



The Last Universal Common Ancestor of Ribosome-Encoding Organisms: Portrait of LUCA

Patrick Forterre¹

Received: 2 April 2024 / Accepted: 25 June 2024 / Published online: 19 August 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

The existence of LUCA in the distant past is the logical consequence of the binary mechanism of cell division. The biosphere in which LUCA and contemporaries were living was the product of a long cellular evolution from the origin of life to the second age of the RNA world. A parsimonious scenario suggests that the molecular fabric of LUCA was much simpler than those of modern organisms, explaining why the evolutionary tempo was faster at the time of LUCA than it was during the diversification of the three domains. Although LUCA was possibly equipped with a RNA genome and most likely lacked an ATP synthase, it was already able to perform basic metabolic functions and to produce efficient proteins. However, the proteome of LUCA and its inferred metabolism remains to be correctly explored by in-depth phylogenomic analyses and updated datasets. LUCA was probably a mesophile or a moderate thermophile since phylogenetic analyses indicate that it lacked reverse gyrase, an enzyme systematically present in all hyperthermophiles. The debate about the position of Eukarya in the tree of life, either sister group to Archaea or descendants of Archaea, has important implications to draw the portrait of LUCA. In the second alternative, one can a priori exclude the presence of specific eukaryotic features in LUCA. In contrast, if Archaea and Eukarya are sister group, some eukaryotic features, such as the spliceosome, might have been present in LUCA and later lost in Archaea and Bacteria. The nature of the LUCA virome is another matter of debate. I suggest here that DNA viruses only originated during the diversification of the three domains from an RNA-based LUCA to explain the odd distribution pattern of DNA viruses in the tree of life.

Keywords LUCA · RNA world · Universal tree of life · Virus origin

Introduction: LUCA, the Story of a Successful Name

All modern organisms (Archaea, Bacteria, and Eukarya) share a last common ancestor that most scientists now call LUCA (the Last Universal Common Ancestor). The word LUCA was first proposed at a meeting entitled “*The last common ancestor and beyond*” that I co-organized in 1996 in France with Antonio Lazcano, Piero Cammarano, and Rudiger Cerff at the Fondation des Treilles in Provence (see http://www-archbac.u-psud.fr/Meetings/LesTreilles/LesTreilles_e.html). Initially, participants started using

the acronym LCA (Last Common Ancestor) until one of them, Jose Castresana, used instead the acronym LUA (Last Universal Ancestor) at an evening session, noticing that the acronym LCA could be used (and is in effect) for the last common ancestor for any groups of organisms. We were not so excited by the acronym LUA, agreeing that it sounded a bit odd. The next morning, another participant, Christos Ouzounis, after a night of reflexion, proposed the acronym LUCA, as a combination of LCA and LUA. This proposal was immediately applauded by all participants who thought that the acronym LUCA, much like the name LUCY, could become a popular name for scientists and the public alike. The acronym LUCA started appearing in the scientific literature in 1999 (Forterre and Philippe 1999a; Kyrpides et al. 1999). This acronym is now widely use and has served as template to design acronyms for the ancestors of each domain, LECA for the Last Eukaryotic Common Ancestor, LACA for the Last Archaeal Common Ancestor, and LBCA for the Last Bacterial Common Ancestor.

Handling editor: Arturo Becerra.

✉ Patrick Forterre
forterre@pasteur.fr; patrick.forterre@i2bc.paris-saclay.fr

¹ Institut Pasteur, 25 rue du Docteur Roux, 75015 Paris, France

A few scientists have been critical of the name LUCA for a variety of reasons. Some astrobiologists were annoyed by the letter U, LUCA being “only” the last common ancestor of terrestrial life; so, what about the last common ancestors of organisms living on other planets? In my opinion, if such extra-terrestrial organisms are indeed discovered in future, only then it will be necessary to explain that we are speaking about the terrestrial “tLUCA” or of another one. Some scientists argue that the name LUCA should not replace the name Cenancestor, previously proposed by Fitch in 1987 (nearly ten years before) (*cen* for common in Greek) (Fitch and Upper 1987). However, the name Cenancestor is plagued by the same problem as the acronym LCA, with all groups of organisms having their own Cenancestor. For purists who insist to apply the rules of taxonomy, LUCA could mean “Last Universal CenAncestor.” If one considers the priority rule, one should in fact remember that the name “progenote” for the last common ancestor of the three domains was proposed ten years before the name Cenancestor (Woese and Fox 1977a). In that case, the problem is that this name implies a particular view of LUCA, an “*organism in which the link between the genotype and the phenotype was not yet firmly established*”. In that sense, the name progenote provides an answer for a question which is still debated among scientists: was LUCA a progenote...or a genote? (Di Giulio 2023 and references therein). Finally, one can also consider that the rules of taxonomy apply to taxon and not to an individual, such as the LUCA (see below for the discussion about this assumption).

The name LUCA was also sometimes entangled in the dispute about the nature—living or not—of viruses. If viruses are “living,” LUCA is clearly not the common ancestor of “all life” since viruses are polyphyletic and are not constrained by the rule of membrane heredity, each realm of viruses having its own ancestor (Koonin et al. 2023). It has been suggested to consider that the C of LUCA means “the Last Universal Cellular Ancestor” or to replace LUCA by LUCELLA for the “Last Universal CELLular Ancestor” (Nasir et al. 2012). I suggest here for simplicity to consider that the C of LUCA means both Common and Cellular and I will discuss later in this paper the relationships between LUCA and viruses, a complicated and partly unresolved story. In fact, viruses, defined as capsid-encoding organisms, can be themselves considered to be cellular during their virocell stage (Raoult and Forterre 2008, Forterre, 2010, 2016). LUCA can be therefore more precisely defined as the Last Universal Common Ancestor of ribosome encoding organisms (REO) (Raoult and Forterre 2008). Anyway, since a viral LUCA does not exist, it does not seem necessary to change LUCA into LUCAREO!

Following the 1996 meeting in which LUCA was baptized, I co-organized with different colleagues two more LUCA meetings to celebrate the tenth (2006) and twentieth

(2016) anniversaries of LUCA. The 1996 meeting coincided with the publication of the first complete genome sequence that of the bacterium *Haemophilus influenzae*. In 2006, dozens of genomes from the three domains were already available and thousands in 2016, a number that is still exponentially increasing if one now adds partial or complete metagenome assembled genomes (MAGs). This avalanche of data has provided scientists with multiple opportunities to revisit the putative nature of LUCA from comparative genomic analyses, but the controversies surrounding this nature are still going on and were vividly discussed at each of these anniversary meetings with diverse groups and generations of scientists. At the 2016 meeting, the nature of the LUCA virome was also on the table. My own view has dramatically changed during that period; whereas I used to imagine LUCA as an already complex DNA-based organism (the opposite of Carl Woese’s progenote) infected by DNA viruses (Forterre 1992a,b), now, I imagine LUCA as an RNA-based organism much simpler than modern cells, although still more elaborated than the progenote proposed by Carl Woese and George Fox. Recently, I also started becoming skeptical about the existence of DNA viruses at the time of LUCA. I will detail my own view of LUCA in this paper. However, before discussing the nature of LUCA based on comparative genomics and considerations about the tree of life, I will first address some more theoretical questions about the concept of LUCA itself.

Was LUCA an Individual?

It is sometimes suggested that LUCA never really existed as a real individual, but only as a concept. To clarify this question, it is useful to make the analogy between LUCA and the African Eve. All modern women share a last common ancestor that once lived in Africa. Eve corresponds to the junction points (coalescence) of all modern women lineages when one goes back in time in direct filiation. Consequently, she was a real person who once lived on our planet. Similarly, the existence of a real individual corresponding to LUCA is the logical consequence of the mechanism of cell division (one cell produce two or more cells): the coalescence of all modern cellular lineages when going back in time for each of them from daughter cells to mother cells. The existence of LUCA therefore derives from the principle of membrane heredity (Cavalier-Smith 2001) that posits that membranes are inherited from cell to cell. This conclusion would be true, even if genes present in LUCA have left no descendant, which is hopefully not the case.

The comparison of LUCA with the African Eve has of course limitation since *Homo sapiens* pass from one generation to the next by sex and cell fusion, whereas a priori LUCA had no eukaryotic-like sex and probably originated

by cell division. However, this comparison is useful to clarify the concept of LUCA. It is misleading to believe that LUCA was a lonely cell or that it only shared the world with cells like itself (the communal LUCA). Nor was our African Eve living alone in her village. She shared the planet with many other individuals of *Homo sapiens* and even other hominids who have left no direct descendants. The situation was very similar for LUCA. This bug was not a lonely individual, endowed with unique properties, but an anonymous cell living among myriad of contemporaries that were not so lucky in the evolutionary game. Of course, when we try to reconstruct LUCA, we are not trying to reconstruct this individual, but the facial composite of members of the ancient lineage of organisms to which it belonged.

Carl Woese suggested once that LUCA only existed as a community of very similar organisms freely exchanging their genes in a common pool (Woese 1998, 2000). For him, Darwinian evolution did not take place at that time of this communal LUCA and only started to operate when the biology of the proto-Archaea, proto-Bacteria, and proto-Eukarya became sufficiently different to dramatically reduce the frequencies of border-free horizontal gene transfer (HGT). He defined this transition period as the “Darwinian threshold” (Woese 2002). However, it is unclear how the communal LUCA, as a single unit, could have evolved in the absence of competition/selection between different selection units (Poole 2009; Forterre 2012). It seems more realistic to imagine that billions of cellular lineages originated, cooperated, competed, and died out, with Darwinian evolution (diversification/selection) going on during the evolutionary period between the origin of life and the emergence of LUCA, as a unique individual (Chen et al. 2004; Forterre and Gribaldo 2007; Cantine and Fournier 2018; van der Gulik et al. 2024). The ancestors of LUCAs thus most likely evolved in size, shape, complexity, and molecular diversity and colonized many different environments of the young planet, transforming them into biotopes. They exchanged genes when they shared the same biotopes and had compatible biology, but, obviously, not when they were living in different parts of the Earth or when their biology had already diverged too much. There was probably never a single communal LUCA that would have colonized the planet with a monotonous population of similar entities. An ecosystem with a single type of genetically identical cells cannot exist. At any time in this evolution, from early cells at the origin of life up to LUCA, any given organism had many contemporaries with different histories and most of them evolved into different lineages in different places on early Earth (Cantine and Fournier 2018).

Some ancestors of LUCA became probably so successful that their descendants colonized the entire planet and wiped out those from other competing lineages, much like *Homo sapiens*, one way or the other, wiped out all other *Homo species* (and many other animal lineages!). The emergence of

the ribosome from the association of its two subunits which probably triggered and coincided with the emergence of the decoding mechanism was probably this type of event that produced a dramatic bottleneck in evolution (Petrov et al. 2015; Bowman et al. 2020). If other RNA-based molecular systems to synthesize proteins were once invented (some possibly producing proteins with D-amino acids and/or with less or more than 20 amino acids), organisms with these systems had all probably disappeared at the time of LUCA or were rapidly eliminated during the diversification of LUCA descendants. The emergence of the ribosome was an important milestone, and I suggested once to divide the period between the origin of life and the emergence of DNA into two steps, the first and second ages of the (cellular) RNA world (Forterre 2005) (Fig. 1). This proposal was of course inspired by the first and second ages of the Middle Earth in Tolkien’s saga, the *Silmarillion*. I suggested calling the cells equipped with ribosomes ribocells to distinguish them from RNA cells from the first age of the RNA world functioning with both RNA genomes and ribozymes (Forterre 2010). Some authors have previously used the term ribocell for all cells with an RNA genome. I would suggest here to name RNA cells those cells that thrive during the first age of the RNA world (Fig. 1). Notably, one can conclude that LUCA originated after a rather long period of ribocells evolution, since several paralogous proteins are present in the universal protein set, indicating that important gene duplications had already taken place before LUCA (Zhaxybayeva et al. 2005; Alvarez-Carreno et al. 2021).

The descendants of LUCA certainly did not wipe out their contemporaries instantly! We thus should imagine LUCA sharing the planet not only with its close relatives (belonging to the same “species”) but also with many other lineages—some of which were very similar (like *Homo neanderthalensis* living at the same time as Eve), others very different (as bacteria are from us); some LUCA contemporaries were living as single cells of different sizes, others perhaps as colonial multicellular organisms. Genome analyses have taught us that we have inherited quite a lot of genes from our close relatives, such as *Homo neanderthalensis* or Denisovien. Similarly, modern genomes of REO should harbor genes that have not been inherited directly from LUCA but from evolutionary lineages that have co-existed for some time with its descendants (Zhaxybayeva and Gogarten 2004; Fournier et al. 2011). They are also full of genes that originated in the genomes of viruses that infected descendants of LUCA and co-evolved with proto-Archaea, proto-Bacteria, and proto-Eukarya (Forterre 2005, 2006). This does not mean that we should play down the importance of LUCA. Instead, we should have a clear view of what it was (and what it was not) to prevent unnecessary debates about its existence or its dissolution in a cloudy web of life. To conclude this section, let’s remember that thinking

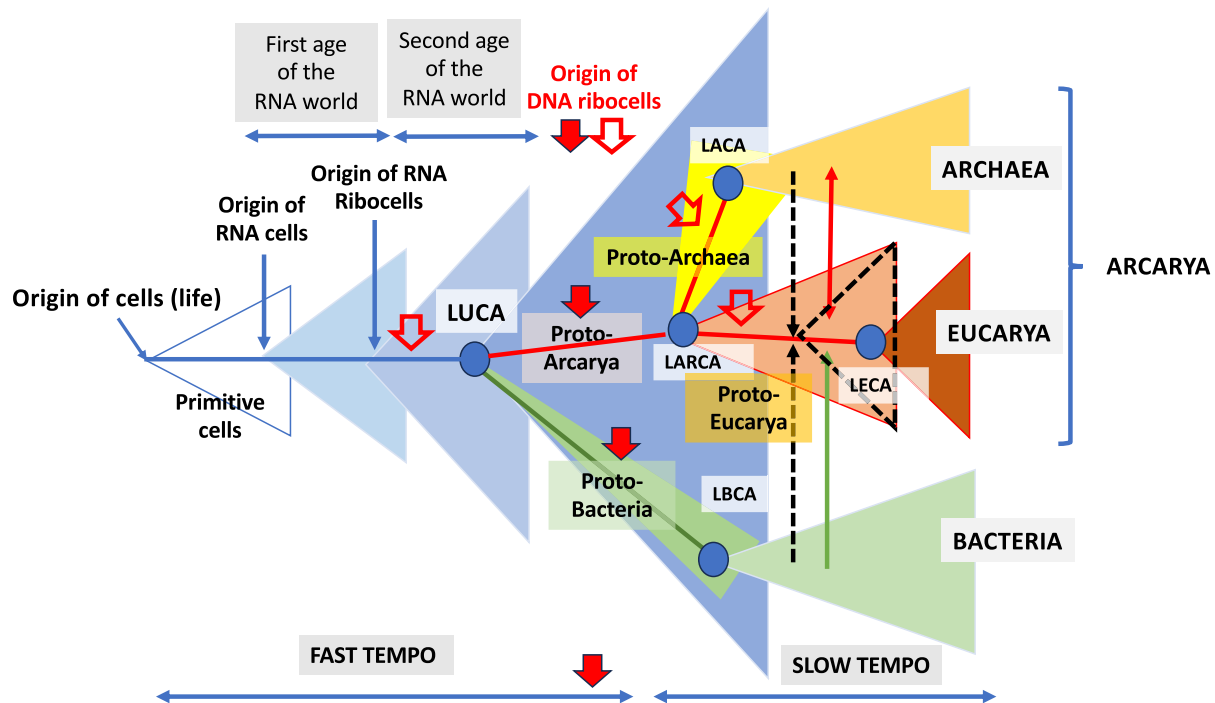


Fig. 1 A schematic tree of life (viruses and other parasitic organisms are included in the triangles that symbolize the diversification of major lineages). Filled large red arrows symbolize the origin of DNA according to my presently favored hypothesis: two transfers from viruses to ribocells, one in proto-Bacteria, and one in proto-Arcarya. Open large red arrows symbolize the origin of DNA according to alternative hypotheses, before LUCA or three independent transfers, post-LUCA in proto-Bacteria, proto-Archaea, and proto-Eukarya.

Double red thin arrows represent HGT between proto-Eukarya and Asgards, whereas the double green thin arrow represent the endosymbiosis that led to mitochondria. The tree is based on a 3D tree scenario. The alternative 2D tree scenario is suggested by black-dotted arrows corresponding to the combination of an archaeon and a bacterium. The dotted triangle remains us that 2D scenarios also involves the existence of proto-Eukarya (Color figure online)

of LUCA is fascinating. Once upon a time, one organism gave birth to two progenies at the origin of two lineages that finally produced completely different organisms. Bacteria emerged from one of these two lineages, whereas Archaea and Eukarya emerged from the other. The evolutionary period that took place between LUCA and the ancestors of the three modern domains (Fig. 1) is often underestimated but we will see in this essay that many essential evolutionary events take place during this period.

The Position of LUCA in the Tree of Life

A most important question regarding LUCA is its position in the universal tree of life (uTol) since it corresponds to the root of the tree or more correctly to the tip of its trunk (Becerra et al. 2007a) (Fig. 1). Depending on this root (the term that I will use thereafter for simplification), the conclusion drawn about the nature of LUCA from comparative genomics will be different. The root of the uTol was first tentatively determined by phylogenetic analyses of paralogous proteins that have diverged before LUCA: the

elongation factors EFTu/EF1 and EFG/EF2 and the catalytic and regulatory subunits of the ATP synthases (Iwabe et al. 1989, Gogarten et al. 1989). A few other duplicated proteins were analyzed during the following decades providing similar results (reviewed and analyzed in Philippe and Forterre 1999). These analyses produced duplicated uTol in which the root of one tree can be determined using the other as the outgroup. In both analyses, the root of the tree turned out to be in the branch leading from the tripartition point to the LBCA (called thereafter the bacterial branch). In such rooted tree, the position of LUCA divides the uTol in two main branches, one leading to Bacteria and the other to Archaea and Eukarya. The methodology used to determine the rooting using paralogous proteins was criticized because the bacterial branch is much longer than the two others in the phylogenies of both elongation factors and ATP synthases. This suggested that the long bacterial branch could have been attracted in both trees by the even longer branches of the paralogous proteins used as the outgroups (Philippe and Forterre 1999; Forterre and Philippe 1999b). A rooting was even obtained in the eukaryotic branch, using the so-called slow-fast method to remove fast evolving positions from the

alignments (Brinkmann and Philippe 1999). However, the bacterial rooting was again recovered by Gogarten and colleagues by analyzing the nature of amino acid conserved along the different branches of a universal ribosomal protein tree (Fournier and Gogarten 2010). These authors detected in the bacterial branch a strong bias for several amino acids, supposed to be signature of a more primitive genetic code. They hypothesized that these amino acids were overrepresented in LUCA compared to those that were introduced later in the genetic code. It would be interesting now to resume such analyses with updated uTol including non-ribosomal proteins and using the enriched species dataset now available.

Notably, the bacterial rooting is the most parsimonious when looking at the distribution pattern of the ribosomal proteins (r-proteins) among the three domains of life (Forterre 2015) (Fig. 2). Ribosomes of all REO share 34 homologous r-proteins (Lecompte et al. 2002; Bowman et al. 2020). In addition to these universal proteins, Archaea and Eukarya share 33 homologous r-proteins that are absent in Bacteria, whereas the bacterial ribosome contain 22 specific bacterial r-proteins. These proteins are located on the rRNA at similar positions to those of their non-homologous counterparts in Archaea and Eukarya. Strikingly, there are no ribosomal proteins shared by Bacteria and Archaea and absent in Eukarya or vice versa. The most parsimonious rooting explaining this

unique pattern is clearly the bacterial one. In that case, the proteins specific to Archaea and Eukarya were not present in LUCA and were added to the ribosome in the lineages leading to these two domains. With alternative roots (either in the archaeal or the eukaryotic branches), one should imagine that the proteins common to Archaea and Eukarya were already present in LUCA and replaced in Bacteria by the non-homologous ribosomal proteins or vice versa (Forterre 2015). However, there is no obvious selection pressure to explain such massive and unidirectional replacements implying multiple losses followed by multiple gains.

From the biochemical work in my former laboratory in Orsay, we obtained similar data favoring for the bacterial rooting in studying the complexes responsible for the universal N6-threonylcarbamoyl adenosine (t6A) tRNA modification that is essential for the correct reading of ANN codons (Thiaville et al. 2014a, 2014b; Forterre 2015). This reaction is performed by protein complexes called TsABDE in Bacteria and KEOPS in Archaea and Eukarya (Thiaville et al. 2014b, Missouri et al. 2019). These complexes include two universal proteins, Kae1/TsaD and Sua5/Qri7 (in Archaea/Bacteria, respectively), but also several accessory proteins that are homologous in Archaea and Eukarya, but not between Bacteria and the two other domains. In modern ribocells, the two universal proteins cannot synthesize t6A without the help of

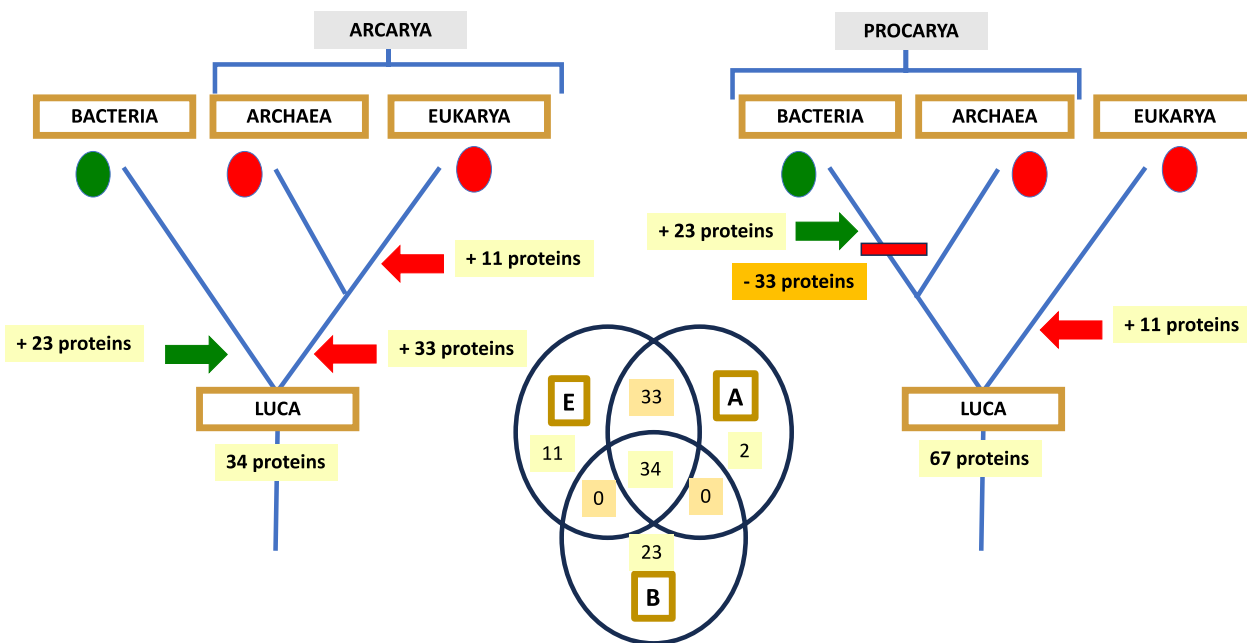


Fig. 2 The distribution of ribosomal proteins helps to root the universal tree of ribocells. When the tree is rooted between Bacteria and Arcarya (left panel) the present distribution of ribosomal proteins in the three domains only requires addition of new ribosomal proteins in the different proto-lineages. If the tree is rooted in the branch leading to Eukarya (right panel) both subtraction and addition of ribosomal

proteins are required in the bacterial branch, which is less parsimonious. All other possible roots (in the branch leading to Archaea or within one of the three domains) are also less parsimonious since they also require subtraction and addition of ribosomal proteins (adapted from Forterre 2015, 2022b)

these additional subunits (Perrochia et al. 2013). However, the homologues of the Kae1 and Qri7 proteins present in mitochondria are sufficient to perform tRNA modification (Thiaville et al. 2014a), suggesting that their ancestors were already capable of performing this tRNA modification in LUCA (Forterre 2015). In the framework of the bacterial rooting, this suggests that the accessory subunits were added independently post-LUCA in proto-Archaea and proto-Bacteria. If one considers another rooting, one should explain why the specific bacterial proteins present in LUCA were replaced by different ones in Archaea and Eukarya or vice versa, which is again less parsimonious.

