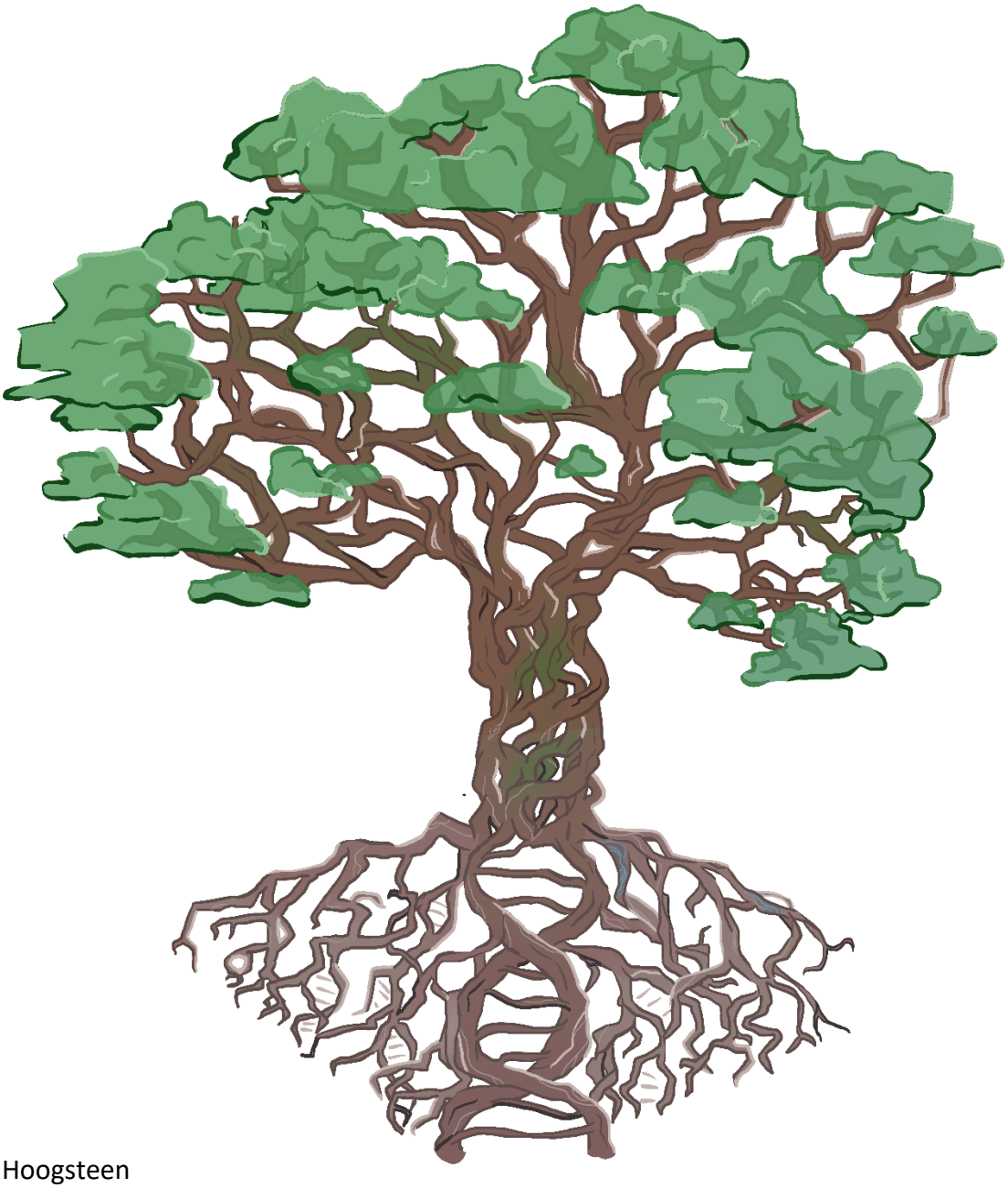
Exploring different methods of determining the properties of the Last Universal Common Ancestor.



D. Hoogsteen

s2839946

Pre-master Biology; Ecology & Evolution

MolGen University of Groningen

Prof. Dr. Oscar P. Kuipers

24-01-2023

Cover art by Artemis Hoogsteen

Abstract

The tree of life shows how all species are related. If it's traced back far enough, it shows that all organisms are related to a single common ancestor, which stands at the root of the tree of life. This organism is referred to as LUCA, the last universal common ancestor. Understanding LUCA is important for understanding the origin of life and early evolution. To unravel LUCA's properties, several approaches to constructing the tree of life were discussed and the results were compared to determine LUCA's most likely properties. I concluded that LUCA was a community of progenotes, that used both RNA and DNA to transcribe several proteins and store genetic information. It lived in alkaline hydrothermal vents, which provided a suitable temperature and enough chemical potential to produce energy without the need to evolve a complex metabolism. It was an anaerobic autotroph with a metabolism that used the acetyl-CoA pathway and relied on its environmental conditions to produce ATP.

