

Be introduced to the First Universal Common Ancestor (FUCA): the great-grandmother of LUCA (Last Universal Common Ancestor)

Francisco Prosdocimi^{1*}, Marco V José² and Sávio Torres de Farias^{3*}

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 Theoretical Biology Group, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Ciudad Universitaria, 04510 CDMX, Mexico.

³ Laboratório de Genética Evolutiva Paulo Leminsk, Departamento de Biologia Molecular, Universidade Federal da Paraíba, João Pessoa, Paraíba, Brasil.

* Correspondence: prosdocimi@bioqmed.ufrj.br; stfarias@yahoo.com.br

Abstract

The existence of a common ancestor to all living organisms in Earth is a necessary corollary of Darwin idea of common ancestry. The Last Universal Common Ancestor (LUCA) has been normally considered as the ancestor of cellular organisms that originated the three domains of life: Bacteria, Archaea and Eukarya. Recent studies about the nature of LUCA indicate that this first organism should present hundreds of genes and a complex metabolism. Trying to bring another of Darwin ideas into the origins of life discussion, we went back into the prebiotic chemistry trying to understand how LUCA could be originated

under gradualist assumptions. Along this line of reasoning, it became clear to us that the definition of another ancestral should be of particular relevance to the understanding about the emergence of biological systems. Together with the view of biology as a language for chemical translation, on which proteins are encoded into nucleic acids polymers, we glimpse a point in the deep past on which this Translation mechanism could have taken place. Thus, we propose the emergence of this process shared by all biological systems as a point of interest and propose the existence of this non-cellular entity named FUCA, as the First Universal Common Ancestor. FUCA was born in the very instant on which RNA-world replicators started to be capable to catalyze the bonding of amino acids into oligopeptides. FUCA has been considered mature when the translation system apparatus has been assembled together with the establishment of a primeval, possibly error-prone genetic code. This is FUCA, the great-grandmother of LUCA.

Running title - Meet FUCA: the ancestor of LUCA

Keywords: origin of life; LUCA; FUCA; RNA World; PTC; Archaea; translation system

1. Historical background

In the mid of the XIXth century, Charles Darwin proposed a point of unification for all living beings through the idea of common ascendance (Darwin, 1859). Concomitantly with the Darwin's postulates, another strong idea was under development: the Cell Theory. The explanation power of cell theory established the paradigm that cells were the most basic unit of life. Therefore, the so-called Last Universal Common Ancestor (LUCA) of life should necessarily be composed of a structured cellular unit.

The comprehension about the nature of LUCA is enlightening and allowed researchers to better understand the constitution of the first cells that most likely emerged in Earth about 3.8 billion years ago. LUCA is currently seen as a cellular organism that presents a lipid membrane and a complex metabolism composed of hundreds of gene families and dozens of biochemical pathways (Penny and Poole, 1999; Delaye et al, 2005; Weiss et al 2016). Although comparative genomics allow us to have a glimpse about the molecular nature of this important cellular ancestral, it is clear that some simpler biological system must have preceded and gave rise to LUCA.

About one century after Darwin, an original and prolific field of biological research started to develop shortly after the discovery of DNA double-stranded structure by Rosalind Franklin, James Watson, and Francis Crick (1953). The development of molecular biology modified our comprehension of the biological world and allowed scientists to perform experiments using genes and proteins inside and outside the cell, bringing a deeper understanding about how biological processes operated in the biochemical level.

In the very same year that Watson and Crick published their paper, a young American researcher named Stanley Miller was inspired by the ideas of Aleksandr Oparin and JBS Haldane in the 1920s. They had demonstrated that the assembly of basic

chemical molecules under specific physicochemical conditions could produce amino acids, the building blocks of proteins. Miller's experiment updated the simulation of primeval Earth conditions and confirmed the production of amino acids using only water (H₂O), methane (CH₄), ammonia (NH₃), and hydrogen (H₂) as input (Miller, 1953). Thus, a point of view about the initial organization of the biological systems started as the studies of prebiotic chemistry provided an innovative way to think about the Origins of life.