This reasoning can be extrapolated to many other systems. The bacterial rooting is also the most parsimonious to explain the independent addition of non-homologous protein components to the translation, transcription, and replication machineries in the lineages leading to Bacteria and those leading to Archaea and Eukarya (see below) (Olsen and Woese 1997). The addition of these proteins during the evolution of proto-Bacteria, proto-Archaea and proto-Eukarya correspond to refinement of the basic molecular mechanisms that became more complex and probably more performant than they were in LUCA. This evolutionary trend was precisely predicted by Woese and Fox decades ago (from the first observations of comparative biochemistry) when they concluded that the molecular fabric of each domain became independently refined after their divergence from the progenote (Woese and Fox, 1997a, 1997b). These independent refinements make the basic molecular fabrics more and more integrated within domain and incompatible between domains. As a consequence, the core molecular biology of each domain remained remarkably stable and domain specific. This explains why Carl Woese was a strong opponent of scenarios in which Eukarya originated from a combination of Archaea and Bacteria, for him: “modern cells are sufficiently complex, integrated, and ‘individualized’ that further major change in their designs does not appear possible” (Woese 2000).

Notably, the bacterial rooting validates a clade including both Archaea and Eukarya, at least in the framework of scenarios in which the three domains are all monophyletic (Woese et al. 1990) (Fig. 1). In 2015, I suggested calling this clade Arcarya (Fig. 1 and 2) and I will use this name therein, which also avoid repeating too often “*in Archaea and Eukarya*” (Forterre 2015). Accordingly, the last common ancestor of Archaea and Eukarya could be named LARCA, the Last ARcaryal Common Ancestor. The name Arcarya has not been used in the literature until now, because most evolutionists presently favor a two primary domains scenario in which Eukarya are a subgroup of Archaea (Lopez-Garcia and Moreira 2023). However, our re-analysis of the data supporting two primary domains (2D) scenarios suggests that the three domains topology (3D) is most likely the correct

one (review in Da Cunha et al. 2022a, b). I will come back to this point at the end of this paper.

LUCA was Probably a Mesophile or a Moderate Thermophile

When the 16S rRNA uTol was rooted in the bacterial branch at the end of the eighties, based on the analysis of paralogous proteins, Karl Stetter, the father of hyperthermophiles, concluded that LUCA lived at very high temperature, because it was surrounded by short branches leading to hyperthermophilic archaea and bacteria (Stetter 1996). However, this clustering may have also resulted from the enrichment of the hyperthermophiles rRNA in GC base pairs to increase their stability. It was suggested that this enrichment, reducing the available sequence landscape available for the evolution of RNA, produces artificially short branches in the rRNA tree (Forterre 1996). In a series of elegant studies, Gouy and colleagues tried to determine if LUCA was a thermophile or not by reconstructing the putative sequences of its rRNA and universal proteins (Galtier et al 1999; Bousseau et al. 2008). The optimal growth temperature of modern organisms indeed correlates with the GC base composition of their rRNA and with the amino acid composition of their proteins (proteins from hyperthermophiles being specifically enriched in certain amino acids and depleted in others). In reconstructing the putative sequences of the LUCA rRNA and of some LUCA universal proteins, Gouy and colleagues obtained results suggesting a mesophilic LUCA (Galtier et al 1999; Boussau et al. 2008). It would be important now to confirm, or not, these results using update species datasets. Other arguments have been used to challenge the hypothesis of a hot LUCA. Lazcano and colleagues argued that LUCA was not a hyperthermophile because it lacked the protein disulfide oxidoreductase (PDOs) superfamily, which include proteins involved in the formation of disulfide bridges during protein folding (Becerra et al. 2007b). Again, the phylogenetic analyses that supported this conclusion would need to be updated. Furthermore, this argument is not very strong since, even if disulfide bridges are part of the strategy used to stabilize proteins at high temperature (Ladenstein and Ren 2006), they are also present in mesophilic proteins and not essential for protein stabilization.

A major argument against the idea that LUCA lived in a very hot biotope is based on the phylogeny of reverse gyrase. This fascinating enzyme, which is formed by the fusion of a helicase and a type I DNA topoisomerase (of the A family), was first discovered in the hyperthermophilic archaeon *Sulfolobus* by its capacity to introduce positive supercoiling into a covalently closed circular DNA (Kikuchi and Asai 1984; Forterre et al. 1985). This was the opposite of the reaction catalyzed by the well-known bacterial DNA

gyrase which produces negative supercoiling. Early comparative genomic analyses revealed that reverse gyrase is the only protein specific for hyperthermophiles, i.e., it was encoded in the genomes of all hyperthermophiles known at that time (organisms with optimal growth temperatures equal or superior to 80 °C) and absent in mesophiles (Forterre 2002a). Later analyses, based on an increasing number of genomes, have confirmed that reverse gyrase is always present in hyperthermophiles but is also sometimes present in moderate thermophiles (organisms with optimal growth temperatures between 50 and 80 °C), whereas it is never present in mesophiles (Brochier-Armanet and Forterre 2007; Catchpole and Forterre 2019). The fact that not one genome from mesophiles encodes reverse gyrase is especially striking considering the huge number of mesophilic genomes now present in databases. The exact role of reverse gyrase in vivo remains unknown, which is quite frustrating, but its genomic distribution pattern clearly indicates that this enzyme is essential for life in very hot environment.

The first two published phylogenetic analyses of reverse gyrase indicated that the archaeal and bacterial enzymes were very similar and mixed in phylogenetic trees, suggesting that this enzyme was not present in LUCA (Forterre et al. 2000; Brochier-Armanet and Forterre 2008). In contrast, Martin and his colleagues more recently published a reverse gyrase tree in which archaeal and bacterial reverse gyrases form two monophyletic groups and thus concluded that LUCA was a hyperthermophile (Weiss et al. 2016a, b). The reverse gyrase tree can be found in the supplementary

Figure 1 in Catchpole and Forterre (2019) in which colors distinguish archaeal and bacterial sequences. Notably, the branch that separates archaeal and bacterial reverse gyrases in the tree published by Martin and colleague is very short. This is problematic since the branches that separate Archaea and Bacteria in universal protein trees are always rather long (Fig. 3) (Figure S2 in Da Cunha et al. 2017, Berkemer and McGlynn 2020, Moody et al. 2022). As originally stated by Carl Woese and colleagues, “the interdomain differences between the characteristic archaeal and bacterial proteins” that diverged from LUCA “must far outweigh any intradomain difference” (Woese et al. 2000). This is probably because protein evolved faster during the period between LUCA and the specific common ancestors of each modern domain (see below the discussion about the evolutionary tempo at the time of LUCA). In a more recent phylogeny of reverse gyrase, including 376 sequences (instead of 97 in Weiss et al. 2016a, b), the archaeal and bacterial reverse gyrases do not form anymore two monophyletic groups. Several groups of archaeal and bacterial reverse gyrases separated by very short branches are intermixed, suggesting anew that reverse gyrase was not present in LUCA (Catchpole and Forterre 2019) (Fig. 3). This result confirms that LUCA was most likely not a hyperthermophile. However, we cannot exclude the possibility that LUCA was a moderate thermophile since some modern ones lack reverse gyrase.

A surprising and interesting observation made by Gouy and his collaborators when they tried to determine the temperature at which LUCA was living was that, in contrast to

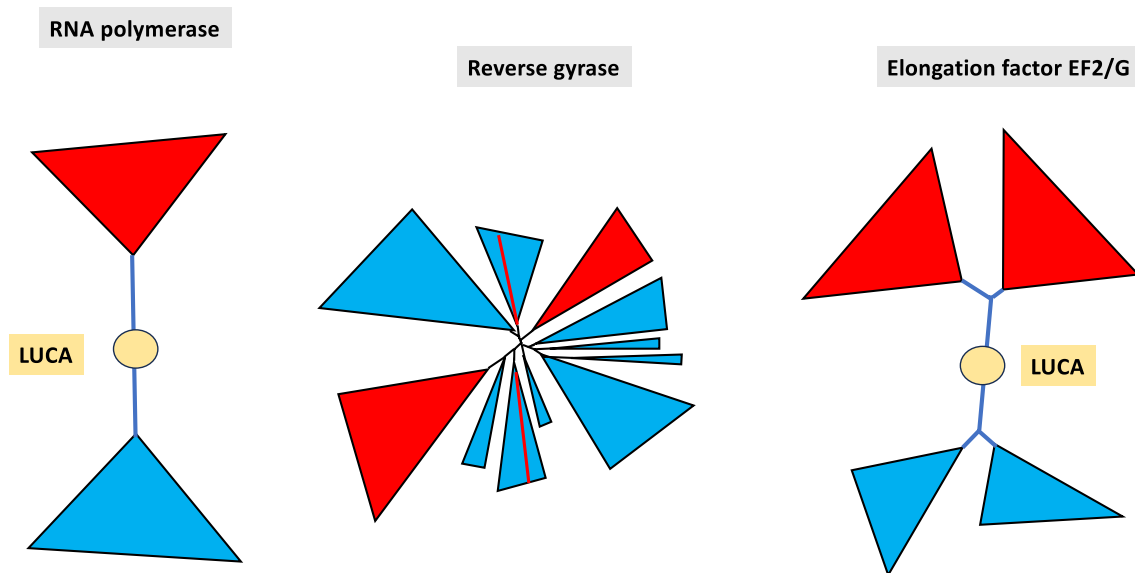


Fig. 3 Comparison of the reverse gyrase phylogeny with those of two universal proteins: the RNA polymerase β subunit and the elongation factor EF2/G. These schematic phylogenies are adapted from Figs. 2 and 3 in Catchpole and Forterre 2019. Bacteria are in Red and Archaea in blue. The original phylogenies included all species encod-

ing a reverse gyrase. The position of LUCA is arbitrary. The long branches between the monophyletic clades of archaeal and bacterial RNA polymerases and elongation factors EFG/2 were typical of universal proteins (Berkemer and McGlynn 2020; Moody et al. 2022) (Color figure online)

LUCA, the LACA and LBCA were probably thermophiles (or even hyperthermophiles in the case of LACA) (Boussau et al. 2008; Groussin and Gouy 2011). If this inference is correct, it means that adaptation to life at high temperature occurred independently in proto-Archaea and proto-Bacteria. Several hypotheses can explain this observation. It was suggested for instance that hyperthermophilic archaea and bacteria were the only LUCA offspring that survived the late heavy bombardment 3.9 billion years ago (Gogarten-boekels et al. 1995). Another possibility is that adaptation to high temperature selected proto-Archaea and proto-Bacteria because they were the first lineages with DNA genomes, a scenario supporting the hypothesis of an RNA-based LUCA (see below) (Boussau et al. 2008). A covalently closed circular DNA is indeed extraordinary resistant to thermodegradation, at least up to 107 °C (Marguet and Forterre 2005). More generally, it is possible that adaptation of life to high temperature favored the emergence of the “prokaryotic phenotype” (including a covalently closed circular DNA genome) in agreement with the thermoreduction hypothesis that I proposed thirty years ago (Forterre 1995). Compared to Eukaryotes, Archaea and Bacteria are characterized by the rapid turnover of their macromolecules counteracting their degradation at high temperature. Moreover, the coupling of translation and transcription, making possible the existence of short-live mRNA, is an important advantage for life at high temperature, considering the susceptibility of mRNA to thermodegradation especially in the presence of magnesium at physiological concentrations (Eigner et al. 1961; Forterre 1992a, 1995; Hethke et al. 1999). RNA is more susceptible than DNA to heat-induced hydrolysis, because the oxygen in 2' of the ribose can attack the phosphodiester bond at high temperature provoking the breakage of the link between the ribonucleotides (Ginoza et al. 1964). The mRNA stability required by the eukaryotic type of molecular biology could explain why Eukarya are missing from biotopes with temperatures between 60 and 110 °C that are only populated by Archaea and Bacteria (Forterre 1995).

The hypothesis of a hot LUCA is often favored by proponents of a hot origin of life who assume a direct link between this hot origin and LUCA (Weiss et al. 2016a, b). However, independently of the data previously discussed which refute the hot LUCA hypothesis, it is difficult to imagine that living organisms thrived in a very high-temperature environment during the two ages of the RNA world, considering that RNA is rapidly degraded at high temperature (see below). Moreover, the conclusion that LUCA was probably a mesophile does not rule out the possibility of a hot cradle for life followed by the emergence RNA-based cells in a milder environment (Miller and Lazcano 1995). Finally, one cannot exclude that, beside a mesophilic LUCA, other lineages were living at that time in hot environments but left no extant descendants (Glansdorff et al. 2008).

The Elusive Biotope and Timing of LUCA

The biotope of LUCA cannot be determined with our present knowledge. It has been suggested that life originated in a potassium rich environment, possibly close to some terrestrial hot springs, to explain the major role played by potassium ion in all modern organisms (Mulkiđjanian et al. 2012). Notably, potassium is also the best ion to protect mRNA against degradation at high temperature (Hethke et al. 1999). Potassium is rare in the environment, especially in water bodies, and all ribocells need an efficient transport system to pump potassium into their cytoplasm and expel sodium out of the cell. Such systems were probably already operational in LUCA, suggesting that it was capable of thriving in potassium-poor and sodium-rich environments (Mulkiđjanian et al. 2009, 2012). Another challenging topic is the age of LUCA. The first reasonable traces of life are microfossils dating from 3.4 to 3.5 Gyr (Schopf et al. 2018; Knoll and Nowak 2017) and it is currently assumed that Archaea and Bacteria were already thriving on our planet at that time or earlier (Schopf et al. 2018; Fournier et al. 2021). Older putative microfossils, such as filamentous structure resembling those of modern bacteria from hydrothermal environment, have been observed in rock from 3.7 to 4.3 Gyr old (Papineau et al. 2022), but they remain controversial. Recently, using phylogenetic approaches, it has been suggested that LUCA lived between 4.32 and 4.53 Gyr (Mahendrarajah et al. 2023). This would mean that LUCA emerged immediately after the formation of the Earth (4.54 Gyr) and even possibly before the moon creating impact (4.4–4.5 Gyr)! These odd results indicates that the methodology used by the authors may not be reliable. In fact, the phylogenies used were based on too limited datasets (a small set of ribosomal proteins in one study and the two catalytic subunits of the A- and F-type ATP synthases in the other). In my opinion, it is doubtful that phylogenetic data, even with better datasets, could provide more than very rough estimate of the age of LUCA, since we cannot seriously determine the tempo of protein evolution at that time. It is still debated if life could have survived the late heavy bombardment of 3.9 Gyr. In the affirmative, LUCA might have lived around 4 Gyr ago; however, if this bombardment drastically eliminated all forms of putative earlier life, one should conclude that LUCA was probably living around 3.7 to 3.8 Gyr ago.

The Nature of LUCA

Very different views about the nature of LUCA have been proposed in the scientific literature. In opposition to the “progenote hypothesis,” it is sometimes assumed

that LUCA was very similar to modern prokaryotes. This idea was proposed by a few scientists who rooted the tree within Bacteria or who were inspired by the superficial phenotypic resemblance between Archaea and Bacteria (see, for instance, Cavalier-Smith 2021). The assumption that LUCA was a “prokaryote” was sometimes a consequence of the ambiguity introduced in the scientific literature by the term “prokaryote” itself (Pace 2009). Since LUCA most likely lacked a nucleus, it was of course a “prokaryote” *stricto sensu*, i.e., an organism that preceded those with the eukaryotic nucleus. In that sense, all cells that thrived on our planet, from the origin of life to the emergence of the eukaryotic nucleus were “prokaryotes,” including all RNA base cells. However, in the literature, the term prokaryote is often synonym of an organism resembling Archaea and/or Bacteria. Rooting the tree between Bacteria and Archaea has also favored a “prokaryotic view,” in which LUCA exhibited all traits now present in Archaea and Bacteria. However, many of these common traits were possibly acquired by convergent evolution toward the modern prokaryote” phenotype (small genome size, coupling between transcription and translation, genes grouped in operon, etc.) as suggested, for instance, in the thermoreduction hypothesis (Forterre 1992a, 1995) and do not necessarily testify for a prokaryotic LUCA. We will see below from comparative genomic analysis that LUCA was most likely very different from Archaea and Bacteria, i.e., from modern “*prokaryotes*.”

Several authors, including myself, once suggested that LUCA was in fact more complex than modern prokaryotes and exhibited some features that are now only present in Eukarya (Forterre 1992a, see Mariscal and Doolittle 2015 for a review of the early “*Eukaryotic first hypothesis*” and Staley 2017, Staley and Fuerst 2017, for the “*compartment commonality hypothesis*,” which posit that LUCA was a nucleated cell). The portrait of LUCA as a kind of proto-eukaryote has now been abandoned by most evolutionists who stick to the idea that Eukarya originated from Archaea. However, we will see later that the situation is more open than often thought and that one cannot exclude that some specific eukaryal features were already present in LUCA and later lost in Archaea and Bacteria.

At the opposite side of the spectrum, it was proposed that LUCA was not even a “prokaryote” but an “acellular organism” (Koga et al. 1998; Martin and Russell 2003; Martin and Koonin 2006). The acellular LUCA was proposed to explain the so-called “lipid divide,” i.e., the dramatic differences between the chemistry and the stereochemistry of archaeal and bacterial lipids. Bacterial phospholipids are indeed made of fatty acid esters linked to sn-glycerol-3-phosphate, whereas those of Archaea are made of isoprenoid ethers linked to sn-glycerol-1-phosphate. This lipid divide suggested to some authors that cellularization

occurred twice independently each time using one of these two lipid types. A very elaborated scenario was proposed, in which LUCA was portrayed as loose complexes of macromolecules enclosed within mineralized compartments inside an expanding hydrothermal chimney at the bottom of the ocean (Koonin and Martin 2005). This hypothesis was supposed to directly link LUCA to an origin of life based on the geochemistry of hydrothermal systems. The authors suggested that cellularization occurred twice independently with different lipids in the chimney that served as a cradle for LUCA. The two cellular lineages then emerged at the tip of this chimney, corresponding to the archaeal and bacterial lineages, respectively (Koonin and Martin 2005). The acellular LUCA hypothesis can be easily refuted by remaining the presence in the universal protein set of proteins whose activity is associated with the presence of a membrane (Delaye et al. 2005). One can cite the factors involved in protein secretion (SecE, SecY) and the complex that directs ribosomes producing membrane proteins to the inner membrane surface (the SRP complex and its associated RNA) (Harris and Godman, 2021).

In fact, the emergence of closed cell-like structures most likely occurred very early on in the Earth’s history, probably as a prerequisite for the origin of life itself (reviewed in Forterre and Gribaldo 2007, Pohorille and Deamer 2009, Schrum et al. 2010, Gill and Forterre 2015, Joyce and Szostak 2018, Cantine and Fournier 2018) (Fig. 1). All modern life is cellular (including viruses since they replicate their genomes in the virocell, Forterre 2010) and one can argue that acellular “life” never existed. All living organisms are individuals whose physical integrity is maintained by the membrane that divides the universe between an inside world (the living organism) and an outside world (its environment), creating an open thermodynamic system, in which the entropy can be locally reduced by an oriented flow of matter and information. Confinement into cellular structures was also required for the concentration of organic molecules and macromolecules and to maintain proximity and linkage between substrates and products in metabolic pathways as well as between the genotype and its phenotypic expression.

In discussing the nature of LUCA and its predecessors, cellular or not, many authors used the term protocell. It was claimed for instance that LUCA was a protocell because it was a progenote, whereas modern cells are “genote” (Di Giulio 2021 and references therein). This term has introduced some confusion in the literature, because it sometimes designates acellular organisms (before cells) and sometimes primitive cellular organisms, the term cell being reserved to “prokaryotic-like” cells. The term protocell is not meaningful either since, as previously discussed, even the first organisms at the origin of life were already likely cellular. I will use here the simple term “RNA cell” to designate cells with RNA genomes before the emergence of the ribosome.

These cells used RNA both as genetic material and enzymatic resource with ribozymes being the main catalysts of this time (Fig. 1). One could simply call “primitive cells” those elusive cellular entities that existed before the emergence of RNA.

The Translation and Transcription Machineries of LUCA

Rooting the tree between Bacteria and Arcarya allows to make some critical predictions about the nature of LUCA. First, it suggests that the ribosomes of LUCA were much simpler than those of modern organisms, with around 30–40 proteins (about half the content of modern ribosomes) (Fig. 2). Nevertheless, the universality of the genetic code, of the three rRNAs and many tRNAs, and of the main initiation and elongation factors indicates that LUCA had probably rather elaborate protein-synthesizing machinery-producing proteins using the modern optimized genetic code (Vestigian et al. 2006, Fer et al. 2022). Notably, 90% of the rRNA structure is conserved between Archaea and Bacteria, indicating that this universal structural rRNA core was already established in LUCA (Bernier et al. 2018). Nevertheless, for Carl Woese, the translation apparatus of LUCA was still rudimentary, and translation was far less accurate in LUCA than it is today. He supposed that the ribosome produced a collection of closely related sequences from a single gene and that LUCA could only produce small proteins, writing that “*most, if not all modern type proteins could not be produced*” (Woese 1998). However, in contradiction with this statement, the universal protein set also includes a few enzymes involved in tRNA modifications essential for increase translation fidelity, such as the previously discussed tRNA modification t6A, and the RNase P involved in tRNA maturation (Czerwoniec et al. 2009; Phan et al. 2021). Van der Gulick and Hoff suggested from comparative genomics of the anticodon modification machinery in the three domains that LUCA contained a set of 44 or 45 tRNAs containing 2 or 3 modifications while reading 59 or 60 of the 61 sense codons (Gulick and Hoff 2016). This strongly suggests that the ribosome of LUCA was already capable of synthesizing *bona fide* proteins with good accuracy, in contradiction with the progenote hypothesis *stricto sensu* (Woese and Fox 1977a; Woese 1998). However, this does not mean that the translation apparatus of LUCA was as efficient as the modern one. Many tRNA and RNA modifications are domain specific, indicating that the fidelity of translation improved during the diversification of the three domains in parallel with the increase in the number of ribosomal proteins and translation initiation factors. Moreover, it seems that the frequency of some amino acids increased since the time of LUCA, indicating that modern proteins

are probably somehow more complex than LUCA proteins (Brooks and Fresco 2002).