Table of Contents

Introduction

Last Universal Common Ancestor

The tree of life shows how all species are related. If it's traced back far enough, it shows that all organisms are related to a single common ancestor, which stands at the root of the tree of life. This organism is referred to as LUCA, the last universal common ancestor. [1]

LUCA is a theoretical concept describing the most recent population of which all known life forms share common descent. Thoughts about what LUCA could have looked like are quite diverse. It could have been a simple cell, a complex unicellular organism, or simply a chemical reaction. It's an essential subject to the study of the origin of life and early evolution.

An important distinction to make here is that while LUCA is the last common ancestor, this does not necessarily mean it is also the first life that existed on earth. It's currently unknown how closely related LUCA is to the first organisms on earth, but they likely lived in a similar environment that traces back far in the history of the earth. [1, 2]

Geological history

The emergence of life and evolution most likely happened early on in the earth's history. The earth formed around 4.54 billion years ago, after the collision of two planets. This collision resulted in the formation of the earth and the moon about 4.54 to 4 billion years ago, known as the Hadean period. The collision of the two planets initially resulted in an extremely hot surface of the planet. As the surface cooled down, it passed through periods of varying temperatures, including around 100C, which is the temperature known to be suitable for current-day thermophilic organisms. [3]

A method used to infer past temperatures on earth is the analysis of oxygen isotope compositions, $\delta^{18}O$, in rocks from those periods. Rocks identified to be from the Hadean period are incredibly rare. The only rocks known to be older than 4 billion years are certain zircons, located in west Australia. These zircons are the only direct evidence of the environmental conditions during this period.

The data derived from these zircons, and other rocks between 4 and 2,6 billion years ago, suggest that the surface temperatures on earth were below 200C and relatively constant between 4.4 and 2.6 billion years ago. These temperatures are suitable for the existence of

liquid water and low enough for condensation to cause the formation of oceans. [4] This suggests that large bodies of liquid water existed for longer periods of time between 4.4 and 4 billion years ago, resulting in an ocean-dominated earth. Conditions during the Hadean period are therefore considered suitable for the emergence of life. [3, 4]

One could use this geological information to look further into what type the environment LUCA lived in. LUCA's genome and metabolism could have had several options depending on the environment. As discussed before, with temperatures ranging between 150 to 200 C and pH values around 10, alkaline hydrothermal vents contain the most likely conditions for LUCA to have lived in. Including the availability of several necessary chemicals for sustained sources of redox potential for energy production, these vents constituted of suitable geochemically active conditions. Therefore LUCA seemed to have been a phototroph or chemo-autotroph, using an anoxygenic photosynthetic metabolism. [5, 6, 7]

Even though a geological perspective can give rise to context regarding LUCA's environment, there is still little solid evidence to determine if early life on earth existed in a pre-RNA, RNA, RNA/protein, or DNA/RNA/protein world. As a consequence, this makes it harder to estimate LUCA's genomic makeup and its metabolism, as it could have used only RNA, or both RNA and DNA. It might have been quite similar to prokaryotes, or it could even have been similar to a virus. [5, 8]

Genetic history

While geology is one important way to trace back in time, another possibility of doing so is through genetics, as genomes record history as well. Even though this, just like geology, has difficulties tracing back in history due to most of the records having been erased over time. In geology, this is due to place tectonics, in genomics different evolutionary processes have erased and changed the evolutionary record. The genetic information that has been preserved in prokaryotic genomes is often difficult to read. Out of the sequenced genes, it's difficult to determine which are ancient and thus it's difficult to predict what genes LUCA could have possessed. [2]

One of the most common evolutionary processes that have changed the evolutionary record is lateral gene transfer (LGT). That is because in early evolution LGT, moving genetic

information between diverse organisms would be the primary mechanism for evolution. This genetic mixing makes it difficult to trace back the origin of certain genes. For example, as discussed by Woese [9], early cells, called progenotes, could be quite different from modern-day cells. While these progenotes possessed the necessary components of gene expression, replication, and cell division, they had very small genomes, and thus a simple metabolism. These metabolisms however could have been diverse. With these cells likely not having a cell wall, and with LGT as the primary evolutionary mechanism, any new improvement in metabolism could be easily spread and shared, resulting more in a communal way of evolution. Therefore LUCA may not have been a singular species. This differs substantially from modern-day evolution, where a distinct species often evolves as a single lineage through vertical descent. Here LGT plays a much smaller role, often only significant in bacterial evolution. [9]

Despite the challenges that arise using this genomic approach, it still has many different approaches for tracing back history. The most common method of studying the history of evolution through genetics is called phylogeny. Hence, phylogeny is what is used to trace back genes and construct the tree of life. Therefore phylogeny could be a useful tool in order to trace back the nature of LUCA and how it gave rise to the domains of life**.** [7, 10] After all, LUCA stands at the very root of this phylogenetic tree of life, stressing the importance of resolving this tree in order to elucidate its properties and better understand early evolution.

