 doi:10.20944/preprints202104.0743.v1

The misleading history of VIRUSes

Francisco Prosdocimi[1*], Sávio Torres de Farias[2,3*], Juliana Reis Cortines[4] and Marco V. José[5]

[1] Laboratório de Biologia Teórica e de Sistemas, Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.
[2] Laboratório de Genética Evolutiva Paulo Leminsk, Departamento de Biologia Molecular, Universidade Federal da Paraíba, João Pessoa, Paraíba, Brasil.
[3] Network of Researchers on the Chemical Evolution of Life (NoRCEL), Leeds, LS7 3RB, UK.
[4] Departamento de Virologia, Instituto de Microbiologia Paulo de Góes, Universidade Federal do Rio de Janeiro, Brazil.
[5] Theoretical Biology Group, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Ciudad Universitaria, 04510 CDMX, Mexico.

* Corresponding authors: prosdocimi@bioqmed.ufrj.br (FP); stfarias@yahoo.com.br.

Abstract

Viruses were classically named after the very same Latin word *virus*, originally meaning poison or venom. Public understanding of viruses reinforces their "malign" aspects, especially nowadays under the COVID-19 global pandemic. It is our aim here to propose a new way to view viruses and understand their origins and evolution. First, viruses are the most abundant biological systems found on Earth. They can be found almost everywhere and form a subtle biological layer named virosphere. Second, viruses are probably the most important drivers of molecular evolution and they are active agents of ecosystems maintenance and homeostasis, allowing and driving their dynamic modification. A significant number of eukaryotic genomes are composed by genome elements similar to viruses and these endogenous viruses are continuously acting for our equilibrium and fitness. They are responsible for the origin of species-specific orphan genes that allow adaptation through the development of

 doi:10.20944/preprints202104.0743.v1

specific traits in separate lineages of eukaryotes. Accumulated evidence indicate that a viral infection was responsible to create the eukaryotic nucleus and, also, it is a syncytium structure caused by viral replication that allows the formation of the placenta. Therefore, viruses were fundamental for the evolutionary fate of eukaryotes and mammals. The presence of virus-specific genes that are absent in cellular organisms indicates that viruses existed before cells. Besides, such as progenotes, viruses are simply ribonucleoproteic entities and their capsids are orders of magnitude simpler than proteolipidic membranes. Here, we (i) propose a complete scenario to describe the major transitions in prebiotic evolution, (ii) present the possibility that viruses emerged before LUCA, and (iii) suggest that viruses originated at the age of progenotes. However, viruses do not form a monophyletic clade. They should be seen as an evolutionary stable strategy recurrently achieved by biological systems to survive. We propose that the word "VIRUS", known as venom, is historically mistaken and introduce a new interpretation for their name as an acronym for "Very Important Replicator Unit and Symbiont". But more than being "very important", viruses are of "Utmost" relevance for the maintenance of life in biosphere, by which reason we suggest referring to them as "UIRUS" to reinforce their incredible role in symbiosis and their beneficial characteristics over the infectious ones.

Background

In a recent article discussing the concept and origin of viruses, Nasir and collaborators inform us that the question "what is a virus?" has reached peak popularity in March 2020 following the advent of COVID-19 pandemic (Nasir et al., 2020). More than one year later, we are still suffering from the problems caused by this greater enemy of contemporaneous society and global economy, the virus SARS-CoV-2 and its new variants. Therefore, at this time on which public interest in virology has never been so pronounced, it is compelling to acknowledge that most people do not even glimpse how viruses originated and evolved, as also the biomedical community fails to recognize and acknowledge their pivotal role in the ecology and evolution of organisms on Earth.

Since its early beginnings (until today), virology has been mostly founded in the study of infectious diseases transmitted by viruses. Actually, most virologists study the epidemiology of viral infections, immunology and the relationship between infected cells and viruses. Thus, it is not surprising that viruses have been understood by society as these selfish, malignant entities that hijack the cells of innocent organisms and use them as a fabric to replicate their own genetic information. Even if it is clear that these entities do not present any sort of conscient behavior, neither any sort of feelings nor desires to hurt, their own nature evolved to promote that terrible action. Under the evolutionary point of view, the way viruses reproduce portrait a well succeeded strategy as the high number and fitness of these biological systems is a self-evident proof of their great Darwinian adaptability.

 doi:10.20944/preprints202104.0743.v1

Our aim here is to explain clearly why the paradigm that considers viruses as venoms, pathogens, selfish or scary entities is mistaken. In order to make our point, we will need to travel into the origins of life and provide a tentative though highly parsimonious scenario that will describe the origin of viruses from the open biological systems that inaugurated life in Earth. Albeit we must recognize that viruses may cause diseases, they cannot be solely understood as mere pathogens because they do not have this single role in nature. Far from that, their infectious capabilities should be seen as a fortuitous, contingent issues that happened along the evolution of life on Earth. Considering the greatness of virosphere, only a tiny fraction of viruses can be considered infectious agents. From the last decade to now, it is emerging a new view about viruses, on which their origin, role in ecology and evolution are finally being reviewed and recognized. We are now starting to see viruses as ecologists rather than seeing them as epidemiologists.

Are viruses alive?

It is still not clear at this moment whether viruses are living beings or not. This is a controversial issue and, although viruses are normally considered to be non-living entities, many researchers argue in the contrary. For example, the French virologist Jean-Michel Claverie suggested that the infectious viral particles (referred as virions) cannot be considered to be the viruses themselves (Claverie, 2006; Nasir et al., 2020). This confusion between virions and viruses is at the base of the idea of viruses being non-living creatures. As Claverie arguments, virions should be understood as analogous to viral gametes and the world virus must acknowledge these organisms along their complex life cycles. In that case, viruses should be considered living entities. As we shall see later, if we also consider the idea from code biology, where biological organisms should be understood as entities capable to understand and process a biological coding (Emmeche, 1998; Barbieri, 2003, 2014; Farias et al., 2020), viruses should be indeed considered as living.

Plus, under the current view of symbiotic biology, the traditional argument that viruses are non-autonomous systems loses its force. All organisms on Earth live in symbiosis and now we know that we have at least the same number of bacterial cells in our body as human cells (Sender et al., 2016). Therefore, we might ask: are humans autonomous systems? It has been shown that microbial metabolites directly affect our health and shape our immune system (Postler and Ghosh, 2017). No organism can live (strive) by itself. Autonomy, therefore, cannot be used as a characteristic to separate living from non-living. More than that, Farias and collaborators have argued that viruses do not actually need a cell to replicate, they merely need a ribosome (de Farias et al., 2019). The fact that ribosomes can only be found nowadays inside cells make the dependency relationship between viruses and cells. But growing evidence suggest that it was not like this in the early days of life in our planet.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 April 2021 doi:10.20944/preprints202104.0743.v1

Therefore, what are viruses? What are their roles in nature? What are their roles in human health? What are their roles in evolution? In order to understand these issues, we propose starting from a historical perspective.

How viruses have been discovered and named?

The view that virus are non-living infectious agents is probably a historical mistake. If we take the etymological route to understand what they are, we will find that the word virus came from the very same Latin word *virus* that meant venom or poisonous secretion obtained from biological creatures (please check https://www.online-latin-dictionary.com/latin-english-dictionary.php?parola=virus and other dictionaries). This word has been defined to refer to viral entities in 1898 by the Dutch microbiologist and botanist Martinus Beijerinck (1851-1931). It is also known, though, that the renowned French microbiologist Louis Pasteur (1822-1895) used the term "virus" to refer to *any causative agent of infectious diseases*. Both Pasteur and Edward Jenner (1749-1823) developed the first vaccines to protect from viral diseases without knowing that viruses even existed. Jenner developed the architype of the smallpox vaccine (1794), which helped to eliminate the first human viral disease, and Pasteur predicted that rabies was caused by a pathogen too small to be seen in a microscope.