This pathway toward sophistication has taken place in all aspects of cellular biology. For example, in modern ribosomes, the mechanism of ribosome biogenesis involves multiple protein factors, but only one of them, the rRNA dimethyl transferase KsgA/Dim1 is present in the three domains, indicating that ribosome biogenesis was probably much simpler at the time of LUCA (Birikmen et al. 2021, Juttner and Ferreira-Cerca 2022). Further sophistications thus take place independently in the lineages leading to Archaea and Bacteria. The number of new factors involved now in ribosome biogenesis is especially high in Arcarya. Birikmen and colleagues identified 156 ribosome biogenesis factors common to Archaea and Eukarya and many more that are Eukaryal specific! Interestingly, whereas most factors common to Arcarya are conserved in all Eukarya, very few are consistently found throughout the archaeal domain (Birikmen et al. 2021). This patchy distribution possibly suggests that the mechanism of ribosome biogenesis was more elaborated in LACA and was streamlined during the evolution of Archaea and/or that some factors were specifically transferred to some archaeal lineages by HGT from proto-Eukarya. In-depth phylogenomic analyses of all these factors is now required to distinguish between these two hypotheses. Notably, a trend toward reductive evolution in Archaea has been proposed for the ribosome itself (Lecompte et al. 2002) and for their genomes in earlier studies (Csurös and Miklos 2009). If reductive evolution was already at work during the evolution of proto-Archaea, the mechanism of ribosome biogenesis of LARCA may have been even more complex, resembling more that of Eukarya. The present situation with simpler ribosomes in Archaea and more complex ones in Eukarya probably testifies for two opposite modes of evolution in the branches leading to LACA and LECA, driven by reduction and complexification, respectively (Forterre 2013a).

In the case of transcription, seven subunits were specifically added in the arcaryal lineage to the four core RNA polymerase subunits that are homologous in all domains (Werner and Grohmann 2011). There is a single universal transcription factor called NusG in Bacteria and Spt5 in Arcarya (Werner 2012). The domain conserved between these proteins is involved in the stimulation of transcription processivity. According to Finn Werner, this protein “*may have played a crucial role in the expression of long genes and, during evolution even permitted an increase in gene or operon length*” (Werner 2012). The various initiation sigma factors in Bacteria and the basal transcription factors in Arcarya (the TATA binding protein and the associated factor TFIIB) are non-homologous to each other, suggesting that they were independently added to the transcription machinery in the branches leading

to Bacteria and Arcarya. This raises the intriguing possibility that LUCA lacked precise transcription initiation mechanism. This was indeed proposed by Finn Werner and Dina Grohman who suggested the “elongation first hypothesis,” in which, in the absence of initiation factors, the RNA polymerase of LUCA started transcription non-specifically by directly associating with the template DNA (Werner 2008, Werner and Grohman 2011). Notably, such scenario is made even more reasonable if the template was RNA, as suggested below. The bacterial sigma factors have many homologues encoded by head-and-tail bacteriophages of the class *Caudoviricetes*, suggesting that proto-Bacteria could have acquired these proteins from viruses, whereas the TATA binding protein (TBP) of Archaea and Eukaryotes includes a domain associated with proteins of diverse functions in the three domains of life (Brindelfalk et al. 2013). This TBP domain was probably already present in LUCA as a stand-alone protein or associated with other protein domains, but its function at that time of such TPB domain protein cannot be determined (Brindelfalk et al. 2013). Beside the basal transcription factors present in all Arcarya, a plethora of additional factors and macromolecular machines are required for gene expression in Eukarya, such as the mediator, again testifying for the extreme complexification that occurred during the evolution of proto-Eukarya. As in the case of the initiation factors, the factors increasing the fidelity of transcription during the elongation step by stimulating the proof-reading activity of the RNA polymerase (GreA and GreB in Bacteria and TFS/TFIIX in Arcarya) are not homologous in Bacteria and Arcarya. This strongly suggests that these factors were added independently in proto-Bacteria and proto-Arcarya and, consequently, that transcription was less faithful in LUCA than it is in modern ribocells.

Finally, the mechanisms that regulate translation and transcription became probably more and more complex during the evolutionary pathways leading to the three modern domains, increasing the efficiency of gene regulatory networks. Proteins involved in these regulatory pathways are very different from one domain to the other and even highly diversified within domains. An interesting study focusing on RNA families (mostly involved in gene regulation and anti-viral defense) has shown that these families were specific for each domain, except for universal families involved in basic mechanism of translation and snoRNA common to Archaea and Eukarya (Hoepfner et al. 2012). The regulation of gene expression of LUCA was thus probably much simpler than in modern organisms, in agreement with our conclusion that LUCA was a ribocell very different from members of the three domains that we can explore today.

The Genome of LUCA, RNA, or DNA?

It is often assumed that LUCA had a DNA genome since DNA is the universal depository of the genetic material in all modern ribocells. However, in conflict with this assumption, the five major proteins involved in DNA replication: the replicative polymerase (replicase), the primase, which initiates the synthesis of Okazaki fragments, the DNA ligase, which links these fragments to nascent DNA strands, the helicase, which opens the double helix in front of the replication fork, and the type II DNA topoisomerase (Topo II), which resolves the topological problems raised by the double-stranded structure of DNA, all belong to different protein superfamilies in Bacteria and Arcarya (Olsen and Woese 1997; Forterre 1999, 2013b; Leipe et al. 1999; Forterre and Gabelle 2009).

The same observation can be made for DNA repair and recombination (Eisen and Hanawalt 1999; White and Allers 2018). With few exceptions, most proteins involved in these processes are specific for either Bacteria or Arcarya. For instance, the proteins involved in nucleotide excision repair in Bacteria (UvrABC) are not homologous to the XP proteins involved in this process in Eukarya. Archaea encode several homologues of eukaryal XP proteins whose function remains partly elusive (White and Allers 2018). One can also mention the existence of two completely different mismatch repair systems, the EndoMS system widespread in Archaea (and possibly acquired by some Bacteria via HGT) and the MutL/S system, ubiquitous in Bacteria and in Eukarya (probably of bacterial origin) and rare in Archaea, possibly acquired from Bacteria via HGT.

The paucity of proteins involved in DNA metabolism in the universal protein set strikingly contrasts with the predominance of enzymes involved in RNA metabolism, such as RNA polymerases, RNA helicases, and RNA-binding proteins (Anantharaman et al. 2002; Delaye et al. 2005). The most parsimonious scenario to explain this observation is that the DNA replication and repair machineries were introduced independently in proto-Archaea and proto-Bacteria (large red arrows in Fig. 1). A corollary is that DNA itself might have been introduced independently in the two proto-lineages, implying that LUCA was thriving in the second age of the RNA world (thereafter called the RNA-LUCA hypothesis). In particular, the probable absence of a Topo II in LUCA is a strong argument against LUCA already having a double-stranded DNA genome, since Topo II are essential to solve topological problems raised by the intertwining of the two DNA strands. Topo II have sometimes been included in the set of universal proteins, because Topo II activities are present in the three domains (Becerra et al. 2007a). This does not consider

the existence of two families of non-homologous Topo II: Topo IIA and Topo IIB (Bergerat et al; 1997; Forterre and Gabelle 2009). The B subunits of Topo IIA and B are distantly related ATPases, but their A subunit, involved in DNA cleavage, are completely unrelated. Phylogenomic analyses have shown that LACA and possibly LARCA only contained Topo IIB, whereas the LBCA only contained Topo IIA. The LECA encoded a Topo IIA, but this enzyme was recruited from viruses of the kingdom *Nucleocyotoviricota* and not from bacterial Topo IIA (Guglielmini et al. 2022). DNA gyrase, a subclass of Topo IIA that introduces negative supercoiling in DNA, has been sometimes attributed to LUCA because it is present in all Bacteria and several groups of Archaea. However, phylogenetic analyses have shown that these archaeal DNA gyrases were recruited by HGT from Bacteria (Villain et al. 2022).

If the genome of LUCA was already made of DNA (the DNA-LUCA hypothesis), one should imagine that DNA replication and repair proteins were systematically replaced by non-homologous ones, either in proto-Bacteria or in proto-Arcarya (Olsen and Woese 1997; Forterre 1999, 2002a, b; Koonin et al. 2020). I once proposed myself that LUCA had a DNA genome replicated by an archaeal-like DNA replication machinery that was replaced in proto-Bacteria by the replication machinery of some *Caudoviricetes* (Forterre 1999) (Open red arrows in Fig. 1). Koonin and colleagues recently updated this hypothesis, suggesting that the DNA genome of LUCA was replicated by a DNA polymerase of the family D (Pol D), presently only known in Archaea, because Pol D is a distant homologue of cellular RNA polymerases (Koonin et al. 2020). In their scenario, the Pol D inherited from LUCA was later replaced in proto-Bacteria and proto-Eukarya by non-homologous DNA polymerases of the C and B families, respectively. Notably, these authors suggest that all DNA replication proteins, except Pol D, were transferred from viruses to cellular lineages post-LUCA. I previously suggested that all cellular DNA replication proteins indeed have a viral origin because DNA itself possibly emerged in an ancient virosphere (Forterre 2002b, 2005, 2006). One of the arguments supporting this “out of viruses” hypothesis was that chemical genome modification is a classical viral strategy to bypass host defenses targeting viral genomes. In fact, in the framework of the “out of viruses” hypothesis, there is no good reason to make an exception for Pol D. It is more parsimonious to suggest that DNA was transferred independently to proto-Bacteria and proto-Arcarya with progressively two complete sets of non-homologous viral proteins involved in DNA replication and repair (Forterre 2002b). I even suggested once that Archaea and Eukarya also got their DNA from two different funder DNA viruses to explain why DNA replication enzymes, such as Pol D and Topo IIB, are specific to Archaea (Forterre

2006) (open red arrow in Fig. 1). Of course, one cannot completely exclude a replacement scenario to save the DNA-LUCA hypothesis since, for instance, the ancestral bacterial replication proteins have been replaced in mitochondria by non-homologous proteins of viral origin (Filée and Forterre 2005). However, this replacement was the result of the dramatic reductive evolutionary pathway of an endosymbiont in its host, a situation probably very different from what’s happened during the evolution of proto-Bacteria and proto-Arcarya. Notably, the RNA-LUCA hypothesis agrees well with our previous conclusion that LUCA was not a variation of modern ribocells, but a simpler organism, with much less sophisticated translation and transcription machineries.

The RNA-LUCA hypothesis has sometimes been refuted because the set of universal protein includes a few proteins involved in DNA metabolism or in the synthesis of DNA precursors (dNTPs) (Leipe et al. 1999, Becerra et al. 2007a, Cantine and Fournier 2018, Koonin et al. 2020). The evolutionary trajectories of some of these proteins involved in DNA metabolism are indeed compatible with the DNA-LUCA hypothesis, i.e., their bacterial version is very divergent from their arcaryal version. However, this is not the case for other proteins involved in DNA repair or in the synthesis of DNA precursors, such as photolyase, thymidylate synthase, or else ribonucleotide reductases, for which it is not possible to identify bacterial *versus* arcaryal versions. These proteins are divided in several families that are sometimes evolutionary unrelated and exhibit a distribution pattern between domains that does not overlap with the uTol topology. These families exhibit complex phylogenies, suggesting multiple cases of HGT between and within domains (Kanai et al. 1997, Filée et al. 2003; Lundin et al. 2010, Kanai et al. 1997, Becerra et al. 2007a; Cantine and Fournier 2018, Vechtomova et al (2020).

Two hypotheses can be proposed to reconcile the RNA-LUCA hypothesis with the existence of universal proteins involved in DNA metabolism or in the synthesis of DNA precursors. First, some of these proteins might have been involved in RNA instead of DNA manipulations. This may be the case for the DNA-dependent RNA polymerases, since the *E. coli* RNA polymerase can use RNA as template (Pelchat and Perreault 2002; Wettich and Biebricher 2001) and the genomes of viroids and of some RNA viruses are replicated by eukaryal RNA polymerase II (Fels et al. 2001; Moraleda and Taylor 2001, Mac Naughton et al. 2002). Topo IA might have been also involved in RNA manipulation in LUCA since Topo IA from all domains of life can act as RNA topoisomerase (Xu et al. 2013, Ahman et al. 2014, 2016). Notably, it could be significant that Topo IA, which is the only universal DNA topoisomerase, is also the only one that can use RNA as substrate (DiGate and Marians 1992; Sekiguchi and Shuman 1997, Rani et al. 2010). Interestingly,

Nagajara and colleagues have shown that the Topo IA of a mycobacterium is involved in rRNA processing, indicating that if LUCA contained a Topo IA, this enzyme might have function in a similar process (Rani et al. 2010). The single-stranded DNA-binding proteins SSB and RPA are very divergent and only share a common OB-fold domain. Proteins containing this motif are very diverse and some of them can bind single-stranded RNA (Theobald et al. 2003). Photolyases can also act on both DNA and RNA (Gordon et al. 1976; Kim and Sancar 1991). Notably, if confirmed, the presence of an RNA photolyase activity in LUCA would suggest that this organism lived exposed to UV irradiation at the surface of the Earth.

Another hypothesis to explain why some proteins acting on DNA are universal is that these proteins were transferred independently from viruses into proto-Archaea and proto-Bacteria. Most of these proteins indeed have homologues encoded by DNA viruses or plasmids. The co-evolution of ribocells with their mobilome would explain why the phylogenies of some enzymes involved in DNA metabolism overlap with the uTol, whereas multiple HGT between cells and viruses in both directions would explain why others, such as thymidylate synthases and ribonucleotide reductases exhibit a complex evolutionary history (Filée et al. 2003; Lundin et al. 2010, 2015). There are two families of thymidylate synthases, ThyA and Thy, and three classes of ribonucleotide reductases (RNR I, II, and III). ThyA and ThyX are non-homologous, suggesting that present-day DNA containing thymidine (T-DNA) might have originated twice independently from DNA containing uracil (U-DNA), which still form the genome of some viruses (Forterre et al 2004). It is even possible that U-DNA itself was “invented” twice independently. The three classes of ribonucleotide reductases share a homologous core, and it is usually assumed that the ribonucleotide reductase activity originated only once. However, this common core is shared by all proteins of the 10-stranded β/α barrel superfamily, such as pyruvate formate lyase, and the three classes of ribonucleotide reductase require completely different subunit components and co-factors to synthesize dNTPs (Lundin et al. 2015). Consequently, the mechanism to generate the radical involved in removal of the 2' oxygen of the ribose differs between the three classes. Lundin and colleagues have proposed an ad hoc scenario in which they both evolve from a primitive ribonucleotide reductase (Lundin et al. 2015). However, whereas class I most likely evolved from class II, one cannot exclude that class II and III originated independently. Although these two classes are present in Archaea and Bacteria, their complex phylogenies do not support their presence in LUCA (Filée et al. 2003; Lundin et al. 2010, 2015). The history of these proteins has been indeed characterized by frequent HGT between Archaea and Bacteria, probably because strong pressure for environmental adaptation, some of them being

aerobic, while others are strictly anaerobic (Filée et al. 2003; Lundin et al. 2010, 2015).

Another argument frequently used against the RNA-LUCA hypothesis is that RNA cannot be replicated with sufficient accuracy to support the existence of a genome encoding the set of genes (a few hundred) supposed to be present in LUCA (Takeuchi et al. 2011, Martin and Koonin 2006). Such assumption seems a priori justified by the maximum genome sizes of most modern RNA viruses which is around 40 kb (for coronaviruses). This argument can be refuted by thinking about the type of cell (either a ribocell or a virocell) that was required to support the RNA to DNA transition. The genome of this RNA cell should have encoded for all enzymes required for the biosynthesis of amino acid and nucleobases or their transport into the cell and for the metabolic and energetic pathways required to produce ATP and GTP. In addition, the genome of this RNA cell should have encoded several sophisticated protein-enzymes, such as an RNA replicase, a ribonucleotide reductase, and a reverse transcriptase. This means that this RNA cell was already equipped with efficient ribosomes producing elaborated proteins. This seems impossible with a genome of 40 kb or less, teaching us that RNA cells with larger genomes have necessarily once existed.

The comparison between ancestral RNA cells and modern RNA viruses is thus certainly misleading. Modern RNA viruses probably represent only a minute fraction of the diversity of the ancestral RNA virosphere, those which managed to survive the transition from RNA to DNA cells. The present genome size limit of RNA viruses may be thus strongly biased by a sampling effect. Interesting observations can nevertheless be made when looking at modern RNA viruses. Despite their small genomes, RNA viruses encode proteins as large and sophisticated as those of DNA viruses or ribocells. These proteins can manipulate cellular membranes to produce cytoplasmic viral factories. A striking example of a large viral protein encoded by RNA virus protein is the nsp3 protein (222 kDa) from murine hepatitis coronaviruses that can build a nuclear pore allowing the exit of the viral RNA from cytoplasmic viral factories (Wolff et al. 2020).

RNA replicases have indeed a rather low fidelity with an error rate of around 1×10^{-4} to 1×10^{-6} (Sanjuan et al. 2010). However, it has been shown that high fidelity mutant of polyomavirus RNA replicase can emerge from a single point mutation (Pfeiffer and Kirkegaard 2003) and that the error rate of the RNA polymerase from yellow fever viruses that have accumulated clusters of beneficial mutation was as low as 1.9×10^{-7} to 2.3×10^{-7} (Pugachev et al. 2004). The fidelity of viral RNA replicase can be also increased by additional factors. For example, coronaviruses encode an exoribonuclease whose activity can increase replication fidelity (Denison et al. 2011).

Notably, there is a general negative correlation between mutation rate and genome size among RNA viruses, i.e., larger genomes are replicated more faithfully, suggesting that larger genomes in the second age of the RNA world could have been replicated even more faithfully (Sanjuan et al. 2010).

Moreover, one should consider that the replicative RNA polymerase in LUCA was not an ancestor of modern viral RNA replicases, but of the universal cellular DNA-dependent RNA polymerases that are now only involved in transcription in modern DNA ribocells. Modern RNA polymerases in the three domains exhibit intrinsic proof-reading activities that increase the transcription fidelity and could have been used to improve the replication fidelity of the LUCA genome if this genome was made of RNA (Poole and Logan 2005). The fidelity of the LUCA RNA polymerase/replicase was also possibly increased by its association with the ancestor of the universal elongation factor NusG/Spt5 (Werner 2012).

The current idea that RNA would be too labile to support the genome of LUCA can be also challenged by the existence of several biochemical pathways for RNA repair in modern ribocells (nicely reviewed in Poole and Logan 2005). As previously discussed, RNA is much more sensitive than DNA to thermodegradation (Ginoza et al. 1964), but this might not be a problem for LUCA if it was indeed living in a rather low-temperature environment. Topological constraints produced by RNA-binding proteins preventing the free rotation of the two RNA strands could have also increased the stability of the RNA double helix which is already intrinsically slightly more stable than the DNA double helix (Wienken et al. 2011).

Finally, when discussing the genome of LUCA, it is not necessary to imagine RNA genomes of ancient RNA/protein ribocells as simply a mimic of modern DNA genomes. One can imagine multiple RNA redundant (multi-copy) linear chromosomes with sizes between 50 and 100 kb that segregated using mitotic-like devices anchored in the membrane (Woese 1998). Such small linear genomes would be less sensitive to mutation error and gene loss and immune to topological problems that require topoisomerase activities (Woese 1998). They could encode clusters of genes coding for related activities and function somehow more like modern mobile elements (Woese 1998). The transcripts of the few universal operons encoding ribosomal protein genes could be relics of this time. Cells harboring such a genome could have divided by simple “mechanical” cell division mechanisms promoted by lipid biosynthesis (Koonin and Mulikidjanian 2013) and/or by a simple system based on an ancestor of the FtsZ/tubulin superfamily (Pende et al. 2021; Santana-Molina et al. 2023).

The Evolutionary Tempo at the Time of LUCA

The RNA-LUCA hypothesis was already proposed by Carl Woese when he discussed the nature of the progenote (Woese 1983, 1987, 1998). Woese suggested that the low fidelity of its RNA genome replication, associated with the low fidelity of its translation apparatus, explains why the evolutionary tempo was much higher at the time of the progenote (LUCA) than it is now. I, myself, later proposed that three independent viral-promoted transitions from RNA to DNA genomes were at the origin of the formation of the three domains by dramatically reducing the rate of protein evolution in their proto-lineages (Forterre 2006). The evolutionary tempo was indeed necessarily much faster in the relatively short period between the origin of life and the emergence of the three domains (possibly a few hundred million years) than it was during the evolution of the three domains from their respective ancestors (possibly more than 3.5 billion years) (Woese and Fox 1977b) (Fig. 1). In the first short period, life evolved from scratch to ribocells, LUCA, and to the respective ancestors of Bacteria and Archaea, whereas in the second, much longer period, the basic fabrics of DNA ribocells have remained stable in their respective domains. This reduction of the evolutionary tempo would explain why the evolution of modern organisms is now strongly constrained by their previous history (bacteria can only evolve into different bacteria, archaea into different archaea, eukarya into different eukarya). The three versions of universal proteins remained indeed strikingly similar within each domain, despite (around) 3 billion years of evolution (Woese and Fox 1977b). In contrast, a fast evolving LUCA had the capacity to produce descendants that became either Bacteria, Archaea, or Eukarya, very different from their common ancestor. The first proto-Bacteria and proto-Archaea that retained an RNA genome for a while would have also evolved more rapidly than modern organisms, explaining the long branches that separate the bacterial and archaeal universal proteins in phylogenetic trees (Da Cunha et al. 2017, Catchpole and Forterre 2019, Berkemer and McGlyll 2020, Moody et al. 2022).

The idea that the RNA to DNA transition played a major role in a dramatic reduction of the evolutionary tempo rests on the assumption that organisms with RNA genomes evolve more rapidly than those with a DNA genome. We have seen that RNA can be replicated more faithfully than usually assumed (but not as faithfully as DNA, see below) and that dsRNA is as stable as dsDNA. The advantage of DNA over RNA in terms of genome stability and reproduction was therefore not immediate for the first organisms with DNA genomes. However, the advantages became important following the emergence of

specific mechanisms increasing the faithful transmission of the genetic information in DNA. Two major such mechanisms can be identified: the emergence of independent mismatch repair systems in Archaea (the EndoMS system) and in Bacteria (the MutL/S system) (White and Allers 2018) allowing to reach mutation rates as low as 1×10^{-10} and the emergence of DNA repair systems to remove uracil from DNA, preventing the mutational effect cytosine deamination. In the case of mismatch repair, one can still imagine that such a system existed during the second age of the RNA world to increase the fidelity of dsRNA replication, but such systems are presently unknown and were probably not required at the time of LUCA for rather small RNA genomes. In the case of cytosine deamination producing uracil, the advantage of DNA over RNA is obvious since uracil can be detected in DNA but not in RNA. The transition from RNA to DNA occurred necessarily in two steps, with first the emergence of U-DNA followed by the emergence of T-DNA (Forterre et al. 2004). The transition from U-DNA to T-DNA was a major step in increasing the stability of the genetic information once mechanisms to detect uracil in T-DNA and repair the modified sequence emerged in some DNA ribocells and/or virocells. Finally, DNA is not only more resistant than RNA to thermodegradation, as already mentioned but also less sensitive than RNA to cleavage by metal ions, which can be another factor that allowed an increase of the evolutionary tempo after the RNA to DNA transition (Butzow et al. 1975). Besides DNA replication, transcription itself became more accurate after the independent acquisition of the GreA/B system in Bacteria and TFS/TFIIS system in Arcarya. All these improvements in the genetic make-up of the proto-Bacteria and proto-Arcarya likely produced a dramatic slowdown in the genome mutation rate in the LBCA and LARCA lineages.