Phylogenetic methods

Many different approaches to constructing this tree of life exist, all of which encounter several challenges and produce results leading to different conclusions. For example, a classical approach to do so is to look for genes that are common in all domains of life. If certain genes are equally distributed, they are more likely to be ancient genes and thus might have been present in LUCA as well. If these genes are used to construct phylogenetic gene trees, they might elucidate LUCA's genetic makeup**.** [2]

However, even using these common genes to attempt to resolve the tree leads to debate. For a long time, the tree of life was thought to consist of three major domains of life, prokaryotes, archaea, and eukaryotes, known as either the universal tree of life, or the 3D tree. [11, 12, 13, 14] More recent studies have challenged the idea of a 3D tree, claiming that a tree consisting

of two major domains, prokaryotes and archaea with eukaryotes evolving within archaea, may be more accurate. This tree is known as the eocyte tree, or the 2D tree. [15, 16, 17] Others have challenged the method of using the 30 common genes to construct the tree, exploring different options for resolving the tree of life. [1, 2]

The objective of this thesis is to unravel LUCA's potential characteristics. To do so, I will compare several approaches to constructing the tree of life, discuss important research and debates within these options, and interpret the results to determine LUCA's most likely properties.

Determining the characteristics of LUCA and pointing out the controversies

Common genes

One way of attempting to construct a universal tree has been to look for genes that are common in nearly all genomes, the genetic core. This is because the genes that are commonly shared across many organisms may trace back to LUCA. Generally, these genes are shared amongst archaea and bacteria but are not found in any eukaryotes, as they came about later on. [9]

A selection of around 30 to 40 genes, mostly consisting of ribosomal proteins, have been used to make trees for the past 20 years. [1, 12] The selection of which genes are included in this genetic core varies greatly. For example, Charlebois [18] has analyzed 147 different prokaryotic genomes and has shown that about 34 mostly identical genes are found in all of these prokaryotes. Most of these genes are those involved in translation, and not in metabolic pathways. Other genes seem to be missing, like other subunits for RNA polymerase and multiple ribosomal proteins.

This method of using common genes also has some issues to consider. Hansmann [12], for example, has raised concerns about the way that available genomic data is used to align genomes and construct phylogenetic trees. She has discussed about 35 genes, consisting of mostly ribosomal proteins. These proteins however were not well conserved, resulting in alignments with large gaps over the entire length. However, as most of these 35 proteins did contain one or two regions that are well conserved, the proteins appeared to be alignable to some decree. Depending on which parts of the poorly aligned sites have been excluded the results of the phylogenetic analysis and bootstrap support for its branches vary greatly.

Another issue is that it could be debated whether any of these common genes evolved through vertical descent or LGT. This is important to note, as using these genomes, and consequent gene trees to reconstruct organismal phylogenetic trees assumes that they have undergone little to no LGT, which may not have been true.

If they were to have been inherited vertically, Woese [9] concludes that LUCA would have had at least, ribosomal RNA, tRNA synthetase, aminoacyl–tRNA synthetases (AARS) and that LUCA would likely have been a prototroph capable of nitrogen fixation, sulfur oxidation and

reduction, using the tricarboxylic acid cycle, a polysaccharide metabolism and using flagella to move around. [9]

On the other hand, these genes could have been shared through LGT. This makes tracing these genes back much more complicated. Using them to infer species trees may result in inaccurate trees.

The domains of life

Within this method, there is debate about whether the tree of life consists of two or three domains. On one hand, the tree has been regarded as a tree consisting of three domains. These domains would have originated from the two lineages stemming from LUCA. One lineage leads to bacteria, the second leads to a common ancestor of archaea and eukaryotes. This is the 3D tree, a hypothesis described by Woese.