The history of our understanding about viruses passed by the first identification of such entity as the causative agent of the *tobacco mosaic disease* (TMD), described by the German agricultural chemist Adolf Mayer (1843-1942), in 1876. Few years later, in 1884, the French microbiologist Charles Chamberland developed a porcelain water filter capable to produce bacteria-free liquids. The Chamberland filter was a success, and it was used, in 1892, by the Russian botanist Dmitri Ivanovsky in the case of TMD. Ivanovsky discovered that, even after filtered, the tobacco sap was still capable to transmit TMD. He hypothesized that the infection was caused by some toxin produced by the bacteria present in tobacco.

And then we get to Beijerinck again. In 1898, he was the researcher who repeated Mayer's experiments. Nevertheless, he had the intuition that the tobacco's filtrate contained some form of soluble living germ (*contagium vivum fluidum*) and used the term *virus* to refer to it. He thought that viruses had a liquid nature.

In fact, virus particles had already been seen more than a decade earlier than the Beijerinck experiments. In 1886, the Scottish researcher Dr. J. Buist of Edinburgh assumed that these particles were of bacterial origin. But it was only after the invention of electron microscopy, in 1931, that the German engineers Ernst Ruska (1906–1988) and Max Knoll (1887–1969) could actually see the marvelous complexity of bacteriophages' molecular structures and identify the existence of this new sort of biological system.

 doi:10.20944/preprints202104.0743.v1

Viruses are the "new bacteria"

It is now time to understand how our view about viruses are changing. In order to do that, we will take profit of a pop metaphor often used as memes in social networks. It is commonly said nowadays while the population gets old, that the "thirties are the new twenties", "the forties are the new thirties" and so on. We take profit of this metaphor to indicate, likewise, that "viruses are the new bacteria". This means that viruses now start to be understood as bacteria were once.

As viruses, bacteria were originally seen as malignant entities that operated "to cause diseases in humans". And then, after a deeper comprehension was acquired by both biologists and the whole society, bacteria were proven to be key to the regulation of ecosystems and also to the correct functioning of animal bodies (Beasley et al., 2015; Russell, 2019; Afzal et al., 2019; Togo et al., 2019). Researchers have discovered astonished (Sender et al., 2016) that we present at least as much bacterial cells in our bodies as human cells. These new findings evidence that biological organisms function much more like a community of different species than as single individuals. The concept of holobiont has been proposed to cope with the idea that we could be understood as ecosystems over which different microbes live in harmony (Gilbert et al, 2012; Guerrero et al., 2013; Salvucci, 2016; Sánchez-Cañizares et al., 2017). All animals are enormous symbiotic entities. As individuals living in society, it is also emerging the view that we need to have healthy habits to be colonized by healthy microbes and constitute a healthy holobiont (Greer et al., 2016). Microbiome studies have shown the relevance of presenting these healthy bacteria in our guts, mouth, skin, vagina and all over our body. Bacteria can no longer be seen as malignant entities. They proved to help us to live better and help ecosystems to acquire equilibrium and homeostasis.

Although this is still an emerging view, the same can now be said for viruses (García-Lopez et al., 2019). In a classical work from the renown American evolutionary biologist Stephen Jay Gould, he said in the 4th part of his book *Full House* that the most abundant organism on Earth was bacteria, proposing the argument of the "*modal bacteria*" (Gould, 1996). As we know in statistics the mode is the most frequent value that can be observed in any dataset. Two and a half decades in the past, he argued that bacteria were the most abundant organisms on Earth. Therefore, the life phenomenon could be "resumed" as bacterial life and bacterial metabolism could be understood as a valid generalization for any living metabolism. This actually occurs as we often learn molecular mechanisms in undergraduate courses by reading textbooks that present the simplified version of them observed in Bacteria.

However, now we know that the number of viruses in the planet is many orders of magnitude bigger than the number of bacteria (Bergh et al., 1989). On that matter, a researcher from the University of British Columbia suggested that the number of virion particles present in the oceans could be compared to the number of stars in the whole Universe, estimated as 10^{23} (Suttle, 2013)! Even if this may be an overestimation, it is now clear that the number of virus on Earth is tantalizingly enormous.

 doi:10.20944/preprints202104.0743.v1

Researchers have given the name virosphere to this great amount and diversity of viruses that can be found almost everywhere in our planet (Comeau et al., 2008; Culley, 2018). As with bacteria before, we are still starting to discover the relevance of virus in the oceans, in the earth and in the air (Paez-Espino et al., 2016; Moniruzzaman et al., 2020; Schulz et al.; 2020). They are probably responsible to fine-tune homeostasis in ecosystems (Middelboe e Brussaard, 2017). For example, in the oceans, viruses function mainly by: (i) controlling the bacteria/host population and, (ii) as a result of host lysis, the viral particles themselves aid in cycling nutrients such as carbon, nitrogen, among others (Weitz et al., 2014; Jover et al., 2014).

In addition, we are also discovering that our bodies are full of viruses. The human genome project discovered we have a great number of viral elements embodied in our own DNA (Lander et al., 2000). Together, genomic elements like LINES, SINES, transposons, retrotransposons and others make up to 42% of our genomes (Lander et al., 2001). Besides, genome biology researchers were astonished to find that any genome present significant number of species-specific, orphan genes that do not show homology to known genes in related species (Tautz and Domazet-Lošo, 2011; Arendsee et al., 2014). And now, it is emerging the notion that those genes are most likely produced by viral recombinases that shuffle our genomes to produce these new and unique genes (Enard et al., 2016), causing the phenomenon of gene shuffling. The human endogenous retroviruses (named HERVs) maintain our genomes active and capable to produce new genes that may further influence our adaptation as individuals, populations and species. This phenomenon is also happening in most animals and plants. We should remember that transposons were originally found in corn, by the pioneer work of Barbara McClintock.

As with bacteria before, we are now proving that viruses are key not only inside the bodies of cellular organisms but also as regulators of ecological environments (Sutle, 2005; Scarpellini et al., 2015; Danovaro et al., 2011; Thurber et al., 2017; Brown et al., 2019). In that sense too, viruses can be metaphorically understood as the "new bacteria".

However, we cannot say that viruses are the modal organisms on Earth due to the fact that viruses cannot be considered organisms. As we saw, this conceptual issue has always tricked virology researchers: what are viruses? How to deal with viruses? Which words should we describe to refer to viruses? As viruses are not considered to be alive by current scientific understanding, we cannot say they are neither organisms nor microbes. Plus, they cannot be said to be "obligatory cellular pathogens" as the great majority of them do not cause any disease. Considering the tantalizing amount of them that exist in Earth, the life phenomenon would probably be unviable in the case that most of them were actually infectious. Thus, many researchers have been using the term "biological entities". Here we would like to make a clearer and more concise definition: we aim to propose that viruses were the first organisms to arise on Earth. Viruses are very important organisms capable to replicate and live in symbiotic relations to other viruses and cells as well as to infect cellular organisms too. Without viruses we would have neither eukaryotes nor mammals as accumulated evidence

 doi:10.20944/preprints202104.0743.v1

sustain that they were of utmost relevance to the origin of both the nuclear membrane (Forterre, 2005; 2006) and the placenta (Cornelis et al., 2012; Lavialle et al, 2013; Denner, 2016). Thus, without them we would not be here to investigate the origin of these… these what?

What are viruses?

When working with the origins of life, we have faced many conceptual and theoretical questions to deal with (Prosdocimi et al., 2018). Possibly the most challenging was the problem about "What is life?" (Farias et al., 2020), a huge and old question in biology. If we assume the model proposed by the Nobel prize winner Erwin Schrodinger in his classical book from 1944, life could be characterized by the presence of "aperiodic crystals" that stored genetic information (Schrodinger, 1944). The aperiodic crystal was later shown to be the nucleic acid, with information written in their sequences of nucleotidic bases, evidenced in one of the most relevant biological works of the XXth century (Watson and Crick, 1953). But biologists refused to accept that the mere presence of a nucleic acid was enough to define a living being. For them, the chemical molecule of a nucleic acid has no life at all.

