Journal of Molecular Evolution
https://doi.org/10.1007/s00239-024-10199-4

REVIEW

On the Nature of the Last Common Ancestor: A Story from its Translation Machinery

Mario Rivas[1] · **George E. Fox**[1]

Received: 26 February 2024 / Accepted: 22 August 2024

Abstract

The Last Common Ancestor (LCA) is understood as a hypothetical population of organisms from which all extant living creatures are thought to have descended. Its biology and environment have been and continue to be the subject of discussions within the scientific community. Since the first bacterial genomes were obtained, multiple attempts to reconstruct the genetic content of the LCA have been made. In this review, we compare 10 of the most extensive reconstructions of the gene content possessed by the LCA as they relate to aspects of the translation machinery. Although each reconstruction has its own methodological biases and many disagree in the metabolic nature of the LCA all, to some extent, indicate that several components of the translation machinery are among the most conserved genetic elements. The datasets from each reconstruction clearly show that the LCA already had a largely complete translational system with a genetic code already in place and therefore was not a *progenote*. Among these features several ribosomal proteins, transcription factors like IF2, EF-G, and EF-Tu and both class I and class II aminoacyl tRNA synthetases were found in essentially all reconstructions. Due to the limitations of the various methodologies, some features such as the occurrence of rRNA posttranscriptional modified bases are not fully addressed. However, conserved as it is, non-universal ribosomal features found in various reconstructions indicate that LCA's translation machinery was still evolving, thereby acquiring the domain specific features in the process. Although progenotes from the pre-LCA likely no longer exist recent results obtained by unraveling the early history of the ribosome and other genetic processes can provide insight to the nature of the pre-LCA world.

Keywords Progenote · LUCA · LCA · Ribosome · Translation

> "Therefore, I should infer from analogy that probably all the organic beings which have ever lived on this Earth have descended from some one primordial form…".
>
> —Chales R. Darwin, on The Origin of species, first British edition, 1859—

From the *Progenote* to the Last Common Ancestor

It was Carl Woese and George E. Fox who first proposed that all extant organisms which inhabit Earth can be grouped into one of the three major domains, Bacteria, Archaea, and Eukarya (Woese and Fox 1977a). Their subsequent trifurcated, explicitly unrooted ribosomal RNA (rRNA) tree suggests that all organisms within these domains derived from a common ancestral life form (Fox et al. 1980). In this regard, all modern organisms share the central dogma. This includes the translation machinery, the genetic code, the essential features of genome replication and gene expression, as well as many essential metabolic reactions and basic ATP energy production. Variation from these essential features is usually attributed to environmental adaptations posterior to the divergence of the three major biological domains (Becerra et al. 2007). Nevertheless, there are differences like the exclusive membrane lipid composition between

Handling Editor: Arturo Becerra.

✉ Mario Rivas
mrivasmedrano@uh.edu; mario_rivas@ciencias.unam.mx

[1] Department of Biology and Biochemistry, University of Houston, Houston, TX 77204-5001, USA

Published online: 11 September 2024

Bacteria and Archaea that remain an unsolved mystery (Wächtershäuser 2003; Peretó et al. 2004).

With the discovery of the Archaea Domain, Woese and Fox (1977a) recognized the existence of what seemed like significant differences in the translation machinery, such as the larger size of the large subunit in Eukaryotes and Archaea. They envisioned that "the basic cell type" would necessarily be on a level of complexity far simpler than what is seen in modern prokaryotes. Such entities would be in the process of evolving the genotype–phenotype relationship and might appropriately be called *progenotes* (Woese and Fox 1977b). One may notice that the hypothetical *progenote* was not initially envisioned as the ancestral population at the trifurcation of the emerging 16S rRNA phylogeny, since a comprehensive 16S rRNA tree including the Archaea was not available until 1980 (Fox et al. 1980). Over the years as described in detail by Gogarten and Olendzenski (1999), the *progenote* has been envisioned as either an organizational level that preceded prokaryotes or as the last common ancestor of extant life (LCA). We strongly encourage the community to instead use the term "*progenote*" as it was initially envisioned by Woese and Fox (1977b).

