Is it possible that cells have had more than one origin?

Sávio Torres de Farias[1,2], Marco V. Jose[3] and Francisco Prosdocimi[4]

[1]Laboratório de Genética Evolutiva Paulo Leminski, Centro de Ciências Exatas e da Natureza, Universidade Federal da Paraíba, João Pessoa, Paraíba, Brazil. https://orcid.org/0000-0002-2997-3629
[2]Network of Researchers on the Chemical Evolution of Life (NoRCEL), Leeds LS7 3RB, UK
[3]Theoretical Biology Group, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, México, D.F. C.P. 04510
[4] Laboratório de Biologia Teórica e de Sistemas, Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil. https://orcid.org/0000-0002-6761-3069

Abstract

Cells occupy a prominent place in the history of life on planet Earth. The central role of cellular organization is observed by the fact that "cellular life" is often used as a synonym for life itself. Thus, most characteristics used to define cells overlap with the ones used to define life. Notwithstanding, new scenarios about the origin of life are bringing alternative views to describe how cells may have evolved from the open biological systems named progenotes. Here, using a logical and conceptual analysis, we re-evaluate the characteristics used to infer a single origin for cells. We argue that some evidences used to support cell monophyly, such as the presence of elements from both the translation mechanism and the universality of the genetic code, actually indicate a unique origin for all "biological systems", a term used to define not only cells, but also virus and progenotes. Besides, we present evidence that at least two biochemical pathways as important as (i) DNA replication and (ii) lipid biosynthesis may not homologous between Bacteria and Archaea. The identities observed between the proteins involved in those pathways along representatives of these two ancestral Domains are too low to indicate common genic ancestry. Altogether these facts can be seen as an indication that cellular organization has possibly evolved two or more times and that LUCA (the Last Universal Common Ancestor) might not have existed as a cellular entity. Thus, we aim to consider the possibility that different strategies acquired by biological systems to exist, such as viral, bacterial and archaeal were originated independently from the evolution of different progenote populations.

Keywords: monophyly, progenotes, membrane evolution, origin of biological systems.

Introduction

In the first half of the 17th century, an innovative instrument opened new perspectives into the microbiological world. Although lenses are known to be used since old Greece and magnifying glasses were already used in the 13th century, the first microscopes began to be designed in Europe about 1620. Those instruments enhanced our power to observe small sized phenomena and biological structures that were previously impossible to be observed in detail. It was 1665 when Robert Hooke (1635-1703), the prestigious English scientist sometimes referred as the "England's Leonardo" (Chapman 1996), published his book called *Micrographia*. In that book, Hook made the first descriptions of self-contained microscopic structures that seemed "small rooms" therefore naming those structures as cells, in reference to the Latin word *cella*. At that time, though, it was not clear whether the observed structure consisted of a living entity or not. An important step to solve this issue was taken in 1676 by the Dutch scientist Antonie Von Leeuwenhoek (1632-1723). He observed that the structures described as cells could move, an important feature to differentiate living and non-living entities at that time (Dunn and Jones 2004). However, it would took more than a century for scientists to acknowledge the actual relevance of cells. Finally, in the 19th century, Mathias Schleiden (1804-1881), Theodor Schwann (1810-1882) and Rudolf Wirchow (1821-1902) laid the foundations of Cell Theory (Ribatti 2018). The German botanist Mathias Schleiden (1804-1881) studied plants under the microscope and proposed, in a book entitled "*Contributions to our knowledge in Phytogenesis*" (1838) that all plants were composed of cells. He also pointed out that the nucleus was an important component present in all plant cells and discovered the existence of the nucleolus. In that same year, Theodor Schwann (1810-1882) published three articles describing: (i) that organs were composed of cells; (ii) providing evidence for the existence of the nucleus in different cell types; and (iii) describing the cellular structure of cartilage (Baker 1955). In the following year (1939), Schwann published an even more consistent work in the form of a book. It was there the first mention about Cell Theory through the use of the German word "*Zellentheorie*" and also the description that the nucleus was present in both plant and animal cells. Besides, he recognized and named the nucleolus in both cell types.

Since those times, cell theory has advanced significantly, being one of the most fundamental principles of biology nowadays. It currently assumes a broad and well-founded set of principles, such as: (i) all living beings are composed of cells; (ii) all cells

2

arise from pre-existing cells; (iii) the cell is the fundamental unit of structure and function in living beings; (iv) the activities of organisms depend on the overall activities of their cells; (v) the flow of energy occurs within cells; (vi) cells have DNA as hereditary material; (vii) all cells have basically the same constitution. Among the points presented above, the characteristics that (ii) all cells arise from pre-existing ones, (vi) present DNA and (vii) have basically the same constitution suggest that cells have had a unique evolutionary origin and, therefore, should be taken as monophyletic entities.

In 1977, the American microbiologists Carl Woese and George Fox were astonished to discover an important division in the molecular diversity of prokaryotes when analyzing RNA sequences from the smaller ribosomal subunit. In that study, they proposed to split the prokaryotic group (known as the Monera kingdom) into two main groups named (i) eubacteria and (ii) archaebacteria. In the following decade, with the advancement of both molecular biology techniques and cell biology studies, Woese et al. (1990) proposed a new classification for life. They therefore created the taxonomic category of Domain, which would be superior to the kingdom classically proposed by the Swedish father of taxonomy Carolus Linaeus in his book *Systema Naturae* (1735). This highest-level category would represent the first major division between basic cellular groups and, this way, life would be composed of three major domains, namely: Bacteria, Archaea and Eukarya. This classification reinforced the idea that cells have had a single and common origin in the deep past. They further suggested that it would be possible to predict the existence of an ancestor for these clades, therefore defining the existence of a last common ancestor named LUCA, the Last Universal Common Ancestor. Since then, the search for characteristics of LUCA and explanations about how it originated have been guiding important works in the origin of life research program (Forterre et al. 2005, Mat et al. 2008, Gogarten and Deamer 2016, Martin et al. 2016). Today, however, we know that the domain Eukarya evolved from a specific group of archaebacteria from the phylum Lokiarchaeota (subphylum Asgard) that was involved in symbiotic relationships with bacterial clades (Spang et al. 2015, Spang et al. 2019). Eukaryotes are therefore a derivate clade and were not involved in the constitution of LUCA.