Thus, the current understanding in biology and microbiology often considers that life should be understood as "cellular life". Therefore, the general view is that only cells can be considered as living entities once they present a couple of important characteristics that make them capable to use this "honorific title". Cells present: (i) metabolism; (ii) they are autonomous, and (iii) they are capable to evolve and reproduce. If we take viral entities, we see that, although they can evolve and reproduce, they do not present metabolism and they are not autonomous. (On the other hand, we have already seen that no holobiont can be said to be actually autonomous as they make symbiosis with other organisms such as bacteria and endogenous retroviruses). In any case, researchers are right to say that viral particles (virions) are much more like crystals and cannot be said to be living entities. The question is that viruses do not present neither (i) any colloidal form of cytoplasm on which chemical reaction can occur, nor (ii) any sort of semi-permeable proteolipidic membrane that allows an active chemical interchange between their inside and the outside media. It is clear that viruses cannot be seen as autonomous as they depend on cells to reproduce their genetic information. Actually, a new understanding on the nature of viruses is emerging by the comprehension that they do not need cells to replicate, they merely need a ribosome (Farias et al., 2019). Today, unfortunately, it seems that ribosomes can only be found inside cells. Therefore, viruses need to ask this favor to cells, *i. e.*, to access their ribosomes so that they can be able to reproduce.

 doi:10.20944/preprints202104.0743.v1

Viruses speak the language of biology

Even if viruses are not autonomous and do not present metabolism, they actually can process the most basic biological code that is the genetic code. According to the viewpoint of code and semantic biology, it is more relevant to have an embodied biological code than having any other features to be considered alive (Emmeche, 1998; Barbieri, 2003; Barbieri, 2014). In that sense, if we change our concept of life, viruses can be considered living entities. Although we do not aim to dispute the concept of life here as we did elsewhere (Farias et al., 2020), it is clear that viruses can decode the language of biology. But which language are we talking about?

First, viruses present proteins encoded in the form of nucleic acids. Their highly packed genomes are arranged in genes and those genes are arranged in codons, even if some of their open-reading frames may overlap others. Those codons are written in the form of a messenger-like RNA that can be read by the translation machinery present inside cells, being capable to produce dozens of proteins that will allow (i) the replication of their genetic content, (ii) the building and assembly of their protein capsids, together with (iii) their capacity to bind further cells to exploit their translational machinery.

In that sense, even if one prefers not to accept the concept of life of code biology, it is clear that virus may not be living, but they can be considered biological systems once they understand and make use of this language of biology as disposed in the genetic code.

Viruses are biological systems

We must now define what we do mean exactly by the concept of "*biological systems*". In our particular definition, a "biological system can be defined as any system capable to (i) operate chemical codes of multiple layers and (ii) persist over time". One of the major exponents of code biology, the Italian researcher Marcello Barbieri, argue that a relevant issue in an organic code regards the fact that the coding rules cannot be merely dictated by the strict laws of physics and chemistry (Barbieri, 2014). Under an organic code, the coding system is stablished by self-organized "arbitrary" rules that allow the correspondence between two independent molecular worlds (Barbieri, 2003). The organic code must comprise a semantic coding system containing (i) a signal and its biological sense, (ii) an adaptor molecule and (iii) a code itself. In biological systems, the signal is presented in the form of genes as codon-organized sequence of nucleotidic bases. The biological sense should be considered as the organized proteins produced from the information encoded on that nucleic acid. The adaptor molecules work in the system of encoding and/or decoding, that is, the translation system. More specifically, they represent the tRNAs and the tRNA-aminoacyl synthetases, together with the mechanism on which each amino acid is bound to a specific tRNA containing a given anti-codon (decoder), and other signals

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 April 2021 doi:10.20944/preprints202104.0743.v1

(Zamudio et al., 2020). This latter mechanism consists in the way on which the genetic code is mechanistically processed inside cells (Barbieri, 2014). Over this first order code that inaugurates biology since the origin of the First Universal Common Ancestor (FUCA; Prosdocimi et al., 2020), other higher-level codes were self-organized (Farias et al., 2020). For example, in viruses there is the code that specifies how a capsid protein is assembled together with others to form the whole capsid that will integrate the nucleic acid. In cells, the control of transcription and translation is also precisely regulated by accessory proteins that form a new layer of order and a specific, complex code of operation.

One of the greatest mysteries in biology is how the genetic code has emerged and evolved. Although there are many competing theories describing its origins and early evolution (Szathmáry, 1999; Guimarães et al., 2008; Rodin et al., 2011; Koonin and Novozhilov, 2017), the most relevant issue for us here is to understand that any entity capable to operate this code and other overlapping codes should be understood as a biological system. But which are those entities? Well, it is clear that cells and viruses are among them. However, there is still another striking relevant biological system that was missing in the discussion until this moment. Now, we need to present them in order to explain the history of organicity on Earth. These biological systems were classically named **progenotes** by Carl Woese and Sidney Fox (Woese and Fox, 1977). As we shall see, they will prove to be of particular relevance in the evolution of viruses and cells from those ancestral biological systems derived from FUCA.

Major transitions in pre-biotic evolution

Although science has advanced enormously along the last century and keeps advancing at a higher pace in the XXIst century, the quest for the origin of life is still unsolved. And although many gaps remain in the history of early life in Earth, there are also many consensual issues among researchers that work in the field. For example, most of them agree that the translation machinery evolved very early in the origin of life (Agmon, 2009; Davidovich et al., 2009; Belousoff et al., 2010; Farias et al., 2014; Petrov et al., 2015; Farias et al., 2017, Farias and Jose, 2020). Otherwise, it would not make sense to have nucleic acids encoding proteins for biochemical pathways in the case you have none system to translate them. Also, we cannot break the central dogma of molecular biology, *i.e.*, there has never been a way to produce nucleic acids from proteins. The dogma functions with a precise direction. DNA makes RNA that makes protein. Even if we know that the enzyme reverse transcriptase can produce DNA from RNA, being probably involved in the origin of the DNA itself, the important issue in the central dogma is that protein information cannot travels back to nucleic acid information. Besides, we have seen that DNA had no role in the origin of life and that RNA was the protagonist (Gilbert, 1986; Di Giulio, 2021). Another important consensus we urge to consider is that biological systems started with *open*,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 April 2021 doi:10.20944/preprints202104.0743.v1

non-encapsulated systems with molecules that interacted directly to each other without the presence of any sort of capsid or envelope.

In that sense, we have hypothesized previously about the existence of a *First Universal Common Ancestor* (FUCA; Prosdocimi et al., 2019; Figure 1c) that originated from a catalytic RNA capable to bind amino acids together (Figure 1b), the early version of a peptidyl-transferase center (PTC) often named proto-PTC (Prosdocimi et al., 2020). FUCA matured when a first form of genetic code has emerged. As we saw, when thinking about genetic information, there is a need that the genetic code had been created very early because, otherwise, no other gene would have any meaning. Thus, the origin of the genetic code is actually the very origin of genetic information and it is a first and necessary step to the origin of all other genes. FUCA therefore inaugurates the existence of organic entities made of proteins encoded in nucleic acids information.