On the other hand, some researchers view the tree as a 2D tree, known as the eocyte hypothesis. In this hypothesis, it is suggested that eukaryotes stem from within a sub-group of archaea. This results in a tree where one lineage still leads to prokaryotes, and the other leads not to a common ancestor of archaea and eukaryotes, but directly to archaea. [11]

The most widely accepted hypothesis is that of the 3D tree, suggested by Woese and colleagues. This tree was constructed around 1980, by comparing sequences of ribosomal RNA. The usage of rRNA is considered ideal because of several reasons. Its sequence only slowly changes, it's experimentally tractable and also quite resistant to LGT. This rRNA-based tree is therefore considered to be the universal tree, shown in Fig. 1. [10]

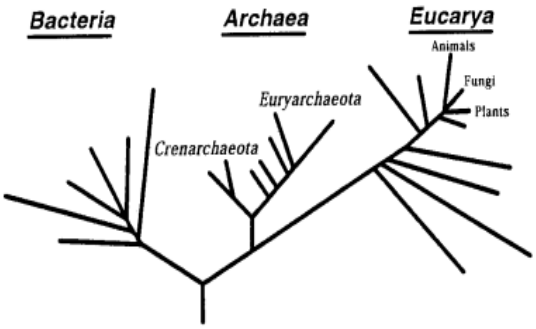


Fig. 1. The rRNA-based universal tree, consisting of 3 domains; Bacteria, Archaea and Eucaryotes. [10]

Later on, sequences of many more genes became available. However, the trees inferred from them often resulted in differences compared to the rRNA tree, most likely because of LGT. To help understand the impact of LGT on phylogenetic trees, two types of trees were compared. One being the rRNA tree, barely affected by LGT and the other being gene trees from AARSs, which have been heavily affected by LGT. Despite their differences, the majority of the AARSs gene trees showed a similar underlying pattern as that of the rRNA tree. [10]

Another argument made by Woese to support the hypothesis of the original rRNA tree is that the impact of LGT may not be so important. On a smaller species level LGT can have a big impact on blurring the lines between species. Here LGT results in a species tree that is more chimeric history, rather than consisting of clear taxonomic distinctions. However on higher levels of taxonomy, like domains, where species are grouped together, LGT has less effect on blurring the lines between domains. (Fig. 2) [10]

a) Taxonomic species tree b) Chimeric species tree c) Chimeric domain tree

Fig. 2. Illustration of the potential impact of LGT on phylogenetic trees. a) shows the traditional taxonomic species tree, where each arrow represent a species that is distinct from the others. B) shows a species tree where LGT between 'species' is common, blurring the lines between them. C) shows a higher level of taxonomy, domains. Despite LGT, the tree still shows three distinct domains. [19]

For these reasons, the universal rRNA-based tree can be regarded as an accurate portrayal of organismal genealogy. Despite the jumbling of genetic history from primitive cells engaging in much LGT, the tree still accurately represents the first stage of distinction between lineages. That's because, at some point, sub-populations emerged from the ancestor community. Within these sub-populations, communal evolution through LGT was still ubiquitous. However, between the two communities LGT became more and more unusual, thus resulting in the tree with two distinct groups; prokaryotes and the ancestor of eukaryotes and archaea, resulting in the 3D tree of life. [10]

Over time, phylogenetic methods have improved. Using these improved methods, it seems that eukaryotic core genes are placed within the Archaea. This supports a new hypothesis, namely that only 2 domains of life exist, prokaryotes and archaea. (Fig. 3) Eukaryotes either arose from within archaea or through partnership of archaea with prokaryotes.

One way that phylogenetic methods have improved is that they can mitigate some of the problems that arise when constructing trees. Testing the core genes with improved methods is therefore critical to the debate about the 3D - 2D tree. Some of the older methods carried

unrealistic assumptions. It was assumed that the base composition, the GC-content, across lineages was homogenous, meaning that they were comparable and that the evolutionary rate across sites was constant. However for some genes, like sub-units of rRNA, these assumptions do not apply. Some have sites that are both slow and fast evolving and where the GC-content between the domains differ greatly. Reconstructing trees based on these assumptions can lead to false or misleading results. [15]

Some studies tried to reduce the effects of these problems. By using methods that are less affected by these problems, the resulting tree tends to favour the 2D tree. [20, 21] For example, Foster [28] also showed that rRNA analyzed with standard methods supported the 3D tree. These results were lost or less valid when using models that accounted for a non-homogenous composition, instead increasing the support for the 2D tree.

Another problem is called long branch association (LBA). Here distantly related lineages are assumed to be more closely related than they actually are. This arises when the lineages have undergone a large number of changes that causes them to appear similar, rather than them being similar by descent. Long branches may not share the same evolutionary history, but still cluster together. This is especially the case with parsimony methods. The genes used for analysis for the tree of life often have long branches.