The confusion stems in part because *progenotes* are defined as entities still 'evolving the relationship between genotype and phenotype.' But what does that mean? Are there any likely *progenotes* out there that we can study? Recent efforts to understand the evolutionary origin of the translation system might provide a window to these earlier times (Hsiao et al. 2009; Petrov et al. 2014, 2015). The extant ribosome consists of a large and small subunit. The large subunit is responsible for the synthesis of the peptide bonds, whereas the small subunit implements the machinery for coded synthesis (Steitz 2008; Fox 2010; Wilson and Doudna Cate 2012). Recent studies have shown that a 67-nucleotide RNA derived from the current large subunit can in fact catalyze peptide bond formation without coding (Bose et al. 2021, 2022). Over the years, this entity was named the protoribosome by Yonath and her co-workers (Bashan et al. 2003; Agmon et al. 2005, 2009; Agmon 2009, 2016; Davidovich et al. 2010; Krupkin et al. 2011; Huang et al. 2013; Yonath 2017). We suggest that the protoribosome can appropriately be considered to be an example of a *progenote*. Later, if this *progenote* became increasingly complex it may have partnered with a second *progenote*, which would perhaps be an ancestor of the small subunit. The resulting entity could now have sequence preference and coding too. A timeline which leads from a purely chemical stage to an RNA-World or an RNA-peptide world to the LCA now makes sense (Fig. 1). These larger populations would likely have included subpopulations of *progenotes* that provided aspects of the central dogma. The existence of *progenotes*, able to catalyze the peptide bond, which is fossilized within extant rRNA, implies that we are now able to study further the early evolution of the code and even its very origin.

The origin and early evolution of the code remain elusive despite the code itself being deciphered over 50 years ago. The evidence suggests that it has evolved into a code that minimizes the effects of point mutations and mistranslation, in a sense, "*the genetic code is one in a million*" (Freeland and Hurst 1998). It has also been proposed that the extant code arose from stereochemical interactions between RNA and the amino acids. Then, it expanded by biosynthetic modification and finally was optimized through codon reassignment. Three complementary forces of its evolution that most likely fixed the code in the LCA of modern organism (Knight et al. 1999). A genetic code fully in place within the LCA is a common conclusion that arises independently from different lines of evidence, like the compositional analysis of ribosomal proteins made by Fournier and Gogarten (2007). There, they identified a subset of amino acids that are most likely the most recent additions to the code and suggested that the expansion of the code may have enhanced the transition from an RNA-based to a protein-based life prior to LCA's times. The implications of the emergence and posterior assembly of two different families of aminoacyl tRNA synthetases that may have

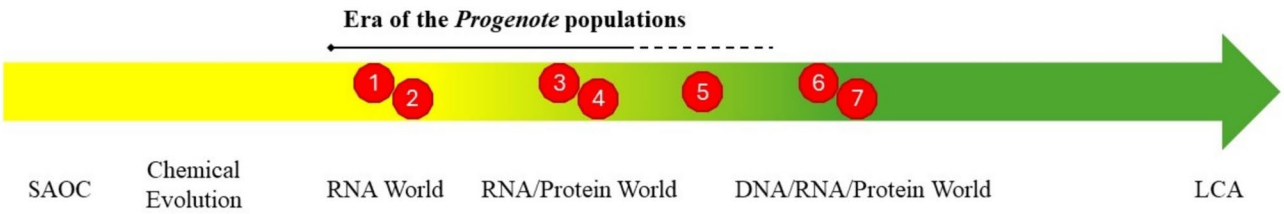

Fig. 1 Timeline from the synthesis and accumulation of organic compounds (SAOC) to the chemical evolution to an RNA-World containing *progenotes* to the LCA. (1) Emergence and early evolution of the PTC (protoribosome) (LSU). First small uncoded peptides. (2) Emergence and early evolution of the decoding center (SSU). (3) Association of the LSU and the SSU. First proto-tRNA adaptors. (4) Early evolution of the genetic code. Early evolution of the aminoacyl tRNA synthetases. Class I first and then Class II. (5) Transition to DNA as genetic material. (6) Consolidation of the tRNAs. (7) Consolidation of the aminoacyl tRNA synthetases. Class I and II

significantly affected the code will be discussed at length later in the text.