We like to think of FUCA as a predecessor of an entire age of open biological systems that we named as **the progenote era** (de Farias et al., 2021). At this age of progenotes, following the self-organization of the genetic code, protein coding genes started to evolve. Possibly the most successful genes to evolve were the ones capable to bind back nucleic acids polymers and other abundant molecules existing in their surrounding environment. This way they could structurally stabilize both themselves and those molecules. Following this first stabilization period, the evolution of the first biochemical pathways happened (Figure 1d). Some of the first pathways to evolve were probably related to amino acid biosynthesis and/or degradation since these molecules were known to be found in high amounts according to Urey-Miller like experiments (Parker et al, 2011a; Parker et al., 2011b) and they were already in use by FUCA-like systems. Also, pathways related to nucleotide production and degradation must be primitive as the genetic information has been carried out by those molecules. For the ones worried about bioenergetics issues related to the energetic requirement necessary for that metabolism to happen, we are pleased to remember that ATP is actually the adenine nucleotide. Our hypothesis suggests that ATP was firstly used as the RNA nucleotide and produced for that reason and eventually become an important molecule for energy interchange by exaptation. In any case, the failure to find evidence in favor of the antiquity of these pathways for nucleotide and amino acid processing might indicate that the very primordial pathways could have been further replaced by more efficient and modern mechanisms. Even if it seems to be an *ad hoc* defense, the pivotal relevance of pathways involving nucleic acids and amino acids for the early metabolism are unquestionable. Besides, it is well known that, at the progenote age, these open biological systems evolved by an intense interexchange of genetic information by lateral transference. As they presented neither capsids nor envelopes, their *naked nucleic acids* were capable of either breaking or fusing easily. The fusion of nucleic acid pieces allowed the initial assembly of the first protogenomes (Figure 1d). Those protogenomes got together genetic information for dealing with interchangeable cycles of chemical molecules present in their media and started to maturate the biochemical pathways.

 doi:10.20944/preprints202104.0743.v1

Due to the fact that the origin of life was ribonucleoproteic, we believe that viral capsids evolved earlier than proteolipidic bilayers. Both the viral capsids are orders of magnitude simpler than lipid membranes and they are made of proteins. At some point, a gene happened to be capable to produce proteins that join, fit in or box together. Those proteins were capable therefore to create some sort of protection box for the nucleic acids (Figure 1e) probably by binding together in symmetric arrangements. This was the origin of primitive capsids capable of storing, protecting and maintaining the genetic information unchanged along time. This would help to avoid the error catastrophe (Eigen, 1971), *i. e.*, the possibility that some catastrophic event might destroy all the genetic information acquired at a given time. Anyhow, this protection allowed nucleic acids to endure more in time and to be transported to other geographic sites on which naked molecules may have had more difficulty to endure.

The hypothesis proposed in the last paragraph describes a putative origin for virus much before than cells existed. At that time, different progenote *quasi-species* (Woese and Fox, 1977; Biebricher and Eigen, 2006) that happened to interchange and join together genes that act on similar metabolites were capable to self-assemble biochemical pathways. Those *quasi-species* could interact more directly to the environment and were selected over other *quasi-species* uncapable to promote that interaction. Therefore, those progenote populations that became capable to create protein-made capsid-like protections endured more, stored their information for much longer and could travel through different environments to seed sterile nearby geographic sites with the breath of organicity.

It is intriguing to know that no existing virus was found to present the genetic information responsible to encode features from the protein synthesis apparatus. Considering the vastness of virosphere, we believe to be possible to find, in the future, viruses that will contain information to produce ribosomes, tRNAs and aminoacyl tRNA-synthetases. Actually, a mimivirus named Tupanvirus was found to present up to 70 tRNAs, 20 aminoacyl-tRNA synthetases, 11 translation factors and an intron of the 18S ribosomal RNA (Abrahão et al., 2018). Even if the evolution of mimiviruses is nowadays better explained by the reduction hypothesis (Patil and Kondabagil, 2021), it is perfectly possible to suppose that some progenotes *quasi-species* could eventually co-opt the information to make ribosomes and the translational apparatus inside viral-like protein capsids. This would be important for seeding nearby geographic regions with the protein fabrics. Further, these capsids may have evolved to get more and more sophisticated; and different lineages of these viral-like *quasi-species* may have acquired and transported different gene sets to farther distances (Figure 1f).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 April 2021 doi:10.20944/preprints202104.0743.v1

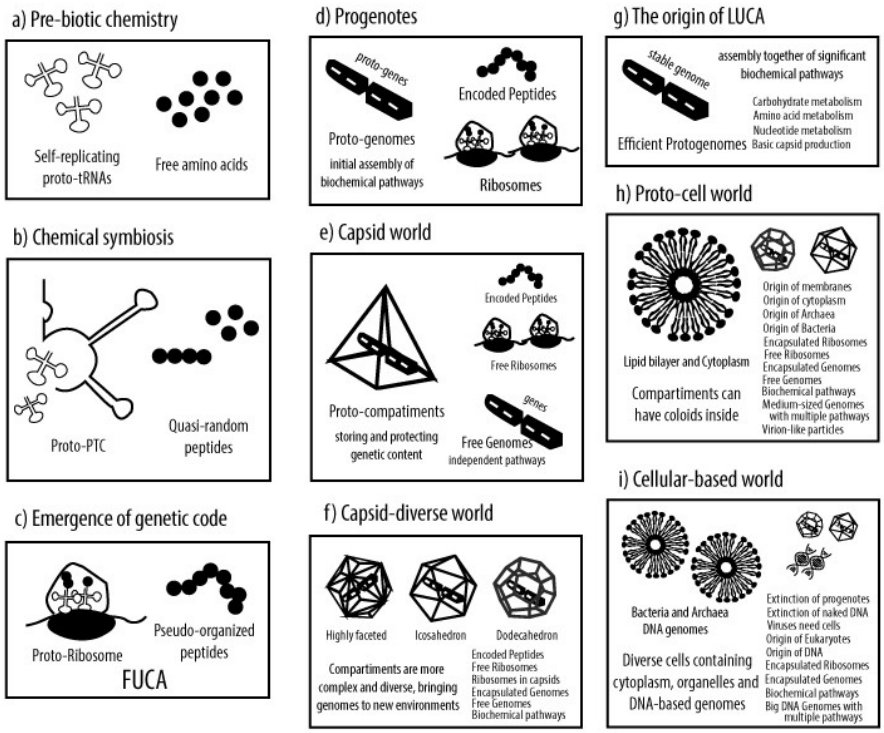

Figure 1: Putative major steps in the early evolution of life. When (a) an RNA-world and a peptide-world come into contact, a (b) proto-PTC originates as an RNA folded in such a way that it is capable of catalyzing the random synthesis of small peptides. Then, (c) an organic code emerged as the genetic code giving origin to FUCA. From FUCA, (d) the progenote world evolves with intense interchange and accumulation of genetic material that existed in "naked" form (non-encapsulated). At that time, different populations of progenotes started to bind different molecules in the environment and produce the first genomes and biochemical pathways. (e) Proteins capable of building simple capsids originated and allowed the protection of the genomes, avoiding error catastrophes. (f) Capsid proteins evolved to more complex forms and allowed RNA genomes to conquer different environments. (g) At some point, one lineage of progenote becomes better fitted as it was capable of assembling some important biochemical routes into its genome, giving origin to LUCA (as a progenote). (h) The evolution of at least two alternative pathways both for lipid-binding proteins and DNA biosynthesis allowed the origin of the two basal prokaryotic groups. (i) Finally, the progenote world of naked nucleic acids was extinct and gave rise to a viral and cellular, organismic world. Encapsulated progenotes capable of interacting with cells became viruses. DNA evolved as the most stable molecule to make genomes. Viruses became a strategy of life that could also be achieved by simplification and genome reduction.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 April 2021 doi:10.20944/preprints202104.0743.v1