In any case, the existence of a conserved genetic code shared amongst all cellular entities has been identified as a strong evidence for the monophyly of cells and it is used to corroborate the idea that cells appeared only once (Weiss et al. 2018). However, new conceptual scenarios about the evolutionary history of life on Earth, as well as new perspectives on what we should understand for life (Barbieri 2016, Farias et al. 2020)

inspired the current discussion about whether cells actually originated one single time, and which are the evidences used to support such inference. Although we do not aim to present any definitive view, we would like to raise arguments that may lead to a reassessment of this issue and their implications on our understanding about the early evolution of life. Along that line of reasoning, we will review historic and contemporary evidence to account for the evolutionary history of life in a tentative to contribute with a new outlook on how we should understand the importance of cells in that extraordinary journey of the phenomenon of Life in Earth along the last ~4 billion years.

Are cells monophyletic: which evidences support this view?

The claim that cells are the basic units of life is circular to the very definition of life once life is usually assumed to be cellular life. In that sense, many characters used to support the unique origin of life in Earth are also used as evidence of a unique origin for cells. Therefore, the definition of life as cell is a tautology. The main characteristics emphasized to propose a single origin for cells are: (a) all cells have DNA as informational molecule, (b) all cells possess a nearly identical genetic code together with a ribosome-centered translation machinery, and (c) they present a semipermeable membrane that delimits their internal space and organizes the communication with the external environment. Although DNA is used by viruses, both the mechanism for translating biological information and the semipermeable membrane are exclusive features of cellular entities (Weiss et al. 2018).

In order to consider the characters that actually confirm the monophyletic status of cells, it is necessary to discern which of them (i) are exclusive elements of cells and therefore should be understood as actual evidence of their unique origin and, on the other side, (i') which of them are elements that belong to cells due to the fact that they are shared by other biological systems and configure actually an evidence for monophyly of life, being older than cells. In that latter case, those characters were probably acquired before the appearance of cells and were inherited by them. We define a biological system as any entity that presents an organic code (Farias et al. 2020) and that is capable to produce proteins based on nucleic acid information. All biological systems are a product of a chemical symbiosis (Prosdocimi et al. 2020a) and are descendants of FUCA, the First Universal Common Ancestor (Prosdocimi et al. 2019).

A) DNA as an informational molecule

An important issue that needs to be better understood is what happened between the emergence of biological systems and the emergence of cells, that is, the period between FUCA and LUCA (Prosdocimi et al. 2018; Prosdocimi and Farias 2020). Even if we agree that cells are the basic units of life, several evidences point to the fact that life began to be organized and emerged long before the appearance of those compartmentalized biological entities (Woese 1998, Robertson and Joyce 2012, Higgs and Lehman 2015, Prosdocimi et al 2018, Joyce and Szostak 2018). The most accepted model that describes a "pre-cellular era" postulates that life was initially organized in an RNA-world (Gilbert 1986). At that time, primitive RNA molecules performed both the function of information storage and the establishment of a proto-metabolism (Poole et al. 1998, Dworkin et al. 2003, Robertson and Joyce 2012, Vazquez-Salazar and Lazcano 2018). After an initial period on which all catalytic activities were operated by ribozymes, this major biological function was gradually replaced by the activities of peptidic enzymes made of chemically diverse amino acids (Cech 2009). It is well known that the RNA world theory presents theoretical difficulties, mainly related to its disregard of the importance and abundance of amino acids in early prebiotic contexts, but also due to difficulties in providing convincing scenarios for the abiotic production of nucleotides. Therefore, alternative scenarios started to be discussed in the scientific community (Nashimoto 2001, Farias et al. 2014a, Farias et al. 2016), mainly those postulating the origin of life as an initial organization on which both nucleic acids and peptidic elements acted in coordination (Prosdocimi et al 2018, Prosdocimi et al. 2020a). In any case, in every scenario there is a consensus that the initial informational molecule was RNA, and not the double-helix DNA present nowadays in cells. Between FUCA and LUCA, under a pre-cellular period, Carl Woese (1998) suggested the existence of an era of semi-open biological entities capable to perform an intense exchange of genetic information. Woese named those entities as progenotes. Prosdocimi et al. (2019) suggested that the process between the initial organization of biological systems and the maturation of the progenote community was organized by the First Common Ancestral Universal (FUCA), making a distinction with the last Common Ancestral Universal (LUCA), often referred as the first organism that presented cellular organization.

The consensus reached by scientific community that the first informational molecule was RNA-based is followed by the necessary conclusion that DNA has been

5

further selected as the main nucleic acid for cells, possibly due to its higher stability and double helix structure that provides an easy way to allow replication and a backup informational structure. But then, how these facts relate to the monophyly of cells? Well, once all cells today use DNA for information storage, the simplest conclusion is probably to propose that cells have had a common evolutionary history with the first forms that co-opted DNA as their main informational molecule. On that matters, it is important to notice that the general mechanism of replication is similar between cellular organisms, being based on: (i) semiconservative replication, (ii) bidirectional replication, (iii) the need for an RNA primer to start the process, and (iv) the presence of nucleases, polymerases and ligases that participate in the process of removing RNA primers. On the other hand, it should be noted that the proteins involved in those processes have a low level of identity and similarity (Leipe et al. 1999) between the basal cell groups, *i. e.*, Bacteria and Archaea. This observation led to the alternative proposal that the last common universal ancestor (LUCA) might have had an RNA-based genome. In that case, it has been proposed that the basal domains might have independently acquired the capacity to store biological information in the form of DNA (Leipe et al. 1999, Forterre 2005, Forterre 2006).

