cell populations, as mentioned above, they foster the evolution of genes and genomes and mobilize genes across lineages. Indeed, contrary to recent claims asserting that “viruses have been neglected by evolutionary biologists” (Raoult and Forterre 2008), viruses have served for decades as models in population genetics—often from an epidemiology standpoint—because they evolve fast and have large population sizes. Their increased evolutionary rate is due partly to the fact that many viral polymerases are error prone but also to the large number of generations that can occur in very short time spans, for instance, as a consequence of the successive infection of cells in the same population or organism. Viral models thus allow testing predictions made by different hypotheses in population genetics (e.g., Gojobori et al. 1990; Lauring and Andino 2010).

In addition, comparative genomics and molecular phylogenetic analyses clearly show that viruses are active vehicles for horizontal gene transfer (HGT). Viral genomes or genome fragments (DNA or retrotranscribed RNA) can recombine with host DNA, for instance during lysogenic stages. Fragments of viral DNA can also be incorporated in prokaryotic genomes between short palindromic sequences in regions known as CRISPRs (clustered regularly interspaced short palindromic repeats), which provide immunity to bacteria and archaea against specific viruses (Horvath and Barrangou 2010). Genes from the host can be incorporated in viral genomes during recombination or conversely, genes of foreign origin (either from viruses or from distant cellular donors transported in viral genomes) can get inserted into cell genomes. In this way, viruses may promote host evolution by mediating the transfer of genes among cellular lines, which is an extensive phenomenon in evolution (Gogarten and Townsend 2005); or by promoting recombination of cellular genes (Zeidner et al. 2005). Some cellular genes, occasionally leading to innovations, appear to come from viruses, such as the genes encoding the telomerase enzyme (Eickbush 1997; Nakamura et al. 1997); or syncytin (of possible retroviral origin), found in placental mammals (Dupressoir et al. 2009). However, as we have seen, genes evolve fast in viral genomes, so determining whether viral genes are truly of viral origin or whether they are cellular genes which have

evolved beyond recognition is extremely difficult. On the contrary, there is no doubt that cellular genes are captured by viral genomes. Cellular genes incorporated into viral genomes through recombination may have been acquired accidentally and gotten lost over generations, but they may be transferred to other cells during their residence in the viral genome. However, cellular genes captured by viral genomes may confer an adaptive advantage, for instance, during infection. A clear example is genes encoding elements of photosystems II (Sullivan et al. 2006) and I (Sharon et al. 2009) found in many cyanophages (viruses infecting the photosynthetic cyanobacteria), which are expressed during phage infection, providing some type of benefit.

The increasing recognition that viruses are important in ecology and evolution and especially the discovery of giant viruses having very large genomes encoding hundreds of proteins have revived a long-standing debate about the role of viruses in biology and the origin of life, attributing to viruses a key or primeval role. However, this debate is fed by many confounding elements. In the following, we will try to clarify them to distinguish facts and concepts from hypotheses and from unfounded speculation. But first, what is the debate about?

Viruses in Origin-of-Life Thinking

Historically, viruses were discovered at a time when the conditions for a scientific investigation into the origin of life on Earth were met. Louis Pasteur had refuted the idea of continuous spontaneous generation in the 1860s and Charles Darwin had published his *On the Origin of Species* in 1859, which inevitably led to questions of how the first life forms had emerged. At the onset of the twentieth century, immense progress was made in organic chemistry, biochemistry, and cytology. It now seemed possible to approach a question that had remained too complex and intractable until then (Fry 2006). Very soon, two clear ideological currents emerged in the field. For some scientists, self-replication (making copies of itself) was the crucial starting point for life; for others, it was self-maintenance or, in other words, metabolism. This chicken-and-egg dichotomy has persisted since then, although formulated in different terms. In the terminology of the time, it came down to a nucleus versus cytoplasm debate. Accordingly, hypotheses on the origin of life subscribed either to nucleocentric (self-replication first) or cytoplasmic (self-maintenance first) views of life.

When viruses were discovered, they were very mysterious, and due to their small sizes and their infective capacities, many scientists explicably considered them as the simplest living entities. As a consequence, they were incorporated in the debate about the origin of life for several decades, although with various different connotations: the

virus as a metaphor for the simplest form of life, the virus as a functional model (an independent existing gene), and collectively, the viruses as a phylogenetic lineage with historical continuity that could be placed between the chemical world and the first cells (Podolsky 1996). As reviewed by Podolsky (1996), the importance of viruses for models of the origin of life and of the concept of what life is has varied over time, with different successive tendencies.