Tourasse [23] showed that, by accounting for a variable substitution rate, the tree consistently changed from a 3D tree, where archaea are monophyletic, to an eocyte-like tree. Using trees that assume a constant rate across sites resulted in an underestimation of the branch lengths as fast-evolving sites are overlooked. Therefore these trees may be biased towards the 3D tree, as recovering a monophyletic archaeal group may be due to underestimating the branch length.

Overall it seems that the models that don't assume homogeneous base composition and constant substitution rates across sites fit data much better than simpler models. This may make them less prone to LBA. However, not many analyses have been done on the core genes using these models, though they do tend to recover the 2D tree with variable support.

Due to these arguments Williams [15] claims that the eocyte tree is now the best-supported hypothesis. If this tree is correct, the closest relatives of the eukaryotic nuclear lineage may hold more clues in order to figure out the tree of life. These relatives were determined using eukaryotic signature proteins (ESPs) and the group containing most of them is called the TACK-

archaea, a superphylum group consisting of the Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota. These TACK-archaea also suggest the existence of only two primary domains (Fig. 3.), Archaea and Bacteria, and it suggests that eukaryotes may be chimeric in nature. [15]

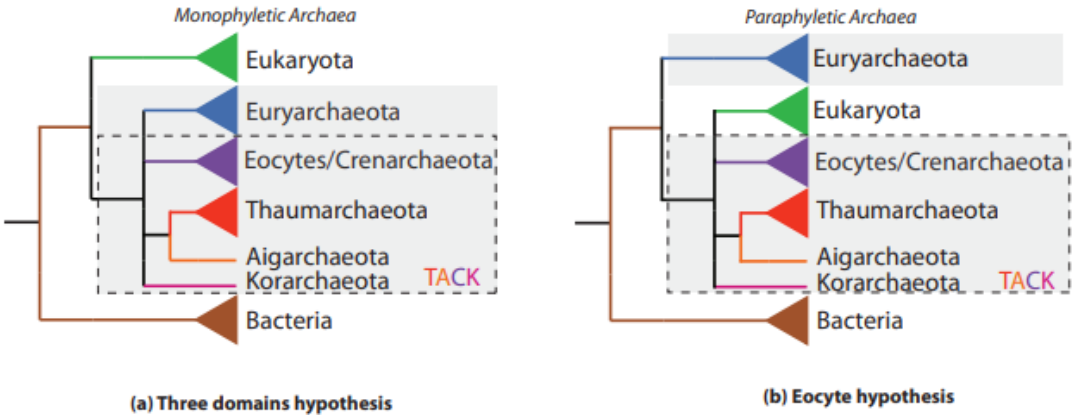


Fig. 3. Side by side comparison of the 3D-tree (a), where eukaryotes are placed outside of the TACK archaea, and the 2D-tree (b), where eukaryotes are placed within the TACK-archaea. [15]

Lasek-Nesselquist [24] supports the claim of eukaryotes emerging from within the TACK superphylum. Using different strategies to mitigate the assumptions of homogenous base composition and constant evolutionary rate, their data showed overall support for the 2D topology for the tree of life, where eukaryotes emerged from within archaea as a sister group to the TK or TCK clade for all models as long as at least one of the strategies was used.

Spang and colleagues [16, 17] have discovered and analyzed new archaeal lineages related to the TACK-archaea, called Loki-archaea. These were the first known lineages within the later-called Asgard superphylum, a sister clade to the TACK archaea. These were discovered close to Loki's Castle, a field of active hydrothermal vents in the Arctic Mid-Ocean Ridge. Their analyzes, accounting for the before mentioned assumptions, showed that eukaryotes may have emerged from within these archaea and that they contained relatively more ESPs compared to other archaeal lineages, thus supporting the 2D tree.

Da Cunha [11, 13] argues against these results using several claims. The first claim is that the genomic data that was used for the analysis may contain contaminations from organisms that are distantly related as they used environmental DNA rather than DNA taken directly from cells, though Spang refutes the idea that the data is contaminated and instead insists that the sequences are from closely related strains as they compared the proteins from the dataset against environmental genomes from NCBI to show it's not contaminated.

The second claim is that the relation between Loki-archaea and eukaryotes becomes unlikely after including data from fast-evolving species and after removing one protein, elongation factor 2. However, even after removing this protein from the dataset, they found that some trees were still in favour of the 2D tree, just not the specific relation between eukaryotes and Loki-archaea. Other trees, like the one from RNA polymerase, recovered the 3D tree. The other claim is that the 2D tree may be affected by LBA, but Spang argues that the opposite is the case.