Although the topic is still a subject of debate in the scientific community, there is available evidence to consider that the storage of information in DNA molecules cannot be used as a reliable indicator for the monophyly of all cells. The main arguments for the independent origin are (i) the low similarity of proteins from the replication machinery but also the fact that (ii) it is possible to find DNA as an informational molecule in several clades of viruses. Therefore, we cannot refute the idea that Archaea and Bacteria could have acquired their replication systems from different progenote populations. Those populations could even share a common ancestor that would explain the overall similarity in the mechanism, but not a very recent one, so to justify the dissimilarity observed in their sequences.

B) Ribosomes, genetic code and the translation machinery

The (i) presence of ribosomes, the (ii) existence of a conserved genetic code and the (iii) sharing of the translation machinery are characters used as strong indicators that cells have had a single and common origin. These features are surely found in all known organisms. The ribosome and the genetic code are fundamental elements of the translation

system and they are intimately related to cell autonomy, another character of extreme relevance in any attempt to characterize cells (Varela et al. 1974). In recent years, several studies have tried to understand the process of origin and evolution of the ribosome (Petrov et al. 2015, Bowman et al. 2020). In all of them, there is a consensus about the fundamental relevance of this complex in the initial organization of biological systems (Yonath 2009, Tamura 2011, Petrov et al. 2014, Farias etal. 2014a, Farias et al. 2014b, Farias et al. 2017, Root-Bernstein and Root-Bernstein 2015, Prosdocimi et al., 2020b).

Even if many of these works still dispute alternative models and scenarios for ribosomal evolution, it is currently accepted that the ribosomal RNA appeared in very primordial stages of the origin of life. That is to say that rRNA emerged when biological systems were still operating through a purely molecular and open progenote-like systems, without individualized structures or wraps as we find today in cells or viruses. Further, there are suggestions that the two ribosomal subunits emerged independently at a later stage, after their activities had been minimally stabilized (Petrov et al. 2015; Farias et al. 2019). The 23S rRNA, is the main molecule of the larger ribosomal subunit and possibly acted as a ribozyme that catalyzed the quasi-random synthesis of peptides very early (Prosdocimi et al. 2020a). On the other hand, the 16S rRNA that composes the smaller subunit acted by selecting RNAs that increased its stability when coupled (Farias et al. 2017). Under different evolutionary analyses carried out using multiple methodologies, the rRNA of the larger subunit had its catalytic site (PTC) identified as the first part of the structure to stabilize (Tamura 2011, Farias et al 2014a, Petrov et al. 2015, Prosdocimi et al. 2020b). Regarding the rRNA 16S, it is suggested that the upper 3' domain was probably the first part of the structure to stabilize, being the most fundamental region for the decoding process of biological information (Harish and Caetano-Anolles 2012, Farias et al. 2020).

When we consider that ribosomes were among the first molecular structures to self-organize, it becomes clear that the emergence of the genetic code must have occurred immediately after or concomitantly with ribosomal maturation. Thus, the genetic code emerged when the biological systems were still expressed in the form of purely molecular systems; that is, *before the appearance of cells* (Hartman and Smith 2019). In addition to evidence for the antiquity of ribosomes (Yonath 2009), some studies point to the fact that tRNAs, which are molecules responsible for the connection between the information contained in nucleic acids and the amino acids, were among the oldest biological molecules we know (Caetano-Anolles and Caetano-Anolles 2016, Farias et al. 2014a;

Farias et al. 2016). The antiquity of tRNA molecules is gaining prominence in the quest for the initial organization of biological systems once they consist on the simplest nucleic acid structures used by such systems. Some models suggest that both the larger subunit (Prosdocimi et al. 2020b) and the smaller ribosomal subunits, as well as the first mRNAs, were most likely formed by the junction of proto-tRNAs (Farias and Jose 2020). All of these ideas reinforce the fact that the translation system was one of the first systems to self-organize, much earlier than the time on which cells appeared, still at the progenote era. This organization of the protein synthesis apparatus have had place long before the emergence of the cells and therefore cannot be used as evidence for the monophyly of cells.

The origin of viruses and its relationship with the translation system

When considering the possibility that the mechanism of protein synthesis could not be used to define cell monophyly, we must also consider contemporary models for the origin and evolution of viruses. Viruses are very particular biological systems because, although they do not have ribosomes, they present biological information encoded in either RNA or DNA sequences, single or double stranded, and they use the same genetic code present in cells. For many years, the similarity between the cellular and viral genetic codes has been explained by the escape model, in which viruses would have been molecular particles that escaped from the cellular context and acquired their own "independence", even though they remained dependent on cells for reproduction (Hendrix et al., 2000). However, with the increase in availability of genomic data, new scenarios have been proposed (Nasir and Caetano-Anollés 2015, Koonin et al. 2015). Among the scenarios that have been gaining popularity, one considers the origin of some viral groups before the emergence of cells (Nasir and Caetano-Anollés 2015, Farias et al. 2019). It is worth noting that no current proposal for the origin of viruses suggests a single origin for the whole group. On that matters, some of them suggest that the viral form should be seen as a life-sustaining strategy that may have appeared several times throughout evolutionary history (Farias et al. 2019). However, when one considers that some viral groups must have been contemporary with the progenotes, the fact that these biological entities have the same genetic code used by the cells must be reinterpreted. In that context, the presence of a nearly universal genetic code must be seen as evidence for the monophyly of all biological systems.

8

It becomes therefore necessary to clearly understand the fact that the origin of biological systems happened much before than the origin of cells. Thus, the genetic code present in both viruses and cells cannot be used as evidence for the unique origin of those latter entities. So, when scientific evidence will be able to confirm that at least one group of viruses appeared before cells, then the present of a genetic code will definitely be seen as evidence for the monophyly among all biological systems. Thus, many groups of viruses and cells probably descended from populations of progenotes living as open systems that shared the translation machinery to express themselves and reproduce.