First, until the mid-1930s, the idea of a virus-centered origin-of-life was on the rise. In 1914, the American psychophysiological Leonard Troland proposed that the first life form might have been an “enzyme or organic catalyst” (Troland 1914) although little later, he spoke of a “genetic enzyme” and identified it with nucleic acids and proteins in the nucleus (Troland 1917). Herman Muller simplified Troland’s ideas and coined the term “gene” to refer to Troland’s “genetic enzyme” (Fry 2006). In 1922, Muller drew a clear conceptual link between virus and gene, saying that “there is no distinction between genes and them [viruses]” (Muller 1922); and in 1929, he openly proposed that the first living organism was a primitive gene (Muller 1929). John B. S. Haldane, in his essay on the origin of life (Haldane 1929), extended that operational view to a more phylogenetic view, asserting that “life may have remained in the virus stage for many millions of years before a suitable assemblage of elementary units was brought together in the first cell.” Similarly, Alexander and Bridges conceived of self-copying entities such as genes and viruses as the simplest components necessary to life and divided living beings into two taxonomic categories: “Cytobiontia” (cellular organisms) and “Ultrabiontia” (viruses) (Alexander and Bridges 1928). The syllogism “smallest = virus, smallest = first, so that virus = first” (Beutner 1938) was generally accepted during those early years; the nucleocentric view of the origin of life was also virocentric.

Second, during the 1930s and subsequent years, an opposite trend gained favor, largely due to the extensive work on the origin of life by Alexander I. Oparin. Oparin conceived the origin of life from a biochemical or “colloidal chemistry” perspective (Oparin 1938). For Oparin, life was a self-regulating system of catalytic reactions; he was a cytoplasmist for whom metabolism was the primary essence of life. In addition to Oparin’s influential work, several investigators, including Robert G. Green (1932) and most notably, André Lwoff (1943; 1957) questioned the idea that because they are small and simple, viruses are primitive. Viruses could be just products of regressive evolution due to their extreme parasitic nature. This opinion was increasingly accepted, and even Haldane affirmed that “most evolutionary change has been degenerative” (Haldane 1932). Although Haldane identified life with molecular self-reproduction, he also said that in a true living system, the function of any part including genes depended on the cooperation of all other parts (Fry 2006).

Viruses and nucleocentric views of life regained credence again after DNA was shown to be the genetic material (Avery et al. 1944) and after the determination of the DNA structure, which suggested an elegant self-copying mechanism (Watson and Crick 1953). It was also shown that the nucleic acid component of viruses was the infectious component (Hershey and Chase 1952). Consequently, the association of virus–nucleic acid–gene was easy to make. Even so, the initial notion that viruses were phylogenetically the most primitive organisms on Earth was abandoned in the 1950s in favor of an operational view according to which viruses were seen as metaphors for “living genes” (Podolsky 1996).

Finally, virocentric ideas on the origin-of-life were largely abandoned in the 1960s and throughout the remainder of the twentieth century due to two major factors. The first had to do with the nature of viruses themselves, since advances in biochemistry and molecular biology demonstrated that viruses were strict molecular parasites. Viruses were just biological entities able, like genes, to evolve but unable to self-replicate and to self-sustain, a position still held by the ICTV (van Regenmortel 2000, 2008). The second factor had to do with the discovery of a better alternative as primitive living entity: ribozymes. Ribozymes are RNAs having catalytic activity that therefor display dual functions as informative polymers and catalysts. Furthermore, the universally conserved ribosome responsible for protein synthesis was shown to be a ribozyme. These discoveries led to the development of the “RNA world” as a powerful model of early life evolution (for a review, see Orgel 2004; Robertson and Joyce 2010). RNA replaced viruses in origin-of-life thinking: nucleocentric views on the origin of life were no longer virocentric.

Novel Virocentric Hypotheses

At the beginning of the twenty-first century, a renaissance of old virocentric ideas has taken place and materialized in a series of heterogeneous proposals. Curiously, this revival occurs at both an operational level, with viruses seen functionally as living beings (Raoult and Forterre 2008); and at a phylogenetic level, with viruses viewed as descendants of viral lineages that preceded cell evolution (Bamford et al. 2005; Koonin et al. 2006).

Several of these virocentric proposals claimed that viruses had played a fundamental role in cellular evolution by substituting whole cellular systems with genes and proteins of viral origin. Thus, to explain why the bacterial DNA replication system is so different to that of archaea and eukaryotes, Forterre speculated that the whole DNA replication system in bacteria had been replaced by viral-encoded proteins (Forterre 1999).

Villareal speculated something similar, but for eukaryotes (Villarreal and DeFilippis 2000). Subsequently, Forterre further proposed that viruses had “invented” DNA and that the DNA replication machineries in the three domains of life, archaea, bacteria, and eukaryotes, had independent viral origins (Forterre 2006).