Further, these nucleic acids encapsulated by proteins inaugurated the era of closed biological systems, the organismal age. We argued elsewhere that this was the origin of the living beings, as the progenotes cannot be considered as such because they are open systems (Prosdocimi and Farias, submitted). And the following questions should be: how and when cells entered into this history? Well, this is a controversial issue, and some researches suggest that cells may have been originated more than once (Di Giulio, 2020; de Farias et al., 2021). If we take for granted that life emerged as ribonucleoproteic, it becomes necessary to explain how lipids got into the game. Recent theories consider that the origin of cells started with the contingent origin and evolution of lipid-binding proteins (Sojo, 2019). As proteins acquired hydrophobic amino acids, those started to bind hydrophobic molecules present in the surrounding medium. Phospholips were probably among them, possibly produced in some special environments (prebiotic refugia) on which they could be synthesized spontaneously (Prosdocimi et al., 2021b). It is amazing to realize that the membranes present in Bacteria and Archaea are strikingly different, binding different sorts of phospholipids in different carbon molecules of the glycerol backbone (Sojo, 2019). Together with the fact that enzymes involved in the DNA replication mechanism are different between these two basal domains of life (Di Giulio, 2020), we have suggested that Bacteria and Archaea have possibly evolved independently from different progenote *quasi-species* (Figure 1h; de Farias et al., 2021). But it is also noteworthy that many biochemical pathways are homologous between Bacteria and Archaea, such as the whole catabolism and anabolism of carbohydrates, amino acids, and nucleotides, among others. Therefore, we must understand that LUCA, the common ancestral of Bacteria and Archaea, was the entity on which those homologous pathways first aggregate together (Figure 1g). And if LUCA existed previously to the origin of membranes, it was therefore a progenote. The independent origin of these two different types of prokaryotic cells have finally produce the beginning of the organismal age in biology (Figure 1i).

How did viruses originate?

It has been proposed before that viruses should not be understood as a monophyletic clade, but rather as a strategy of life (de Farias et al., 2019). As we saw earlier, there is growing evidence that cells possibly evolved at least two times since both DNA replication mechanisms and lipid biosynthetic pathways are not homologous between Bacteria and Archaea (Di Giulio, 2021; Farias et al., 2021). Therefore, it seems that both cellular and viral architectures are not monophyletic and have originated multiple times in the history of life on Earth. They should therefore be understood as different *strategies* that molecular systems acquired to preserve their organic structure and evolve.

Plus, it seems that the viral strategy and architecture could be achieved by at least three routes (Wessner, 2010): (i) reduction from cellular organisms, (ii)

encapsulation of transposon-like elements or (iii) evolution from progenotes. The hypothesis of reduction is based on the fact that cells that become either parasites or obligatory symbionts often loose great parts of their genomes (López-Madrigal et al., 2011; McCutcheon and Moran, 2012). Also, reduction seems to be the most parsimonious hypothesis to explain the existence of giant viruses from *Mimiviridae* family, which present extremely large genomes (Raoult et al., 2004; Abrahão et al., 2018). There is also the possibility that transposons and retroposon-like elements could be inserted next to capsid proteins and allow the evolution of new viruses (Wessner, 2010). Finally, the evolution of viruses from progenotes in an age before cells even existed seems likely, at least for particular viral clades.

One of the main evidences for a common origin for viruses is the presence of specific non-cellular types of genes and protein structures in many viral clades that are capable to infect both Bacteria and Archaea (Abrescia et al., 2012; Nasir and Caetano-Anollés, 2015). These genes were probably inherited from a common lineage before cells existed. Nasir and collaborators (2020) suggest the existence of a structure named virocell to cope with the fact that virus actually need a cell to reproduce. They suggest that virus originated, therefore, just after the first cells appeared. Our approach is that this is not necessary, as viruses do not need cells to reproduce, they need ribosomes (Farias et al., 2019). And they probably could find "naked ribosomes" in the environment at the age of progenotes. Actually, there exist evidence for the existence of extracellular ribosomes even in the contemporary world (Sadik et al., 2018; Tosar et al., 2020). Though many clades of virus may have been originated in the progenote age of open biological systems, it is nowadays difficult to know which were them. It is often supposed that the older viruses would be single stranded RNA viruses as there is a consensual agreement among researchers that double-stranded nucleic acid molecules and DNA are improvements acquired along the evolution of nucleic acids (Prosdocimi and Farias, 2019; Di Giulio, 2021). Additionally, it is more likely to suppose that some viruses containing single strand RNA molecules should be older than other forms of viruses (Nasir et al., 2020). After that, double stranded RNA viruses should have been originated and, finally, viruses containing DNA genomes.

Concluding remarks

Viruses have mistakenly been understood as mere infectious agents along the history of microbiology. Nowadays, viruses are known to be the most abundant biological systems present on Earth by far. They are present mostly in the oceans, what bring us back to the original application of the latin word *virus* by the Dutch botanist Martinus Beijerinck. Curiously, he named viruses as such and mistakenly identified them as liquid entities. Following the ideas of Jean-Michel Claverie, viruses should not be confused with their viral particles or virions, and those particles are better seen as analogous to the viral gametes (Claverie, 2006; Nasir et al., 2020). Viruses are "the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 April 2021 doi:10.20944/preprints202104.0743.v1

new bacteria" as they are also proving to be important endosymbionts capable to regulate our homeostasis and the natural environment. It is possible to infer that viral infections might erupt only when organisms are out of balance, and such entities could be actually trying to help in bringing the organism back to homeostasis. In that case, there is the possibility that many viral diseases are not actually caused by the viruses. Additionally, we know that about 8% of human genomes are composed of endogenous retrovirus. The role of those to mix genes and produce orphan genes cannot be neglected.

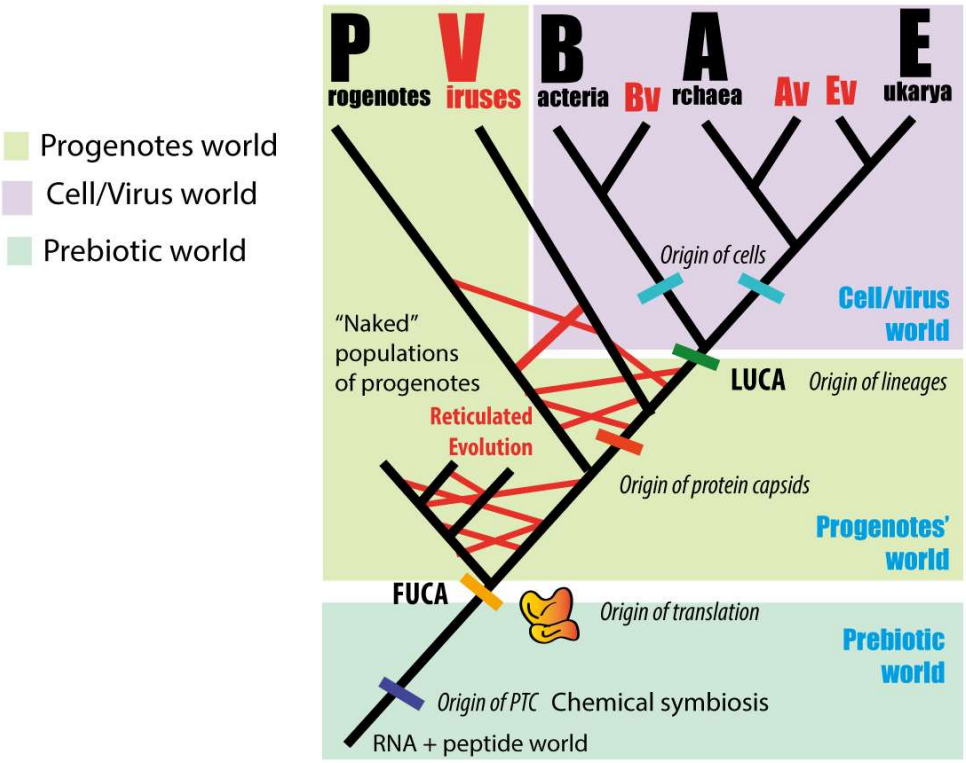

Figure 2: Putative routes for the evolution of viruses. Most viruses have probably evolved from pre-LUCA ages as they share genes absent in cellular organisms. However other viral lineages may have evolved from genome reduction and simplification of bacteria (BV as Bacterioviruses), archaea (AV as Archaeoviruses) and eukaryotes (Eukaryoviruses). Those lineages most likely acquired genes from some pre-LUCA viruses to produce their capsids and performed some reticulated evolution with those (not shown in figure). If we consider them as viruses, the mimiviruses are probably some special type of eukaryoviruses.