However, even though the existence of a universal genetic code cannot be used as an evidence for cell monophyly, we need to consider that ribosomes are nowadays found only within cells. In this way, the encapsulation of a translation system can be considered as an indication of a single origin indeed. However, when we consider a scenario prior to the emergence of cells, it becomes even more difficult to imagine how a translation system could have appeared within a lipid envelope. It is proposed that the evolution of open systems of quasi-species led to the process of compartmentalization of molecular subsystems (Woese 1998). In this sense, the systems that maintained and incorporated the translation system would have gained autonomy and were capable to give rise to what we know as cells (Koonin et al. 2006, Farias et al. 2019). It is possible that some systems have been compartmentalized in simpler wraps without the translation system, not developing autonomy and remaining dependent on the ribosomes present only in cellular systems: such systems would have formed the first viral groups.

C) Could membranes represent definitive evidence for the unique origin of cells?

When considering the scenario above mentioned as possible, a further question is: was the translation system incorporated by one single membrane system or by multiple ones? If we can show that the translation machinery has been incorporated only once, this would be an unequivocal evidence for a single origin for cells. On the other hand, we may wonder if there exist a possibility that the translation system has been co-opted by more than one population of those pre-cellular, progenote subsystems. In the latter case, we would conclude that the entities known as cells may have had at least two origins, and, thus, would not form a monophyletic group.

In order to search answers for this question, we need to analyze the existing compartmentalization systems in detail, that is, the membrane systems that led to the

individualization process in biological systems. Such systems would have operated in the transition between the world of progenotes and the cellular world. In other words, it represents the transition between (i) a molecular world of open biological systems and (ii) a cellular world of closed biological systems. It is also the same to talk about a transition between the era of progenotes and FUCA into the era of cells and LUCA. At this point, we present the still unsolved dispute about whether phospholipids were used for encapsulation once or twice along the initial emergence of cellular systems: this quarrel is referred in scientific literature as the "lipid divide" issue (Villanueva et al. 2017, Jordan et al. 2019, Mencía 2020). When one studies the constitution of the plasmatic membrane along the two basal groups of cellular organisms (Bacteria and Archaea), it is observed that even if they present a lipoprotein membranes, the very constitution of the membranes is radically different between the two groups.

To start with, it has been noticed that the biochemical structure of the bacterial membrane is composed by (i) a lipid bilayer containing phospholipids linked through *ester-like* chemical bonds between an alcohol and the fatty acid. However, in archaeal membranes the chemical bond that bind the alcohol to the fatty acid is of the *ether* type (Lombard et al. 2012). In addition to these differences in the bond type, (ii) the alcohol used in the structure of the phospholipids is different as Bacteria use glycerol-3-phosphate and Archaea use glycerol-1-phosphate. Besides, (iii) the incorporated fatty acids are quite distinct once the fatty acids in Archaea usually present a side chain (Sojo 2019). Finally, (iv) the proteins used in each of the basal groups to metabolize lipids have low similarity and cannot be considered homologous (Koga et al. 1998, Sojo 2015). However, when analyzing the lipid biosynthesis pathways present in Archaea and Bacteria, we can find, in specific groups of bacteria (such as the phylum FCB: Fibrobacteres/Chlorobi/Bacteroidetes), a membrane typical of archaea (Villanueva et al. 2018). However, phylogenetic analyses tend to suggest that the acquisition of specific pathways from one domain to another are better explained by lateral transference and natural selection once the archaeal membrane is known to be more stable in high temperatures (Coleman et al. 2019, Sojo 2019).

Until the moment we write this report, there is no accepted proposals that explain how a transition between the two types of membranes could have happened. Thus, the model that seeks to explain the diversity found in the membrane composition between Archaea and Bacteria suggests that their putative ancestral, LUCA, may have had a mixed or hetero-chiral membrane (Wächtershäuser 2003). Those mixed membranes were proven

to be viable in laboratory (Shimada and Yamagishi 2011, Yokoi et al. 2012). First, it was proposed that LUCA's mixed membrane has probably been replaced by homo-chiral membranes because it was unstable. However, a study that evaluated the stability of the homo- and hetero-chiral membranes, demonstrated that both types have similar stability (Shimada and Yamagishi 2011). Although it is probably too early to state boldly that two independent compartmentalization processes have occurred on the formation of the first cells, many evidences point in this direction.

Did a cellular common ancestor ever exist?

Therefore, we believe to have considered a couple aspects that dispute the very fact that a cellular common ancestor such as LUCA has ever existed as originally proposed. Alternatively, could we consider that 2 ancestors for Bacteria and Archaea evolved independently? If we consider this idea as possible, it is mandatory that the translation system has been inherited by both ancestors, as well as other progenote systems that would form the pathways that operated the basal metabolism of cells.

First, we have seen that the presence of ribosomes in cellular structures should not be used as evidence of a common origin for all cells. On the contrary, it must be seen as evidence for the monophyly of all biological systems, including progenotes, viruses, bacteria and archaea. This differentiation is absolutely necessary to better understand the natural history of cells and differentiate it from the natural history of other biological systems. Following this reasoning and reinterpreting the classification originally proposed by Woese et al. (1990), *we conclude that these authors were **verifying** a common origin for the ribosomal molecule and the translation system; but not for cells*. Thus, it is indeed possible to assume that LUCA may have never existed as an ancestor for all cells under the form of an already cellular organism (Koga et al. 1998). Alternatively, the common ancestor of Bacteria and Archaea may have had the form of a community of open biological systems sharing a translation machinery, reinforcing the monophyly amongst all biological systems and ascendance from a progenote era. FUCA was clearly a first ancestor for entities sharing a chemical and informational code, *i. e.*, those ones on in which the translation system has emerged and matured through the self-organization of an organic code (Farias et al. 2020).