Additional claims asserted that viruses and/or some viral families form a fourth domain of life (Raoult et al. 2004; Raoult and Forterre 2008). Some authors even consider viruses as “capsid-encoding organisms,” a renovated version of the Ultrabiontia, as opposed to “ribosome-encoding organisms” (cells) (Raoult and Forterre 2008). The claim for a fourth domain of life was specifically made for the giant Mimivirus and related nucleocytoplasmic large DNA viruses (NCLDV), which have very large genomes. Some of these genomes exceed the size of some parasitic bacterial genomes and possess some homologues to cellular genes involved in typical cellular processes including translation (Arslan et al. 2011; Boyer et al. 2009; Raoult et al. 2004; Van Etten 2011). Based on simple phylogenetic analyses, the viral homologues of those cellular genes branched at the base of eukaryotes, which was taken as evidence of this viral family forming a fourth domain of life that could be placed in a universal tree of cellular organisms using shared genes (Boyer et al. 2010; Raoult et al. 2004). Furthermore, it was argued that these giant viruses can even be parasitized by other viruses (e.g., Sputnik), which would make them truly living organisms since they could be parasitized themselves (La Scola et al. 2008). Nevertheless, in addition of being a weak argument (e.g., the same reasoning could apply to single genes, which would be alive because they are parasitized by introns, etc.), some authors quickly argued that these were nothing more than satellite viruses (Krupovic and Cvirkaite-Krupovic 2011).

In a different type of proposal, present-day viruses are considered remnants of past virus-like, self-replicating entities that preceded cells and that somehow contributed to the makeup of cellular genomes before cells truly existed. These long-standing ideas have been summarized and revised by Koonin et al. (2006) in their “Virus World” hypothesis.

As we will see, while some of these ideas fall under the two rival views on the nature and origin of life (metabolism versus genetic information) and cannot be properly tested, others can be tested by using appropriate molecular phylogenetic analyses and, in most cases, refuted (López-García and Moreira 2009; Moreira 2000; Moreira and López-García 2009; Williams et al. 2011). At any rate, these different heterogeneous models have stimulated virocentric debates, which can be classed as two types. One debate is conceptual and relates to the definition of life and to whether or not viruses can be considered alive from an operational point of view. The other debate is phylogenetic and relates to the actual possibility of placing viruses in the tree of life.

The Conceptual Debate: Are Viruses Alive?

Defining life is not an easy task. Biologists are most often reluctant to do so. Biology is a positive science that works by describing organisms (units of life) but has difficulties in abstracting concepts and definitions from mere descriptions. This is particularly difficult when it comes to delimiting barriers to some kind of natural continuum, such as microbial species. Defining life, as with defining species, is therefore problematic. However, this should not prevent biologists from attempting the task and reaching a consensus on what life is (Morange 2011). Indeed, when some biologists say that viruses are alive, they implicitly apply a definition of life. There are many definitions of life and/or living organisms that include more or less long lists of properties (Luisi 1998). However, when reduced to essentials, most definitions can be aligned along the two historically opposed views on the origin of life: metabolist (cytoplasmist) versus geneticist (nucleocentric).

For metabolist views, the essential defining property of life is self-maintenance. Obviously, viruses are not alive in such a definition of life because they lack any form of metabolism. Viruses have a “borrowed life,” exploiting the cell’s metabolism and resources for their replication. None of the recent fascinating discoveries about the huge viral diversity and its undeniable role in cellular evolution actually challenge the basic essence of viruses: they are strict molecular parasites unable to transform energy and matter. They are unable to “create order from disorder” and actively keep the system far from thermodynamic equilibrium, as physicists would put it. However, some authors have nonetheless tried to apply this type of definition to viruses by using a conceptual trick. According to them, most people confound virions with viruses; they claim that this is equivalent to confounding fish eggs with fish, or human spermatozooids with humans. For them, viruses would be quasi-autonomous entities whose true state is the “cell factory” or the “virocell,” i.e., the transformed infected cell actively replicating the virus (Claverie 2006; Claverie and Abergel 2010; Forterre 2010). Using this artifice (virus = “virocell”), the properties intrinsic to life, which cells undeniably possess, are transferred to the virus. From a scientific standpoint, it is clear that “a virus cannot be reduced to a virus particle or virion which is only one stage of its replication cycle” (Van Regenmortel 2010). Similarly, a fish species cannot be reduced to only a fish egg, but both the virion and the fish egg are part of their respective replicative cycles (Moreira and López-García 2009; Van Regenmortel 2010). Virions and infecting viruses in a cell are both viruses because they are both part of the viral cycle. However, following the fish egg analogy, if the fish species turns out to be a parasitic species, let’s say a remora, which in its adult phase is attached to a larger fish, let’s say a shark, the

remora is still the remora and not the remora plus the shark it is attached to. This reasoning can be applied to any other parasitic species (e.g., a human parasitic tapeworm is not a “tapeworm-man,” etc.). Defining an entity (a virus) in terms of itself plus a portion of another entity (a cell) is alien to logic and can be viewed as epistemological cheating. Viruses, unlike cells (including obligatory parasitic cells, which are still capable of metabolic activities and derive from more complex, autonomous cells), are devoid of metabolism and are not alive if self-maintenance is considered as the irreducible property of life.