Alternative selection of genes

Rather than using the genes that are universal to trace back which are ancient, one could also look at the phylogeny of the genes themselves to determine which genes are ancient. As discussed before, LGT can lead to false conclusions about the tree of life. Many of these studies tried to determine which genes are common between archaea and prokaryotes, and make a distinction of which of those are common because of LGT and which through VD. If LGT has a big impact, many gene trees will not reflect the universal tree, while for the genes inherited vertically, the trees may match. Using this method means assuming that certain genes are inherited through vertical descent and that the vertically inherited gene trees accurately reflect the universal tree. However this also means that a big part of the genomic data doesn't get used as only around 30 genes are considered universal, (Fig. 4A)

Selecting the core genes also becomes more difficult because sequencing becomes more accurate. As the methods improve, any small difference in the genes between organisms becomes more apparent. This requires the criteria for selecting ancient genes to become more relaxed.

To improve the selection of common genes Weiss [1] used two different methods. One is to select the genes that are present in both domains, but are not per se universal, (Fig. 4B) resulting in about 11.000 genes.

The other is to further narrow these selected genes down by including two more criteria. One is that the genes should be present in at least two phylum-level clades, and the other is that the trees should preserve monophyly. Using these extra criteria makes it more restrictive as the genes would have only been inherited through LGT under very specific conditions, which is quite unlikely. If these genes were inherited through LGT they would have needed to first transfer across a domain, which is quite common, but then also transfer inside the domain to

different clades, without afterwards transferring back across a domain, which is uncommon. With this method, 355 genes remain candidates for LUCA to have possessed, (Fig. 4C).

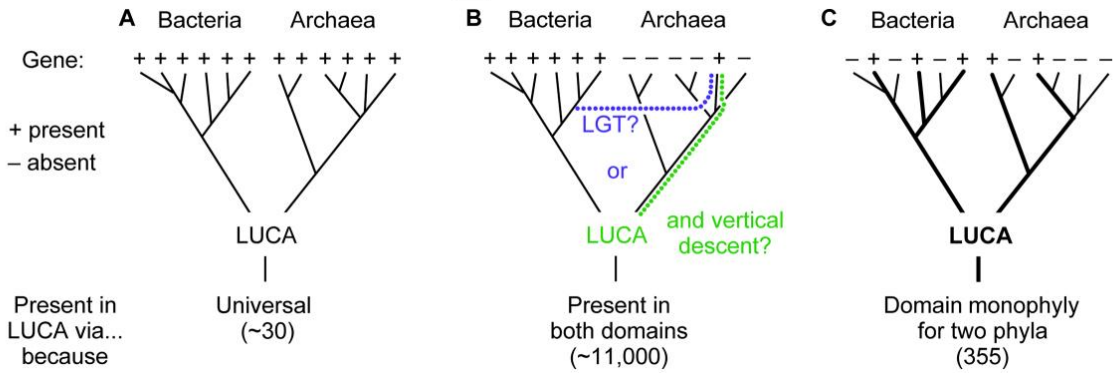


Fig. 4. Side by side comparison of the three methods of determining LUCA genome. 'A' shows the traditional universal genes method, 'B' shows the presence in both domains method, and how it is unclear if a gene is acquired through LGT of VD and 'C' shows the more restrictive version of B where it also needs to be present in two phyla per domain without losing monophyly. [1]

Weiss [1] concludes several things about LUCA's physiology and environment by interpreting the genes found with this method. The first is that LUCA was an anaerobe, as many O2-sensitive enzymes were found in its metabolism. These enzymes would not have functioned properly in the presence of oxygen.

The second was that LUCA was an autotroph. It probably used the Acetyl-CoA pathway to assimilate carbon. This simple chemical reaction is exergonic in nature, which could have been a common occurrence at that time. This pathway reduces CO2 with H2, to CO and a methyl group. Notable here is that the carbon monoxide dehydrogenase (CODH), which synthesizes CO, is similar between archaea and prokaryotes, while the synthetization of the methyl group uses different one-carbon units in archaea and prokaryotes. Autotrophs with CODH can obtain ATP using this reaction, while those without cannot. Inferring from LUCA's lifestyle and the presence of CODH in LUCA, it was probably an autotroph.

The third conclusion was about the way LUCA would have generated ATP. As some relics of ATP synthase were traced back to LUCA, like acetyl phosphate and sub-units of the rotor-stator of ATP synthase, no direct proteins of the proton-pump itself were found to be present. This may not have been a problem as it has been theorized that LUCA may have lived in alkaline hydrothermal vents. [5, 7] In that case, it may have been able to use its environment to its advantage. That's because in these alkaline hydrothermal vents there were naturally occurring pH gradients, which it could have utilized to synthesize ATP.