Together with cells and progenotes, viruses compose both (i) a peculiar strategy of life, and (ii) a biological system. The origin of life on Earth happened when RNAs started to catalyze peptide synthesis producing a chemical symbiosis (Prosdocimi et al., 2021a). After random peptide synthesis, the genetic code has been self-assembled producing the first organic entity and ancestor of all progenotes, the FUCA

(Prosdocimi et al., 2019). FUCA evolved into a population of open-biological systems known as progenotes that consisted of naked complexes of ribonucleoproteins on which lateral transference played an important role, allowing small genomes to assemble into middle-to-large genomes. Biochemical pathways evolved when peptides bound to molecules from the environment producing some sort of cyclical routes of metabolic recycling. At some point, capsid-like proteins evolved and were capable to encapsulate protogenomes, protecting them from degradation and error catastrophe. The encapsulation also allowed these early systems to travel into nearby environments and colonize farther sites containing organic molecules.

It is possible that many viral clades should be descendants from the progenote age, as they present a significant set of virus-specific genes and protein structures that are lacking in all known cellular organisms. Most biochemical pathways were self-organized in that age and formed an advanced sort of progenote that was the most recent common ancestor between Bacteria and Archaea. The genome of this biological system named LUCA was composed of pathways as important as carbohydrate, amino acid and nucleic acid synthesis, among others. In that scenario, LUCA was not cellular. Then, when proteins started to interact with lipids (Sojo, 2019), the evolution of lipid-binding proteins followed along two routes: (a) in one of them, glycerol-1-phosphate would bind irregular lipids and, (b) in another, glycerol-6-phosphate would bind regular ones. Those two populations would be the ancestors of either Bacteria or Archaea domains. The lipid layers that encapsulated LUCA-like progenotes and their surrounding media originated probably when RNA was still the main nucleic acid forming the genome of progenotes. However, it is possible that other progenotes had already discovered DNA synthesis and made symbiosis with those early cells. However, DNA synthesis also followed two main routes: a bacterial and an archaeal one, indicating symbiosis with different populations of DNA-making progenotes. Further, specialized archaea from Asgard clade fused with a Rickettsia-like bacteria and viruses, producing mitochondria and the eukaryotic cell nucleus (Forterre, 2006). We believe to have enough evidence to suppose that many viral clades came from the age of progenotes, being named simply as Viruses or Progenoteviruses (PV).

However, other types of viruses such as mimiviruses were probably derived from genome reduction. Many viral clades were likely originated from reduction from bacterial, archaeal and eukaryotic genomes. Also, it is possible that some viruses have been originated from either (i) genome fusions between any of those viral clades previously described, or from (ii) transposons fused with capsid genes inside cells, due to exon shuffling or other genomic recombination mechanisms.

The relevance of viruses is extreme and once this fact has been clearly understood, we cannot name these entities anymore as *venoms* or *poisons*. Therefore, we propose a new acronym of VIRUS as Very Important Replicator Unit and Symbiont. As very important seems too little to cope with their *Utmost* importance in ecology and evolution in Earth, we suggest that non-infectious viruses should be better referred as the acronym of UIRUS from now on.

Acknowledgements

We would like to thank to Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ grant number CNE E-26/202.780/2018) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq grant number PDE 205072/2018-6) for financing FP. Marco V. José was financially supported by PAPIIT-IN201019, UNAM, México.

References

Abrahão J, Silva L, Silva LS, Khalil JYB, Rodrigues R, Arantes T, Assis F, Boratto P, Andrade M, Kroon EG, Ribeiro B, Bergier I, Seligmann H, Ghigo E, Colson P, Levasseur A, Kroemer G, Raoult D, La Scola B. Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. Nat Commun. 2018 Feb 27;9(1):749. doi: 10.1038/s41467-018-03168-1. PMID: 29487281; PMCID: PMC5829246.

Abrescia NG, Bamford DH, Grimes JM, Stuart DI. Structure unifies the viral universe. Annu Rev Biochem. 2012;81:795-822. doi: 10.1146/annurev-biochem-060910-095130. Epub 2012 Apr 5. PMID: 22482909.

Afzal I, Shinwari ZK, Sikandar S, Shahzad S. Plant beneficial endophytic bacteria: Mechanisms, diversity, host range and genetic determinants. Microbiol Res. 2019 Apr;221:36-49. doi: 10.1016/j.micres.2019.02.001. Epub 2019 Feb 4. PMID: 30825940.

Agmon I. The dimeric proto-ribosome: Structural details and possible implications on the origin of life. Int J Mol Sci. 2009 Jun 30;10(7):2921-34. doi: 10.3390/ijms10072921.

Arendsee ZW, Li L, Wurtele ES. Coming of age: orphan genes in plants. Trends Plant Sci. 2014 Nov;19(11):698-708. doi: 10.1016/j.tplants.2014.07.003. PMID: 25151064.

Barbieri M (2003) The organic codes: an introduction to semantic biology. Cambridge University Press, Cambridge.
Barbieri, M. From Biosemiotics to Code Biology. Biol Theory 9, 239–249 (2014a). https://doi.org/10.1007/s13752-013-0155-6.

Beasley DE, Koltz AM, Lambert JE, Fierer N, Dunn RR. The Evolution of Stomach Acidity and Its Relevance to the Human Microbiome. PLoS One. 2015 Jul 29;10(7):e0134116. doi: 10.1371/journal.pone.0134116. PMID: 26222383; PMCID: PMC4519257.

Belousoff MJ, Davidovich C, Zimmerman E, Caspi Y, Wekselman I, Rozenszajn L, Shapira T, Sade-Falk O, Taha L, Bashan A, Weiss MS, Yonath A. Ancient machinery embedded in the contemporary ribosome. Biochem Soc Trans. 2010 Apr;38(2):422-7. doi: 10.1042/BST0380422.

Biebricher CK, Eigen M. What is a quasispecies? Curr Top Microbiol Immunol. 2006;299:1-31. doi: 10.1007/3-540-26397-7_1. PMID: 16568894.

Brown MR, Baptista JC, Lunn M, Swan DL, Smith SJ, Davenport RJ, Allen BD, Sloan WT, Curtis TP. Coupled virus - bacteria interactions and ecosystem function in an engineered microbial system. Water Res. 2019 Apr 1;152:264-273. doi: 10.1016/j.watres.2019.01.003.

Claverie, J.M. (2006) Viruses take center stage in cellular evolution. Genome Biol. 7, 110.

Comeau AM, Hatfull GF, Krisch HM, Lindell D, Mann NH, Prangishvili D. Exploring the prokaryotic virosphere. Res Microbiol. 2008 Jun;159(5):306-13. doi: 10.1016/j.resmic.2008.05.001. Epub 2008 May 29. PMID: 18639443.

Cornelis G, Heidmann O, Bernard-Stoecklin S, Reynaud K, Véron G, Mulot B, Dupressoir A, Heidmann T. Ancestral capture of syncytin-Car1, a fusogenic endogenous retroviral envelope gene involved in placentation and conserved in Carnivora. Proc Natl Acad Sci U S A. 2012 Feb 14;109(7):E432-41. doi: 10.1073/pnas.1115346109. Epub 2012 Jan 17. PMID: 22308384; PMCID: PMC3289388. Format:

Culley A. New insight into the RNA aquatic virosphere via viromics. Virus Res. 2018 Jan 15;244:84-89. doi: 10.1016/j.virusres.2017.11.008. Epub 2017 Nov 11. PMID: 29138044.