For geneticist views, self-replication and evolution are the defining attributes of life. Evolution is just a consequence of the imperfect replication of informative polymers that generate variants upon which drift or natural selection act. Viruses do evolve but are unable to self-replicate. Consequently, this definition of life, if applied strictly, does not accommodate viruses either. A self-replicating ribozyme would be alive (Robertson and Joyce 2010), but a virus would not because it cannot self-replicate. Yet, because viruses evolve, many biologists consider them alive. However, viruses, like genes, do not evolve by themselves because they cannot self-replicate; they are evolved by cells because they are replicated by cells (Moreira and López-García 2009). Similarly, cultural traits such as language, art, or technology, i.e., memes (Dawkins 1976), do not evolve by themselves; they are evolved by humans. Therefore, to be logical, if we want to consider viruses alive according to a geneticist view, there are only two possibilities. Either we accept that anything that can evolve or be evolved is alive, which would include genes and memes in general; or we extend the definition of the term virus to include a hypothetical capacity for self-replication. The latter option is implicitly adopted by some authors, notably by Koonin et al. in their “Viral World” hypothesis, who maintain that viruses historically predate cells; the viruses of that “Viral World” being loosely considered self-replicating “virus-like” genetic elements from which modern viruses evolved (Koonin et al. 2006). Similarly, other authors speak of self-replicating protoviruses at the origin of cells (Jalasvuori and Bamford 2008). These proposals are ambiguous and entertain confusion because viruses as we know them do not self-replicate. These authors use the term “virus” in a dual manner: for current non-self-reproducing viruses and for more generic genetic elements having self-replicating properties that viruses actually lack. Despite their ambiguity, these models where self-replicating virus-like entities pre-date cells (Bamford et al. 2005; Jalasvuori and Bamford 2008; Koonin et al. 2006) can be perfectly understood under a pure geneticist view of life and its origin.

The Phylogenetic Debate: Can Viruses be Placed in the Tree of Life?

What Is the Tree of Life?

Before we can answer the question of whether or not viruses can be placed in a tree of life, we need to explain what a tree of life means in this context because there is also an active debate on the nature and even the existence of the tree of life itself. A tree of life may be defined as a conceptual representation of organismal history. Since organisms and cells form from pre-existing organisms/cells, this tree would proceed basically by successive branch bifurcation. There is little doubt that extant life derives from a common ancestor, as can be deduced from the universality of biochemistry and the genetic code. Therefore, a universal organismal tree of vertical descent must exist because physically, cells derive from cells. However, how to reconstruct such a tree from contemporary traits is far from trivial because few traits are universally shared. In the late 1970s, Carl R. Woese had the idea of using the universally conserved ribosomal RNA genes to build such a tree, which incidentally resulted in the discovery of the third domain of life, the archaea (Woese and Fox 1977). Molecular phylogenetic analyses of other conserved markers confirmed the tripartite division of organisms in the three domains of life identified by Woese. However, as more genes and genomes became available, additional molecular phylogenetic analyses revealed conflicting tree topologies for different gene markers. This is due to a number of reasons, including loss of signal with time and most importantly, horizontal gene transfer, which seems to be especially important among prokaryotes (Dagan and Martin 2006). Several authors think that, since HGT blurs the signal of vertical descent in trees, a tree of life cannot be reconstructed and should be replaced by a network of genes and genomes (Dagan and Martin 2009; Doolittle and Bapteste 2007), or by a forest of individual gene trees (Puigbo et al. 2010).

This controversy reflects again the dichotomous view of life. From a metabolist perspective, life is seen in terms of organisms whose history can be reconstructed using a core of conserved gene markers and represented in the form of a tree of life. For geneticist views, life is seen in terms of genes, and a tree of life cannot be reconstructed because organisms are perceived as assortments of genes from different origins. Paradoxically, under a pure geneticist view, which would be the only one that could accommodate viruses as living beings (see above), a tree of life does not exist, making the placement of viruses in such a tree artificial.

In addition to that theoretical consideration, can viruses be *actually* placed in the tree of life?

Methodological Problems

Let's assume that reconstructing a tree of life is possible using a core of conserved genes. Can viruses be placed in it? To build a molecular phylogenetic tree, a basic requirement is to compare homologous characters, i.e., characters that were present in the ancestor of the taxa being compared and inherited vertically. If we wanted to reconstruct a universal tree that includes viruses and cells, we should compare genes that are shared universally by all viruses or viral families (a given gene may be lost secondarily in one or a few species) and the three domains of cellular organisms. However, those genes do not exist. Viruses are polyphyletic, that is, they have evolved independently several times (e.g., RNA viruses have evolved independently of DNA viruses, etc.). Therefore, there no single gene actually shared by all viral families, which prevents any attempt to construct a universal phylogeny of all viruses (this is, however, possible at a smaller scale for independent viral families), and even less of all viruses and cells.