Then, the woesian LUCA seen as a cellular ancestor and sometimes better referred as LUCellA (Last Universal Cellular Ancestor; by Nasir, Kim and Caetano-Anollés 2012)

possibly has never existed. In that case, it does not seem possible to define any common ancestor between Bacteria and Archaea that would have existed before cellularization, indicating that the paths through which compartmentalization has been achieved followed from parallel evolutionary routes. Nevertheless, we must account for the fact that the metabolism of carbohydrates, amino acids and other intermediate metabolites were most likely inherited together with the translation machinery from an ancestral clade of progenotes. Thus, it seems more appropriate to forget about a cellular LUCA or, alternatively, to consider the existence of at least two different cellular ancestors, or two LUCAs; one of them which would have given rise to Bacteria, and another that would have given rise to Archaea. Furthermore, we consider the possibility that other progenote populations originated one or more viral clades.

Last considerations

Since the consolidation of Cell Theory as one of the strongest theories in biology, the assumption that cells originate solely from preexisting cells has been established as one of the most fundamental bases for biological knowledge, a true and unshakable fact. However, when we aim to understand the origin of living organisms on our planet and try to understand the diversity observed today, several questions about the origin of cellular systems are missing and lack a definitive explanation. Some particular scenarios for the origin of both biological systems and cells lead us to suggest that the cellular organization may have had more than one origin in the beginning of life.

Hence, when one takes seriously the possibility that cells could be polyphyletic, it becomes necessary to understand that the natural history of cells is different from the natural history of biological systems. In that sense, we use the term biological systems to refer to any system made of chemical molecules with the capacity to form organic codes (Farias et al. 2020), as the genetic code on which amino acids are encoded by codons in a nucleic acid. Those biological systems therefore include not only cells and viruses, seen as different contemporary strategies for maintaining organicity (Farias et al. 2019), but also the pre-biotic open systems known as progenotes. And even though these life enduring strategies may have most likely emerged more than once, they have in common the fact that they are based on the same molecular language, since the organic phenomenon presents strong evidence that appeared only once on our planet. Thus, we

12

presented earlier the proposal that biological systems emerged from a chemical symbiosis (Prosdocimi et al. 2020) that started with an interaction between nucleic acids and amino acids. This interaction stabilized and increased the chance that these heteromolecular systems could be maintained and persisted. From this starting point, the system has probably evolved until the maturation of a ribonucleoprotein machinery that gave rise to the ribosome, generating the first universal common ancestor (FUCA) (Prosdocimi et al. 2019). FUCA therefore was composed by a population of open-systems, self-replicating ribonucleoproteins that inaugurated the progenote era and evolved into maturity when self-organization processes resulted in the creation of a genetic code. This code was, for the first time, capable to organize an ordered interaction between nucleic acids and proteins through the formation of a biological language. From that moment on, the pre-cellular open systems started to accumulate information and self-organize, producing the first genomes by the assembly of biochemical pathways, which probably appeared in different progenote populations evolving independently. Progenotes populations capable to bind and process carbohydrates, amino acids and other intermediated metabolites and co-factors were possibly very well succeeded. At some point, however, possibly due to changing environmental conditions, a wrap would have appeared and been selected, possibly acting to protect these genetic systems and prevent their collapse by error catastrophe. This has probably happened through the rise and evolution of lipid-binding proteins and pathways (Sojo 2019). It is also possible to suppose the existence of transient peptidic wraps that have occurred randomly since the origin of FUCA. Those peptidic covers were probably the ancestral of viral capsids and were selected to store stable genomic information, even if they cannot present any sophisticated form of metabolism due to the absence of cytoplasm and limited information exchange with the outside medium. However, the appearance of cell membranes evolving from lipid-binding proteins may have produced more sophisticated forms for these wraps, evolving to be able to control lipids by biochemical pathways and allowing the emergence of a cytoplasm. The fact that we observe two chemically different pathways to control the link between lipids and the formation of membranes in the basal prokaryotic groups suggests that cells may have appeared more than once. Besides, although today we know only two different pathways capable to produce cell membranes (suggesting that this mechanism may have appeared twice), there is the possibility that other less efficient pathways to produce proteolipidic envelopes could have been originated and further extinguished. It is also possible to assume that different viral clades may have appeared at that time

13

through the formation of different types of protein capsids. Because they are simpler than membranes, it is possible to suppose that such biological systems were even more common in the early days of life in Earth. This last fact parallels to the extreme abundance of viruses seen today. Thus, it is possible to suppose that different strains of organisms with a viral-type strategy co-existed before the emergence of cells.