The Metabolism and Lifestyle of LUCA

The metabolism of LUCA cannot be easily determined, because metabolic enzymes are rare or absent in the sets of 50 to 100 strictly universal proteins conserved in all members of the three domains. This is because metabolic traits have been frequently lost and/or acquired by HGT during evolution, especially between Archaea and Bacteria. However, the real number of metabolic enzymes present in LUCA was certainly rather high. LUCA and its contemporaries needed to produce ATP, amino acids, and nucleotides to support RNA and protein production, as well as the phospholipids required for membrane synthesis. It was also suggested that LUCAs and its contemporaries were most likely genetically redundant for many catalytic activities, with many paralogues and functional analogs

already established in various lineages (Glansdorff et al. 2008). Ancestors of all metabolic pathways present at the time of LUCA were not necessarily present in LUCA itself. Some of them were probably “invented” in other lineages but were later transferred to some descendant of the LUCA before these lineages disappeared (although this supposes that LUCA possessed efficient uptake mechanisms to ingest those synthesized by its contemporaries).

Several authors have tried various strategies to determine the metabolic pathways of LUCA using less restrictive criteria than their presence in all members of each domain, looking for proteins that are not truly universal but present in diverse phyla of each domain, for universal protein folds present in modern metabolic enzymes or else on the distribution patterns of biosynthetic pathways, and metabolic enzymes in the three domains. For instance, it was concluded in one study that LUCA was able to synthesize at least 16 out of the 20 standard amino acids (Hernandez-Montes et al. 2008) and that both salvage and de novo pathway for purine and pyrimidine biosynthesis were already present in LUCA (Armanta-Medina et al. 2014). An in-depth study of the Histidine biosynthesis pathway concluded that most enzymes involved in this pathway were already present in LUCA and possibly organized in operon (Fondi et al. 2009). Several universal enzymatic reactions that were described often require enzymes containing ancestral domains involved in the manipulation of phosphate groups (Escobar-Turriza et al. 2019). Eight such studies have been recently reviewed by Goldman and colleagues (Crapito et al. 2022). These authors inferred from these analyses a consensus LUCA proteome including 366 proteins present in at least four out of the eight previous studies. Their analysis concludes that the genome of LUCA encoded, as expected, proteins involved in amino acid and nucleotide metabolism and use common nucleotide-derived organic co-factors.

I will discuss in more detail here the work of Martin and colleagues, because it has been widely publicized by scientific journalists (Weiss et al. 2016a, review in Weiss et al. 2018, Cooper 2017, see also the chapter on the last common universal ancestor in Wikipedia) and still used to describe the metabolism of LUCA in a recent review (Bozdag et al. 2024). These authors focused on proteins shared by Archaea and Bacteria to determine the proteome of LUCA. These authors used two criteria to discriminate between proteins that were inherited from LUCA and those that were transferred between Archaea and Bacteria post-LUCA: the protein should be present in several groups of each domain, and the archaeal and bacterial proteins should form two monophyletic clades in phylogenetic analyses. Using this strategy, they identified 338 proteins that were supposed to be present in LUCA (117 being present in the data set of Goldman and colleagues). They deduced from this list that LUCA was an autotrophic anaerobe thriving in a hydrothermal vent. This

reconstructed metabolism turned out to be fully compatible with previous origin of life hypotheses proposed by these authors, assuming a direct link between the geochemistry of the life cradle and the physiology of LUCA (Martin and Russell 2003; Weiss et al. 2018). This work was criticized by other authors who identified several pitfalls in the datasets used to build the 338 phylogenetic trees (Gogarten and Deamer 2016; Berkemer and McGlynn 2020). Gogarten and Deamer noticed that many trees only included a small number of closely related groups of Archaea or Bacteria, indicating that a single HGT from one domain to the other could have been sufficient to fulfill the criterion of “presence in at least two groups” (false positive). They also noticed several “false negatives,” i.e., well-known universal proteins such as the A- and F-type ATPases catalytic subunits and many amino-acyl tRNA synthetases that were missing from the 338 proteins dataset. In their reply (Weiss et al. 2016b), Martin and colleagues did not discuss why some of their trees only included closely related Archaea, missing the diversity of the domain. They briefly suggest that some universal proteins were missing in their reconstituted LUCA because the monophyly of Archaea and Bacteria was blurred by HGT, which is clearly not the case for the 25 missing ribosomal proteins. Berkemer and McGlynn undertook a more detailed re-analysis of the 338 trees and noticed that many of them were undersampled in term of species, resulting in phylogenies that do not reflect the evolution of the corresponding proteins. They completed the species dataset for each protein and showed that phylogenies based on more sequences rejects the LUCA hypothesis for 82% of the 338 proteins identified as LUCA proteins!

Alerted by the discrepancy previously discussed between our result and those of Martin and colleagues concerning reverse gyrase, I have looked myself at the individual 338 trees (accessible in Weiss et al. 2018). They were retrieved and colored with the Archaeal branches in blue and the bacterial ones in red for easy interpretations (all colored trees are available at <https://osf.io/ypszh/> DOI: <https://doi.org/10.17605/OSF.IO/YPSZH>). As previously noticed by Berkemer and McGlynn, the distribution of Archaeal and Bacterial species within domains was often limited, with sometime only two closely related orders of the same phylum. This was problematic since a modern protein already present in LUCA should have been present in both the LBCA and in the LACA. Therefore, despite possible multiple losses during the diversification of these domains, they are expected to be present in several distantly related lineages in both domains. In most trees, this condition was not satisfied, making it impossible to conclude that this protein was present in LUCA or not. Notably, the number of species was dramatically different from one tree to another, from several hundred for the elongation factor EFTu/EF1 to only 9 for a methyltransferase of the FkBM

family! Moreover, the number of species analyzed was often very limited: 44 trees have fewer than 20 species and 15 have fewer than 10 species. Most trees are unbalanced in terms of domain composition with often many much more bacterial than archaeal species. Finally, the number of trees that exhibited a reasonably long branch compatible with a presence in LUCA was very limited (around 12–18 trees out of 336). These proteins correspond to previously recognized universal markers, such as 9 ribosomal proteins and the two elongation factors EF1/Tu and EF2/G. In all other trees, the branch between Archaea and Bacteria was very short. Considering the poor sampling of many proteins and using the branch length criterion, I also conclude that only about 80% of proteins attributed to LUCA by Martin and colleagues were a false positive, in agreement with the result of Berkemer and McGyll. This implies that many of the 366 proteins retrieved by Coleman and colleagues from the comparative analysis are also most likely false positive since they included 117 from the 338 proteins of the Martin and colleague’s dataset.

I was also surprised that Martin and colleagues only recovered in their analysis 20 out of the 50–60 proteins of the universal protein set defined by strict criteria. For instance, beside the examples already noticed by Gogarten and Deamer they missed the two large subunits of the DNA-dependent RNA polymerase, and 25 of the 34 universal ribosomal proteins. A probable explanation is the use of a very strict threshold (25% identity) in the first step of their analysis to recover proteins present in both Archaea and Bacteria. This threshold most likely counter selected *bona fide* LUCA proteins that diverged after LUCA to produce distinct versions (*sensu* Woese) of archaeal and bacterial proteins. On the contrary, this threshold probably enriched their dataset of proteins that have been transferred between the two domains.

Many studies that were performed during the first two decades of this century are now somewhat outdated considering the huge expansion of genomic databases that occurred in the recent years. These studies need to be updated, especially considering that the diversity within each domain has exploded with the expansion of metagenomic analyses. The sporadic distribution of a protein in the three domains testifies for its presence in LUCA only if its phylogeny fits with the topology of the Tol and if the branch between the Archaea and Bacteria is reasonably long, as previously discussed for reverse gyrase. A good example might be the recent analysis of proteins involved in the mechanism of Fe–S cluster assembly (Garcia et al. 2022). In modern ribocells, hundreds of proteins depend on the presence of Fe–S clusters for redox chemistry and Lewis acid-type catalysis. Fe–S clusters dependent proteins were thus probably already present in LUCA. Barras and colleagues propose that two mechanisms for Fe–S clusters assembly could be traced back to LUCA (Garcia et al. 2022). They published phylogenetic

trees of cysteine desulfurase MisS + MisU and SmsCB, in which Archaea and Bacteria are indeed separated by reasonable long branches.

In conclusion, the definition of the LUCA proteome remains to be robustly predicted and the physiology of LUCA remains unknown. It will be important in future to resume this type of work using a broad and update representation of species covering the diversity of each domain and a lower threshold to select proteins common to Archaea and Bacteria. A criterion to define this threshold should be its ability to recover all proteins already known to be present in the strict universal set. The species dataset for each domain should have to exclude fast evolving species, such as DPANN archaea and CPR bacteria, which are known to introduce bias in phylogenetic analyses (see below). It should be also very useful to know where are located the roots of the trees for each of the three domains—something still controversial—to determine with more accuracy if a protein was present in the LBCA, the LACA, and/or the LECA.

The Energetics of LUCA

It is often claimed in the literature that LUCA contained an ATP synthase, because archaeal A-type ATP synthase and bacterial F-type ATP synthase are homologous (Lane et al. 2010, Ducluzeau et al. 2014, see recent examples in Goldman et al. 2023, Mahendrarajah et al. 2023). Their catalytic and regulatory subunits and their membrane-anchored subunits are indeed homologous, but this is not the case for the central stalk which connects the cytoplasmic catalytic and regulatory subunits to the membrane-anchored subunits, as indicated by the presence of dissimilar structural folds (Mulkiđjanian et al. 2007, 2009). This observation is critical since the central stalk is essential for the rotary mechanism responsible for the ATP synthase activity. Mulkiđjanian and colleagues thus suggested that the ancestor of the A- and F-type ATP synthases in LUCA had no ATP synthase activity but functions as an ATP-dependent protein translocase, in which the translocated protein itself occupied the place of the central stalk. This hypothesis implies that LUCA probably only used fermentation pathways for ATP production.

Spang and colleagues suggested that an archaeal-like A-type ATP synthase could have been present in LUCA, because many bacterial genomes also encode this enzyme, suggesting that this protein was possibly already present in the LBCA (Mahendrarajah et al. 2023). However, this proposal is not supported by the phylogenies of their A and B subunits, where the bacterial A-type ATP synthases are dispersed into several clusters, and most of them are close or branch within their archaeal homologues, suggesting that an A-type ATP synthases was probably not present in LUCA.

It seems more likely that these enzymes originated in proto-Archaea and that some of them were later transferred from Archaea to Bacteria shortly before or after the emergence of the LCBA.

Notably, the ATP synthase activity is not essential for life, even at the “prokaryotic” stage. It has been known for a long time that a bacterium cultivated in conditions that inhibit the activity of the F-ATPase is viable, using fermentative pathways for ATP production (Harold and Van Brunt 1977). This is the lifestyle of Eukarya lacking mitochondria, since the eukaryotic orthologue of the archaeal ATP synthase, the V-type ATPase, functions as an ATPase. If LUCA indeed lack an ATP synthase powered by a rotary mechanism, the ATP production mechanisms of LUCA could have been reminiscent of those of proto-Eukarya before the emergence of mitochondria. The two independent inventions of the rotatory mechanism associated to the A and F-type ATP synthases were probably critical events that take place in the lineages of proto-Archaea and proto-Bacteria, providing a dramatic selective advantage to the first proto-archaeon and proto-bacterium with an ATP synthase.

Notably, ATP synthases in modern organisms are supported by a variety of electron transport chains involving many components. It has been suggested that some ancestors of these components were already present in LUCA (reviewed in Ducluzeau et al. 2014; Goldman et al. 2023). However, there is no clear phylogenetic evidence in the literature to support this claim. LUCA is often described in the scientific literature as an autotroph a priori, because it fits well with the hypothesis of an autotrophic origin of life. These autotrophic scenarios are proposed in opposition to “primitive soup” scenarios, in which the first organisms feed on carbon-rich chemicals that first accumulate in their primitive setting from non-biological pathways. An autotrophic LUCA supposes that the modern biological mechanisms of carbon fixation, such as the reductive tricarboxylic acid (TCA) cycle and/or the reductive acetyl CoA pathway, were already present in LUCA. This was first refuted by Pereto and colleagues who analyzed the phylogenies of the two main enzymes involved in these two pathways, the citryl-CoA synthase and citryl-CoA lyase and the CO dehydrogenase/acetyl CoA synthase, and concluded that the genes encoding these enzymes have been frequently affected by HGT and were probably absent in LUCA (Becerra et al. 2007a, 2007b). In another study, based on a much larger number of sequences, Gribaldo and colleagues also identified many HGT in the evolution of the CO dehydrogenase/acetyl CoA synthase CODH/ACS between and within domains (Adam et al. 2018). They nevertheless suggested that this enzyme was present in LUCA because, once considering these HGT, they concluded that this enzyme was probably present in the LACA and LBCA. Unfortunately, they did not consider the branch length between Archaea and

Bacteria in their analysis. Accordingly, one cannot exclude early transfers between proto-Archaea and proto-Bacteria. In any case, Gribaldo and colleagues concluded that the presence of this enzyme in LUCA cannot be an argument in favor of an autotrophic LUCA, since the ancestral CO dehydrogenase/acetyl CoA synthase “*might have been originally unable to fix carbon and operate only catabolically, consistent with a heterotrophic LUCA*” (Adam et al. 2018). Finally, it is important to remember that a heterotrophic LUCA can be reconciled with an autotrophic origin of life (as a mesophilic LUCA can be reconcile with a hot origin of life) considering the large evolutionary distance between the first cell and LUCA. One cannot exclude that both heterotrophs and autotrophs were thriving on or planet at the time of LUCA.

The Membrane of LUCA

Very different types of enzymatically synthesized phospholipids with different chemistries and stereochemistry probably originated before LUCA. The archaeal and bacterial/eukaryal types of phospholipids are only those that were present in the membranes of the successful ancestors of modern ribocells. There is no consensus today on the nature of the lipids present in LUCA. We do not know if they resembled those of Archaea, Bacteria, or a mixture of the two. Authors who have carried out phylogenetic analyzes of the enzymes involved in phospholipid biosynthesis have reached opposite conclusions (Lombard et al. 2012; Yokobori et al. 2016; Coleman et al. 2019). The history of these enzymes has seen numerous HGT between the three domains and certain enzymes involved in lipid biosynthesis in Archaea are present in many bacteria where they seem involved in other mechanisms and vice versa. The phylogenies obtained are therefore difficult to interpret. Interestingly, in discussing the evolution of primordial membranes and membrane proteins, Koonin and co-workers suggested that the membrane of LUCA and its early descendants might have been in fact more permeable to protons than modern ribocells but already impermeable to sodium, explaining why, according to their scenario, the ancestors of the ATP synthase used sodium and not proton gradients to sustain ATP synthesis (Mulkidjanian et al. 2009). This suggests that phospholipids in LUCA membrane could have somehow differed from modern ones.

Several authors have suggested that LUCA contained both types of lipids found in modern organisms and that the loss of one of them triggered the divergence between the archaeal and bacterial lineage, because membranes containing a single type of lipids should have been more stable (Wächtershäuser 2003; Koga et al. 1998). This does not seem to be the case since heterochiral hybrid liposomes made of bacterial and archaeal polar lipids are no less stable

than homochiral liposomes (Shimada and Yamagishi 2011). Indeed, an engineered *E. coli* with 20–30% of archaeal lipids grows as well as the wild type and is even slightly more resistant to stress (Caforio et al. 2018). If LUCA had only archaeal-type lipids, it is therefore unclear which type of selection pressure could explain the replacement of the more stress-resistant archaeal type by the less resistant bacterial type in the bacterial lineage (Forterre et al. 2019). In contrast, if LUCA had bacterial-type lipids or similar ones, archaeal lipids might have been selected during the adaptation of the proto-archaeal lineage to high temperature (Glansdorff et al. 2008; Groussin and Gouy 2011). Indeed, membranes made of archaeal phospholipids are more stable to heat exposure and much less permeable to protons and ions than those made of bacterial or eukaryal phospholipids (Choquet et al. 1996; Konings et al. 2002). This property is especially important at high temperatures, when lipid membranes became more permeable to protons and small inorganic ions. The failure to prevent their passive diffusion would abolish the production of ATP via the ATP synthase. This would ultimately lead to cell death at high temperature if fermentative pathways for ATP production are not sufficient to counteract the effect of high temperature on macromolecule stability and integrity. Notably, all known hyperthermophiles harbor an ATP synthase activity. It would be interesting to test if they can live without this enzyme, as demonstrated in the of mesophilic bacteria (Harold and Van Brunt 1977) or if they absolutely require an active ATP synthase. In the second case, the independent acquisition of an ATP synthase activity in the proto-lineages of Archaea and Bacteria could have been selected during the process of their adaptation to high-temperature biotopes.

The cytoplasmic membranes of modern ribocells are surrounded by various types of cell envelopes. Most archaeal and eukaryotic membrane surfaces and those of some bacteria are covered by glycoproteins forming the so-called S-layer in Archaea and Bacteria and glycocalyx in Eukarya. Lombard suggested the presence of a S-layer-like envelope in LUCA, because it probably already harbored the Z-IPTase (Lombard 2016), one of the most characteristic enzyme involved in the synthesis of precursors of the glycosylation pathways in the three domains. Examination of the unrooted tree of the Z-IPTase phylogeny (Fig. 1 in Lombard 2016) supports this claim if one removes a group of eukaryotic Z-IPTases that branches between Bacteria and other Arcarya and if one roots the tree in the bacterial branch. This produces a 3D tree with a rather long branch between Bacteria and Arcarya. It would be important now to update the phylogenies of this enzymes and of others possibly involved in glycosylation pathways. In addition to envelopes made of glycoproteins, the cells of nearly all Bacteria (with or without S-layers), a few Archaea, and a few Eukarya are surrounded by rigid cell walls (such as the peptidoglycan

layer in Bacteria) that strengthen their stability and probably protect them against attack by some viral lineages. These cell walls are made of non-homologous components in the three domains, suggesting that LUCA was probably devoid of cell wall. Nevertheless, one cannot completely exclude that LUCA harbored a cell wall—of a forgotten type—that was lost thereafter, since cell walls have been lost many times independently in the three domains.

Independently of their lipid types, all modern ribocells, including those with cell walls, have the property to produce membrane bound extracellular vesicles, (EVs), suggesting that this property was probably already present in LUCA (Gill et al. 2019). Among proteins found in EVs, interesting candidates to the title of universal proteins are members of the SPFH (stomatin, prohibitin, flotillin, and HflK/C) superfamily (Tavernakis et al. 1999, Hinderhofer et al. 2009; Marguet et al. 2013; Yokoyama and Matsui 2020). These proteins, that are known to facilitate membrane curvature and cell fusion (Browman et al. 2007) have been detected in both archaeal and eukaryal EVs (Salzer et al. 2008, Ellen et al. 2009, Gaudin et al. 2013, Skryabin et al. 2021 and references therein). However, stomatin and related proteins are small with multiple paralogs, especially in Eukarya and their presence in LUCA is difficult to ascertain. A small GTPase has been recently involved in the production of EV by some Archaea (Mills et al. 2014). Although orthologues of the protein detected seem to be restricted to some groups of Archaea, multiple families of small GTPases are present in the three domains and this family of proteins (also related to elongation factors) was most likely already present in LUCA and could have been also involved in EV production at that time. Preliminary studies suggest in fact that various mechanisms of EV production coexist in the modern biosphere, even in members of the same domain (Gill et al. 2019; Liu et al. 2021a, b, c) and it will be difficult to identify possible mechanism(s) already present in LUCA.

The possibility that LUCA and/or its contemporaries had internal membrane systems is rarely discussed. It could seem strange to recall this possibility when favoring the RNA-LUCA hypothesis. However, many RNA viruses can manipulate the endoplasmic reticulum (ER) to produce internal membrane structures, such as viral factories (reviewed in Deb Boon et al. 2010, Stelitano et al. 2023). The formation of nuclear-like compartment occurred at least three times independently by convergent evolution in the history of life, once in proto-Eukarya, once in some PVC Bacteria (a clade grouping Planctomycetes, Verrucomicrobiales, and Chlamydia), and once in “jumbo bacteriophages” of the class *Caudoviricetes* (Fuerst 2013, Riva-Marin and Devos 2018, Niewegłowska 2023). In Eukarya, a closed nuclear compartment is formed by the invagination of the endoplasmic reticulum and covered by a specific nuclear envelope (lamina); in some PVC Bacteria, an open nuclear compartment is

produced by the invagination of the cytoplasmic membrane, whereas in giant head and tailed *Caudoviricetes*, the viral nucleus is enclosed by a viral encoded membrane protein. One thus cannot dismiss that possibility that LUCA was a synaryote, a nucleated organism (Forterre 1992a, b; Forterre and Gribaldo 2010; Staley and Fuerst 2017; Niewegłowska et al. 2023).

The synkaryotic LUCA hypothesis was supported once by the discovery in PVC bacteria of proteins with predicted secondary structures and domain arrangement resembling those typical of eukaryal membrane coat proteins (Santarella-Mellwig et al. 2010; Forterre and Gribaldo 2010). These proteins are formed by a combination of beta propeller domains followed by a stacked pair of alpha helices. One of these proteins co-localizes with intracellular membrane vesicles present in one of the two PVC cellular compartments. Phylogenetic analyses have suggested that LECA and the last common ancestor of PVC bacteria both contained already four divergent versions of proteins structurally analogous to modern coat proteins (Santarella-Mellwig et al. 2010). However, a recent updated analysis failed to recover similar proteins in Archaea and only found a few of them in Bacteria outside the PVC superphylum (Ferrelli et al. 2023). It seems thus likely that the resemblance between the bacterial and eukaryal membrane coat proteins reflects convergent evolution or HGT between Bacteria and proto-Eukarya.

A more likely hypothesis for the origin of the nucleus is that this unique organelle originated via the interaction between proto-Eukarya and the viral factories of diverse giant viruses from the phylum *nucleocytoviricota*. This scenario, first proposed at the beginning of this century (Bell 2001; Takemura 2001) has been recently boosted by the discovery that viruses can produce nucleus and nuclear pores and by in-depth phylogenetic analyses of several critical eukaryal proteins, such as actin, RNA polymerase, and TopIIA DNA topoisomerases (Guglielmini et al. 2019, 2022, Da Cunha et al. 2022b, review in Gaia and Forterre 2023).