Several authors would argue that there are, nonetheless, many genes that are shared between cells and viruses. However, this is not proof of ancestral common origin because we know that HGT is a very important process in evolution and that furthermore, viruses are active vectors of HGT between cells (see above). In a number of hypotheses, while HGT is acknowledged, it is proposed that cellular genes were transferred to cells from viruses and not the other way around based on the phylogenetic analysis of a few proteins involved in replication that seemed to appear at the base of the cellular domains (Forterre 1999; 2006; Villarreal and DeFilippis 2000). However, in these studies we face other important problems that can affect molecular phylogenies very seriously, namely long-branch attraction. To make meaningful phylogenetic analyses, it is very important to apply appropriate models of sequence evolution because viral genes evolve fast and are particularly prone to phylogenetic reconstruction artifacts such as long-branch attraction, which would tend to place them at the base of the phylogenetic tree (Philippe et al. 2000). Indeed, the reanalysis of the same genes that served Forterre, using improved models of evolution and an adequate taxon sampling, to suggest that the replication machinery in bacteria was of viral origin (Forterre 1999) did not place those genes at the base of the tree but very closely related to the homologous genes of the respective host cells, showing the occurrence of multiple independent transfers from the host cell to the viruses (Moreira 2000).

A certain category of virocentric hypotheses recognizes that viruses cannot be placed in a tree of life because among other things, they are polyphyletic, evolve fast, and the genes that they share with cells were mostly acquired by HGT from cells. This is, for instance, the case of the “Viral

World” (Koonin et al. 2006). Yet they propose that there is a series of largely distributed “hallmark viral” genes and/or protein folds that are not present in cells and that could attest to ancient virus-specific genes that remain present in part of the virosphere (Bamford et al. 2005; Koonin et al. 2006). However, the conclusion that particular viral capsid folds attest to a common origin ignores alternative explanations of the presence of common viral features in very different viral families, or of viruses infecting very distant organisms. These include not only HGT, which is extremely frequent in viruses, but also host shifts (the capacity to infect distant hosts) and convergence, which is relatively easy for simple structures such as icosahedral capsid proteins (Moreira and López-García 2009). Therefore, that similar capsid protein folds occur in distant viral families infecting distant hosts is not compelling evidence for those specific folds and viruses to actually predate cells because it does not eliminate the several alternative explanations that are equally likely. Nevertheless, current data do not allow verifying or falsifying this hypothesis and hence that a world or virus-like self-replicating entities existed remains valid but untestable.

Do Some Giant Viruses Form a Fourth Domain of Life?

A particular case of how methodological problems can lead to wrong conclusions is illustrated by the recent claim that the giant Mimivirus and related nucleo-cytoplasmic large DNA viruses form a fourth domain of life (Boyer et al. 2010; Raoult et al. 2004). This assertion is based on the phylogenetic analysis of homologues of cellular genes involved in typical cellular housekeeping functions such as translation, which according to the authors, would situate NCLDV as an independent domain at the base of eukaryotes. As we have seen above with hypotheses stating that viruses were at the origin of the replication machinery of the cellular domains, this can be tested using proper phylogenetic analyses.

Because genes evolve fast in viruses, it is essential to use appropriate models of sequence evolution to avoid long-branch attraction artifacts that place artificially fast-evolving sequences at the base of the tree (Philippe et al. 2000). It is also very important to include sequences representative of good taxonomic sampling. Obviously, if sequences of the taxon to which a given sequence belongs are not present in the analysis, that sequence risks being situated in the wrong place in a phylogenetic tree. It is therefore essential to include sequences from the virus' hosts or host family to see whether or not the viral sequence places close to them. Using better taxon sampling and adequate models of sequence evolution, it can be shown that the vast majority of homologues to cellular genes present in giant viruses correspond to transfers from the host (or from

bacteria co-infecting the same host) to its virus; and that NCLDV genes do not form a monophyletic group at the base of eukaryotes, but appear dispersed within the eukaryotic tree close to their respective hosts (López-García and Moreira 2009; Moreira 2000; Moreira and Brochier-Armanet 2008; Moreira and Lopez-Garcia 2005; Moreira and López-García 2009; Williams et al. 2011). Therefore, the hypothesis that NCLDV giant viruses form a fourth domain of life that can be placed in an organismal tree of life (Boyer et al. 2010; Raoult et al. 2004) can be refuted.

After criticism based on in-depth phylogenetic analyses, D. Raoult and J.M. Claverie, two of the proponents of the hypothesis that giant viruses form a fourth domain of life (Raoult et al. 2004) have displayed ambivalent positions, which may be taken as proof for the inconsistency of such a proposal. Raoult, after having proposed a fourth domain for viruses based on the reconstruction of an organismal tree of life using conserved genes, first contradictorily denied the existence of a tree of life “from which viruses were out” (Raoult 2009); and later, made an ambiguous proposal for a fuzzy “rhizome of life” (Raoult 2010). In other words, to escape criticism Raoult shifted from a metabolist view, to a purely geneticist view, to something undefined in between, thus adding even more confusion to the debate. Claverie changed his opinion to say that giant viruses are degenerated cells (Arslan et al. 2011). This implies that these viruses would not constitute a fourth domain on the same footing as cellular domains, as previously claimed, because now they would derive from within a cellular domain.