Thus, when we consider that life can be defined through a process carried out by overlapping codes in multiple layers (Farias and Prosdocimi, 2019, Farias et al. 2020), we realize that some arguments often used to define what cells, such as the presence elements of the translation system and the genetic code, should be more correctly used to define the monophyly of the biological systems. Therefore, both cells and viruses possibly originated in the early days of life on our planet from the evolution of quasi-species entities known as progenotes (Figure 1).
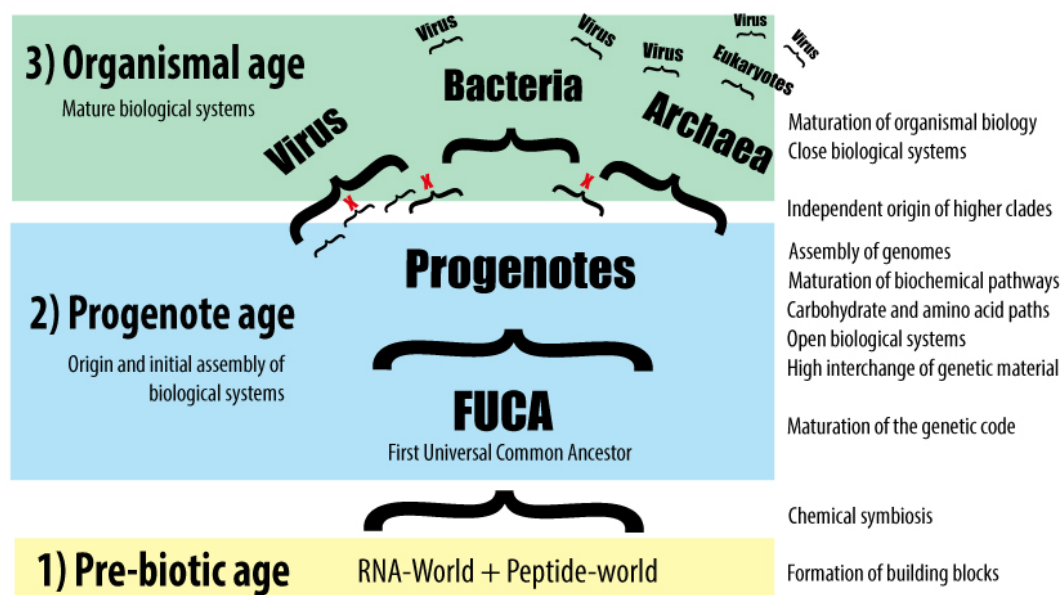
Figure 1: From pre-biotic to organismal age. The formation of the building blocks for life was produced in a pre-biotic age. After a chemical symbiosis, nucleic acids and amino acids started to interact for the benefit of both, creating stable structures that would evolve into the ribosome. The first universal common ancestor is born from this chemical symbiosis and maturates when the genetic code is initially formed. Open-biological systems evolve from FUCA creating progenote populations. Each population mature different biochemical pathways independently. Some populations develop the capability to produce peptidic wraps capable to maintain, save and store genetic material. Progenotes capable to bind and process carbohydrates, amino acids and co-factors evolve. Nucleic acid binding proteins organize the formation of two different types of replication machineries. Lipid binding proteins organize the formation of two different types of membrane and self-organize to the formation of the different basal groups of prokaryotes. Viruses are inherited from both progenotes but also are produced by escaping modules from bacterial and/or archaeal groups. Eukaryotes evolve from the symbiosis between archaeal (Lokyarchaeota) and bacterial groups. Viral strategy is also produced by escaping modules from eukaryotic genomes.

Finally, due to (i) the differences observed in the basic constitution of bacterial and archaeal membranes, to (ii) the differences observed in the constitution of viral capsids and considering (iii) the differences observed in the replication of nucleic acids by bacteria, archaea, and viruses, we suggest that these biological systems architectures (viral, bacterial, and archaeal) may have originated independently from the evolution of different progenotes' populations. Nevertheless, they were descendants of other sophisticated progenote populations capable to translated biological information and to manipulate carbohydrates, amino acids and other metabolites and co-factors.

Acknowledgements

References

- Baker, J. R. 1955. The cell-theory: a restatement, history, and critique. Part V. The multiplication of nuclei. Q. J. Microsc. Sci . 96, 449–481.

- Barbieri, M. 2016. From the common ancestor to the first cells: the code theory. Biol. Theory. 11: 102-112.

- Bowman, J.C., Petrov, A.S., Frenkel-Pinter, M., Penev, P.I., Williams, L.D. 2020. Root of the Tree: The Significance, Evolution, and Origins of the Ribosome. Chem Rev.120(11):4848-4878. doi: 10.1021/acs.chemrev.9b00742.

- Caetano-Anollés, D., Caetano-Anollés, G. 2016. Piecemeal Buildup of the Genetic Code, Ribosomes, and Genomes from Primordial tRNA Building Blocks. Life (Basel) 2;6(4). pii: E43.

- Cech, T.R. 2009. Crawling out of the RNA world. Cell, 136(4), 599–602. https://doi.org/10.1016/j.cell.2009.02.002

- Chapman, A. 1996. "England's Leonardo: Robert Hooke (1635–1703) and the art of experiment in Restoration England". Proc. R. Inst. G. B. 67: 239–275.

- Coleman, G.A., Pancost, R.D., Williams, T.A. 2019. Investigating the Origins of Membrane Phospholipid Biosynthesis Genes Using Outgroup-Free Rooting. Gen Biol Evol. 11(3), 883–898. https://doi.org/10.1093/gbe/evz034

- Dunn, G.A, Jones, G.E. 2004. Cell motility under the microscope: Vorsprung durch Technik. Nat. Rev. Mol. Cell. Biol. 5(8):667-672, 2004. doi:10.1038/nrm1439

- Dworkin, J.P., Lazcano, A., Miller, S. L. 2003. The roads to and from the RNA world. J Theor. Biol. 222(1), 127–134. https://doi.org/10.1016/s0022-5193(03)00020-1
- Farias, S.T., José, M.V. 2020. Transfer RNA: The molecular demiurge in the origin of biological systems. Prog. Biophys. Mol. Biol. 153, 28–34. https://doi.org/10.1016/j.pbiomolbio.2020.02.006

- Farias, S.T., Rêgo, T.G., José, M.V. 2016. tRNA Core Hypothesis for the Transition from the RNA World to the Ribonucleoprotein World. Life (Basel, Switzerland), 6(2), 15. https://doi.org/10.3390/life6020015

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

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

- Farias, S.T., Rêgo, T.G, José, M.V. 2017. Peptidyl Transferase Center and the Emergence of the Translation System. *Life (Basel)*. 7(2):21. doi:10.3390/life7020021

- Farias, S.T., Rêgo, T.G., José, M.V. 2014a. Evolution of transfer RNA and the origin of the translation system. Front. Gen. 5:303. doi:10.3389/fgene.2014.00303

- Farias, S.T., Rêgo, T.G., José, M.V. 2014b. Origin and evolution of the Peptidyl Transferase Center from proto-tRNAs. FEBS Open Bio. 4:175-178. doi:10.1016/j.fob.2014.01.010

- Farias, S.T., Rêgo, T.G., José, M.V. 2020. Origin of the 16S Ribosomal Molecule from Aancestor tRNAs. Preprints. 2020110075 (doi: 10.3390/sci1010008.v1).