The Controversial Relationships Between LUCA and Eukarya

If Eukarya emerged within Archaea, as in 2D scenarios, eukarya-specific proteins or proteins only present in Eukarya and Bacteria cannot be traced to the proteome of LUCA. This explains why many authors now only consider Archaea and Bacteria when they try to reconstruct the portrait of LUCA. In contrast, if Archaea and Eukarya are sister group, as in the 3D scenario, some of these proteins might have been present in LUCA and later lost in proto-Archaea. Others might have been even lost in both Bacteria and Archaea (Fig. 4). It is thus important to know which model is correct when discussing the nature of LUCA. Opinions in favor of

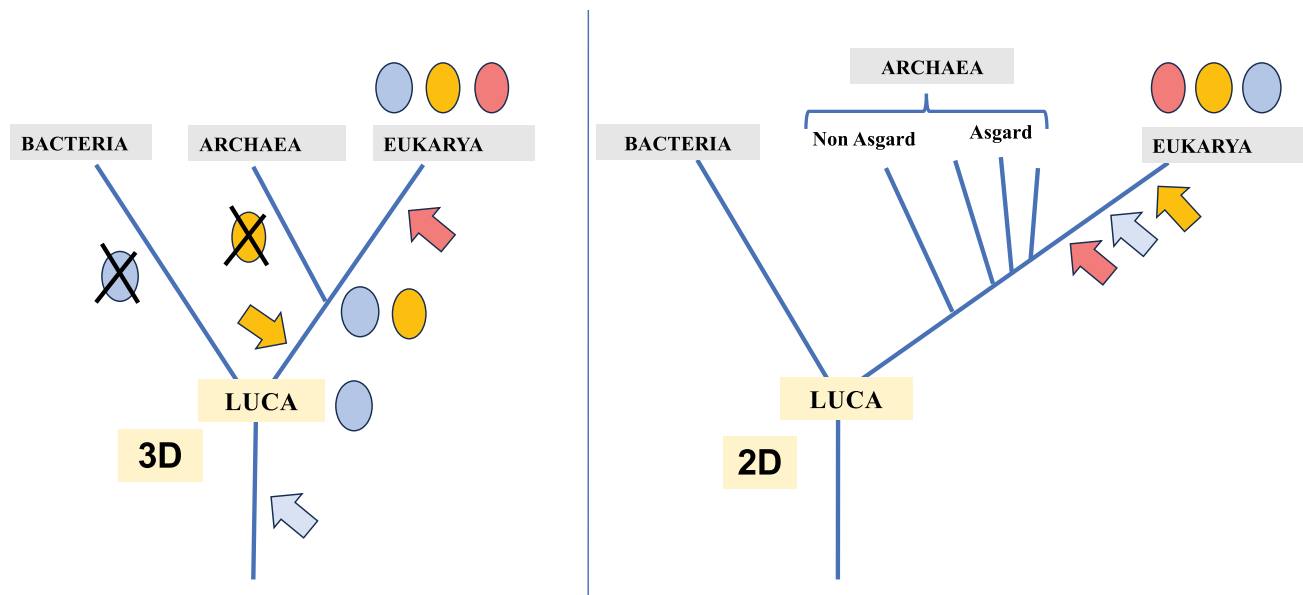


Fig. 4 Possible origins of eukarya-specific proteins (ESPs) depending of the 3D or 2D scenarios. In 3D scenario some of ESPs were possibly present in LUCA (blue circle) and/or in LARCA (blue and yellow circles) and later lost in proto-Bacteria and/or in proto-Archaea, other originated in proto-Eukarya (red circles). In 2D scenarios, all ESPs

originated in proto-Eukarya. ESP here should not be confused with eukaryotic-signature proteins (also currently named ESP) that are not specific to Eukarya since they are also present in some Archaea (Color figure online)

the 2D model have been strongly boosted by the discovery of Asgard Archaea from metagenomic analyses (thereafter called Asgard for simplicity) (Spang et al. 2015; Zaremba-Niedzwiedzka et al. 2017). During the last ten years, the number of Asgard lineages has exploded with around 20 distinct lineages now recognized, covering a huge number of MAGs (Metagenomes Associated Genome) (Liu et al. 2021a, b, c; Da Cunha et al. 2022a, b; Eme et al. 2023). In recent published phylogenetic analyses based on concatenation of different subsets of universal protein sequences, Eukarya branch either as a sister group to Asgard or, more frequently, as sister group to one of the many Asgard lineages presently known (Liu et al. 2021a, b, c; Xie et al. 2021).

The Asgard are now systematically introduced in the scientific literature as “*the closest prokaryotic relatives of eukaryotes.*” This specific relationships between Eukarya and Asgard was first observed in a phylogenetic analysis based on the concatenation of 36 universal proteins (Spang et al. 2015). However, in-depth examination of the 36 individual trees revealed that these close relationships resulted from a combination of several biases in the species and protein datasets (Da Cunha et al. 2017; Gaia et al. 2018; Nasir et al. 2021). The 2D trees were favored by the presence of fast evolving species, such as DPANN, *Methanopyrus kandleri*, or Korarchaeota, in the species dataset and of small proteins in the universal protein dataset. In another re-analysis, it was shown that 2D trees were favored by unbalanced species datasets in which Archaea

are overrepresented compared to Bacteria and Eukarya and that several proteins sequences used as phylogenetic markers were possibly misaligned (Nasir et al. 2016, 2021). These biases have been present in all studies published during the last nine years, even though different authors used different subsets of universal proteins and species (review in Da Cunha et al. 2022a, b, see also Caetano-Anollés and Mughal 2021). The recovery of the 2D topology in all these analyses can be explained by the shortness of the branch testifying for the monophyly of Archaea in 3D trees, especially compared to the long branch of Bacteria. The long bacterial branch tends to attract fast evolving Archaea, whereas the signal corresponding to the short archaeal branch is usually missing in short proteins. This hypothesis is in accordance with recent analyses, based on simulation experiments, that have shown that oversampling of some groups or removing fast evolving positions in the alignment prevents recovery of short internal bipartitions (Hernandez and Ryan 2021; Rangel and Fournier 2023). These results can also explain why studies using an oversampling of archaeal sequences and/or methods that remove fast evolving position failed to recover the 3D tree.

A recurrent argument used to support the close relationships between Asgard and Eukarya is the presence in Asgard of eukaryal-like proteins that are not present in other archaeal lineages, such as actin, tubulin, and many others. However, these so-called eukaryal-signature proteins (ESPs) exhibit a very patchy distribution between

the different Asgard lineages, which is difficult to explain, except if they testify for ancient HGT between Archaea and proto-Eukarya (Da Cunha et al. 2022a). In the case of actin, an exhaustive phylogenetic analysis published together with the discovery of actin in giant viruses revealed that various clades of Asgard actins originated during the diversification of proto-Eukarya, together with the various clades of eukaryal actin-related proteins, (ARPs) (Da Cunha et al. 2022b). The topology of this actin tree refutes the idea that eukaryal actin originated from Asgard ones and is better explained by HGT between Archaea and proto-Eukarya. The same situation is observed in the case of tubulin, except that Asgard tubulin is only present in one of the many Asgard lineages, the Odinararchaea. This Asgard tubulin branches within the clades of eukaryal tubulin paralogues, as sister group to α and β tubulins, suggesting again transfer from proto-Eukarya to Asgard (Rodrigues-Oliveira et al. 2023). Notably, transfer of proto-eukaryal actin and tubulin to some Bacteria has been previously well documented (Schlieper et al. 2005, Guljamow et al. 2007, Martin-Galliano et al. 2011, Shiratori et al. 2019). In-depth analysis of other ESPs remains to be done to test if the HGT hypothesis can be generalized. Unfortunately, the analyses ESPs are now systematically interpreted by most authors in the framework of the 2D scenario, without considering the alternative HGT hypothesis.

Interestingly, it seems that HGT between Asgard and proto-Eukarya can not only explain the patchy distribution of ESP, but also some odd observations that we made in re-analyzing the 36 single trees of the first publication describing the discovery of Asgards; whereas in some trees, Eukarya and the three Asgards known at that time (two Lokiarchaea and one Hodarchaeon, formerly Loki 3) branched far from each other; in other trees, one, two, or all three Asgards are sister group to Eukarya (Da Cunha et al. 2017). We noticed that only one of the 36 trees, corresponding to the EF2/G elongation factor, exhibited the same topology as the concatenated tree, with the monophyly of the three Asgards and the sisterhood of Hodarchaea and Eukarya. Remarkably, removing the EF2/G from the Hodarchaeon (formerly Loki 3) was sufficient to break the sisterhood between Asgard and Eukarya (Fig. 6 in Da Cunha et al. 2017). We initially suggested that the remarkable mimicry between the EF2/G tree and the 36 proteins tree could be due to the contamination of the Asgard MAGs, especially the MAG of the Hodarchaeon, by eukaryal sequences. In favor of this hypothesis, we noticed the presence in the sequence of the Hodarchaeal EF2/G factor of specific insertions typical of Eukarya that were missing in other Archaea, including the two other Asgards (formerly Loki 1 and 2). We think now that the mimicry between the EF2/G tree and the 36 proteins tree do not testify for contamination but most likely for HGT between proto-Eukarya and Asgard. This is supported by phylogenetic analyses in which Hodarchaea branch as

sister group to Eukaryotes, whereas all other Asgard branch between diverse Archaeal clades, far from Eukarya (Narrowe et al. 2018, Cunha et al. 2017, Eme et al. 2023) (Fig. 5).

The EF2/G case does not seem to be an isolated one. We have identified another striking example in looking at the tree of the universal protein Kae1/TsaD published by Li and colleagues (Liu et al. 2021a). In this tree, two of the 11 Asgard lineages, the Kariarchaea and Heimdallarchaea, are sister group to Eukarya, whereas the other 9 Asgard lineages branch between other Archaeal clades. This strongly suggests that the ancestral Kae1/TsaD present in a common ancestor of Kariarchaea and Heimdallarchaea was displaced by the Kae1/TsaD from a proto-Eukarya (Da Cunha et al. 2022a) (Fig. 5). Removal of these two lineages transforms the 2D tree obtained by Li and colleagues into a 3D tree (Da Cunha et al. 2022a). Examination of the 131 individual trees of the archaeal and bacterial proteins recently concatenated by Ettema and colleagues to build a tree in which Eukarya are sister group to Hodarchaea (Eme et al. 2023) reveal more cases of possibly HGT, in both directions, between proto-Eukarya and Asgard (unpublished observations). Notably, Hodarchaea are only sister group to Eukarya in about 10% of the trees, whereas in most other ones, they branch most of the time very distant from Eukarya.

The concatenation of the two large subunits of the RNA polymerase, the two largest universal proteins, produced a 3D tree, using a balanced dataset of 50 species for each domain in which fast evolving species have been removed (Da Cunha et al. 2017). In this Bayesian tree with a non-homogeneous model, Asgard are located deep into the archaeal tree. In this analysis, we used the nuclear RNA polymerase II for Eukarya, but we obtained later again the 3D topology after addition of the eukaryal RNA polymerases I and II (Da Cunha et al. 2022a), as well as viral RNA polymerases (Guglielmini et al., 2019). The long eukaryotic branch was shortened by these additions, limiting the possibility that the 3D topology obtained was due to an attraction between Eukarya and Bacteria. Interestingly, Martinez-Gutierrez and Aylward have shown that the two large RNA polymerases subunits are the best proteins to recover a correct phylogenetic signal out of 41 proteins conserved in Archaea and Bacteria (Martinez-Gutierrez and Aylward 2021). Embley and colleagues also recover the 3D topology of the RNA polymerase tree using our dataset (Supplementary Figure 6 in Williams et al. 2020). To obtain a 2D tree, they had to perform amino acid recoding, reducing the number of amino acids to four. However, Asgard were still far from Eukarya in this 2D tree, with Eukarya becoming sister group to Crenarchaea. Since amino acid recoding reduces the phylogenetic signal (Hernandez and Ryan 2021) it is likely that this strategy prevents the recovery of the specific archaeal branch of the 3D tree.

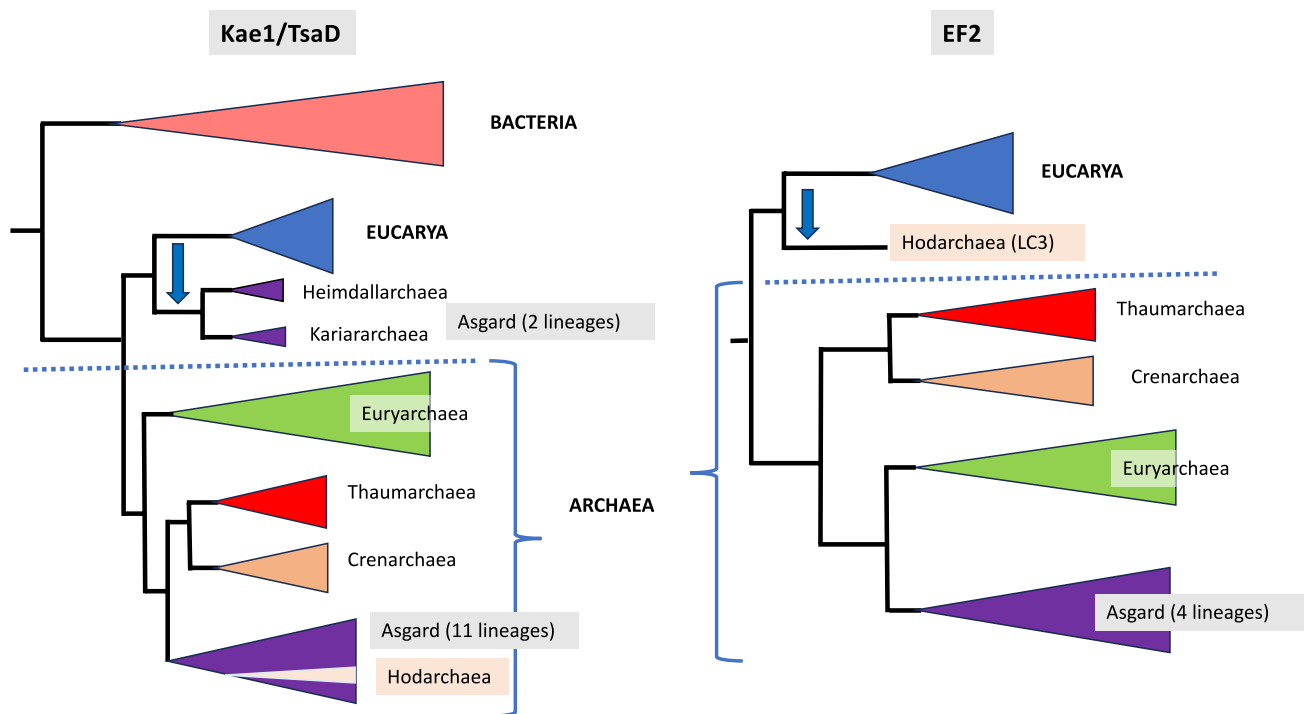


Fig. 5 Two examples of putative horizontal gene transfer of universal proteins between Asgard and proto-Archaea. In the Kae1/TsaD tree (left panel) two Asgard lineages are sister group to Eukarya, whereas the 11 other Asgard lineages (including Hodarchaea) branch between Euryarchaea and a clade grouping Crenarchaea and Thaumarchaea (adapted from the Fig. 6 in Da Cunha et al. 2022a, b). In the EF2 tree, Hodarchaea branch as sister group to Eukarya, whereas the other Asgard lineages branch again between large archaeal clades (adapted from the Fig. 1 in Cunha et al. 2017). Phylogenies of EF2/EF3 with a similar topology (Hodarchaea as sister group to Eukarya,

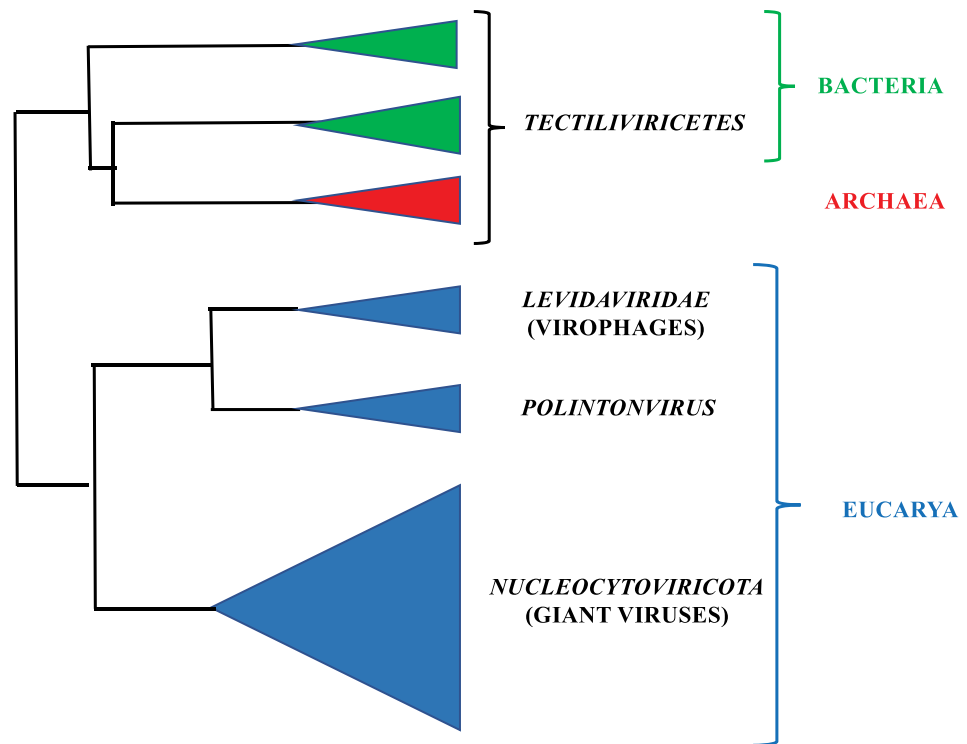
far from other Asgards) have been published by Ettema and colleagues, see Fig. 2 in Narrowe et al. 2018, and M037_b.blo30b3g07.bmge.treefile, in Eme et al. 2023). In the first two original trees, the Hodarchaeon was named Heimdallarchaea LC3. These phylogenies are best explained by horizontal gene transfers from proto-Archaea to some Asgard lineages (blue arrows). Other phylogenies suggesting gene transfers between proto-Archaea and some Asgard lineages in both directions can be found in the individual trees of Eme et al. 2023 (Color figure online)

All these observations strongly support the idea that the 3D topology is the correct one and that the strong eukaryal flavor of Asgard could be the result of several biases in phylogenetic analyses that support a 2D tree, combined with the probable co-evolution of Asgard and proto-Eukarya in similar environments, favoring HGT. Notably, the first two Asgard successfully cultivated (Imachi et al. 2020; Rodrigues-Oliveira et al. 2023) live in symbiotic relationships with other organisms (in that case archaeal methanogens). One could imagine that some ancient Asgard thrived as ectosymbionts of protists and that some modern ones possibly still live in symbiotic association with modern Eukarya (Da Cunha et al. 2022a).

If the 3D topology is correct, how can we determine if some traits common to Bacteria and Eukarya or specific to Eukarya were present in LUCA? The eukarya-specific branch of the Tol being much smaller than the branch linking Eukarya to Bacteria via LUCA, one can argue that the presence of a long branch between Bacteria and Eukarya in a phylogenetic tree could be a good criteria to distinguish traits that were present in LUCA from those introduced in

Eukarya by the bacterium at the origin of mitochondria or from other ancient bacteria that colonized some proto-Eukarya. The evolution of the tubulin superfamily possibly provides such an example: if the archaeal FtsZ/CetZ and artubulin originated from HGT via Bacteria and proto-Eukarya, respectively, one can imagine that the ancestor of bacterial FtsZ and eukaryal tubulin was present in LUCA but lost in proto-Archaea. In the case of eukarya-specific traits, there is of course no possibility to deduce their presence in LUCA from phylogenomic analysis. It is often assumed that these traits were acquired in proto-Eukarya, because evolution is supposed to go from simple to complex. This is a prejudice, since it is well known that evolution runs in both directions, from simple to complex and back again. Unicellular organisms are a priori simpler than multicellular organisms; however, we know that unicellular yeasts evolved several times independently from multicellular fungi (Dujon 2010). Eukaryotic-specific traits are often considered to be derived simply because Eukarya are still considered by most molecular and cell biologists to be “higher” organisms, although evolutionists are (usually) aware that there

Fig. 6 Phylogeny of DNA viruses from the phylum *Bamfordvirae*. This schematic unrooted phylogeny is adapted from Fig. 3 of Woo et al. 2021. The topology of this “universal tree of viruses” is the opposite of the topology of the universal tree of ribocells



is no such thing as “lower” or “higher” organisms in the real world. The specific features of Eukarya are so complex that it is also often assumed that they cannot be lost during evolution. This is misleading since, for instance, fission yeast cells can undergo nuclear division in the absence of spindle microtubules (Castagnetti et al. 2010) and once *bona fide* eukaryotic genomes full of introns can lose all of them as well as genes encoding spliceosomal components (Lane et al. 2007). If eukaryal features can be lost in modern eukaryal lineages, one can imagine that some eukarya-specific features were present in LUCA and later lost in the proto-archaeal and proto-bacterial lineages. This possibility is especially appealing if Archaea and Bacteria indeed originated by reductive evolution (Forterre 1992a, 1995, 2013a, b; Penny and Poole 1999; Kurland et al. 2006, 2007; Glansdorff et al. 2008). Such reductive evolution can be explained by the thermoreduction hypothesis previously discussed (Forterre 1995) or by the raptor hypothesis (Kurland et al. 2006). In the later, the streamlining in Archaea and Bacteria resulted from an adaptation to rapid growth and/or minimal resources to escape predation by phagotrophic proto-Eukarya (Kurland et al. 2006). Both hypotheses can be combined if Archaea and Bacteria evolved toward the “prokaryotic phenotype” by adapting to extremely hot environments to avoid proto-eukaryal predators, since the upper temperature limit of life for Eukarya is around 60 °C.

Eukarya-specific features that can be tentatively linked to the second age of the RNA world are good candidates to be ancient features already present in LUCA (Jeffares et al.

1998; Penny and Poole 1999; Collins et al. 2009; Forterre 2013a, b). This is possibly the case of some lineages of simple RNA viruses that are only present in Eukarya (see the following chapter). It might be significant that retroviruses and more generally retroelements that could be witnesses of the RNA to DNA transition are either specific (retroviruses) or very abundant in Eukarya. Another eukarya-specific feature worth discussing is the spliceosome, a ribozyme even more complex than the ribosome, with a huge number of proteins and five RNA molecules (Jeffares et al. 1998; Collins and Penny 2005; Roy and Gilbert 2006). Spliceosomes might have been wonderful devices in the early RNA–protein cells to create larger proteins by combining small RNA genes (ancestors of exons) to produce bigger ones (Doolittle 1978; Reaney 1984). Importantly, the discovery of nucleomorphs (residual eukaryotic nuclei present in secondary endosymbionts) whose genomes have lost all introns and all genes encoding spliceosome components (Lane et al. 2007) has now made credible the possibility that LUCA harbored a primitive spliceosome. According to this scenario, ancestral split genes in LUCAs were later retro-transcribed to produce non-split genes in Archaea and Bacteria. This could have occurred in the framework of the viral origin of DNA, since this hypothesis involves a retro-transcription step (Forterre 2005, 2006). The spliceosome would be a relic of times when, besides the evolving ribosomes, multiple and diverse types of spliceosomes contributed to the diversification of proteins in the second age of the RNA world.