Many Names, Too Many Interpretations

Fitch and Upper (1987) coined the term Cenancestor to name the ancestral organism from which Archaea, Bacteria, and Eukaryotes descend. The last universal common ancestor (LUCA) was, for several years, the most commonly used term by which such entities were known. It was initially used in the first report that reconstructed LUCA's genetic content that included genomic data from an Archaea (Kyrpides et al. 1999). There were other proposals like the last universal cellular ancestor made by Philippe and Forterre (1999), universal ancestor by Doolittle (2000), last common community by Line (2002), and urancestor by Kim and Caetano-Anollés (2011), among others. These terms are of course not synonyms because they reflect the particular vision of the authors and the ongoing controversies about its metabolic nature, origin, and posterior evolution. As of today, the simplest commonly used term is the last common ancestor (LCA). This entity is currently understood as the ancestral population from which all extant living creatures descend. Although strictly speaking, the LCA is an inventory of the genetic characteristics that are shared among extant organisms (Becerra et al. 2007).

The constantly increasing number of completely sequenced genomes has made it possible to apply new approaches and techniques to improve the estimations of the LCA genetic content and from the latter derive its nature and environment. Several studies have used clever bioinformatic approaches to characterize the minimal set of genes present in the LCA. This has included the search for gene families instead of individual genes (Harris et al. 2003; Mirkin et al. 2003; Weiss et al. 2016b), the search of protein architectures (Yang et al. 2005; Ranea et al. 2006), as well as individual protein domains and motifs (Yang et al. 2005; Kim and Caetano-Anollés, 2011). Such reconstructions and their techniques exploited the intrinsic features within the primary sequences to improve the search of homologous sequences along the phylogeny. Despite the universality, centrality, and antiquity of the noncoding RNA genes, such as the rRNA and/or the tRNA, they are not the subject of these types of homology searches due to the technical challenges, such as their small sizes and their 4 letter alphabet. Instead, other approaches have been made like the comparison of atomic-resolution structures of ribosomes from distant phyla. Thanks to the latter, it was suggested that approximately 90% of the extant prokaryotic rRNA forms an ancestral conserved core, which is the structural and functional unit of all known cytoplasmatic ribosomes (Bernier et al. 2018).

LCA's Genetic Content Reconstructions

It was long before fully sequenced genomes were available that people started to wonder about the nature of the last common ancestor (LCA). The very first attempt to reconstruct the genetic elements most likely present in the LCA was done more than 30 years ago by Lazcano et al. (1992). It was an exhaustive and comprehensive review of the literature available at the time. There, it was suggested that the machineries of DNA replication, gene expression, and basic biosynthetic pathways are essentially the same among Archaea, Bacteria, and Eukaryotes. Thus, concluding that "*the LCA was very much like a contemporary prokaryote at its fundamental level of biological complexity*."

The release of the first completely sequenced genomes started a new era of comparisons and searches of sequences from genes and proteins among different organisms. The comparison of parasitic bacterial genomes, from *Haemophilus influenzae* and *Mycoplasma genitalium*, resulted in the first estimation of the minimal gene set necessary to sustain essential cellular functions. Unfortunately, the absence of homologous genes from several key proteins involved in DNA replication led the authors to a likely faulty conclusion; they suggested that the LCA had an RNA genome (Mushegian and Koonin 1996). This interpretation can be attributed to an underestimation of the gene content due to the parasitic lifestyle of *H. influenza* and *M. genitalium*. Later, *Methanococcus jannaschii* was the first Archaea whose genome was completely sequenced (Bult et al. 1996). This allowed the first estimation of the genome content of the LCA that included archaeal genes for comparison against bacterial and eukaryotic genes (Kyrpides et al. 1999). As a result, the authors infer that the LCA was an organism with several biochemical functions and genetic machineries similar to extant unicellular organisms.