Danovaro R, Corinaldesi C, Dell'anno A, Fuhrman JA, Middelburg JJ, Noble RT, Suttle CA. Marine viruses and global climate change. FEMS Microbiol Rev. 2011 Nov;35(6):993-1034. doi: 10.1111/j.1574-6976.2010.00258.x.

Davidovich C, Belousoff M, Bashan A, Yonath A. The evolving ribosome: from non-coded peptide bond formation to sophisticated translation machinery. Res Microbiol. 2009 Sep;160(7):487-92. doi: 10.1016/j.resmic.2009.07.004.

de Farias ST, Jheeta S, Prosdocimi F. Viruses as a survival strategy in the armory of life. Hist Philos Life Sci. 2019 Oct 14;41(4):45. doi: 10.1007/s40656-019-0287-5. PMID: 31612293.

de Farias ST, Jose MV, Prosdocimi F. Is it possible that cells have had more than one origin? Biosystems. 2021 Jan 30;202:104371. doi: 10.1016/j.biosystems.2021.104371. Epub ahead of print. PMID: 33524470.

de Farias ST, Rêgo TG, José MV. tRNA Core Hypothesis for the Transition from the RNA World to the Ribonucleoprotein World. Life (Basel). 2016 Mar 23;6(2):15. doi: 10.3390/life6020015. PMID: 27023615; PMCID: PMC4931452.

Di Giulio M. The late appearance of DNA, the nature of the LUCA and ancestors of the domains of life. Biosystems. 2021 Apr;202:104330. doi: 10.1016/j.biosystems.2020.104330. Epub 2020 Dec 19. PMID: 33352234.

Eigen, M. Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften 58, 465–523 (1971). https://doi.org/10.1007/BF00623322

Emmeche C (1998). Defining life as a semiotic phenomenon. Cybernetics And Human Knowing, Vol.5, no.1, 1998, pp. 3–17.

Farias ST, Gaudêncio Rêgo T, José MV. Peptidyl Transferase Center and the Emergence of the Translation System. Life (Basel). 2017 Apr 25;7(2):21. doi: 10.3390/life7020021.

Farias ST, José MV. Transfer RNA: The molecular demiurge in the origin of biological systems. Prog Biophys Mol Biol. 2020 Jul;153:28-34. doi: 10.1016/j.pbiomolbio.2020.02.006.

Farias S, Prosdocimi F, Caponi G. Organic Codes: A Unifying Concept to Life? Preprints 2020, 2020110412 (doi: 10.20944/preprints202011.0412.v1).

Farias ST, Rêgo TG, José MV. Origin and evolution of the Peptidyl Transferase Center from proto-tRNAs. FEBS Open Bio. 2014 Feb 8;4:175-8. doi: 10.1016/j.fob.2014.01.010.

Forterre P. The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. Biochimie. 2005 Sep-Oct;87(9-10):793-803. doi: 10.1016/j.biochi.2005.03.015. Epub 2005 Apr 12. PMID: 16164990.

Forterre P. The origin of viruses and their possible roles in major evolutionary transitions. Virus Res. 2006 Apr;117(1):5-16. doi: 10.1016/j.virusres.2006.01.010. Epub 2006 Feb 14. PMID: 16476498.

García-López R, Pérez-Brocal V, Moya A. Beyond cells - The virome in the human holobiont. Microb Cell. 2019 Jul 1;6(9):373-396. doi: 10.15698/mic2019.09.689. PMID: 31528630; PMCID: PMC6717880.

Gilbert SF, Sapp J, Tauber AI. A symbiotic view of life: we have never been individuals. Q Rev Biol. 2012 Dec;87(4):325-41. doi: 10.1086/668166. PMID: 23397797.

Gilbert, W. Origin of life: The RNA world. Nature 319, 618 (1986). https://doi.org/10.1038/319618a0

Gould, Stephen Jay. Full House. Harvard University Press, 1996. JSTOR, www.jstor.org/stable/j.ctt2jbtw4. Accessed 22 Feb. 2021.

Greer R, Dong X, Morgun A, Shulzhenko N. Investigating a holobiont: Microbiota perturbations and transkingdom networks. Gut Microbes. 2016;7(2):126-35. doi: 10.1080/19490976.2015.1128625. Epub 2016 Mar 16. PMID: 26979110; PMCID: PMC4856449.

Guerrero R, Margulis L, Berlanga M. Symbiogenesis: the holobiont as a unit of evolution. Int Microbiol. 2013 Sep;16(3):133-43. doi: 10.2436/20.1501.01.188. PMID: 24568029.

Guimarães RC, Moreira CH, de Farias ST. A self-referential model for the formation of the genetic code. Theory Biosci. 2008 Aug;127(3):249-70. doi: 10.1007/s12064-008-0043-y.

Jover, L., Effler, T., Buchan, A. *et al.* The elemental composition of virus particles: implications for marine biogeochemical cycles. *Nat Rev Microbiol* **12,** 519–528 (2014). https://doi.org/10.1038/nrmicro3289

Koonin EV, Novozhilov AS. Origin and Evolution of the Universal Genetic Code. Annu Rev Genet. 2017 Nov 27;51:45-62. doi: 10.1146/annurev-genet-120116-024713.

Lander ES, Linton LM, et al., International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature. 2001 Feb 15;409(6822):860-921. doi: 10.1038/35057062. Erratum in: Nature 2001 Aug 2;412(6846):565. Erratum in: Nature 2001 Jun 7;411(6838):720. Szustakowki, J [corrected to Szustakowski, J]. PMID: 11237011.

Lavialle C, Cornelis G, Dupressoir A, Esnault C, Heidmann O, Vernochet C, Heidmann T. Paleovirology of 'syncytins', retroviral env genes exapted for a role in placentation. Philos Trans R Soc Lond B Biol Sci. 2013 Aug 12;368(1626):20120507. doi: 10.1098/rstb.2012.0507. PMID: 23938756; PMCID: PMC3758191.

López-Madrigal S, Latorre A, Porcar M, Moya A, Gil R. Complete genome sequence of "Candidatus Tremblaya princeps" strain PCVAL, an intriguing translational machine below the living-cell status. J Bacteriol. 2011 Oct;193(19):5587-8. doi: 10.1128/JB.05749-11. PMID: 21914892; PMCID: PMC3187454.

McCutcheon, J., Moran, N. Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol 10, 13–26 (2012). https://doi.org/10.1038/nrmicro2670

Moniruzzaman, M. et al. (2020) Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. Nat. Commun. 11, 1–11

Nasir A, Caetano-Anollés G. A phylogenomic data-driven exploration of viral origins and evolution. Sci Adv. 2015 Sep 25;1(8):e1500527. doi: 10.1126/sciadv.1500527. PMID: 26601271; PMCID: PMC4643759.

Parker ET, Cleaves HJ, Callahan MP, Dworkin JP, Glavin DP, Lazcano A, Bada JL. Enhanced synthesis of alkyl amino acids in Miller's 1958 H2S experiment. Orig Life Evol Biosph. 2011 Dec;41(6):569-74. doi: 10.1007/s11084-011-9253-2. Epub 2011 Dec 3. PMID: 22139514.

Parker ET, Cleaves HJ, Dworkin JP, Glavin DP, Callahan M, Aubrey A, Lazcano A, Bada JL. Primordial synthesis of amines and amino acids in a 1958 Miller H2S-rich spark discharge experiment. Proc Natl Acad Sci U S A. 2011 Apr 5;108(14):5526-31. doi: 10.1073/pnas.1019191108. Epub 2011 Mar 21. PMID: 21422282; PMCID: PMC3078417.