Then, by the early years of 1980s, the discovery of catalytic properties of RNA molecules introduced another element through which the initial organization of biological systems could be understood (Kruger et al, 1982, Guerrier-Takada et al. 1983). For the first time, these findings made possible the proposition of hypotheses capable to describe biological entities that did not need cellular structures. This idea culminated with the proposition of a molecular-based *RNA world* on which self-replicative and catalytic molecules of nucleic acids could interact and be target of natural selection, pushing forward the path into the origins of life (Gilbert, 1986).

1.1. The lineage of LUCA

It has been shortly before the proposal of the RNA world theory, in the late 1970s, that the American microbiologist Carl Woese started to produce sequences for a specific RNA molecule known as 16S rRNA, a constituent of the small ribosomal subunit (Woese and Fox, 1977a). Containing about ~1500 nucleotides, Woese found specific oligonucleotide signatures in the sequences of the rRNAs that were capable to differentiate molecules coming from either Bacteria or Eukarya organisms.

Woese's interest in the 16S rRNA was associated to the fact that ribosomes were abundant in cells and could be easily separated and radiolabeled by the molecular techniques available. However, when he sequenced for the first time the rRNA from a methanogen organism named *Methanobacterium thermoautotrophicum*, he was unable to find those expected signatures... But what should this result mean: did he found an organism that was neither Bacteria nor Eukarya?

By the end of 1976, Woese and George Fox were capable to sequence other methanogen organisms that could not be classified in these two groups based on their 16S rRNA sequences, confirming their previous results. On the other hand, this new group presented a different type of oligonucleotide signature that was observed to be common among them. Therefore, the 16S rRNA molecule revealed to present a bipartite diversity in the prokaryotic world. And it then became clear that a completely unknown group of organisms existed: a new, different, and unexpected clade of unicellular organisms.

Due to the fact that most members of this new group of prokaryotes lived in extreme environment, Woese and Fox believed them to be the closest relatives of LUCA, the most ancient form of life that lived in Earth. In the early days of our

planet, it was known that the atmosphere was completely different, most likely based on a reductive environment, hot and possibly very dry. They discovered that the 16S rRNA of a bacteria living in high-salt environments (an halophile) clustered also into this new group of organisms. The results suggested that most members of the third group inhabited extreme, prebiotic-like environments. This led the authors to name the new clade as the Archaeobacteria by the use of the Greek word *archae-*, meaning ancient. Thus, a new urkingdom of microbes was defined and it has probably participated closely to the heritage of LUCA.

However, further research on Archaeobacteria conducted by Woese himself and collaborators changed the picture initially proposed (Woese et al, 1990). Actually, the Archaeobacteria (or Archaea) could be found not only in extreme but also in typical mesophile environments. Studying other phenotypes of this group, it became clear that they were actually a missing link between Bacteria and Eukarya. It has been found that Archaea presented characteristics shared with Bacteria, such as (i) prokaryotic ribosomes; (ii) circular chromosomes and (iii) lack of membrane-enclosed organelles; but also, others shared with Eukaryotes, such as a (i') DNA associated to histones; (ii') several types of RNA polymerases; and (iii') the use of methionine as the initiator amino acid from protein synthesis; among others.

These observations now seemed to indicate that the early cellular ancestor (LUCA) was probably a population of ancestral organisms that presented more similarities to Bacteria than to Archaea. Ever since, the origin of life and the origin of LUCA were coincidentally placed in the tree of life as the middle point of a trichotomy in-between Bacteria, Archaea and Eukarya domains. More recently, however, LUCA is been considered as the ancestor of only Bacteria and Archaea, while Eukarya are being considered as a group that evolved from a lineage inside the archaean superphylum Asgard (Williams et al, 2012; Raymann et al. 2015; Zaremba-Niedzwiedzka et al. 2017).

The most recent genomic study on the deep nature of LUCA's genome has found that this ancestral organism should present at least 355 gene families, being thermophilic, anaerobic, capable to fix CO₂ and N, and possibly H₂-dependent (Weiss et al., 2016).

1.2. LUCA and the Viruses

The questions about the deep lineage of LUCA are complex to approach, once non-cellular living organisms do not exist in current days. The sole non-cellular biological systems that exist today are viruses. Viruses are frequently considered as non-living organisms because they are not free-living entities, as they need to hijack cells to be able to manifest their metabolism and reproduce. Besides, viruses have been excluded from deep trees of life by the simple reason that they do not have ribosomes; and without 16S rRNA molecules they cannot figure in those trees.