- Forterre, P., Gribaldo, S., Brochier, C. 2005. Luca: à la recherche du plus proche ancêtre commun universel [Luca: the last universal common ancestor]. Med Sci (Paris). 21(10):860-5. French. doi: 10.1051/medsci/20052110860.

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

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

- Gilbert, W. 1986. Origin of life: The RNA world. Nature. 319 (6055): 618–618.

- Gogarten, J.P., Deamer, D. 2016. Is LUCA a thermophilic progenote? Nat Microbiol. 25;1:16229. doi: 10.1038/nmicrobiol.2016.229. PMID: 27886195.

- Harish, A., Caetano-Anollés, G. 2012. Ribosomal history reveals origins of modern protein synthesis. PLoS One 7(3):e32776. doi: 10.1371/journal.pone.0032776.

- Hartman, H., Smith, T.F. 2019. Origin of the Genetic Code Is Found at the Transition between a Thioester World of Peptides and the Phosphoester World of Polynucleotides. Life (Basel*)*. 9(3):69. doi:10.3390/life9030069

- Hendrix, R.W., Lawrence, J.G., Hatfull, G.F., Casjens, S. 2000. The origins and ongoing evolution of viruses. Trends Microbiol. 8(11):504-8.

- Higgs, P.G., Lehman, N. 2015. The RNA World: molecular cooperation at the origins of life. Nat Rev Gen, 16(1), 7–17. https://doi.org/10.1038/nrg3841

- Jordan, S.F., Nee, E., Lane, N. 2019. Isoprenoids enhance the stability of fatty acid membranes at the emergence of life potentially leading to an early lipid divide. Interface Focus. 6;9(6):20190067.

- Joyce, G. F., Szostak, J. W. 2018. Protocells and RNA Self-Replication. Cold Spring Harb. Perspect Biol. 10(9), a034801. https://doi.org/10.1101/cshperspect.a034801

- Koga, Y., Kyuragi, T., Nishihara, M., Sone, N. 1998. Did archaeal and bacterial cells arise independently from noncellular precursors? A hypothesis stating that the advent of membrane phospholipid with enantiomeric glycerophosphate backbones caused the separation of the two lines of descent. J Mol Evol. 46(1):54-63.
- Koga, Y., Kyuragi, T., Nishihara, M., Sone, N. 1998. Did archaeal and bacterial cells arise independently from noncellular precursors? A hypothesis stating that the advent of membrane phospholipid with enantiomeric glycerophosphate backbones caused the separation of the two lines of descent. J. Mol Evol. 46(1), 54–63. https://doi.org/10.1007/pl00006283

- Koonin, E.V., Dolja, V.V., Krupovic, M. 2015. Origins and evolution of viruses of eukaryotes: The ultimate modularity. Virology. 479-480:2-25. doi: 10.1016/j.virol.2015.02.039.

- Koonin, E.V., Senkevich, T.G., Dolja, V.V. 2006. The ancient Virus World and evolution of cells. Biol. Direct. 19;1:29. doi: 10.1186/1745-6150-1-29.

- Leipe, D.D., Aravind, L., Koonin, E.V. 1999. Did DNA replication evolve twice independently? Nucleic Acids Res. 1;27(17):3389-401. doi: 10.1093/nar/27.17.3389.

- Lombard, J., López-García, P., Moreira, D. 2012. The early evolution of lipid membranes and the three domains of life. Nat. Rev. Microbiol. 10(7):507-515, 2012. doi:10.1038/nrmicro2815

- Martin, W.F., Weiss, M.C., Neukirchen, S., Nelson-Sathi, S., Sousa, F.L. 2016. Physiology, phylogeny, and LUCA. Microb Cell. 25;3(12):582-587. doi: 10.15698/mic2016.12.545. PMID: 28357330; PMCID: PMC5348977.

- Mat, W.K., Xue, H., Wong, J.T. 2008. The genomics of LUCA. Front Biosci. 1;13:5605-13. doi: 10.2741/3103. PMID: 18508609.

- Mencía, M. 2020. The archaeal-bacterial lipid divide, could a distinct lateral proton route hold the answer? Biol Direct. 21;15(1):7.

- Nashimoto, M. 2001. The RNA/protein symmetry hypothesis: experimental support for reverse translation of primitive proteins. J. Theor Biol. 209(2), 181–187. https://doi.org/10.1006/jtbi.2000.2253

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

- Nasir, A., Kim, K.M., Caetano-Anollés, G. 2012. Viral evolution: Primordial cellular origins and late adaptation to parasitism. Mob Genet Elements. 2(5):247-252. doi:10.4161/mge.22797

- Petrov, A.S., Bernier, C.R., Hsiao, C., Norris, A.M., Kovacs, N.A., Waterbury, C.C., Stepanov, V.G., Harvey, S.C., Fox, G.E., Wartell, R.M., Hud, N.V., Williams, L.D. 2014. Evolution of the ribosome at atomic resolution. Proc. Natl. Acad. Sci. U. S. A.. 111(28):10251-6, 2014. doi: 10.1073/pnas.1407205111.

- Petrov, A.S., Gulen, B., Norris, A.M., Kovacs, N.A., Bernier, C.R., Lanier, K.A., Fox, G.E., Harvey, S.C., Wartell, R.M., Hud, N.V., Williams, L.D. 2015. History of the ribosome and the origin of translation. Proc. Natl. Acad. Sci. U. S. A. 112(50):15396-401. doi: 10.1073/pnas.1509761112.

