Trends in Microbiology

CellPress
REVIEWS

Opinion

Investigating the Concept and Origin of Viruses

Arshan Nasir [1,*] Ethan Romero-Severson,[1] and Jean-Michel Claverie[2]

The ongoing COVID-19 pandemic has piqued public interest in the properties, evolution, and emergence of viruses. Here, we discuss how these basic questions have surprisingly remained disputed despite being increasingly within the reach of scientific analysis. We review recent data-driven efforts that shed light into the origin and evolution of viruses and explain factors that resist the widespread acceptance of new views and insights. We propose a new definition of viruses that is not restricted to the presence or absence of any genetic or physical feature, detail a scenario for how viruses likely originated from ancient cells, and explain technical and conceptual biases that limit our understanding of virus evolution. We note that the philosophical aspects of virus evolution also impact the way we might prepare for future outbreaks.

Highlights

The distinctions between virions and viruses and modern and ancient cells are crucial to understand virus origins and evolution.

Viruses can be better defined by their generic features of genome propagation and dissemination rather than physical or biological properties of their virions or hosts.

Virus genomes are characterized by the abundance of virus-specific genes that lack detectable cellular homologs. Despite their abundance, virus-specific genes are rarely discussed in the models of virus origin and evolution.

The alignment-based methods are ill-suited for the origins of life research, especially when the objective is to place fast-evolving organisms or viruses in the tree of life.

Protein structures may provide a better alternative to resolve the very deep branches in the tree of life.

The Need to Redefine Viruses

The COVID-19 pandemic exemplifies the constant threat and pressure exerted by viruses on human health and the global economy. The pandemic has triggered an aggressive international response to contain virus spread, cure the disease, and prevent future infections. In parallel, it has rekindled public curiosity in virus definitions, origins, evolution, and their various modes of emergence. For example, Google search for '*what is a virus'* reached peak popularity in March 2020 coinciding with the global rise in COVID-19 cases. Surprisingly, such fundamental questions have remained unsettled even among evolutionary virologists [1–5] and cause confusion in the media portrayal and public perception. For instance, despite overwhelming scientific evidence supporting a natural zoonotic transmission of SARS-CoV-2 from animals to humans [6], many still suspect that the virus was purposefully engineered in laboratories. Similarly, viruses are generalized as noxious pathogens in common discussions and this focus greatly underestimates the many beneficial roles they play in the biosphere [7,8] and as mutualistic symbionts of many hosts (reviewed in [9,10]). In this article, we revisit fundamental questions about the nature, origins, and evolution of viruses during a time when public interest in virus biology is at its peak. We emphasize the need to rethink viruses in the light of new discoveries [2] and call for broader acceptance of new views that are resisted by (sometimes) century-old concepts established in early virology research (reviewed in [11]).

What Is a Virus?

Defining viruses is surprisingly controversial. This is largely because of the seemingly split nature of the virus reproduction cycle into two distinct stages: (i) an intracellular stage during which the virus reprograms the infected cell to produce viral particles or **virions** (see Glossary), and (ii) an extracellular stage during which virions escape the infected cells and persist in the external environment (similar to plant seeds [12]).[1] Both stages, when considered separately, provide dramatically contrasting views about the nature and roles of viruses. For example, virions are metabolically

[1]Theoretical Biology and Biophysics (T-6), Los Alamos National Laboratory, Los Alamos, NM, USA
[2]Aix Marseille University, CNRS, IGS, Structural and Genomic Information Laboratory (UMR7256), Mediterranean Institute of Microbiology (FR3479), Marseille, France

*Correspondence:
anasir@lanl.gov (A. Nasir).

[1] In addition, a third stage may exist if the virus genome either integrates into the host DNA or becomes part of the host cytoplasm. Such examples may not lead to virion production or diseases. Because classical signs of virus infection (e.g., virion production, cell rupture) may not be obvious, it is possible that we have massively underestimated nonharmful virus–cell interactions involving virus genome endogenization and domestication by cells [83].

Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

inert infectious particles that do not meet any of the criteria we may use to define 'life' or living organisms [2]. However, since they can be purified, counted, and visualized under the microscope, their physical and biochemical properties (e.g., size, shape, metabolic capabilities, capsid) along with host/tissue specificity have become popular in the description, illustration, and naming of viruses (e.g., human immunodeficiency virus). These, in turn, have shaped our perceptions about viruses as nonliving inanimate biological objects that are, paradoxically, infectious.

Treating virions as viruses is a conceptual mistake [2,12–16] that overlooks the dramatic changes viruses introduce inside infected cells. A virus-infected cell can effectively be transformed into a 'hot spot' for virion production [17] and can practically lose its identity (i.e., it now produces virions rather than two daughter cells) [18]. In some viral infections, large cell-like '**virion factories**' are clearly visible [19]. This remarkable transformation is due to the virus-mediated manipulation and alteration of host metabolism and defenses [7]. The intracellular stage therefore involves substantial viral activity and is often the target of antiviral drugs to combat virus infection (e.g., antivirals that target HIV polymerase). Despite its immense role in establishing virus infection and existence inside the infected cells, it has unfortunately been referred to as the 'eclipse' or 'vegetative' phase [20,21] to indicate lack of hallmark signs of virus infection (e.g., virion production, plaques, and cell rupture) and ignored in the definitions and descriptions of viruses. As suggested by Jean-Michel Claverie, virion factory better represents the 'virus self' and virions are simply means to disseminate genetic information much like human gametes and plant seeds [12]. In other words, we should depart from the established usage of the word 'virus' as being synonymous to 'virion'. The term 'virus' should refer to the process encompassing all phases of the virus infection cycle [3]. In this context, questioning the origin of 'viruses' takes a completely different and much broader meaning than simply questioning the origin of the virus particles [2,11,13,16].

Avoid the Presence/Absence Criteria to Define Viruses

The virion- and host-centric virus definitions can cause ambiguities in distinguishing different viral lineages and even viruses from cellular organisms. For example, Forterre recently proposed to redefine viruses as 'capsid-encoding organisms' [22] and later as 'virion-encoding organisms' [2]. Both definitions recognize viruses as 'organisms' that produce capsids/virions and rightly put emphasis back on the intracellular stage of virus infection cycle. However, these views suffer from our 'human' habit of classifying biological entities based on the presence/absence or contrast of physical and genetic features. As we discuss later, such definitions rarely withstand the test of time and are vulnerable to change with new discoveries. For example, viruses were long considered tiny and submicroscopic biological entities (properties that describe virions not viruses!) before the discovery of 'giant viruses' with genomes and virions bigger than the genomes and sizes of many parasitic cells [23–25]. In fact, holding onto the century-old size/shape virion-centric definitions delayed the discovery of giant viruses by more than a decade.[2] Similarly, some scientists consider viruses 'non-living' because they do not encode metabolism-related genes [1]. However, this feature is neither unique nor common to all viruses. Many **endosymbiotic** cellular organisms are also characterized by extremely reduced metabolic and translational machineries [26–28], and recent metagenomic surveys have verified the existence of several, and likely very ancient, metabolic genes in the genomes of giant viruses [7]. These genes likely help reconfigure the metabolism of infected cells during virus infection [7].

Glossary

Endosymbiosis: the intimate existence of organisms inside the cells or body of other organisms. Notable examples include endosymbiosis of the ancestors of mitochondria and chloroplasts by proto-eukaryotes or the ancestors of eukaryotes.
Last universal common ancestor (LUCA): the common ancestor of modern cells, Archaea, Bacteria, and Eukarya. LUCA was not the first cell. It was the last population of cells that diversified into modern cells.
Orthologous: refers to genes that diverged from the common ancestor as a result of speciation.
Tree of life: a diagram that describes the evolutionary history among modern species using the metaphors of branching patterns, roots, and leaves to represent evolutionary relationships, ancestors, and modern species, respectively. The topology of the tree of life and the place of viruses in the tree are hotly debated topics.
Virion: virus particle that can be purified and visualized. The core of a virion comprises the virus nucleic acid (DNA or RNA) enclosed inside a protein shell called a capsid.
Virion factories: intracellular compartments, formed inside virus-infected cells, that increase virus replication.