This “early spliceosome” view that was popular for a while has now been abandoned by most scientists, since it is not compatible with the 2D scenario. Since the spliceosome share the same splicing mechanism as type II introns (Cech 1986), it is now currently assumed that the spliceosome originated from bacterial group II introns present in one of the bacteria at the origin of Eukarya (Martin and Koonin 2006, see Poole 2006 for an early criticism). This hypothesis seems difficult to reconciled with the fact that both the major and minor spliceosome were already present and fully evolved in LECA (Collins and Penny 2005; Hoepfner et al. 2012), the genome of which was full of introns (Csuros et al. 2011). This means that a simple group II intron (a single RNA molecule) was transformed into two highly sophisticated molecular machines in the timespan between LARCA and LECA (Rogozin et al. 2012).

One can conclude from this chapter that the wide support for 2D scenarios, limiting the discussion about LUCA to the comparison between Archaea and Bacteria, misleads us into putting Eukarya—especially the eukaryotic RNA world—out of the picture. It brings us back to pre-Woesian time, when the prokaryote first paradigm was already dominant among molecular biologists. This led me to write that Carl Woese is still “*ahead of our time*” (Forterre 2022a). One can hope that when evolutionists will look more seriously at data in favor of the 3D scenario, it will be again possible to think about the portrait of LUCA with an open mind, free from the prokaryote first prejudice.

The Virome of LUCA

The billions upon billions of cells that predated LUCA certainly did not live in perfect harmony, but competed, killed each other, parasitized each other, ate each other. The world has always been full of predators and preys, as pointed out by Penny and colleagues: “*there was no garden of Eden*” (de Nooijer et al. 2009). The modern biosphere is dominated by the conflict between cells, viruses, and virus-derived elements, such as plasmids, transposons, and retrotransposons (Forterre and Prangishvili 2009a). This was probably already the case at the time of LUCA and much earlier. Jalasvuori and Bamford suggested that production of RNA-containing lipid vesicles by primitive cells and fusion of these RNA-filled vesicles with empty ones was the first mode of genome propagation (Jalasvuori and Bamford 2008). They proposed that modern viruses evolved via this mechanism. Indeed, at some point in the above scenario, one could imagine that lipid (or lipid/peptide) vesicles should have delivered their RNA into other vesicles already containing RNA, leading to competition between the different RNA genomes. After the invention of the ribosome, RNA viruses might have appeared in the second age of the cellular RNA world, when

proteins were used to stabilize and/or facilitate the fusion/interaction with the “host” of RNA-containing vesicles produced by RNA cells, leading to the emergence of the first (true) virions (Forterre and Prangishvili 2009b; Forterre and Krupovic 2013). LUCA and its contemporaries were thus certainly already infected by a variety of *bona fide* viruses producing protein-based virions. These viruses (first RNA viruses, later retroviral-like elements and finally DNA viruses and plasmids) then evolved by association/recombination with various RNA and DNA replicons, including plasmids, transposons, and evolutionarily unrelated viruses (Krupovic et al. 2019). Modern viruses can be defined as “capsid-encoding organisms” (Raoult and Forterre 2008), the smallest viruses encoding at least one protein that helps to protect and disseminate the viral genome (capsid or nucleocapsid) (Krupovic and Bamford 2010).

Several lineages of viruses originated first independently, as indicated by the non-homologous relationship between capsid proteins from different viral lineages (see below). The term “realm” was proposed recently by the ICTV (International Committee for the Taxonomy of Viruses, 2020) to name proposed monophyletic viral lineages defined by their major capsid proteins or replicative enzyme. All RNA viruses have been grouped with retroviruses in a single realm, the *Riboviria*, because they share a homologous RNA replicase, even though they exhibit a great diversity of capsid proteins. In the modern virosphere, RNA viruses infecting Eukarya (RNA eukaryoviruses) are especially abundant and diverse, whereas RNA viruses infecting Bacteria (RNA bacterioviruses) are less diverse and less abundant and RNA archaeoviruses are yet unknown (Nasir et al. 2014). Notably, several lineages of RNA viruses are only present in Eukarya (Wolf et al. 2018; Koonin et al. 2024).

In the framework of the 2D scenario, it has been recently suggested that Eukarya originated from a bacterium that engulfed an Asgardarchaeon and that all eukaryoviruses, including RNA ones, originated from viruses that infected this bacterium (Krupovic et al. 2023). This scenario seems unrealistic to me since it supposes that this bacterium and/or its early descendants (the first proto-Eukarya) were infected by ancestors of all lineages of eukaryoviruses. Moreover, it implies that RNA eukaryoviruses, especially the simplest ones, have no direct evolutionary link with ancestral RNA viruses that predated LUCA but originated much later from RNA bacterioviruses in the proto-eucaryotic lineage. It has been proposed indeed that all RNA eukaryoviruses evolved from bacterioviruses of the *Leviviridae* family, because this family branches at the base of the RNA replicase tree of *Riboviria* rooted with reverse transcriptases (Wolf et al. 2018; Koonin et al. 2024). However, this rooting is arbitrary and rooting the tree within eukaryotic *Riboviria* with ssRNA genomes make more sense to me since the transition from RNA to DNA genomes suggests that reverse transcriptases

derived from RNA replicases and not the other way around (Forterre and Gaia, 2021). In my opinion, the odd distribution of RNA viruses between the three domains is better explained in the framework of the 3D scenario. One can imagine that most ancestors of modern RNA viruses were already present at the time of LUCA and its contemporaries and that only a subset was able to co-evolve successfully with proto-Bacteria, whereas all RNA virus lineages disappeared during the adaptation of proto-Archaea to high-temperature biotopes. Considering the instability of RNA at high temperature, especially single-stranded RNA, the reduction and elimination of RNA viruses in the lineages leading to Bacteria and Archaea, respectively, was thus possibly related to the thermophilic and/or hyperthermophilic phenotypes of the LBCA and of the LACA (Forterre 1995).

It is generally supposed that LUCA and its contemporaries were already infected by DNA viruses because of the existence of evolutionarily related DNA viruses infecting members of different domains, the so-called cosmopolitan viruses (Bamford 2003; Krupovic et al. 2020). Two major lineages of cosmopolitan DNA viruses are presently known, corresponding to the phylum *Bamfordvirae* and the realm *Duplodnaviria*. DNA viruses from these two lineages utilize structurally unrelated major capsid proteins (MCPs) and packaging ATPases (pATPases) (Krupovic and Bamford 2010; Koonin et al. 2023). Although the universality of these two viral lineages suggests a priori that ancestors of these viruses already infected LUCA, this is not really supported by their evolutionary relationships between domains. Indeed, since viruses usually co-evolved with their hosts, one would have expected a closer resemblance between *Bamfordvirae* and *Duplodnaviria* infecting Archaea than between those infecting Archaea and Bacteria if the ancestors of these viruses already infected LUCA and/or its close relatives. Instead, one observes the opposite situation. In the case of *Bamfordvirae*, a phylogeny based on their MCP and pATPases produces a tree in which archaeoviruses branch within bacteriophages, and not as sister group to eukaryoviruses (Fig. 6) (Woo et al. 2021). Whereas, all archaeal and bacterial *Bamfordvirae* have small genomes and produce small virions, eukaryoviruses of this phylum include viruses producing virions with extremely different sizes, from small ones (Polinton-like viruses, *Lavidnaviria*) to huge ones (*Nucleocytoviricota*). Among *Duplodnaviria*, archaeal and bacterial viruses produce very similar head and tailed virions and their MCPs only exhibit the so-called HK90 fold platform, whereas in eukaryoviruses of the *Duplodnaviria* realm, this platform is decorated by “towers” of different sizes but homologous between *Herpesvirae* and *Mirusviricota* (Gaia et al. 2023). In several recent phylogenies of *Duplodnaviria*, archaeoviruses branch again within bacteriophages (Low et al. 2019; Liu et al. 2021b; Evseev et al. 2023). Strikingly, 47 of the 50 families of *Duplodnaviria*

approved by the ICTV in September 2022 contained both archaeal and bacterial members (Evseev et al. 2023).

The similarity between DNA viruses of the archaeal and bacterial virosphere compared to DNA viruses of the Eukaryal virosphere is difficult to explain in the framework of both the 2D and 3D scenarios if their ancestors were already present at the time of LUCA. In the 2D scenarios, one must suppose that *Bamfordvirae* and *Duplodnaviria* were already diversified before LUCA, to explain the branching of archaeal groups within a greater diversity of bacterial groups. One should then suppose that these viruses remained very similar (inside each group) during the evolution of proto-Archaea and proto-Bacteria and later on, during the diversification of Archaea and Bacteria. In opposition to this three billion years stasis, one should assume that *Bamfordvirae* and *Duplodnaviria* evolved very rapidly during eukaryogenesis to explain why they are so different from their relatives infecting Archaea or Bacteria today. One should argue that the dramatic evolution of DNA eukaryoviruses was due to their adaptation to the “eukaryotic phenotype” that can be seen as an ad hoc hypothesis.

In the framework of the 3D scenario, a tempting hypothesis is that DNA viruses only originated post-LUCA, i.e., during the diversification of the three domains, in agreement with the RNA-LUCA hypothesis. Notably, this would explain why many lineages of DNA viruses are specific to one domain. These domain-specific lineages are rare in the bacterial virosphere, possibly because the emergence of the peptidoglycan in proto-Bacteria prevented their infection by most viral lineages (Forterre and Prangishvili 2009a; Prangishvili 2013). On the contrary, many archaeoviruses are domain specific: one can mention the case of the realm *Adnaviria* that includes viruses packaging their DNA in the A form or viruses producing tailed or tail-less lemon-shaped virions (Krupovic et al. 2021; Wang et al. 2022). In Eukarya, most families of RNA viruses are eukaryotic specific as well as several lineages of DNA viruses, such as *Hepadnaviridae* whose genome is retro-transcribed during their life cycle, mimicking the transition from RNA to DNA genomes, or else *Baculoviridae*, that encode DNA-dependent RNA polymerases homologous but very divergent from those of ribocells and *Nucleocytoviricota*.

If DNA viruses emerged in the proto-lineages of the three domains, the high similarity between archaeal and bacterial cosmopolitan DNA viruses could be due to the exchange of these viruses and/or some of their genetic materials between these two domains and/or between proto-Archaea and proto-Bacteria. This would explain why archaeal *Varidnaviria* and *Duplodnaviria* branch within bacterial ones in “universal tree of viruses” (Woo et al. 2021, Evseev et al. 2023). The fact that Archaea and Bacteria have exchanged mobile genetic elements has been well documented in the case of conjugative plasmids (Catchpole et al. 2023). Notably, in

addition to cosmopolitan viruses, plasmids and other mobile genetic elements of the DNA world are very similar between Archaea and Bacteria. Their anti-DNA viral defense systems (CRISPR, restriction-modification systems) are also strikingly identical. In contrast, the DNA mobilome of Eukarya is characterized by very different families of transposons and IS elements that have no close relatives in Archaea and Bacteria and most of their anti-viral defense systems are specific and primarily directed against RNA viruses.

Interestingly, cosmopolitan DNA viruses are much more diverse in Eukarya than in the two other domains. This suggests that they first appeared in proto-Eukarya and that some of them were later transferred to proto-Bacteria and/or proto-Archaea before being exchanged between these two domains that share very similar lifestyles and prokaryotic phenotypes. This challenging hypothesis implies that the diverse lineages of cosmopolitan DNA eukaryoviruses and their parents infecting Archaea and Bacteria diverged during the evolution of proto-Eukarya. Notably, phylogenetic analyses have already shown that all present-day major lineages of *Duplodnaviria*, *Bamfordviria*, and other eukaryal *Varidnaviria* diverged before LECA (Guglielmini et al. 2019; Woo et al. 2021, Gaia et al. 2023). Much clearly remains to be done to fully understand the evolutionary trajectory of viral lineages in relation to the uTol.

Conclusion

Carl Woese wrote several times that the nature of LUCA was “one of the more interesting biological problems” (Woese 1983). The development of molecular biology and phylogenomic analyses has provided us with a wealth of information that can be tentatively used to solve this problem. However, the task remains challenging, and the portrait of LUCA is still controversial. Unfortunately, we will never have a time machine to check if our favorite hypothesis is correct and this portrait will remain fuzzy. One thing that we can take for granted from comparison and analysis of the three modern domains is that, even if LUCA did reach some level of complexity, it was very different from modern organisms, explaining why it had a much greater evolutionary potential.

A major problem in studying LUCA and early evolution seems to be an underestimation of the number of biological innovations that took place during the diversification of ribocells between LUCA and the ancestors of the three modern domains. Many scientists are reluctant to consider the elusive entities that populated these ancestral lineages that have now disappeared because, by definition, they will always remain unknown to us. This induces a preference for scenarios that only include a combination of modern organisms that we can fully describe. However, *Homo sapiens* is not born from an intercourse between a gorilla and a

chimpanzee, and the common ancestor of these three great Apes was none of them. Thankfully in that case, we know about individual proto-lineages from the fossil record, something that we unfortunately lack to draw the portrait of LUCA. The reluctance of considering extinct lineages has probably facilitated the acceptance of 2D scenario in which Eukarya seem to emerge directly from the association of modern species. This is an illusion since proto-Eukarya necessarily once thrived on our planet, even in the 2D scenarios.

It is now currently believed that the debate about the position of Eukarya in the uTol is closed and that the 2D model has been definitively validated by the discovery of Asgard. The Asgard origin of Eukarya is becoming a paradigm since it is now accepted as truth by nearly all biologists without considering the data that contradict this view. Unfortunately, the number of teams working on this topic remains very limited and their studies have all been affected by the same biases (Da Cunha et al. 2022a, b). I have argued here and elsewhere with my co-workers in favor of the classical Woese’s uTol that can be recovered when these various biases are taken into consideration (Da Cunha et al. 2017, 2022a, 2022b). The current 2D uTol paradigm can be viewed as a major bottleneck, preventing consideration of Eukarya when drawing the portrait of LUCA. It will be important that a new generation of scientists starts to consider that the debate about the topology of the uTol is not closed and attack this problem with an open mind, free from the “prokaryotic prejudice” (Forterre 2022b).

Finally, one can hope that some of the hypotheses about the nature of LUCA, even if they remain speculative, will provide food for experimentation by future generations of biologists. Studies of membranes of engineered cells with mixed archaeal and bacterial lipids are first steps in that direction. I have argued here in favor of a LUCA equipped with an RNA genome. This is a disputed opinion, and it has been regularly argued that rather complex cells with an RNA genome cannot exist. Hopefully, it will be possible in future to synthesize artificial cells with large RNA genomes (or multiple small ones) to test experimentally the “viability” of such RNA cells. Another exciting avenue would be the reproduction of the hypothetical LUCA ribosome, containing the 34 universal ribosomal proteins and their associated RNA to test the translation fidelity and the viability of such a reduced ribosome. It would be also worth experimentally reconstructing the RNA polymerase of LUCA to test its ability to faithfully replicate RNA molecules. More studies on ATP synthase are urgently needed to understand the transition between their ATPases and ATP synthase activities; it should be especially important to determine in what direction this transition occurred in the proto-eukaryal lineage.

Speculations about the origin of life has provided many incentives to initiate experimental work leading to the creation of new scientific fields, such as prebiotic chemistry, with

practical implications for chemistry in general. Similarly, one can hope that speculations about the portrait of LUCA will provide more incentives for experimental work that could tell us more about the history of the major molecular mechanisms still operating in modern ribocells.

In a recent paper, Donoghue and colleagues found that reverse gyrase was present in LUCA, and conclude that LUCA was similar to modern procaryotes (Moody et al. 2024). However, their analysis did not take into account differences in the molecular biology of Archaea and Bacteria, the branch lengths in protein trees and the higher evolutionary tempo at the time of LUCA.

Acknowledgements I am grateful to Arturo Becerra for the invitation to publish in this special issue and Anthony Poole who made me aware long ago that rooting the tree of life in the bacterial branch was not incompatible with the presence of eukarya-specific features in LUCA. I thank Ryan Catchpole, Morgan Gaia, and Violette Da Cunha for helpful discussions and their help in the analysis of single trees from several published analyses. I am also grateful to members of the Fondation des Treilles scientific committee who supported the organization of several meetings devoted to the portrait of LUCA.

Declarations

Conflict of interest The author has no competing interests to declare that are relevant to the content of this article.

References

- Adam PS, Borrel G, Gribaldo S (2018) Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *Proc Natl Acad Sci USA*. <https://doi.org/10.1073/pnas.1716667115>
- Ahmad M, Xu D, Wang W (2014) Type IA topoisomerases can be “magicians” for both DNA and RNA in all domains of life. *RNA* 14:854–864
- Ahmad M, Xue Y, Lee SK, Martindale JL, Shen W, Li W, Zou S, Ciaramella M, Debat H, Nadal M, Leng F, Zhang H, Wang Q, Siaw GE, Niu H, Pommier Y, Gorospe M, Hsieh TS, Tse-Dinh YC, Xu D, Wang W (2016) RNA topoisomerase is prevalent in all domains of life and 538 associates with polyribosomes in animals. *Nucleic Acids Res* 44:6335–6349
- Alvarez-Carreno C, Penev PI, Petrov AS, Williams LD (2021) Fold evolution before LUCA: common ancestry of SH3 domains and OB domains. *Mol Biol Evol* 38:5134–5143
- Anantharaman V, Koonin EV, Aravind L (2002) Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res* 30:1427–1464
- Armenta-Medina D, Segovia L, Perez-Rueda E (2014) Comparative genomics of nucleotide metabolism: a tour to the past of the three cellular domains of life. *MC Genomics*. <https://doi.org/10.1186/1471-2164-15-800>
- Bamford DH (2003) Do viruses form lineages across different domains of life? *Res Microbiol* 154:231–236
- Becerra A, Delaye L, Islas S, Lazcano A (2007a) The very early stage of biological evolution and the nature of the last common ancestor of the three major cell domains. *Annu Rev Ecol Evol Syst* 38:371–379
- Becerra A, Delaye L, Lazcano A, Orgel LE (2007b) Protein disulfide oxidoreductases and the evolution of thermophily: was the last common ancestor a heat-loving microbe? *J Mol Evol* 65:296–303
- Bell PJ (2001) Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? *J Mol Evol* 53:251–256
- Bergerat A, de Massy B, Gabelle D, Varoutas PC, Nicolas A, Forterre P (1997) An atypical topoisomerase II from archaea with implications for meiotic recombination. *Nature* 386:414–417
- Berkemer SJ, McGlynn SE (2020) A new analysis of archaea-bacteria domain separation: variable phylogenetic distance and the tempo of early evolution. *Mol Biol Evol* 37:2332–2340
- Bernier CR, Petrov AS, Kovacs NA, Penev PI, Williams LD (2018) Translation: the universal structural core of life. *Mol Biol Evol* 35:2065–2076
- Birikmen M, Bohnsack KE, Tran V, Somayaji S, Bohnsack MT, Ebersberger I (2021) Tracing eukaryotic ribosome biogenesis factors into the archaeal domain sheds light on the evolution of functional complexity. *Front Microbiol*. <https://doi.org/10.3389/fmicb.2021.739000>
- Boussau B, Blanquart S, Necsulea A, Lartillot N, Gouy M (2008) Parallel adaptations to high temperatures in the archaean eon. *Nature* 456:942–945
- Bowman JC, Petrov AS, Frenkel-Pinter M, Penev PI, Williams LD (2020) Root of the tree: the significance, evolution, and origins of the ribosome. *Chem Rev* 120:4848–4878
- Bozdogan GO, Szeinbaum N, Conlin PL, Chen K, Fos MS, Garcia A, Penev PI, Schaible GA, Trubl G (2024) Major biological innovations in the history of life. *Astrobiology*. <https://doi.org/10.1089/ast.2021.0119>
- Brindefalk B, Dessailly BH, Yeats C, Orengo C, Werner F, Poole AM (2013) Evolutionary history of the TBP-domain superfamily. *Nucleic Acids Res* 41:2832–2845
- Brinkmann H, Philippe H (1999) Archaea sister group of bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. *Mol Biol Evol* 16:817–825
- Brochier-Armanet C, Forterre P (2007) Widespread distribution of archaeal reverse gyrase in thermophilic bacteria suggests a complex history of vertical inheritance and lateral gene transfers. *Archaea* 2:83–93
- Brochier-Armanet C, Talla E, Gribaldo S (2008) The multiple evolutionary histories of dioxygen reductases: Implications for the origin and evolution of aerobic respiration. *Mol Biol Evol* 26:285–297
- Brooks DJ, Fresco JR (2002) Increased frequency of cysteine, tyrosine, and phenylalanine residues since the last universal ancestor. *Mol Cell Proteomics* 1(2):125–131
- Browman DT, Hoegg MB, Robbins SM (2007) The SPFH domain-containing proteins: more than lipid raft markers. *Trends Cell Biol* 17:394–402
- Butzow JJ, Eichhorn GL (1975) Different susceptibility of DNA and RNA to cleavage by metal-ions. *Nature* 254:358–359
- Caetano-Anollés G, Mughal F (2021) The Tree of Life describes a tripartite cellular world: neglected support from genome structure and codon usage and the fallacy of alignment-dependent phylogenetic interpretations. *Bioessays*. <https://doi.org/10.1002/bies.202100130>
- Caforio A, Siliakus MF, Exterkate M, Jain S, Jumde VR, Andringa RLH, Kengen SWM, Minnaard AJ, Driessen AJM, van der Oost J (2018) Converting *Escherichia coli* into an archaeobacterium with a hybrid heterochiral membrane. *Proc Natl Acad Sci U S A* 115:3704–3709
- Cantine MD, Fournier GP (2018) Environmental adaptation from the origin of life to the last universal common ancestor. *Orig Life Evol Biosph* 48:35–54