As the completely sequenced genomes of more and more organisms became available, many research groups tried to improve the estimation of the LCA genetic content. Take, for instance, Norman Pace's group that used the Clusters of Orthologous Genes (COGs) database

to search for the universally conserved genes exclusively within fully sequenced organisms (Harris et al. 2003). This study required that highly conserved genes exhibit the same phylogenetic signal as the ribosomal RNA. The result being that most of such universal genes are related to the ribosome. Even further, such an approach oversees the effect of horizontal gene transfer (HGT), a phenomenon whose degree of impact in the reconstruction of ancestral life forms is still under debate (Doolittle 1999, 2000).

To deal with the fact that almost 90% of the COGs are incompatible with the rRNA universal tree and to reconcile gene loss, gene emergence and events of HGT several algorithms that compute parsimonious evolutionary scenarios for genome evolution were developed (Mirkin et al. 2003). The authors concluded that gene loss and HGT are major aspects that shape prokaryotic evolution with almost equal frequency. They also concluded that if LCA was a minimal free-living entity it would necessarily benefit from HGT and in a lesser way from the invention of new genes. And on the other hand, if LCA was a complex entity it would eventually benefit from differential gene loss.

A separate approach with a biological perspective was then developed. Instead of using every completely sequenced organism available, a representative sample of well-known and well-characterized organisms from Archaea, Bacteria, and Eukaryotes was chosen. This biologically curated sample also tried to exclude endosymbionts and parasites (Delaye et al. 2005). The gene complement of the LCA that was presented there is more compatible with a cellular entity that emerged prior to the divergence of the three cellular domains of life.

By taking advantage of the Structural Classification of Proteins database (SCOP), a strategy that uses the presence or absence of protein domain architecture was used to construct the phylogeny of 174 complete genomes (Yang et al. 2005). This study was grounded in the well-accepted notion that protein tertiary structure is more conserved than primary sequence and that it allows one to see deeper into the past. They reported 49 super family folds common to all genomes under scrutiny, suggesting a LCA with a sophisticated genetic inventory and gene products far beyond those from just the translation machinery. This conclusion was supported by Ouzounis et al. (2006) who suggested that the LCA possessed a complex genome similar to extant free-living prokaryotes. They implemented a search strategy based on primary sequence that suggests functional capabilities like metabolism, information processing, active membrane transport, and complex regulatory functions were among the capacities of the LCA.

The notion that three-dimensional structure comparison is more sensitive than conventional primary sequence methodologies in detecting remote homology has also been used to identify a set of ancestral protein domains most likely present in the LCA (Ranea et al. 2006). A functional analysis of such ancestral domains again reveals a genetically complex LCA, with all essential functional cellular systems in place. The latter conclusion supports previous proposals suggesting that life acquired its modern cellular characteristics before the divergence of the three domains (Doolittle 2000).

A more recent proposal suggests that the Urancestor (≈LCA) was similar to modern organisms in terms of gene content. It is also grounded in a phylogenomic study of protein domain structure and their classification into highly conserved fold super-families (Kim and Caetano-Anollés, 2011). The authors argue that despite its possession of advanced metabolic capabilities, being especially rich in nucleotide metabolism, the Urancestor had pathways for membrane synthesis and crucial elements of translation. However, it lacked fundamental elements for transcription and extracellular communication, as well as for the synthesis of deoxyribonucleotides. Therefore, its proteomic history suggests that the Urancestor is closer to a simple *progenote* population that harbored a set of modern molecular functions.