How did these fascinating viruses and their large genomes evolve? The available data do not support their origin as an independent domain of life or as a line of degenerated cells. However, thorough genome comparative analyses suggest that these NCLDVs evolved from an ancestral virus having a core of proteins involved in viral replication and capsid formation after the recruitment of many eukaryotic and some bacterial genes via HGT, as well as many mobile genetic elements followed by massive lineage-specific duplications (Filee 2009; Koonin and Yutin 2010). Thus, these giant viruses result from the action of various mechanisms of genome evolution.

Conclusion

In recent years, the extensive development of comparative genomics and metagenomics have shown that viruses are extremely diverse and that they play an important role in organismal ecology and evolution by exerting different selective pressures on cell populations, by serving as vehicles for horizontal gene transfer among cells, and by allowing an increased acceleration of the evolutionary rate of genes that can be transferred back to cells. Viruses constitute a large

reservoir of genes, many of which have no homologues in cells and whose origin is unclear. They might constitute present or past cell genes that have evolved beyond recognition. These recent discoveries have resuscitated old virocentric debates that considered viruses (1) alive and/or (2) predating cells and/or (3) forming a group of organisms on the same footing as cellular organisms.

However, despite their extraordinary diversity, no novel properties can be attributed to viruses, which remain strict genetic parasites lacking carbon and energy metabolism. Therefore, from a metabolist perspective for which self-maintenance is the primeval property of life, viruses are not alive. Although it is conceivable that viruses emerged very early in evolution (i.e., right after cells or during cell evolution), they could not have evolved prior to cells or pre-cellular stages that could be parasitized by them. Claims that viruses are alive and predate cellular evolution can be understood only within the framework of a strict geneticist view of life based exclusively on the property of evolution and using the definition of the term “virus” as a relaxed metaphor for self-replicating entities—something which viruses, strictly speaking, are not.

By contrast, claims that viruses form a group of organisms at the same level as cells and most specifically, that giant viruses form a fourth domain of organisms that can be placed in the tree of life based on genes shared by cells involved in typical cellular functions can be clearly refuted by proper molecular phylogenetic analyses. These show without ambiguity that homologous genes shared by these giant viruses and cells were acquired by the viruses from their hosts or from bacteria co-infecting their hosts. Therefore, whereas the hypothesis that virus-like self-replicating entities predated cells is scientifically valid, although not testable, the discussion about a fourth domain of life corresponding to giant viruses, despite the intrinsic interest that these viruses have as products of evolution, has no scientific legitimacy.

Acknowledgments This work was supported by the French Centre National de la Recherche Scientifique (CNRS).

References

- Alexander J, Bridges CB. Some physico-chemical aspects of life, mutation and evolution. In: Alexander J, editor. Colloid chemistry, theoretical and applied. New York: Reinhold; 1928. p. 17.
- Arslan D, Legendre M, Seltzer V, Abergel C, Claverie JM. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. *Proc Natl Acad Sci U S A*. 2011;108:17486–91.
- Avery OT, MacLeod CM, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic

- acid fraction isolated from *Pneumococcus* type III. *J Exper Med.* 1944;79:137–58.
- Bamford DH, Grimes JM, Stuart DI. What does structure tell us about virus evolution? *Curr Opin Struct Biol.* 2005;15:655–63.
- Beutner R. Life's beginning on the Earth. Baltimore: Williams & Wilkins; 1938.
- Boyer M, Yutin N, Pagnier I, Barrassi L, Fournous G, Espinosa L, et al. Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proc Natl Acad Sci U S A.* 2009;106:21848–53.
- Boyer M, Madoui MA, Gimenez G, La Scola B, Raoult D. Phylogenetic and phyletic studies of informational genes in genomes highlight existence of a 4 domain of life including giant viruses. *PLoS ONE.* 2010;5:e15530.
- Claverie JM. Viruses take center stage in cellular evolution. *Genome Biol.* 2006;7:110.
- Claverie JM, Abergel C. Mimivirus: the emerging paradox of quasi-autonomous viruses. *Trends Genet.* 2010;26:431–7.
- Culley AI, Lang AS, Suttle CA. High diversity of unknown picorna-like viruses in the sea. *Nature.* 2003;424:1054–7.
- Dagan T, Martin W. The tree of one percent. *Genome Biol.* 2006;7:118.
- Dagan T, Martin W. Getting a better picture of microbial evolution en route to a network of genomes. *Philos Trans R Soc Lond B Biol Sci.* 2009;364:2187–96.
- Danovaro R, Corinaldesi C, Dell'anno A, Fuhrman JA, Middelburg JJ, Noble RT, et al. Marine viruses and global climate change. *FEMS Microbiol Rev.* 2011;35:993–1034.
- Dawkins R. The selfish gene. New York City: Oxford University Press; 1976. p. 224.
- Doolittle WF, Bapteste E. Pattern pluralism and the Tree of Life hypothesis. *Proc Natl Acad Sci U S A.* 2007;104:2043–9.
- Dupressoir A, Vernochet C, Bawa O, Harper F, Pierron G, Opolon P, et al. Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. *Proc Natl Acad Sci U S A.* 2009;106:12127–32.
- Edwards RA, Rohwer F. Viral metagenomics. *Nat Rev Microbiol.* 2005;3:504–10.
- Eickbush TH. Telomerase and retrotransposons: which came first? *Science.* 1997;277:911–2.
- Filee J. Lateral gene transfer, lineage-specific gene expansion and the evolution of nucleocytoplasmic large DNA viruses. *J Invertebr Pathol.* 2009;101:169–71.
- Forterre P. Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins. *Mol Microbiol.* 1999;33:457–65.
- Forterre P. Three RNA, cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. *Proc Natl Acad Sci U S A.* 2006;103:3669–74.
- Forterre P. Defining life: the virus viewpoint. *Orig Life Evol Biosph.* 2010;40:151–60.
- Frada M, Probert I, Allen MJ, Wilson WH, de Vargas C. The "Cheshire Cat" escape strategy of the coccolithophore *Emiliania huxleyi* in response to viral infection. *Proc Natl Acad Sci U S A.* 2008;105:15944–9.
- Fry I. The origins of research into the origins of life. *Endeavour.* 2006;30:24–8.
- Fuhrman JA. Marine viruses and their biogeochemical and ecological effects. *Nature.* 1999;399:541–8.
- Gogarten JP, Townsend JP. Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol.* 2005;3:679–87.
- Gojbori T, Moriyama EN, Kimura M. Molecular clock of viral evolution, and the neutral theory. *Proc Natl Acad Sci USA.* 1990;87:10015–8.
- Green RG. On the nature of the filterable viruses. *Science.* 1932;82:444.
- Haldane JBS. The causes of evolution. London: Harper; 1932.
- Haldane JBS. The origin of life. *Rationalist Ann.* 1929;3–10.
- Hershey AD, Chase M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol.* 1952;36:39–56.
- Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science.* 2010;327:167–70.
- Jalasuuri M, Bamford JK. Structural co-evolution of viruses and cells in the primordial world. *Orig Life Evol Biosph.* 2008;38:165–81.
- Koonin EV, Yutin N. Origin and evolution of eukaryotic large nucleocytoplasmic DNA viruses. *Intervirology.* 2010;53:284–92.
- Koonin EV, Senkevich TG, Dolja VV. The ancient virus world and evolution of cells. *Biol Direct.* 2006;1:29.
- Kristensen DM, Mushegian AR, Dolja VV, Koonin EV. New dimensions of the virus world discovered through metagenomics. *Trends Microbiol.* 2010;18:11–9.
- Krupovic M, Cvirkaite-Krupovic V. Virophages or satellite viruses? *Nat Rev Microbiol.* 2011;9:762–3.
- La Scola B, Desnues C, Pagnier I, Robert C, Barrassi L, Fournous G, et al. The virophage as a unique parasite of the giant mimivirus. *Nature.* 2008;455:100–4.
- Lauring AS, Andino R. Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog.* 2010;6:e1001005.
- López-García P, Moreira D. Yet viruses cannot be included in the tree of life. *Nature Rev Microbiol.* 2009;7:615–7.
- Luisi PL. About various definitions of life. *Orig Life Evol Biosph.* 1998;28:613–22.
- Lwoff A. L'évolution physiologique: étude des pertes de fonctions chez les microorganismes. Paris: Hermann et Cie; 1943. p. 308.
- Lwoff A. The concept of virus. *J Gen Microbiol.* 1957;17:239–53.
- Morange M. Problems raised by a definition of life. In: Gargaud M, López-García P, Martin H, editors. Origins and evolution of life. New York: Cambridge University Press; 2011. p. 3–13.
- Moreira D. Multiple independent horizontal transfers of informational genes from bacteria to plasmids and phages: implications for the origin of bacterial replication machinery. *Mol Microbiol.* 2000;35:1–5.
- Moreira D, Brochier-Armanet C. Giant viruses, giant chimeras: the multiple evolutionary histories of Mimivirus genes. *BMC Evol Biol.* 2008;8:e12.
- Moreira D, Lopez-Garcia P. Comment on "The 1.2-megabase genome sequence of Mimivirus". *Science.* 2005;308:1114.
- Moreira D, López-García P. Ten reasons to exclude viruses from the tree of life. *Nat Rev Microbiol.* 2009;7:306–11.
- Muller HJ. Variation due to change in the individual gene. *Am Nat.* 1922;56:32–50.
- Muller H. The gene as the basis of life. In: Duggar BM, editor. Proceedings of the International Congress of Plant Sciences. Menasha, WI, USA: George Banta; 1929. 917–918.
- Murphy FA. The foundations of Virology. West Conshohocken: Infinity Publishing; 2011.
- Nakamura TM, Morin GB, Chapman KB, Weinrich SL, Andrews WH, Lingner J, et al. Telomerase catalytic subunit homologs from fission yeast and human. *Science.* 1997;277:955–9.
- Oparin AI. The origin of life. New York: Mac Millan; 1938.
- Orgel LE. Prebiotic chemistry and the origin of the RNA world. *Crit Rev Biochem Mol Biol.* 2004;39:99–123.
- Philippe H, Lopez P, Brinkmann H, Budin K, Germot A, Laurent J, et al. Early branching or fast-evolving eukaryotes? *Proc R Soc Lond B Biol Sci.* 2000;267:1213–21.
- Podolsky S. The role of the virus in origin-of-life theorizing. *J Hist Biol.* 1996;29:79–126.
- Prangishvili D, Vestergaard G, Haring M, Aramayo R, Basta T, Rachel R, et al. Structural and genomic properties of the hyperthermophilic archaeal virus ATV with an extracellular stage of the reproductive cycle. *J Mol Biol.* 2006;359:1203–16.
- Puigbo P, Wolf YI, Koonin EV. The tree and net components of prokaryote evolution. *Genome Biol Evol.* 2010;2:745–56.