As such, the last conclusion that Weiss [1] made using the new selecting method is that LUCA lived in alkaline hydrothermal vents, rich in sulfur. Three other things support this claim. One is that LUCA had an enzyme called reverse gyrase. Several archaea and all thermophilic bacteria possess this enzyme, supporting the claim that it lived in a hot place and was likely a thermophilic organism. It also had several proteins related to sulfur, for example sulfur transferases, supporting the claim that its environment was rich in sulfur. The second thing supporting this claim is that high-temperature environments are more conducive to chemical reactions than lower-temperature environments. The third is that the closest archaea, methanogens, and the closest bacteria, clostridia, share many of the same features. Both are anaerobic and make use of the acetyl-COA pathway. In modern times these both live in hydrothermal environments, supporting the hypothesis that LUCA lived in alkaline hydrothermal vents.

Conclusions

Many approaches to elucidating LUCA's properties exist. This includes the phylogenetic methods used to build the tree of life, and the criteria used to select ancient genes. I will shortly summarize the findings presented and discuss which were most likely, and what consequences this would have on what LUCA would have been like.

One of the main controversies discussed is about the tree of life. On one side there is the 3D tree, which has been extensively researched. It was based on ribosomal RNA, for which the effects of LGT are quite low. On the other side, however, there is the 2D tree. The tree tends to take this shape if at least some of the effects of LBA, base composition, and evolutionary rate are accounted for. In that case, the picture often changes to favour the 2D tree of life. Though not yet as extensively researched as the 3D tree of life, these methods are considered to be more accurate, leading me to believe that the 2D tree is more plausible than the 3D tree.

Another obstacle to understanding LUCA is the method of selecting which genes to use for making these trees. The criteria for selecting core genes have been quite varied and many different studies use slightly different variations for which genes to select. As the sequencing and phylogenetic methods continue to improve, the criteria for selecting ancient genes will have to change as well. If these criteria stay the same, LUCA would have no genes left as most would not be considered common genes anymore due to minor differences becoming apparent as sequencing improves.

Selecting only the 30 common genes that are shared between all doesn't account for LGT as much and it would exclude genes that may have been lost along the way, thus not being common everywhere. Using only the genes that are common in all also ignores a lot of genetic data that could give more information about LUCA. Therefore, Weiss has suggested a new way of selecting core genes that aren't as affected by LGT as the 30 common genes method. It broadens the scope and gives an image of LUCA that seems more accurate to what an organism would actually look like. This way of selecting ancient genes without LGT affecting the results much is favourable to me. It would be interesting to see future research that may have other methods of selecting genes than the two that were discussed.

Using the 30 common genes method to analyze LUCA results in it possessing several ribosomal proteins, AARS, and tRNA synthetase. The results commonly suggest that LUCA was an anaerobic phototroph capable of nitrogen fixation, sulfur oxidation, and reduction and that it may have used the tricarboxylic acid cycle and a polysaccharide metabolism. This method also suggests LUCA to have been a phototroph

Weiss' method, however, suggests that LUCA was a chemo-autotroph, producing energy from chemical reactions, without relying on sunlight. Both methods claim LUCA to be anaerobic and capable of using sulfur in its metabolism. Weiss also suggests that LUCA's metabolism would have been a relatively simple metabolism, as some experiments show that the end products or the intermediate products of the acetyl-CoA pathway synthesize spontaneously at temperatures favourable to life. [25] LUCA also didn't need a proton-pump, as it could use the environment instead. This metabolism seems simpler than that of the 30 common genes method which suggests a polysaccharide metabolism.

Overall the pictures sketched with these methods do have a few differences but don't majorly contradict each other. Weiss's method gives rise to new insights into LUCA that should be further investigated. This new selection of genes suggests many arguments that happen to align with the geological predictions as well, supporting this newer method for selecting ancient genes.

In addition, Woese's suggestion about the communal evolution of progenotes in the early stages of life seems quite plausible as well. This suggestion isn't contradicted in the newer methods either as these progenotes would have a simple metabolism as well.

Based on all of this, I concluded that LUCA was a community of progenotes, that used both RNA and DNA to transcribe several proteins and store genetic information. It lived in alkaline hydrothermal vents, which were most likely common at the time, and provided a suitable temperature and enough chemical potential to produce energy without the need to evolve a complex metabolism. It was an anaerobic autotroph, that had a metabolism that made use of the acetyl-CoA pathway and relied on its environmental conditions to produce ATP.

References

[1] Weiss M.C., Preiner M., Xavier J.C., Zimorski V., Martin W.F. The last universal common ancestor between ancient Earth chemistry and the onset of genetics. (2018). PLoS Genet 14(8): e1007518.