- Castagnetti S, Olfierenko S, Nurse P (2010) Fission yeast cells undergo nuclear division in the absence of spindle microtubules. *PLoS Biol*. <https://doi.org/10.1371/journal.pbio.1000512>
- Catchpole RJ, Forterre P (2019) The evolution of reverse gyrase suggests a nonhyperthermophilic last universal common ancestor. *Mol Biol Evol* 36:2737–2747
- Catchpole RJ, Barbe V, Magdelenat G, Marguet E, Terns M, Oberto J, Forterre P, Da Cunha VA (2023) Self-transmissible plasmid from a hyperthermophile that facilitates genetic modification of diverse archaea. *Nat Microbiol* 8:1339–1347
- Cavalier-Smith T (2001) Obcells as proto-organisms: membrane heredity, lithophosphorylation, and the origins of the genetic code, the first cells, and photosynthesis. *J Mol Evol* 53:555–595
- Cech TR (1986) The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell* 44:207–210
- Chen IA, Roberts RW, Szostak JW (2004) The emergence of competition between model protocells. *Science* 305:1474–1476
- Choquet CG, Patel GB, Sprott GD (1996) Heat sterilisation of archaeal liposomes. *Can J Microbiol* 42:183–186
- Coleman GA, Pancost RD, Williams TA (2019) Investigating the origins of membrane phospholipid biosynthesis genes using outgroup-free rooting. *Genome Biol Evol* 11:883–898
- Coleman GA, Davin AA, Mahendrarajah TA, Szánthó LL, Spang A, Hugenholtz P, Szöllösi GJ, Williams TA (2021) A rooted phylogeny resolves early bacterial evolution. *Science* 372(6542):eabe0511. <https://doi.org/10.1126/science.abe0511>
- Collins L, Penny D (2005) Complex spliceosomal organization ancestral to extant eukaryotes. *Mol Biol Evol* 22:1053–1066
- Collins LJ, Kurland CG, Biggs P, Penny D (2009) The modern RNP world of eukaryotes. *J Hered*. <https://doi.org/10.1093/jhered/esp064>
- Cooper K (2017) Looking for LUCA, the Last Universal Common Ancestor. In *News and discovery of Astrobiology at NASA*. <https://astrobiology.nasa.gov/news/looking-for-luca-the-last-universal-common-ancestor/>
- Crapitto AJ, Campbell A, Harris AJ, Goldman AD (2022) A consensus view of the proteome of the last universal common ancestor. *Ecol Evol*. <https://doi.org/10.1002/ece3.8930>
- Csuros M, Rogozin IB, Koonin EV (2011) detailed history of intron-rich eukaryotic ancestors inferred from a global survey of 100 complete genomes. *PLoS Comput*. <https://doi.org/10.1371/journal.pcbi.1002150>
- Csurös M, Miklós I (2009) Streamlining and large ancestral genomes in archaea inferred with a phylogenetic birth-and-death model. *Mol Biol Evol* 26:2087–2095
- Czerwonec A, Dunin-Horkawicz S, Purta E, Kaminska KH, Kasprzak JM, Bujnicki JM, Grosjean H, Rother K (2009) MODOMICS: a database of RNA modification pathways. *Nucleic Acids Res*. <https://doi.org/10.1093/nar/gkn710>
- Da Cunha V, Gaia M, Gadelle D, Nasir A, Forterre P (2017) Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *PLoS Genet*. <https://doi.org/10.1371/journal.pgen.1006810>
- Da Cunha V, Gaia M, Forterre P (2022) The expending asgard archaea and their elusive relationships with eukaryotes. *mLife*. <https://doi.org/10.1002/mlf2.12012012>
- Da Cunha V, Gaia M, Ogata H, Jaillon O, Delmont TO, Forterre P (2022) Giant viruses encode actin-related proteins. *Mol Biol Evol*. <https://doi.org/10.1093/molbev/msac022>
- De Nooijer S, Holland BR, Penny D (2009) The emergence of predators in early life: there was no garden of eden. *PLoS One*. <https://doi.org/10.1371/journal.pone.0005507>
- Delaye L, Becerra A, Lazcano A (2005) The last common ancestor: what's in a name? *Orig Life Evol Biosph* 35:537–554
- Den Boon JA, Ahlquist P (2010) Organelle-like membrane compartmentalization of positive-strand RNA virus replication factories. *Annu Rev Microbiol* 64:241–256
- Denison MR, Graham RL, Donaldson EF, Eckerle LD, Baric RS (2011) Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. *RNA Biol* 8:270–279
- Di Giulio M (2021) The phylogenetic distribution of the cell division system would not imply a cellular LUCA but a progenotic LUCA. *Biosystems*. <https://doi.org/10.1016/j.biosystems.2021.104563>
- Di Giulio M (2023) The absence of the evolutionary state of the Prokaryote would imply a polyphyletic origin of proteins and that LUCA, the ancestor of bacteria and that of archaea were progenotes. *Biosystems*. <https://doi.org/10.1016/j.biosystems.2023.105014>
- DiGate RJ, Marians KJ (1992) Escherichia coli topoisomerase III-catalyzed cleavage of RNA. *J Biol Chem* 267:20532–20535
- Doolittle WF (1978) Genes in pieces: were they ever together? *Nature* 272:581–582
- Ducluzeau AL, Schoepp-Cothenet B, Baymann F, Russell MJ (2014) Nitschke W Free energy conversion in the LUCA: Quo vadis? *Biochim Biophys Acta* 1837:982–988
- Dujon B (2010) Yeast evolutionary genomics. *Nat Rev Genet* 11:512–524
- Eigner J, Boedtker H, Michaels G (1961) Thermal degradation of nucleic acids. *Biochim Biophys Acta* 51:165–168
- Eisen JA, Hanawalt PC (1999) A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat Res* 435:171–213
- Eme L, Tamarit D, Caceres EF, Stairs CW, De Anda V, Schön ME, Seitz KW, Dombrowski N, Lewis WH, Homa F, Saw JH, Lombard J, Nunoura T, Li WJ, Hua ZS, Chen LX, Banfield JF, John ES, Reysenbach AL, Stott MB, Schramm A, Kjeldsen KU, Teske AP, Baker BJ, Etema TJG (2023) Inference and reconstruction of the heimdallarchaeal ancestry of eukaryotes. *Nature* 618:992–999
- Escobar-Turrisa P, Hernandez-Guerrero R, Poot-Hernández AC, Rodríguez-Vázquez K, Ramírez-Prado J, Pérez-Rueda E (2019) Identification of functional signatures in the metabolism of the three cellular domains of life. *PLoS One*. <https://doi.org/10.1371/journal.pone.0217083>
- Evseev P, Gutnik D, Shneider M, Miroshnikov K (2023) Use of an integrated approach involving alphafold predictions for the evolutionary taxonomy of duplodnaviria viruses. *Biomolecules*. <https://doi.org/10.3390/biom13010110>
- Fels A, Hu K, Riesner D (2001) Transcription of potato spindle tuber viroid by RNA polymerase II starts predominantly at two specific sites. *Nucleic Acids Res* 29:4589–4597
- Fer E, McGrath KM, Guy L, Hockenberry AJ, Kaçar B (2022) Early divergence of translation initiation and elongation factors. *Protein Sci*. <https://doi.org/10.1002/pro.4393>
- Ferrelli ML, Pidre ML, García-Domínguez R, Alberca LN, Del Saz-Navarro D, Santana-Molina C, Devos DP (2023) Prokaryotic membrane coat - like proteins: An update. *J Struct Biol*. <https://doi.org/10.1016/j.jsb.2023.107987>
- Filée J, Forterre P (2005) Trends Microbiol. Viral proteins functioning in organelles: a cryptic origin? *Trends Microbiol* 13:510–513
- Filée J, Forterre P, Laurent J (2003) The role played by viruses in the evolution of their hosts: a view based on informational protein phylogenies. *Res Microbiol* 154:237–243
- Fitch WM, Upper K (1987) The phylogeny of tRNA sequences provides evidence for ambiguity reduction in the origin of the genetic code. *Cold Spring Harb Symp Quant Biol* 52:759–767
- Fondi M, Emiliani G, Liò P, Gribaldo S, Fani R (2009) The evolution of histidine biosynthesis in archaea: insights into the *his* genes structure and organization in LUCA. *J Mol Evol* 69:512–526

- Forterre P (1992) New hypotheses about the origins of viruses, prokaryotes and eukaryotes. In: V[^]an Tr[^]an Thanh JK, Mounolou JC, Shneider J, Mc Kay C (eds) *Frontiers of Life*. Editions Frontieres, Gif-sur-Yvette, pp 221–234
- Forterre P (1992b) Neutral terms. *Nature*. <https://doi.org/10.1038/355305c0>
- Forterre P (1995) Thermoreduction, a hypothesis for the origin of prokaryotes. *CR Acad Sci III* 318:415–422
- Forterre P (1996) A hot topic: the origin of hyperthermophiles. *Cell* 85:789–792
- Forterre P (1999) Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins. *Mol Microbiol* 33:457–465
- Forterre P (2002a) A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein. *Trends Genet* 18:236–237
- Forterre P (2002b) The origin of DNA genomes and DNA replication proteins. *Curr Opin Microbiol* 5:525–532
- Forterre P (2005) The two ages of the RNA world, and the transition to the DNA world, a story of viruses and cells. *Biochimie* 87:93–803
- Forterre P (2006) Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. *Proc Natl Acad Sci USA* 103:3669–3674
- Forterre P (2010) Manipulation of cellular syntheses and the nature of viruses: the Virocell concept. *CR Chimie*. <https://doi.org/10.1016/j.crci.2010.06.007>
- Forterre P (2012) Darwin's goldmine is still open: variation and selection run the world. *Front Cell Infect Microbiol*. <https://doi.org/10.3389/fcimb.2012.00106>
- Forterre P (2013) The common ancestor of archaea and eukarya was not an archaeon. *Archaea*. <https://doi.org/10.1155/2013/372396>
- Forterre P (2013b) Why are there so many diverse replication machineries? *J Mol Biol* 425:4714–4726
- Forterre P (2015) The universal tree of life: an update. *Front Microbiol*. <https://doi.org/10.3389/fmicb.2015.00717>
- Forterre P (2016) To be or not to be alive: How recent discoveries challenge the traditional definitions of viruses and life. *Stud Hist Philos Biol Biomed Sci* 59:100–108
- Forterre P (2022) Carl Woese, still ahead of our time. *mlife* 1:359–367
- Forterre P (2022) Archaea: a goldmine for molecular biologists and evolutionists. In: Ferreira-Cerca S (ed) *Archaea: methods and protocols, methods in molecular biology*. Springer, New York
- Forterre P, Gabelle, (2009) Phylogenomics of DNA topoisomerases: their origin and putative roles in the emergence of modern organisms. *Nucleic Acids Res* 37:679–692
- Forterre P, Gaïa M (2021) The Origin of Viruses. In: Bamford DH, Zuckerman M (eds) *Encyclopedia of Virology*, 4th edn. Academic Press, Oxford, pp 14–22
- Forterre P, Gribaldo S (2007) The origin of modern terrestrial life. *HFSP J* 1:156–168
- Forterre P, Gribaldo S (2010) Bacteria with a eukaryotic touch: a glimpse of ancient evolution? *Proc Natl Acad Sci USA* 107:12739–12740
- Forterre P, Krupovic M (2013) The origin of virions and virocells, the escape hypothesis revisited. In: Witzany G (ed) *Viruses: essential agents of life*. Springer Science & Business Media, Cham, pp 43–60. <https://doi.org/10.1007/978-94-007-4899-6>
- Forterre P, Philippe H (1999a) The last universal common ancestor (LUCA), simple or complex? *Biol Bull* 196:373–375
- Forterre P, Philippe H (1999b) Where is the root of the universal tree of life? *BioEssays* 21:871–879
- Forterre P, Prangishvili D (2009a) The great billion-year war between ribosome- and capsid-encoding organisms (cells and viruses) as the major source of evolutionary novelties. *Ann N Y Acad Sci* 1178:65–77
- Forterre P, Prangishvili D (2009b) The origin of viruses. *Res Microbiol* 160:466–472
- Forterre P, Mirambeau G, Jaxel C, Nadal M, Duguet M (1985) High positive supercoiling in vitro catalyzed by an ATP and polyethylene glycol-stimulated topoisomerase from *Sulfolobus acidocaldarius*. *EMBO J* 4:2123–2128
- Forterre P, Bouthier De La Tour C, Philippe H, Duguet M (2000) Reverse gyrase from hyperthermophiles: probable transfer of a thermoadaptation trait from archaea to bacteria. *Trends Genet* 16:152–154
- Forterre P, Filée J, Myllykallio H (2004) Origin and evolution of DNA and DNA replication machineries. In: Ribas L (ed) *The genetic code and the origin of life*. Landes Bioscience, Austin, pp 145–168
- Forterre P, Gaia M, Da Cunha V (2019) Engineered bacterium fuels evolution debate. *Nature* 571:326
- Fournier GP, Gogarten JP (2010) Rooting the ribosomal tree of life. *Mol Biol Evol* 27:1792–1801
- Fournier GP, Andam CP, Alm EJ, Gogarten JP (2011) Molecular evolution of aminoacyl tRNA synthetase proteins in the early history of life. *Orig Life Evol Biosph* 41:621–632
- Fournier GP, Moore KR, Rangel LT, Payette JG, Momper L, Bosak T (2021) The Archean origin of oxygenic photosynthesis and extant cyanobacterial lineages. *Proc Biol Sci*. <http://https://doi.org/10.1098/rspb.2021.0675>
- Fuerst JA (2013) The PVC superphylum: exceptions to the bacterial definition? *Antonie Van Leeuwenhoek* 104:451–466
- Gaia M, Da Cunha V, Forterre P (2018) The tree of life. In: Rampelotto PH (ed) *Molecular mechanisms of microbial evolution, grand challenges in biology and biotechnology*. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-69078-0_3
- Gaïa M, Forterre P (2023) From mimivirus to mirusvirus: the quest for hidden giants. *Viruses*. <https://doi.org/10.3390/v15081758>
- Gaïa M, Meng L, Pelletier E, Forterre P, Vanni C, Fernandez-Guerra A, Jaillon O, Wincker P, Ogata H, Krupovic M, Delmont TO (2023) Mirusviruses link herpesviruses to giant viruses. *Nature* 616:783–789
- Galtier N, Tourasse N, Gouy M (1999) A nonhyperthermophilic common ancestor to extant life forms. *Science* 283:220–221
- Garcia PS, D'Angelo F, Ollagnier de Choudens S, Dussouchaud M, Bouveret E, Gribaldo S, Barras F (2022) An early origin of iron-sulfur cluster biosynthesis machineries before Earth oxygenation. *Nat Ecol Evol* 6:1564–1572
- Gaudin M, Gaudiard E, Schouten S, Houel-Renault L, Lenormand P, Marguet E, Forterre P (2013) Hyperthermophilic archaea produce membrane vesicles that can transfer DNA. *Environ Microbiol Rep* 5:109–116
- Gill S, Forterre P (2015) Origin of life: LUCA and extracellular membrane vesicles (EMVs). *Int J Astrobiol* 15:7–15
- Gill S, Catchpole R, Forterre P (2019) Extracellular membrane vesicles in the three domains of life and beyond. *FEMS Microbiol Rev* 43:273–303
- Ginoza W, Hoelle CJ, Vessey KB, Carmack C (1964) Mechanisms of inactivation of single-stranded virus nucleic acid by heat. *Nature* 203:606–609
- Glansdorff N, Xu Y, Labedan B (2008) The Last Universal Common Ancestor: emergence, constitution and genetic legacy of an elusive forerunner. *Biol Direct*. <https://doi.org/10.1186/1745-6150-3-29>
- Gogarten JP, Deamer D (2016) Is LUCA a thermophilic progenote? *Nat Microbiol*. <https://doi.org/10.1038/nmicrobiol.2016.229>
- Gogarten JP, Kibak H, Dittrich P, Taiz L, Bowman EJ, Bowman BJ, Manolson MF, Poole RJ, Date T, Oshima T, Konishi J, Denda K, Yoshida M (1989) Evolution of the vacuolar H⁺-ATPase:

- implications for the origin of eukaryotes. *Proc Natl Acad Sci U S A* 86:6661–6665
- Gogarten-Boekels M, Hilario E, Gogarten JP (1995) The effects of heavy meteorite bombardment on the early evolution—the emergence of the three domains of life. *Orig Life Evol Biosph* 25:251–264
- Goldman AD, Weber JM, LaRowe DE, Barge LM (2023) Electron transport chains as a window into the earliest stages of evolution. *Proc Natl Acad Sci U S A*. <https://doi.org/10.1073/pnas.2210924120>
- Gordon MP, Huang CW, Hurtler (1976). In: Wang SY (ed) *Photochemistry and photobiology of nucleic acids*. Academic Press, New York, pp 265–308
- Groussin M, Gouy M (2011) Adaptation to environmental temperature is a major determinant of molecular evolutionary rates in archaea. *Mol Biol Evol* 28:2661–2674
- Guglielmini J, Woo AC, Krupovic M, Forterre P, Gaia M (2019) Diversification of giant and large eukaryotic dsDNA viruses predated the origin of modern eukaryotes. *Proc Natl Acad Sci USA* 116:19585–19592
- Guglielmini J, Gaia M, Da Cunha V, Criscuolo A, Krupovic M, Forterre P (2022) Viral origin of eukaryotic type IIADNA topoisomerases. *Virus Evol* 8(2):veac097. <https://doi.org/10.1093/ve/veac097>
- Guljamov A, Jenke-Kodama H, Saumweber H, Quillardet P, Frangeul L, Castets AM, Bouchier C, Tandeau de Marsac N, Dittmann E (2007) Horizontal gene transfer of two cytoskeletal elements from a eukaryote to a cyanobacterium. *Curr Biol* 17(17):757–759. <https://doi.org/10.1016/j.cub.2007.06.063>
- Harold FM, Van Brunt J (1977) Circulation of H⁺ and K⁺ across the plasma membrane is not obligatory for bacterial growth. *Science* 197:372–373
- Harris AJ, Goldman AD (2021) The very early evolution of protein translocation across membranes. *PLoS Comput Biol*. <https://doi.org/10.1371/journal.pcbi.1008623>
- Hernandez AM, Ryan JF (2021) Six-state amino acid recoding is not an effective strategy to offset compositional heterogeneity and saturation in phylogenetic analyses. *Syst Biol* 70:1200–1212
- Hernández-Montes G, Díaz-Mejía JJ, Pérez-Rueda E (2019) Segovia L (2008) The hidden universal distribution of amino acid biosynthetic networks: a genomic perspective on their origins and evolution. *Genome Biol*. <https://doi.org/10.1186/gb-2008-9-6-r95>
- Hethke C, Bergerat A, Hausner W, Forterre P, Thomm M (1999) Cell-free transcription at 95 degrees: thermostability of transcriptional components and DNA topology requirements of *Pyrococcus* transcription. *Genetics* 152:1325–1333
- Hinderhofer M, Walker CA, Friemel A, Stuermer CA, Möller HM, Reuter A (2009) Evolution of prokaryotic SPFH proteins. *BMC Evol Biol*. <https://doi.org/10.1186/1471-2148-9-10>
- Hoepfner MPG, Gardner PP, Poole AM (2012) Comparative analysis of RNA families reveals distinct repertoires for each domain of life. *PLoS Comput Biol*. <https://doi.org/10.1371/journal.pcbi.1002752>
- Imachi H, Nobu MK, Nakahara N, Morono Y, Ogawara M, Takaki Y, Takano Y, Uematsu K, Ikuta T, Ito M, Matsui Y, Miyazaki M, Murata K, Saito Y, Sakai S, Song C, Tasumi E, Yamanaka Y, Yamaguchi T, Kamagata Y, Tamaki H, Takai K (2020) Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* 577:519–525
- International Committee on Taxonomy of Viruses Executive Committee (2020) The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. *Nat Microbiol* 5:668–674
- Iwabe N, Kuma K, Hasegawa M, Osawa S, Miyata T (1989) Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc Natl Acad Sci U S A* 86:9355–9359
- Jalasvuori M, Bamford JK (2008) Structural co-evolution of viruses and cells in the primordial world. *Orig Life Evol Biosph* 38:165–181
- Jeffares DC, Poole AM, Penny D (1998) Relics from the RNA world. *J Mol Evol* 46:18–36
- Joyce GF, Szostak JW (2018) Protocells and RNA Self-replication. *Cold Spring Harb Perspect Biol*. <https://doi.org/10.1101/cshperspect.a034801>
- Jüttner M, Ferreira-Cerca S (2022) A comparative perspective on ribosome biogenesis: unity and diversity across the tree of life. *Methods Mol Biol*. https://doi.org/10.1007/978-1-0716-2501-9_1
- Kanai S, Kikuno R, Toh H, Ryo H, Todo T (1997) Molecular evolution of the photolyase-blue-light photoreceptor family. *J Mol Evol* 45:535–548
- Kikuchi A, Asai K (1984) Reverse gyrase—a topoisomerase which introduces positive superhelical turns into DNA. *Nature* 309:677–681
- Kim ST, Sancar A (1991) Effect of base, pentose, and phosphodiester backbone structures on binding and repair of pyrimidine dimers by *Escherichia coli* DNA photolyase. *Biochemistry* 30:8623–8630
- Knoll AH, Nowak MA (2017) The timetable of evolution. *Sci Adv*. <https://doi.org/10.1126/sciadv.1603076>
- 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* 46:54–63
- Konings WN, Albers SV, Koning S, Driessen AJ (2002) The cell membrane plays a crucial role in survival of bacteria and archaea in extreme environments. *Antonie Van Leeuwenhoek* 81:61–72
- Koonin EV, Martin W (2005) On the origin of genomes and cells within inorganic compartments. *Trends Genet* 21:647–654
- Koonin EV, Mulikjanian AY (2013) Evolution of cell division: from shear mechanics to complex molecular machineries. *Cell*. <https://doi.org/10.1016/j.cell.2013.02.008>
- Koonin EV, Krupovic M, Ishino S, Ishino Y (2020) The replication machinery of LUCA: common origin of DNA replication and transcription. *BMC Biol*. <https://doi.org/10.1186/s12915-020-00800-9>
- Koonin EV, Krupovic M, Dolja VV (2023) The global virome: How much diversity and how many independent origins? *Environ Microbiol* 25:40–44
- Koonin EV, Kuhn JH, Dolja VV, Krupovic M (2024) Megataxonomy and global ecology of the virosphere. *ISME J*. <https://doi.org/10.1093/ismej/wrad042>
- Krupovic M, Bamford DH (2010) Order to the Viral Universe. *J Virol* 84:12476–12477
- Krupovic M, Dolja VV, Koonin EV (2019) Origin of viruses: primordial replicators recruiting capsids from hosts. *Nat Rev Microbiol* 17:449–458
- Krupovic M, Dolja VV, Koonin EV (2020) The LUCA and its complex virome. *Nat Rev Microbiol* 18:661–670
- Krupovic M, Kuhn JH, Wang F, Baquero DP, Dolja VV, Egelman EH, Prangishvili D, Koonin EV (2021) *Adnaviria*: a new realm for archaeal filamentous viruses with linear a-form double-stranded DNA genomes. *J Virol*. <https://doi.org/10.1128/JVI.00673-21>
- Krupovic M, Dolja VV, Koonin EV (2023) The virome of the last eukaryotic common ancestor and eukaryogenesis. *Nat Microbiol*. <https://doi.org/10.1038/s41564-023-01378-y>
- Kurland CG, Collins LJ, Penny D (2006) Genomics and the irreducible nature of eukaryote cells. *Science* 312:1011–1014
- Kurland CG, Canbäck B, Berg OG (2007) The origins of modern proteomes. *Biochimie* 89:1454–1463
- Kyrpides N, Overbeek R, Ouzounis C (1999) Universal protein families and the functional content of the last universal common ancestor. *J Mol Evol* 49:413–423