However, nowadays there is an emerging view that virus-like biological systems may have played important roles in pre-cellular living systems (Forterre, 2006). Besides, the evolutionary history of viruses seems to be polyphyletic as there is evidence that some groups were formed by the further simplification of cellular organisms. Thus, virus should be understood much more as *a strategy of life* rather than a monophyletic group that share the same evolutionary origin (Nasir and Caetano-Anolles, 2015). Recent researches with giant mimivirus seem to indicate polyphyly, although there is still much controversy in those grounds (Harish et al, 2016; Forterre and Gaia, 2016). By the way, it is difficult to use the word "life" and "living" to represent viruses and it seems reasonable to consider that virus-like biological systems may have existed long before lipid cellular barriers were coopted to be the basis of life in Earth.

2. Life versus biological Systems

The suggestion of an alternative view to the origins of life requires that we take a special attention to the concepts we are working on. The concept of life, as commonly understood in biology, is normally considered as cellular life. Although there is constant matter and new appreciations and proposals about the concept, we cannot argue against a so established concept. On the borders of our understanding, however, there has always been this question about the nature of viruses. Current considerations under virology often consider that viruses can only be said to be alive when they are practicing their metabolism inside a cell. Out of cells, viruses are considered to be non-living crystals of ribonucleoproteins.

Although viruses do not present cells, they do present something that is strikingly important to any biological system: a well-established genetic code. That code is actually the key to provide a chemical translation and guarantee that their nucleic acid information will produce proteins that will allow their metabolism and reproduction. Viruses may not be cellular, viruses may not be living, but viruses do speak the language of biology. They do present complex proteins encoded in their genomes and they do have a plan of existence chemically written in their nucleic acids (being DNA or RNA).

Though they do not have ways to execute their plans by their own means, it is possible to consider a world made of non-cellular, virus-like entities that might deliver their codes into a Translation system operating outside them. Thus, it is not difficult to imagine pre-cellular virus-like particles capable to exist and to attach into other systems that might translate their information and allow their reproduction.

Even Carl Woese knew that pre-cellular entities were needed to explain the origin of life in Earth before LUCA. He defined the concept of progenotes as being protocells that probably presented

(i) error-prone genetic codes, (ii) high mutation rates and (iii) high exchange of genetic material (Woese and Fox, 1977b). These entities were necessary to accumulate genetic material from different biochemical pathways to be able to form larger genomes.

Thus, we propose the usage of the term *Biological System* as an alternative for life. This decision has proven to be fertile and clarifying in many circumstances, as we shall see. However, to make a better use of this term, we need to clearly define it. For us, *a biological system is such a system on which its molecular nature is centered in the controlled interaction between polymers of nucleotides (nucleic acids) and polymers of amino acids (peptides and/or proteins)*. In general, the biological system can be considered a system based in a process of chemical translation, on which information stored in some chemical polymer can be translated information in another chemical polymer.

This view puts the process of protein synthesis in the center of what we consider Biology to be. All Biology is based on a process of chemical encoding, a system of chemical language translation on which a polymer become another by following specific grammatical rules. Thus, biological systems came to be in the very moment on which those rules emerged.

2.1. Origins versus Emergence

Another important conceptual issue that must be better understood into this new view of what are biological systems is the difference between the terms *origin* and *emergence*. The term *origin* often refers to the first rise of something that has never been before. It suggests the occurrence of a very singular, special event that, for the first time, brings something from non-existence to existence. It is a jump, a rising, an appearance.

On the other hand, the concept of emergence as we use it here should be understood as a more continuous path on which something come to be. It is not a drastic appearance as the concept of origin denotes, but a more subtle process on which a system can be built slowly and persistently over time.

Also, inhering the concept of emergence from chaos theory, we aim to signify a process on which "the whole is greater than the sum of the parts". This is clearly what happened at the emergence of biological systems, on which a polymer of nucleic acids starts to produce peptides initially by nearly random attachments until the emergence of the genetic code.