- Poole, A. M., Jeffares, D. C., Penny, D. 1998. The path from the RNA world. J. Mol Evol. 46(1), 1–17. https://doi.org/10.1007/pl00006275

- Prosdocimi, F., José, M., Farias, S.T. 2019. 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. doi 10.1007/978-3-030-30363-1_3

- Prosdocimi, F., Farias, S. T. 2020. From FUCA To LUCA: A Theoretical Analysis on the Common Descent of Gene Families. Acta Sci Microbiol. doi: 10.31080/ASMI.2020.03.0494

- Prosdocimi, F., Jheeta, S., Farias, S. T. 2018. Conceptual challenges for the emergence of the biological system: Cell theory and self-replication. Med Hypoth*e*sis. 119:79-83, 2018. doi:10.1016/j.mehy.2018.07.029
- Prosdocimi, F., Jose, M. V., Farias, S. T. 2020a. The Theory of Chemical Symbiosis: A Margulian View for the Emergence of Biological Systems (Origin of Life). Acta Biotheor. doi: 10.1007/s10441-020-09388-7

- Prosdocimi, F., Zamudio, G.S., Palacios-Pérez, M., Farias, S.T., José, M.V. 2020b. The Ancient History of Peptidyl Transferase Center Formation as Told by Conservation and Information Analyses. *Life (Basel)*. 10(8):E134. doi:10.3390/life10080134

- Ribatti, D. 2018. An historical note on the cell theory. Exp Cell Res. 364(1):1-4. doi:10.1016/j.yexcr.2018.01.038

- Robertson, M.P., Joyce, G.F. 2012. The origins of the RNA world. Cold Spring Harb Perspect Biol. 4(5), a003608. https://doi.org/10.1101/cshperspect.a003608

- Root-Bernstein, M., Root-Bernstein, R. 2015. The ribosome as a missing link in the evolution of life. J. Theor Biol. 367:130-158, 2015. doi:10.1016/j.jtbi.2014.11.025

- Schleiden, M.J. 1838. Beiträge zur Phytogenesis.Müllers Archiv füŕAnatomie, Physiologie und wissenschaftliche Medizin. 5: 137–177.

- Schwann, T. 1839. Mikroskopische Untersuchungen uber die Ube-reinstimmung in der Struktur und im Wachstum der Thiere und Pflanzen. Berlin: Sandersche Buchhandlung.

- Shimada, H., Yamagishi, A. 2011. Stability of heterochiral hybrid membrane made of bacterial sn-G3P lipids and archaeal sn-G1P lipids. Biochemistry. 17;50(19):4114-20.

- Sojo, V. 2015. On the biogenic origins of homochirality. Orig Life Evol Biosph. 45(1-2):219-24.

- Sojo, V. 2019. Why the Lipid Divide? Membrane Proteins as Drivers of the Split between the Lipids of the Three Domains of Life. Bioessays. 41(5):e1800251.

- Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., Van Eijk, R., Schleper, C., Guy, L., Ettema, T.J.G. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature. 521, 173–179.

- Spang, A., Stairs, C.W., Dombrowski, N., Eme, L., Lombard, J., Caceres, E.F., Greening, C., Baker, B.J., Ettema, T.J.G. 2019. Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. Nat. Microbiol. 4, 1138–1148.

- Tamura, K. 2011. Ribosome evolution: emergence of peptide synthesis machinery. J Biosci. 36(5):921-928. doi:10.1007/s12038-011-9158-2

- Varela, F.G., Maturana, H.R., Uribe, R. 1974. Autopoiesis: the organization of living systems, its characterization and a model. Curr. Mod. Biol. 5(4):187-196. doi:10.1016/0303-2647(74)90031-8

- Vázquez-Salazar, A., Lazcano, A. 2018. Early Life: Embracing the RNA World. Curr. Biol. CB, 28(5), R220–R222. https://doi.org/10.1016/j.cub.2018.01.055

- Villanueva, L., Schouten, S., Damsté, J.S. 2017. Phylogenomic analysis of lipid biosynthetic genes of Archaea shed light on the 'lipid divide'. Environ Microbiol. 19(1):54-69.

- Villanueva, L., von Meijenfeldtm F.B., Westbyem A.B., Hopmans, E.C., Dutilh, B.E., Sinninghe Damsté, J,S. 2018. Bridging the divide: bacteria synthesizing archaeal membrane lipids. BioRvix. doi:10.1101/448035. PPR:PPR59253.

- Wächtershäuser, G. 2003. From pre-cells to Eukarya – a tale of two lipids. Mol. Microbiol. 47: 13-22.

- Weiss, M.C., Preiner, M., Xavier, J.C., Zimorski, V., Martin, W.F. 2018. The last universal common ancestor between ancient Earth chemistry and the onset of genetics. PLoS Genetics. 14(8):e1007518. doi:10.1371/journal.pgen.1007518
- Woese, C. 1998. The universal ancestor. Proc. Natl. Acad. Sci. U. S. A.. 95(12):6854-6859. doi:10.1073/pnas.95.12.6854

- Woese, C.R., Fox, G.E. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. U. S. A.. 74(11):5088-509. doi:10.1073/pnas.74.11.5088

- Woese, C.R., Kandler, O., Wheelis, M.L. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eukarya. Proc. Natl. Acad. Sci. U. S. A. 87(12):4576-4579. doi:10.1073/pnas.87.12.4576

- Yokoi, T., Isobe, K., Yoshimura, T., Hemmi, H. 2012. Archaeal phospholipid biosynthetic pathway reconstructed in Escherichia coli. Archaea. 2012:438931.

- Yonath, A. 2009. Large facilities and the evolving ribosome, the cellular machine for genetic-code translation. J R Soc Interface. 6 Suppl 5(Suppl 5):S575-S585. doi:10.1098/rsif.2009.0167.focus