[2] Martin W.F., Weiss M.C., Neukirchen S., Nelson-Sathi S., Sousa F.L. Physiology, phylogeny, and LUCA. (2006). Microb Cell.;3(12):582-587.

[3] Sleep N.H., Zahnle K., Neuhoff P.S. Initiation of clement surface conditions on the earliest Earth. (2001). Proc Natl Acad Sci U S A.;98(7):3666-72. D

[4] Valley J.W., Peck W.H., King E.M., Wilde S.A. (2002). A cool early Earth. Geology, 30, 351-354.

[5] Lane N., Allen J.F., Martin W.F. (2010). How did LUCA make a living? Chemiosmosis in the origin of life. BioEssays: news and reviews in molecular, cellular and developmental biology, 32 4, 271-80.

[6] Marakushev S.A., Belonogova O.V. Emergence of the chemoautotrophic metabolism in hydrothermal environments and the origin of ancestral bacterial taxa. (2011). Dokl Biochem Biophys 439, 161.

[7] Weiss M.C., Sousa F.L., Mrnjavac N., Neukirchen S., Roettger M., Nelson-Sathi S., Martin W.F. The physiology and habitat of the last universal common ancestor. (2016). Nature Microbiology, 1.

[8] Becerra A., Delaye L, Islas S., Lazcano A. The Very Early Stages of Biological Evolution and the Nature of the Last Common Ancestor of the Three Major Cell Domains. (2007). Annual Review of Ecology, Evolution, and Systematics. 38. 361-379.

[9] Woese C. The universal ancestor. (1998). Proc Natl Acad Sci U S A.;95(12):6854-9.

[10] Woese C.R. Interpreting the universal phylogenetic tree. (2000). Proc Natl Acad Sci U S A.;97(15):8392-6.

[11] 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 13(6): e1006810.

[12] Hansmann S., Martin W. Phylogeny of 33 ribosomal and six other proteins encoded in an ancient gene cluster that is conserved across prokaryotic genomes: influence of excluding poorly alignable sites from analysis. (2000). Int J Syst Evol Microbiol.;50 Pt 4:1655-1663.

[13] Da Cunha V., Gaia M., Nasir A., Forterre P. Asgard archaea do not close the debate about the universal tree of life topology. (2018). PLoS Genet 14(3): e1007215.

[14] Martin W., Russell M.J. On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. (2003). Philos Trans R Soc Lond B Biol Sci.; 358(1429):59-83; discussion 83-5.

[15] Williams T., Foster P., Cox C., et al. An archaeal origin of eukaryotes supports only two primary domains of life. (2013). Nature 504, 231–236.

[16] Spang A., Saw J., Jørgensen S., Zaremba-Niedzwiedzka K., Martijn J., Lind A., Eijk R., Schleper C., Guy L., Ettema T. Complex archaea that bridge the gap between prokaryotes and eukaryotes. (2015). Nature. advance online publication.

[17] Spang A., Eme L., Saw J.H., Caceres E.F., Zaremba-Niedzwiedzka K., Lombard J., et al. Asgard archaea are the closest prokaryotic relatives of eukaryotes. (2018). PLoS Genet 14(3): e1007080.

[18] Charlebois R.L., Doolittle W.F. Computing prokaryotic gene ubiquity: rescuing the core from extinction. (2004). Genome Res.; 14(12):2469-77.

[19] Doolittle W.F., & Zhaxybayeva O. On the origin of prokaryotic species. (2009). Genome research, 19 5, 744-56.

[20] Lake J. A. Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances. (1994). Proc Natl Acad Sci U S A 91, 1455-1459.

[21] Yang Z., Roberts D. On the use of nucleic acid sequences to infer early branchings in the tree of life. (1995). Mol Biol Evol.;12(3):451-8.

[22] Foster P.G., Cox C.J., Embley T.M. The primary divisions of life: a phylogenomic approach employing composition-heterogeneous methods. (2009). Philos Trans R Soc Lond B Biol Sci.; 364(1527):2197-207.

[23] Tourasse N.J., Gouy M. Accounting for evolutionary rate variation among sequence sites consistently changes universal phylogenies deduced from rRNA and protein-coding genes. (1999). Molecular phylogenetics and evolution, 13 1, 159-68.

[24] Lasek-Nesselquist E., Gogarten J.P. The effects of model choice and mitigating bias on the ribosomal tree of life. (2013). Mol Phylogenet Evol.; 69(1):17-38.

[25] Varma S.J., Muchowska K.B., Chatelain P., et al. Native iron reduces CO_2 to intermediates and end-products of the acetyl-CoA pathway. (2018). Nat Ecol Evol 2, 1019–1024.