3. Biological Systems are Chemical translators

Although it is possible to wonder the existence of non-cellular biological systems, it is almost impossible to think in a biological entity harboring a metabolism without the Translation apparatus for protein synthesis. Having (i) proteins encoded in a three-letter genetic code made of nucleic acids and a (ii) metabolism primarily coordinated by proteins and enzymes seems to be a feature that unequivocally defines a biological system.

Both RNA-world advocates and researchers that propose the origin of life by hypercycles (the theory suggesting that biology started with a protein-based protometabolism happening without codification mechanisms) have difficulties to go forward within their theories by adding the other macromolecule into them.

When we consider RNA-world ribozymes capable of both self-replication and catalysis (Higgs and Lehman, 2015), it is difficult to see how a ribozyme-based protometabolism could further become proteic. It seems more parsimonious to think of a biological system that already starts together, with peptides interacting with proto-RNAs or RNA-like polymers. Proteins are the most important molecules to coordinate the cellular metabolism, interacting with compounds and catalyzing chemical reactions, even if the importance of RNAs is also enormous. Even nowadays, the importance of RNAs to the cellular metabolism has been underestimated and new results from the ENCODE project suggest that three-quarters of the human genomic DNA should be transcribed at some moment (Djebali et al., 2012).

The molecule of DNA however seems to have a very limited importance at the emergence of biological systems. Actually, it is somewhat consensual that DNA has been added to the system much time later (Forterre, 2002; Forterre, 2006; Forterre, 2013). Double helix polymers of the desoxyribonucleic acid that

nowadays compose the genomes of most organisms can be seen much more like a *high-security media* to store molecular information that emerged later and allowed an important stabilization for both heredity and the control of metabolism. Although its importance has been inestimable, it was probably absent at the scenario on which the biological systems emerged. The existence of viruses harboring RNA genomes, even in single strand, can be seen as evidence that DNA is not essential to biological systems. Being a secure media to store molecular information, it has been placed in a safe place in the nucleus of eukaryotic cells, though it seems to have no role in the beginnings.

The emergence of a chemical translation process on which RNA-like molecules convert their information into peptides seems to be the main characteristic shared by any biological system and therefore seem to have evolved from a common origin at the First Universal Common Ancestor (FUCA).

Once we suggest that the most realistic scenario for the emergence of biological systems would happen on the interaction between RNAs and proteins, the importance to explain the origins of the ribosome and the genetic code turns key. However, the ribosomes present today in cells are highly specialized and complex molecular machines and it is necessary to understand their old history of development.

3.1. The Code as it is **molecular** Established

Chemically speaking, the genetic code on which codons encode amino acids is molecularly produced by the action of extremely important enzymes named tRNA-aminoacyl synthetases. There are two different families for these proteins and they are responsible to bind each amino acid to the transfer RNAs presenting not only specific anti-codons but also other conserved binding sites on their cloverleaf structure (Zamudio and José, 2018). These enzymes are of paramount importance for the understanding of the origin and evolution of the genetic code. However, in the very early time of biology we should suppose that the translation system did not functioned very well and codon to peptide encoding should have been performed with no specificity, without a clear system of codification.

In that sense, the birth of biological systems are not at the emergence of the code, but at the rise of some sort of ribozyme that was capable to bind together amino acids. FUCA is therefore a process that started with no code, but her point of maturation happened when the code has been completely established. Thus, to understand the initial steps in the emergence of FUCA we need to study deeply the anatomy of ribosomes.

3.2. The anatomy of ribosomes

Ribosomes of any biological system present two subunits. The smaller one binds the messenger RNA while the larger presents three sites for the binding of transfer RNAs bringing specific amino acids to the system.

The exact ribosomal site on which amino acids are bound together is considered as its catalytic center. This catalytic center has been named as Peptidyl Transferase Center, or simply PTC. The PTC is part of the 23S RNA of bacteria and it is known to catalyze the binding of two separate amino acids into a dipeptide using an Adenine as the most important catalytic center. In this synthesis by dehydration, a molecule of water is jumped out while the C-terminal of an amino acid binds to the N-terminal of another, starting with the polymerization. Other amino acids can be further added to this dipeptide, raising the number of amino acids that can be bound together and allowing the formation of oligopeptides.