[2] The first giant virus, *Acanthamoeba polyphaga mimivirusi,* was initially mistaken for a Gram-positive bacterium. It was first discovered in 1992 during a pneumonia outbreak in Bradford, UK and the large size of its virion misled scientists to believe that it must be a bacterium (called 'Bradfordcoccus'). Its virus nature was finally revealed in 2003 [84] and the virus was aptly named 'mimivirus' for 'bacteria-mimicking virus'. This is a famous example where adhering to century-old virion and size-based virus definitions delayed a significant discovery.

Using virion or capsid to distinguish viral lineages and viruses from cells can generate similar confusions. For example, it can complicate classifications for virus-like genetic elements and viruses that either lack virions (e.g., plasmids, viroids [29]) or encode only part of the virion (e.g., polydnaviruses). For example, the genome of polydnaviruses is dispersed within the genome of parasitoid wasps. The polydnavirus-associated wasps encode the virion packaging system and utilize virions as gene delivery vectors to infect caterpillars [30]. Since the virion is encoded by the wasp genome, polydnavirus-associated wasps may better resemble virion-encoding organisms under the virion-centric definition [31]. Similarly, virus-infected cells can excrete vesicles containing the virus nucleic acid, [32] and healthy cells routinely utilize extracellular vesicles for genetic communication [33]. These examples generalize the concept and morphology of a 'virion'. Similarly, capsid-like compartments have been detected in cellular organisms where they perform functions such as storage of enzymes [34], and many viral capsid proteins either evolved directly from cellular proteins [35] or have distant homologs in cellular genomes [36]. These examples blur the separation of viruses from cells (and other parasitic genetic elements) based on the presence/absence of physical or genetic descriptors.

In sum, we discourage the use of any virus definition based on the presence or absence of any subset of genes or physical features (e.g., size, morphology, capsid proteins) because such definitions are often ambiguous, not broadly applicable, and more importantly prone to change with new discoveries. We assert that viruses can be better defined by their generic properties of genome dissemination and propagation [11]. Viruses replicate using the macromolecular machinery of other biological entities. This prong establishes absolute parasitism, which is a hallmark of viruses and virus-like genetic elements. Another feature of viruses is the ability to encapsulate and disseminate genomes in metabolically inert structures. This prong can also be generalized, and such structures could be any type of infectious particle without constraints of size, shape, or biochemical composition (e.g., vesicles) [37]. This definition encompasses both the encapsulated and non-encapsulated genomes (e.g., plasmids) and emphasizes the generic feature of how viruses propagate in cells rather than being dependent on the presence/absence of specific biomarkers [11].

Origins of Viruses: Which Hypothesis Is Biologically Plausible?

Under our generic definition, virus origin must mean the origin of parasitism and the subsequent ability of those parasitic entities to propagate via the production of metabolically inert structures. Since all modern-day viruses strictly parasitize cells (with the exception of virophages that parasitize the viral factory of other viruses) [38,39], we can assume that virus-mediated parasitism and propagation originated only after cells appeared in evolution as cells would provide both the resources to parasitize upon and the means for genome dissemination (e.g., capsids/vesicles). We therefore rule out virus existence in a 'pre-cellular' world as it would be incompatible with the proposed virus definition (Figure 1 for comparative scenarios).

The next logical questions are the timings and mechanisms of when and how the first viruses appeared. The former question is relatively straightforward. In our view, viruses originated from 'ancient' cells that existed before the **last universal common ancestor** (**LUCA**) diversified into modern cells (i.e., the three superkingdoms, Archaea, Bacteria, and Eukarya) [40].[3] There are multiple lines of evidence supporting this timing. For example, the genomes of archaeoviruses, bacterioviruses, and eukaryoviruses, are characterized by the abundance of

[3] In the 1970s, Carl Woese pioneered the method of using molecular sequences to study evolution. His work led to the recognition of Archaea [85], then called the 'third domain' of life [86]. Archaea have recently taken center stage in evolutionary debates regarding the origin of eukaryotes. There is great controversy on whether Archaea were the first group of diversified organisms on Earth [87], are a sister group to eukaryotes [88,89], or are our ancestors [90,91].