- Raoult D. The post-Darwinist rhizome of life. *Lancet*. 2010;375:104–5.
- Raoult D. There is no such thing such as a tree of life (and of course viruses are out!). *Nat Rev Microbiol* 2009;7:615.
- Raoult D, Forterre P. Redefining viruses: lessons from Mimivirus. *Nat Rev Microbiol*. 2008;6:315–9.
- Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, et al. The 1.2-megabase genome sequence of Mimivirus. *Science*. 2004;306:1344–50.
- Robertson MP, Joyce GF. The Origins of the RNA World. *Cold Spring Harb Perspect Biol* 2010
- Rodriguez-Valera F, Martin-Cuadrado AB, Rodriguez-Brito B, Pasic L, Thingstad TF, Rohwer F, et al. Explaining microbial population genomics through phage predation. *Nat Rev Microbiol*. 2009;7:828–36.
- Sharon I, Alperovitch A, Rohwer F, Haynes M, Glaser F, Atamna-Ismael N, et al. Photosystem I gene cassettes are present in marine virus genomes. *Nature*. 2009;461:258–62.
- Sullivan MB, Lindell D, Lee JA, Thompson LR, Bielawski JP, Chisholm SW. Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol*. 2006;4:e234.
- Suttle CA. Marine viruses—major players in the global ecosystem. *Nat Rev Microbiol*. 2007;5:801–12.
- Troland LT. The chemical origin and regulation of life. *Monist*. 1914;24:92–133.
- Troland LT. Biological enigmas and the theory of enzyme action. *Am Nat*. 1917;51:321–50.
- Van Etten JL. Another really, really big virus. *Viruses*. 2011;3:32–46.
- Van Regenmortel MH. Logical puzzles and scientific controversies: the nature of species, viruses and living organisms. *Syst Appl Microbiol*. 2010;33:1–6.
- van Regenmortel MHV. Introduction to the species concept in virus taxonomy. In: van Regenmortel MHV, Fauquet CM, Bishop DHL, Carstens EB, Estes MK, Lemon SM, Maniloff J, Mayo MA, McGeoch DJ, Pringle CR, Wickner RB, editors. 7th Report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego; 2000. 3–16.
- van Regenmortel MHV. The nature of viruses. In: Mahy BWJ, van Regenmortel MHV, editors. *Encyclopedia of Virology*. Elsevier/Academic Press; 2008. 398–402.
- Villarreal LP, DeFilippis VR. A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. *J Virol*. 2000;74:7079–84.
- Watson JD, Crick FH. The structure of DNA. *Cold Spring Harb Symp Quant Biol*. 1953;18:123–31.
- Williams TA, Embley TM, Heinz E. Informational gene phylogenies do not support a fourth domain of life for nucleocytoplasmic large DNA viruses. *PLoS ONE*. 2011;6:e21080.
- Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A*. 1977;74:5088–90.
- Zeidner G, Bielawski JP, Shmoish M, Scanlan DJ, Sabehi G, Beja O. Potential photosynthesis gene recombination between *Prochlorococcus* and *Synechococcus* via viral intermediates. *Environ Microbiol*. 2005;7:1505–13.