In our view, the appearance of a nucleic acid molecule in the early Earth that was capable to bind together two amino acids, somehow started a process of chemical symbiosis on which the binding of this nucleic acid to the peptides produced by itself allowed the system to both (i) stabilize under a self-referential perspective and (ii) aggregate complexity in layers, probably through the mechanism known as accretion.

4. FUCA is born at the proto-PTC

Here we propose that the biological systems emerged in the very moment on which a macromolecule of nucleic acid containing dozens of nucleotides were capable to fold in the 3D space and catalyze the junction of two amino acids into a dipeptide. Thus, FUCA was born when a proto-PTC emerged for the first time, allowing already existing self-replicating RNA-like polymers to produce random di- and oligopeptides. These random peptides produced possibly bound back to the RNA polymers and allowed a higher stabilization of the system that got more robust and was further bound to other stabilizing molecules.

Chaos theory advocates might say that a strange attractor should have been formed at that point. Astrophysics uses the term accretion to explain how planets and other bodies may be formed by the aggregation of material due to gravity. Here, we also use the idea of accretion to explain how the ribosome has been assembled from the proto-PTC into a higher-level system. Although the peptides were bound randomly, the process itself cannot be seen as random because there should have been very simple and specific types of amino acids existing in the primordial pool of molecules. These early amino acids were most likely Glycine, Serine, Alanine and others (Miller, 1953, Paker et al. 2011). Their bound together attracted other amino acids and have possibly formed a first layer of peptides that bound to the proto-PTC and stabilized its interaction to the proto-tRNAs. After subsequent layers of complexification, these processes would evolve to the creation of the larger ribosomal subunit with its tRNA sites A, P and E. Many works have already tried to understand how this subunit has been clearly formed; some considering that PTC were at the beginning (Petrov et al, 2015) and others presenting evidences that the ribosome structure started elsewhere (Caetano-Anolles, 2015). It is consensual, however, that the process of ribosome assembly took a long time

until it could become functional and efficient in its task of amino acid binding.

5. The maturation of FUCA

The maturation of FUCA happened when the basis for what we call today the genetic code has been achieved. With the basis for the code, we aim to describe the 3 types of ancestral RNAs working as a system for encoding (mRNA), decoding (tRNA) and binding amino acids (rRNA). Recent studies seem to indicate that the same module of 30-70 nucleotides should have operated to produce initially a tRNA-like molecule. Specific duplications and concatenations of this initial polymeric module might have produced the proto-PTC by one side and the proto-messenger by the other (Farias et al, 2014, 2016). And by the further attachment of specific proto-peptides produced by the own system some formed each time more stable linear molecules, cloverleaf structures or ribozymes.

The First Universal Common Ancestor (FUCA) is therefore an ancestor of LUCA's lineage. It was born when self-replicating polymers of RNA-like nucleotides started to bind amino acids and its maturation happened with the establishment of the genetic code.

As a molecule, FUCA has no gender and sex would only arise in biology billions of years later. We decided to describe FUCA as a woman, the great-grandmother of LUCA (referring as "it" but sometimes as "she") to homage the bravery of women in our patriarchal societies. As we understand the birth of FUCA as a (i) process of chemical symbiosis and as a (ii) revelation about the importance of symbiotic processes to the emergence of most fundamental biological process, the feminine allusion is also a tribute to the memory of Lynn Margulis. The idea of a great-

grandmother is important for the understanding that FUCA also emerged much earlier than LUCA.

The emergence of a proto-PTC has been a contingent moment of enlighten to the creativity of the universe and to the union of macromolecules. At that time, molecules could only collaborate by interacting but never before by building other collaborator molecules. When these ancient nucleic acids were capable to domesticate the abundant amino acids and interact with them achieving more stabilization than each one by itself, a new age has risen and FUCA has been born.

The molecular nature of the universe has discovered that they could collaborate and help themselves while helping others. A moral molecular imperative has become truth. Instead of competing and destructing each other in a RNA-based world on which replicators destroyed each other to get their monomers and build their own copies (Dawkins, 1978), some other replicators found that binding molecules together produced some form of mutual stabilization. A chemical symbiotic relationship emerged as one polymeric macromolecule has now helped the other in a world of chemical war. Together, their stable complex aggregated other ions and molecules, allowing new layers to be produced and augmenting the interaction and balance between themselves.