- Ladenstein R, Ren B (2006) Protein disulfides and protein disulfide oxidoreductases in hyperthermophiles. *FEBS J* 273:4170–4185
- Lane CE, van den Heuvel K, Kozera C, Curtis BA, Parsons BJ, Bowman S, Archibald JM (2007) Nucleomorph genome of *Hemiselmis andersenii* reveals complete intron loss and compaction as a driver of protein structure and function. *Proc Natl Acad Sci USA* 104:19908–19913
- Lane N, Allen JF, Martin W (2010) How did LUCA make a living? Chemiosmosis in the origin of life. *Bioessays* 32:271–280
- Lecompte O, Thierry RR, JC, Moras D, Poch O, (2002) Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res* 30:5382–53890
- Leipe DD, Aravind L, Koonin EV (1999) Did DNA replication evolve twice independently? *Nucleic Acids Res* 27:3389–3401
- Liu Y, Makarova KS, Huang WC, Wolf YI, Nikolskaya AN, Zhang X, Cai M, Zhang CJ, Xu W, Luo Z, Cheng L, Koonin EV, Li M (2021a) Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature* 593:553–557
- Liu Y, Demina TA, Roux S, Aiewsakun P, Kazlauskas D, Simmonds P, Prangishvili D, Oksanen HM, Krupovic M (2021) Diversity, taxonomy, and evolution of archaeal viruses of the class *Caudoviricetes*. *PLoS Biol*. <https://doi.org/10.1371/journal.pbio.3001442>
- Liu J, Soler N, Gorlas A, Cvirkaitė-Krupovic V, Krupovic M, Forterre P (2021) Extracellular membrane vesicles and nanotubes in archaea. *MicroLife*. <https://doi.org/10.1093/femsm/luqab007>
- Lombard J (2016) Early evolution of polyisoprenol biosynthesis and the origin of cell walls. *PeerJ*. <https://doi.org/10.7717/peerj.2626>
- Lombard J, López-García P, Moreira D (2012) The early evolution of lipid membranes and the three domains of life. *Nat Rev Microbiol* 10:507–515
- López-García P, Moreira D (2023) The symbiotic origin of the eukaryotic cell. *C R Biol* 346:55–73
- Low SJ, Džunková M, Chaumeil PA, Parks DH, Hugenholtz P (2019) Evaluation of a concatenated protein phylogeny for classification of tailed double-stranded DNA viruses belonging to the order Caudovirales. *Nat Microbiol* 4:1306–1315
- Lundin D, Gribaldo S, Torrents E, Sjöberg BM, Poole AM (2010) Ribonucleotide reduction - horizontal transfer of a required function spans all three domains. *BMC Evol Biol*. <https://doi.org/10.1186/1471-2148-10-383>
- Lundin D, Berggren G, Logan DT, Sjöberg BM (2015) The origin and evolution of ribonucleotide reduction. *Life* 5:604–636
- MacNaughton TB, Shi ST, Modahl LE, Lai MM (2002) Rolling circle replication of hepatitis delta virus RNA is carried out by two different cellular RNA Polymerases. *J Virol* 76:3920–3927
- Mahendrarajah TA, Moody ERR, Schrepf D, Szánthó LL, Dombrowski N, Davín AA, Pisani D, Donoghue PCJ, Szöllősi GJ, Williams TA, Spang A (2023) ATP synthase evolution on a cross-braced dated tree of life. *Nat Commun*. <https://doi.org/10.1038/s41467-023-42924-w>
- Marguet E, Gaudin M, Gaudiard E, Fourquaux I, le Blond du Plouy S, Matsui I, Forterre P, (2013) Membrane vesicles, nanotubes and/or nanotubes produced by hyperthermophilic archaea of the genus *Thermococcus*. *Biochem Soc Trans* 41:436–442
- Mariscal C, Doolittle WF (2015) Eukaryotes first: how could that be? *Philos Trans R Soc Lond B Biol Sci*. <https://doi.org/10.1098/rstb.2014.0322>
- Martin W, Koonin EV (2006) Introns and the origin of nucleus-cytosol compartmentalization. *Nature* 440:41–45
- Martin W, Russell MI (2003) On the origin of cells: a hypothesis for the evolutionary transition from abiotic geochemistry to the chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philos Trans R Soc B* 358:59–85
- Martinez-Gutierrez CA, Aylward FO (2021) Phylogenetic Signal, Congruence, and Uncertainty across Bacteria and Archaea. *Mol Biol Evol* 38:5514–5527
- Martin-Galiano AJ, Oliva MA, Sanz L, Bhattacharyya A, Serna M, Yebenes H, Valpuesta JM, Andreu JM (2011) Bacterial tubulin distinct loop sequences and primitive assembly properties support its origin from a eukaryotic tubulin ancestor. *J Biol Chem* 286:19789–19803
- Miller SL, Lazcano A (1995) The origin of life—did it occur at high temperatures? *J Mol Evol* 41:689–692
- Mills J, Gebhard LJ, Schubotz F, Shevchenko A, Speth DR, Liao Y, Duggin IG, Marchfelder A, Erdmann S (2014) Extracellular vesicle formation in Euryarchaeota is driven by a small GTPase. *Proc Natl Acad Sci USA*. <https://doi.org/10.1073/pnas.2311321121>
- Missoury S, Plancqueel S, Li de la Sierra-Gallay I, Zhang W, Liger D, Durand D, Dammak R, Collinet B, van Tilbeurgh H (2019) The structure of the Tsab/TsaD/TsaE complex reveals an unexpected mechanism for the bacterial t6A tRNA-modification. *Nucleic Acids Res* 47:9464–9465
- Moody ERR, Mahendrarajah TA, Dombrowski N, Clark JW, Petitjean C, Offre P, Szöllősi GJ, Spang A, Williams TA (2022) An estimate of the deepest branches of the tree of life from ancient vertically evolving genes. *elife*. <https://doi.org/10.7554/eLife.66695>
- Moody ERR, Álvarez-Carretero S, Mahendrarajah TA, Clark JW, Betts HC, Dombrowski N, Szánthó LL, Boyle RA, Daines S, Chen X, Lane N, Yang Z, Shields GA, Szöllősi GJ, Spang A, Pisani D, Williams TA, Lenton TM, Donoghue PCJ (2024) The nature of the last universal common ancestor and its impact on the early Earth system. *Nat Ecol Evol*. <https://doi.org/10.1038/s41559-024-02461-1>
- Moraleda G, Taylor J (2001) Host RNA polymerase requirements for transcription of the human hepatitis delta virus genome. *J Virol* 75:10161–10169
- Mulkiđjanian AY, Makarova KS, Galperin MY, Koonin EV (2007) Inventing the dynamo machine: the evolution of the F-type and V-type ATPases. *Nat Rev Microbiol* 5:892–899
- Mulkiđjanian AY, Galperin MY, Koonin EV (2009) Co-evolution of primordial membranes and membrane proteins. *Trends Biochem Sci*. <https://doi.org/10.1016/j.tibs.2009.01.005>
- Mulkiđjanian AY, Bychkov AY, Dibrova DV, Galperin MY, Koonin EV (2012) Origin of first cells at terrestrial, anoxic geothermal fields. *Proc Natl Acad Sci U S A*. <https://doi.org/10.1073/pnas.1117774109>
- Narrowe AB, Spang A, Stairs CW, Caceres EF, Baker BJ, Miller CS, Ettme TJG (2018) Complex evolutionary history of translation elongation factor 2 and diphthamide biosynthesis in archaea and parabasalids. *Genome Biol Evol* 10:2380–2393
- Nasir A, Forterre P, Kim KM, Caetano-Anollés G (2014) The distribution and impact of viral lineages in domains of life. *Front Microbiol*. <https://doi.org/10.3389/fmicb.2014.00194>
- Nasir A, Kim KM, Da Cunha V, Caetano-Anollés G (2016) Arguments reinforcing the three-domain view of diversified cellular life. *Archaea*. <https://doi.org/10.1155/2016/1851865>
- Nasir A, Mughal F, Caetano-Anollés G (2021) The tree of life describes a tripartite cellular world. *Bioessays*. <https://doi.org/10.1002/bies.202000343>
- Niewegłowska ES, Brilot AF, Méndez-Moran M, Kokontis C, Baek M, Li J, Cheng Y, Baker D, Bondy-Denomy J, Agard DA (2023) The PA3 phage nucleus is enclosed by a self-assembling 2D crystalline lattice. *Nat Commun*. <https://doi.org/10.1038/s41467-023-36526-9>
- Olsen GJ, Woese CR (1997) Archaeal genomics: an overview. *Cell* 89:991–994
- Pace NR (2009) Problems with prokaryote. *J Bacteriol* 191:2008–2010
- Papineau D, She Z, Dodd MS, Iacoviello F, Slack JF, Hauri E, Shearing P, Little CTS (2022) Metabolically diverse primordial microbial

- communities in Earth's oldest seafloor-hydrothermal jasper. *Sci Adv*. <https://doi.org/10.1126/sciadv.abm2296>
- Pelchat M, Grenier C, Perreault JP (2002) Characterization of a viroid-derived RNA promoter for the DNA-dependent RNA polymerase from *Escherichia coli*. *Biochemistry* 41:6561–6571
- Pende N, Sogues A, Megrian D, Sartori-Rupp A, England P, Palabikyan H, Rittmann SKR, Graña M, Wehenkel AM, Alzari PM, Gribaldo S (2021) SepF is the FtsZ anchor in archaea, with features of an ancestral cell division system. *Nat Commun*. <https://doi.org/10.1038/s41467-021-23099-8>
- Penny D, Poole A (1999) The nature of the last universal common ancestor. *Curr Opin Genet Dev* 9:672–677
- Perrochia L, Guetta D, Hecker A, Forterre P, Basta T (2013) Functional assignment of KEOPS/EKC complex subunits in the biosynthesis of the universal t6A tRNA modification. *Nucleic Acids Res* 41:9484–9499
- Petrov AS, Gulen B, Norris AM, Kovacs NA, Bernier CR, Lanier KA, Fox GE, Harvey SC, Wartell RM, Hud NV, Williams LD (2015) History of the ribosome and the origin of translation. *Proc Natl Acad Sci USA* 112:15396–15401
- Pfeiffer JK, Kirkegaard K (2003) A single mutation in poliovirus RNA-dependent RNA polymerase confers resistance to mutagenic nucleotide analogs via increased fidelity. *Proc Natl Acad Sci U S A* 100:7289–7294
- Phan HD, Lai LB, Zahurancik WJ, Gopalan V (2021) The many faces of RNA-based RNase P, an RNA-world relic. *Trends Biochem Sci* 46:976–991
- Philippe H, Forterre P (1999) The rooting of the universal tree of life is not reliable. *J Mol Evol* 49:509–523
- Pohorille A, Deamer D (2009) Self-assembly and function of primitive cell membranes of the archaea: a unifying view. *Nat Rev Microbiol* 160:449–456
- Poole AM (2006) Did group II intron proliferation in an endosymbiont-bearing archaeon create eukaryotes? *Biol Direct*. <https://doi.org/10.1186/1745-6150-1-36>
- Poole AM (2009) Horizontal gene transfer and the earliest stages of the evolution of life. *Res Microbiol* 160:473–480
- Poole AM, Logan DT (2005) Modern mRNA proofreading and repair: clues that the last universal common ancestor possessed an RNA genome? *Mol Biol Evol* 22:1444–1455
- Prangishvili D (2013) The wonderful world of archaeal viruses. *Annu Rev Microbiol* 67:565–585
- Pugachev KV, Guirakhoo F, Ocran SW, Mitchell F, Parsons M, Penal C, Girakhoo S, Pougatcheva SO, Arroyo J, Trent DW, Monath TP (2004) High fidelity of yellow fever virus RNA polymerase. *J Virol* 78:1032–1038
- Rangel LT, Fournier GP (2023) Fast-evolving alignment sites are highly informative for reconstructions of deep tree of life phylogenies. *Microorganisms*. <https://doi.org/10.3390/microorganisms11102499>
- Rani P, Kalladi SM, Bansia H, Rao S, Jha RK, Jain P, Bhaduri T, Nagaraja VA (2010) Type IA DNA/RNA topoisomerase with RNA hydrolysis activity participates in ribosomal RNA processing. *J Mol Biol* 432:5614–5631
- Raoult D, Forterre P (2008) Redefining viruses: lessons from Mimivirus. *Nat Rev Microbiol* 6:315–319
- Reaney DC (1984) RNA splicing as an error-screening mechanism. *J Theor Biol* 110:315–321
- Rivas-Marín E, Devos DP (2018) The paradigms they are a-changin': past, present and future of PVC bacteria research. *Antonie Van Leeuwenhoek* 111:785–799
- Rodrigues-Oliveira T, Wollweber F, Ponce-Toledo RI, Xu J, Rittmann SKR, Klingl A, Pilhofer M, Schleper C (2023) Actin cytoskeleton and complex cell architecture in an Asgard archaeon. *Nature* 613:332–339
- Rogozin IB, Carmel L, Csuros M, Koonin EV (2012) Origin and evolution of spliceosomal introns. *Biol Direct*. <https://doi.org/10.1186/1745-6150-7-11>
- Roy SW, Gilbert W (2006) The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat Rev Genet* 7:211–221
- Salzer U, Zhu R, Luten M, Isobe H, Pastushenko V, Perkmann T, Hinterdorfer P, Bosman GJ (2008) Vesicles generated during storage of red cells are rich in the lipid raft marker stomatin. *Transfusion* 48:451–462
- Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R (2010) Viral mutation rates. *J Virol* 84:9733–9748
- Santana-Molina C, Del Saz-Navarro D, Devos DP (2023) Early origin and evolution of the FtsZ/tubulin protein family. *Front Microbiol*. <https://doi.org/10.3389/fmicb.2022.1100249>
- Santarella-Mellwig R, Franke J, Jaedicke A, Gorjanacz M, Bauer U, Budd A, Mattaj IW, Devos DP (2010) The compartmentalized bacteria of the planctomycetes-verrucomicrobia-chlamydiae superphylum have membrane coat-like proteins. *PLoS Biol*. <https://doi.org/10.1371/journal.pbio.1000281>
- Schlieper D, Oliva MA, Andreu JM, Löwe J (2005) Structure of bacterial tubulin BtubA/B: evidence for horizontal gene transfer. *Proc Natl Acad Sci U S A* 102:9170–9175
- Schopf JW, Kitajima K, Spicuzza MJ, Kudryavtsev AB, Valley JW (2018) SIMS analyses of the oldest known assemblage of microfossils document their taxon-correlated carbon isotope compositions. *Proc Natl Acad Sci USA* 115:53–58
- Schrum JP, Zhu TF, Szostak JW (2010) The origin of cellular life. In: Atkin JF, Gesteland RF, Cech TR (eds) *RNA worlds*. Cold Spring Harbor Laboratory Press, New York, pp 51–62
- Sekiguchi J, Shuman S (1997) Site-specific ribonuclease activity of eukaryotic DNA topoisomerase I. *Mol Cell* 1:89–97
- Shimada H, Yamagishi A (2011) Stability of heterochiral hybrid membrane made of bacterial sn-G3P lipids and archaeal sn-G1P lipids. *Biochemistry* 50:4114–4120
- Shiratori T, Suzuki S, Kakizawa Y, Ishida KI (2019) Phagocytosis-like cell engulfment by a planctomycete bacterium. *Nat Commun* 10(1):5529. <https://doi.org/10.1038/s41467-019-13499-2>
- Skryabin GO, Komelkov AV, Galetsky SA, Bagrov DV, Evtushenko EG, Nikishin II, Zhordaniia KI, Savelyeva EE, Akselrod ME, Paianidi IG, Tchevkina EM (2021) Stomatin is highly expressed in exosomes of different origin and is a promising candidate as an exosomal marker. *J Cell Biochem* 122:100–115
- Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, van Eijk R, Schleper C, Guy L, Ettema TJG (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521:173–179
- Staley JT (2017) Domain cell theory supports the independent evolution of the eukarya bacteria and archaea and the nuclear compartment commonality hypothesis. *Open Biol*. <https://doi.org/10.1098/rsob.170041>
- Staley JT, Fuerst JA (2017) Ancient, highly conserved proteins from a LUCA with complex cell biology provide evidence in support of the nuclear compartment commonality (NuCom) hypothesis. *Res Microbiol* 168:395–412
- Stelitano D, Cortese M (2023) Electron microscopy: the key to resolve RNA virus replication organelles. *Mol Microbiol*. <https://doi.org/10.1111/mmi.15173>
- Stetter KO (1996) Hyperthermophiles in the history of life. *Ciba Found Symp* 202:1–10
- Takemura M (2001) Poxviruses and the origin of the eukaryotic nucleus. *J Mol Evol* 52:419–425
- Takeuch N, Hogeweg P, Koonin EV (2011) On the origin of DNA genomes: evolution of the division of labor between template and catalyst in model replicator systems. *PLoS Comput Biol*. <https://doi.org/10.1371/journal.pcbi.1002024>

- Tavernarakis N, Driscoll M, Kyrpides NC (1999) The SPFH domain: implicated in regulating targeted protein turnover in stomatins and other membrane-associated proteins. *Trends Biochem Sci* 24:425–427
- Theobald DL, Mitton-Fry RM, Wuttke DS (2003) Nucleic acid recognition by OB-fold proteins. *Annu Rev Biophys Biomol Struct* 32:115–133
- Thiaville PC, El Yacoubi B, Perrochia L, Hecker A, Prigent M, Thiaville JJ, Forterre P, Namy O, Basta T, de Crécy-Lagard V (2014a) Cross kingdom functional conservation of the core universally conserved threonylcarbamoyladenine tRNA synthesis enzymes. *Eukaryote Cell* 13:1222–1231
- Thiaville PC, Iwata-Reuyl D, de Crécy-Lagard V (2014b) Diversity of the biosynthesis pathway for threonylcarbamoyladenine (t(6)A), a universal modification of tRNA. *RNA Biol* 11:1529–1539
- van der Gulik PT, Hoff WD (2016) Anticodon modifications in the tRNA Set of LUCA and the fundamental regularity in the standard genetic code. *PLoS One*. <https://doi.org/10.1371/journal.pone.0158342>
- van der Gulik PTS, Hoff WD, Speijer D (2024) The contours of evolution: In defense of Darwin's tree of life paradigm. *Bioessays*. <https://doi.org/10.1002/bies.202400012>
- Vechtomova YL, Telegina TA, Kritsky MS (2020) Evolution of proteins of the DNA photolyase/cryptochrome family. *Biochemistry*. <https://doi.org/10.1134/S0006297920140072>
- Vetsigian K, Woese C, Goldenfeld N (2006) Collective evolution and the genetic. *Code Proc Natl Acad Sci USA* 103:10696–10701
- Villain P, Catchpole R, Forterre P, Oberto J, da Cunha V, Basta T (2022) Expanded dataset reveals the emergence and evolution of DNA gyrase in archaea. *Mol Biol Evol*. <https://doi.org/10.1093/molbev/msac155>
- Wächtershäuser G (2003) From pre-cells to Eukarya—a tale of two lipids. *Mol Microbiol* 47:13–22
- Wang F, Cvirkaite-Krupovic V, Vos M, Beltran LC, Kreutzberger MAB, Winter JM, Su Z, Liu J, Schouten S, Krupovic M, Egelman EH (2022) Spindle-shaped archaeal viruses evolved from rod-shaped ancestors to package a larger genome. *Cell* 185:1297–1307
- Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S, Martin WF (2016) The physiology and habitat of the last universal common ancestor. *Nat Microbiol*. <https://doi.org/10.1038/nmicrobiol.2016.116>
- Weiss MC, Neukirchen S, Roettger M, Mrnjavac N, Nelson-Sathi S, Martin WF, Sousa FL (2016) Reply to 'Is LUCA a thermophilic progenote?' *Nat Microbiol*. <https://doi.org/10.1038/nmicrobiol.2016.230>
- Weiss MC, Preiner M, Xavier JC, Zimorski V, Martin WF (2018) The last universal common ancestor between ancient Earth chemistry and the onset of genetics. *PLoS Genet*. <https://doi.org/10.1371/journal.pgen.1007518>
- Werner F (2008) Structural evolution of multi-subunit RNA polymerases. *Trends Microbiol* 16:247–250
- Werner F (2012) A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. *J Mol Biol*. <https://doi.org/10.1016/j.jmb.2012.01.031>
- Werner F, Grohmann D (2011) Evolution of multi-subunit RNA polymerases in the three domains of life. *Nat Rev Microbiol* 9:85–98
- Wettich A, Biebricher CK (2001) RNA species that replicate with DNA-dependent RNA polymerase from *Escherichia coli*. *Biochemistry* 40:3308–3315
- White MF, Allers T (2018) DNA repair in the archaea—an emerging picture. *FEMS Microbiol Rev* 42:514–526
- Wienken CJ, Baaske P, Duhr S, Braun D (2011) Thermophoretic melting curves quantify the conformation and stability of RNA and DNA. *Nucleic Acids Res*. <https://doi.org/10.1093/nar/gkr035>
- Williams TA, Cox CJ, Foster PG, Szöllösi GJ, Embley TM (2020) Phylogenomics provides robust support for a two-domains tree of life. *Nat Ecol Evol* 4:138–147
- Woese CR (1983) The primary lines of descent and the universal ancestor. In: Bendall DS (ed) *Evolution from molecules to man*. Cambridge University Press, Cambridge, pp 209–229
- Woese CR (1987) Bacterial evolution. *Microbiol Rev* 51:221–271
- Woese CR (1998) The universal ancestor. *Proc Natl Acad Sci USA* 95:6854–6859
- Woese CR (2000) Interpreting the universal phylogenetic tree. *Proc Natl Acad Sci USA* 97:8392–8396
- Woese CR (2002) On the evolution of cells. *Proc Natl Acad Sci USA* 99:8742–8747
- Woese CR, Fox GE (1977a) The concept of cellular evolution. *J Mol Evol* 10:1–6
- Woese CR, Fox GE (1977b) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA* 74:5088–5090
- Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains archaea, bacteria, and Eucarya. *Proc Natl Acad Sci USA* 87:4576–4579
- Woese CR, Olsen GJ, Ibba M, Söll D (2000) Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol Mol Biol Rev* 64:202–236
- Wolf YI, Kazlauskas D, Iranzo J, Lucía-Sanz A, Kuhn JH, Krupovic M, Dolja VV, Koonin EV (2018) Origins and Evolution of the Global RNA Virome. *mBio*. <https://doi.org/10.1128/mBio.02329-18>
- Wolff G, Limpens RWAL, Zevenhoven-Dobbe JC, Laugks U, Zheng S, de Jong AWM, Koning RI, Agard DA, Grünewald K, Koster AJ, Snijder EJ, Bárcena M (2020) A molecular pore spans the double membrane of the coronavirus replication organelle. *Science* 369:1395–1398
- Woo AC, Gaia M, Guglielmini J, Da Cunha V, Forterre P (2021) Phylogeny of the *Varidnaviria* morphogenesis module: congruence and incongruence with the tree of life and viral taxonomy. *Front Microbiol*. <https://doi.org/10.3389/fmicb.2021.704052>
- Xie R, Wang Y, Huang D, Hou J, Li L, Hu H et al (2021) Expanding Asgard members in the domain of archaea sheds new light on the origin of eukaryotes. *Sci China Life Sci*. <https://doi.org/10.1007/s11427-021-1969-6>
- Xu D, Shen W, Guo R, Xue Y, Peng W, Sima J, Yang J, Sharov A, Srikantan S, Yang J, Fox D 3rd, Qian Y, Martindale JL, Piao Y, Machamer J, Joshi SR, Mohanty S, Shaw AC, Lloyd TE, Brown GW, Ko MS, Gorospe M, Zou S, Wang W (2013) Top3beta is an RNA topoisomerase that works with fragile X syndrome 678 protein to promote synapse formation. *Nat Neurosci* 16:1238–1247
- Yokobori SI, Nakajima Y, Akanuma S, Yamagishi A (2016) Birth of archaeal cells: molecular phylogenetic analyses of G1P dehydrogenase, G3P dehydrogenases, and glycerol kinase suggest derived features of archaeal membranes having G1P polar lipids. *Archaea*. <https://doi.org/10.1155/2016/1802675>
- Yokoyama H, Matsui I (2020) The lipid raft markers stomatin, prohibitin, flotillin, and HflK/C (SPFH)-domain proteins form an operon with NfeD proteins and function with apolar polyisoprenoid lipids. *Crit Rev Microbiol* 46:38–48
- 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 (2017) Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541:353–358

- Zhaxybayeva O, Gogarten JP (2004) Cladogenesis, coalescence and the evolution of the three domains of life. *Trends Genet* 20:182–187
- Zhaxybayeva O, Lapierre P, Gogarten JP (2005) Ancient gene duplications and the root(s) of the tree of life. *Protoplasma* 227:53–64

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.