Petrov AS, Gulen B, Norris AM, Kovacs NA, Bernier CR, Lanier KA, Fox GE, Harvey SC, Wartell RM, Hud NV, Williams LD. History of the ribosome and the origin of translation. Proc Natl Acad Sci U S A. 2015 Dec 15;112(50):15396-401. doi: 10.1073/pnas.1509761112.

Postler TS, Ghosh S. Understanding the Holobiont: How Microbial Metabolites Affect Human Health and Shape the Immune System. Cell Metab. 2017 Jul 5;26(1):110-130. doi: 10.1016/j.cmet.2017.05.008. Epub 2017 Jun 15. PMID: 28625867; PMCID: PMC5535818.

Patil S, Kondabagil K. Coevolutionary and phylogenetic analysis of Mimiviral replication machinery suggest the cellular origin of Mimiviruses. Mol Biol Evol. 2021 Feb 11:msab003. doi: 10.1093/molbev/msab003. Epub ahead of print. PMID: 33570580.

Prosdocimi F and Farias S. (2019) A emergência dos sistemas biológicos: uma visão molecular sobre a origem da vida. ArteComCiencia. 1 ed. Rio de Janeiro. https://www.amazon.com.br/dp/B08WTHJPM4

Prosdocimi F, Farias ST, José MV. Prebiotic Chemical Refugia: Multifaceted Scenario for the Formation of Biomolecules in Primitive Earth. Preprints 2021b, 2021030341 (doi: 10.20944/preprints202103.0341.v1).

Prosdocimi F, Jheeta S, Torres de Farias S. Conceptual challenges for the emergence of the biological system: Cell theory and self-replication. Med Hypotheses. 2018 Oct;119:79-83. doi: 10.1016/j.mehy.2018.07.029. Epub 2018 Jul 30. PMID: 30122496.

Prosdocimi F, José M, Farias S. The First Universal Common Ancestor (FUCA) as the Earliest Ancestor of LUCA's (Last UCA) Lineage. Chapter 3. In book: Evolution, Origin of Life, Concepts and Methods. Pierre Pontarotti (Editor). Springer. October 2019. doi 10.1007/978-3-030-30363-1_3

Prosdocimi F, José MV and de Farias ST. The Theory of Chemical Symbiosis: A Margulian View for the Emergence of Biological Systems (Origin of Life). Acta Biotheor 69, 67–78 (2021a). https://doi.org/10.1007/s10441-020-09388-7

Prosdocimi F, Zamudio GS, Palacios-Pérez M, Torres de Farias S, V José M. The Ancient History of Peptidyl Transferase Center Formation as Told by Conservation and Information Analyses. Life (Basel). 2020 Aug 5;10(8):134. doi: 10.3390/life10080134. PMID: 32764248; PMCID: PMC7459865.

Raoult, D. et al. The 1.2-megabase genome sequence of Mimivirus. Science 306, 1344–1350 (2004) doi:10.1126/science.1101485.

Rodin AS, Szathmáry E, Rodin SN. On origin of genetic code and tRNA before translation. Biol Direct. 2011 Feb 22;6:14. doi: 10.1186/1745-6150-6-14.

Russell SL. Transmission mode is associated with environment type and taxa across bacteria-eukaryote symbioses: a systematic review and meta-analysis. FEMS Microbiol Lett. 2019 Feb 1;366(3):fnz013. doi: 10.1093/femsle/fnz013. PMID: 30649338.

Sadik N, Cruz L, Gurtner A, Rodosthenous RS, Dusoswa SA, Ziegler O, Van Solinge TS, Wei Z, Salvador-Garicano AM, Gyorgy B, Broekman M, Balaj L. Extracellular RNAs: A New Awareness of Old Perspectives. Methods Mol Biol. 2018;1740:1-15. doi: 10.1007/978-1-4939-7652-2_1. PMID: 29388131; PMCID: PMC6509047

Sánchez-Cañizares C, Jorrín B, Poole PS, Tkacz A. Understanding the holobiont: the interdependence of plants and their microbiome. Curr Opin Microbiol. 2017 Aug;38:188-196. doi: 10.1016/j.mib.2017.07.001. Epub 2017 Jul 18. PMID: 28732267.
Schrödinger, E. (1944). What is life?: The physical aspect of the living cell ; with Mind and matter ; & Autobiographical sketches. Cambridge: Cambridge University Press.

Scarpellini E, Ianiro G, Attili F, Bassanelli C, De Santis A, Gasbarrini A. The human gut microbiota and virome: Potential therapeutic implications. Dig Liver Dis. 2015 Dec;47(12):1007-12. doi: 10.1016/j.dld.2015.07.008.

Schulz, F. et al. (2020) Giant virus diversity and host interactions through global metagenomics. Nature 578, 432–436

Sender R, Fuchs S, Milo R. Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. Cell. 2016; 164 (3): 337–40. doi:10.1016/j.cell.2016.01.013.

Sojo V. Why the Lipid Divide? Membrane Proteins as Drivers of the Split between the Lipids of the Three Domains of Life. Bioessays. 2019 May;41(5):e1800251. doi: 10.1002/bies.201800251. Epub 2019 Apr 10. PMID: 30970170.

Suttle CA. Viruses: unlocking the greatest biodiversity on Earth. Genome. 2013. 56(10): 542-544. https://doi.org/10.1139/gen-2013-0152

Suttle CA. Viruses in the sea. Nature. 2005 Sep 15;437(7057):356-61. doi: 10.1038/nature04160.

Szathmáry E. The origin of the genetic code: amino acids as cofactors in an RNA world. Trends Genet. 1999 Jun;15(6):223-9. doi: 10.1016/s0168-9525(99)01730-8.

Tautz D, Domazet-Lošo T. The evolutionary origin of orphan genes. Nat Rev Genet. 2011 Aug 31;12(10):692-702. doi: 10.1038/nrg3053. PMID: 21878963.

Thurber RV, Payet JP, Thurber AR, Correa AM. Virus-host interactions and their roles in coral reef health and disease. Nat Rev Microbiol. 2017 Apr;15(4):205-216. doi: 10.1038/nrmicro.2016.176.

Togo A, Dufour JC, Lagier JC, Dubourg G, Raoult D, Million M. Repertoire of human breast and milk microbiota: a systematic review. Future Microbiol. 2019 May;14:623-641. doi: 10.2217/fmb-2018-0317. Epub 2019 Apr 26. PMID: 31025880.

Tosar JP, Segovia M, Castellano M, Gámbaro F, Akiyama Y, Fagúndez P, Olivera Á, Costa B, Possi T, Hill M, Ivanov P, Cayota A. Fragmentation of extracellular ribosomes

and tRNAs shapes the extracellular RNAome. Nucleic Acids Res. 2020 Dec 16;48(22):12874-12888. doi: 10.1093/nar/gkaa674. PMID: 32785615; PMCID: PMC7736827.

Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature. 1953 Apr 25;171(4356):737-8. doi: 10.1038/171737a0. PMID: 13054692.

Weitz JS, Wilhelm SW. Ocean viruses and their effects on microbial communities and biogeochemical cycles. F1000 Biol Rep. 2012;4:17. doi: 10.3410/B4-17. Epub 2012 Sep 5. PMID: 22991582; PMCID: PMC3434959.

Wessner, DR. (2010) The Origins of Viruses. Nature Education 3(9):37

Woese CR, Fox GE. The concept of cellular evolution. J Mol Evol. 1977 Sep 20;10(1):1-6. doi: 10.1007/BF01796132. PMID: 903983.

Zamudio GS, Palacios-Pérez M, José MV. Information theory unveils the evolution of tRNA identity elements in the three domains of life. Theory Biosci. 2020 Feb;139(1):77-85. doi: 10.1007/s12064-019-00301-6. Epub 2019 Sep 18. PMID: 31535331.