This system further specialized to be the very language of biology, the language of chemical interchange that further evolved until the formation of the complete ribosome, the genetic code and the maturation of the molecular translational process.

6. Conclusions

We do not aim here to convince the readers about the ideas presented here, as our intention is only to present interesting and inspiring theoretical interpretations about how the biological systems may have emerged. We still know very few about our deep molecular origins and we felt that we need to focus in the tentative to explain how these wonderful phenomena that gave rise to the biological systems might have been originated. Our exploration surely aims to be scientific and many propositions presented here are being tested in laboratories all over the world under the scrutiny of modern science. Much more hypotheses can be further tested experimentally and also by the use of theoretical models to rebuild ancestral molecules; and empirically resurrect them. In this current essay, our aim was also to propose fundamentally philosophical and epistemological ideas about the deep origins of biological systems that would further produce life and suggest interesting points for research.

The view of (i) life as cellular and the (ii) proposal of the Last Universal Common Ancestor (LUCA) as the branching point on which cellular organisms originated was very important at their time. The very idea of both LUCA and FUCA are clearly to be considered as a corollary of Darwin's core idea regarding the common ancestry among all living organisms. In one of the most important experiments in the history of science, the Darwin contemporary Louis Pasteur has demonstrated that biogenesis could not happen spontaneously and organisms need other organisms to arise. The chain of life is therefore linked together back to LUCA.

There is an agreement among scholars studying the origins of life that LUCA should be considered the ancestral of cellular organisms. However, the most recent proposal about LUCA's genome reveals a highly complex cellular organism with about

355 gene families working together to produce life and a cellular organism.

Being enlightened by Darwinian ideas, it has been our aim to bring gradualism under the theoretical research about the Origins of life. This brought us back to consider the origin of life from the point of prebiotic chemistry and we propose here that biological systems started to emerge when RNA-like molecules started to bind amino acids together. This is the place on time when FUCA has been born. The history of FUCA development however has probably taken much time but we consider her to be matured at the moment on which the system known nowadays as Translation has been completely developed together with a functional genetic code responsible to translate the information present in a nucleic acid into a peptide under an organized form.

Conflicts of Interest: The authors declare no conflict of interest.

7. Acknowledgements

We would like to thank FAPERJ (JCNE E-26/202.810/2015) for funding FP. MVJ was financially supported by PAPIIT-IN224015; UNAM; México.

8. References

1. Caetano-Anollés G. Ancestral Insertions and Expansions of rRNA do not Support an Origin of the Ribosome in Its Peptidyl Transferase Center. *J Mol Evol.* 2015 Apr;80(3-4):162-5. doi: 10.1007/s00239-015-9677-9.
2. Darwin CR, 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. London: John Murray. [1st edition]
3. Dawkins, Richard. (1978, ©1976) *The selfish gene* /New York : Oxford University Press.
4. Delaye L, Becerra A, Lazcano A. The last common ancestor: what's in a name? *Orig Life Evol Biosph.* 2005, 35(6):537-54.
5. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakraborty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Luo OJ, Park E, Persaud K, Preall JB, Ribeca P, Risk B, Robyr D, Sammeth M, Schaffer L, See LH, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Ruan X, Hayashizaki Y, Harrow J, Gerstein M, Hubbard T, Reymond A, Antonarakis SE, Hannon G, Giddings MC, Ruan Y, Wold B, Carninci P, Guigó R, Gingeras TR. Landscape of transcription in human cells. *Nature.* 2012 Sep 6;489(7414):101-8. doi: 10.1038/nature11233. PubMed PMID: 22955620; PubMed Central PMCID: PMC3684276.
6. Farias ST, do Rêgo TG, José MV. Evolution of transfer RNA and the origin of the translation system. *Front Genet.* 2014 Aug 28;5:303. doi: 10.3389/fgene.2014.00303. eCollection 2014.

7. 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). pii: E15. doi: 10.3390/life6020015.
8. Forterre P. The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res.* 2006 Apr;117(1):5-16. Epub 2006 Feb 14. Review. PubMed PMID: 16476498.
9. Forterre P. The origin of DNA genomes and DNA replication proteins. *Curr Opin Microbiol.* 2002 Oct;5(5):525-32. Review. PubMed PMID: 12354562.
10. Forterre P. The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res.* 2006 Apr;117(1):5-16. Epub 2006 Feb 14. Review. PubMed PMID: 16476498.
11. Forterre P. [The great virus comeback]. *Biol Aujourd'hui.* 2013;207(3):153-68. doi: 10.1051/jbio/2013018. Epub 2013 Dec 13. Review. French. PubMed PMID: 24330969.
12. Forterre P, Gaïa M. Giant viruses and the origin of modern eukaryotes. *Curr Opin Microbiol.* 2016 Jun;31:44-49. doi: 10.1016/j.mib.2016.02.001.
13. Gilbert W. The RNA world. *Nature* 1986, v.319, p. 618.
14. Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell.* 1983, 35(3 Pt 2):849-57.
15. Harish A, Abroi A, Gough J, Kurland C. Did Viruses Evolve As a Distinct Supergroup from Common Ancestors of Cells? *Genome Biol Evol.* 2016 Aug 27;8(8):2474-81. doi: 10.1093/gbe/evw175.
16. Higgs PG, Lehman N. The RNA World: molecular cooperation at the origins of life. *Nat Rev Genet.* 2015 Jan;16(1):7-17. doi: 10.1038/nrg3841.
17. Kruger, K.; Grabowski, P.J.; Zaug, A.J.; Sands, J.; Gottschling, D.E.; Cech, T.R. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell.* 1982, 31(1):147-57.
18. Miller SL. A production of amino acids under possible primitive earth conditions. *Science.* 1953 May 15;117(3046):528-9. PubMed PMID: 13056598.
19. 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.
20. 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.
21. Penny D and Poole A. The nature of the last universal common ancestor. *Curr Opin Genet Dev.* 1999, 9(6):672-7.

22. 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.
23. Raymann K, Brochier-Armanet C, Gribaldo S. The two-domain tree of life is linked to a new root for the Archaea. *Proc Natl Acad Sci U S A*. 2015 May 26;112(21):6670-5. doi: 10.1073/pnas.1420858112. Epub 2015 May 11. PubMed PMID: 25964353; PubMed Central PMCID: PMC4450401.
24. Watson JD and Crick FHC. (1953), A structure for Deoxyribose Nucleic Acid. *Nature*, 171 (4356): 737–738.
25. Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S, Martin WF. The physiology and habitat of the last universal common ancestor. *Nat Microbiol*. 2016, 25;1(9):16116. doi: 10.1038/nmicrobiol.2016.116.
26. Williams TA, Foster PG, Nye TM, Cox CJ, Embley TM. A congruent phylogenomic signal places eukaryotes within the Archaea. *Proc Biol Sci*. 2012 Dec 22;279(1749):4870-9. doi: 10.1098/rspb.2012.1795. Epub 2012 Oct 24. PubMed PMID: 23097517; PubMed Central PMCID: PMC3497233.
27. Woese CR and Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A*. 1977a Nov;74(11):5088-90. PubMed PMID: 270744; PubMed Central PMCID: PMC432104.
28. Woese CR and Fox GE. The concept of cellular evolution. *J Mol Evol*. 1977b Sep 20;10(1):1-6. PubMed PMID: 903983.
29. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* 1990, 87(12):4576-9.
30. Zamudio GS, José MV. Identity Elements of tRNA as Derived from Information Analysis. *Orig Life Evol Biosph*. 2018 Mar;48(1):73-81. doi: 10.1007/s11084-017-9541-6. Epub 2017 Jun 28. PubMed PMID: 28660466.
31. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, Juzokaite L, Vancaester E, Seitz KW, Anantharaman K, Starnawski P, Kjeldsen KU, Stott MB, Nunoura T, Banfield JF, Schramm A, Baker BJ, Spang A, Ettema TJ. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature*. 2017 Jan 19;541(7637):353-358. doi: 10.1038/nature21031. Epub 2017 Jan 11. PubMed PMID: 28077874.