The most recent attempt to reconstruct the genetic content of the LCA also tries to derive its physiology and habitat from the defiant premise that non-universal proteins can illustrate LCA's physiology (Weiss et al. 2016b). They also support the recent two-domain tree of life hypothesis, which proposes that Eukaryotes arose from the Archaeal branch of the Prokaryote lineage (Williams et al. 2013; Raymann et al. 2015). Within this study, the authors depict the LCA as an anaerobic autotroph living in a hydrothermal setting, dependent upon geochemistry and therefore "only half-alive." Such a disruptive vision has been the subject of many rigorous revisions and vivid discussions (Gogarten and Deamer 2016; Weiss et al. 2016a). Those of course are not within the objective of the present review, but we encourage the readers to examine them and form their own opinion.

It is evident that there have been a considerable number of attempts to reconstruct the gene content of the LCA. All of them exploit completely sequenced genomes but use different approaches from primary sequence comparisons to phylogenetic strategies, to protein domain architecture, to tertiary structure searches, and even a mixture of them (Table 1). While several arrive to the conclusion that the LCA resembles an extant free-living prokaryote others point to a simpler being perhaps closer to a *progenote*. Nevertheless, they all agree on one feature that must be present in the LCA, the translation machinery. For years, it has been recognized as one ancestral element whose analysis must shed light on the earliest history of life, even predating the LCA (Agmon et al. 2005; Davidovich et al. 2009; Fox 2010; Petrov et al. 2014, 2015; Rivas and Fox 2023). We have searched throughout the ten studies described above and their results to extract their

Table 1 Comparison of the 10 reconstructions of the gene content of the LCA

Study	# of organisms	Main goal of the study	Methodology	Gene content of the LCA
Mushegian and Koonin (1996)	*Mycoplasma genitalium* *Haemophilus influenzae*	Minimal gent set determination	BLAST-P and BLAST-N	256 genes
Kyrpides et al. (1999)	1 Archaea *Methanococcus jannaschii* Unknown # of Bacteria Unknown # of Eukaryotes	*M. jannaschii* functional annotation	BLAST2 (protein)	324 proteins
Harris et al. (2003)	8 Archaea 25 Bacteria 3 Eukaryotes	Genetic core of the LCA	COGs	80 universal/50 rRNA topology
Mirkin et al. (2003)	6 Archaea 19 Bacteria 1 Eukaryote	Genome evolution/Tree topology	COGs	572 genes/COGs
Delaye et al. (2005)	8 Archaea 8 Bacteria 4 Eukaryotes	Gene complement of the LCA	Twice one-way BLAST	145 HCPs in Archaea 245 HCPs in Bacteria 283 HCPs in Eukaryotes
Yang et al. (2005)	19 Archaea 119 Bacteria 36 Eukaryotes	Determination of genome phylogenies	Protein Domain Architecture to Fold Super-Families	49 FSF
Ouzounis et al. (2005)	184 entire genomes	Reconstruction of the gene content of the LCA	Reciprocal best hits BLAST	1511 Ortholog Families
Ranea et al. (2006)	15 Archaea 85 Bacteria 14 Eukaryotes	Define the ancestral protein domains in the LCA	ORFs > HMM > CATH > SSAP	140 Ancestral Protein Domains
Kim and Caetano-Anollés (2011)	48 Archaea + 1 AOP 239 Bacteria + 71 BP + 111 BOP 133 Eukaryotes + 22 EP + 20 EOP	Reconstruct the FSF repertoires of the urancestral proteome	Phylogenetic distribution of FFs, FSFs, and Fs	max_set 152 FSFs and min_set 70 FSFs
Weiss et al. (2016b)	1847 Bacteria 134 Archaea	Reconstruct the microbial ecology of LUCA	MCL algorithm for protein family clusterization	355 Clusters of Proteins

The first column mentions the reference of each study. The second column indicates the number and the type of organisms that were included in their respective methodologies. The third column mentions the main goal of each study since not all studies have the determination of the LCA's genetic content as their main objective. The fourth column mentions the major methodology that each study followed to determine the genetic content. The fifth column indicates the number of genes/proteins/COGs/FSFs that were concluded as part of the genetic content of the LCA. Abbreviations are as follows: Archaeal obligate parasitic (AOP). Bacterial parasitic (BP). Bacterial obligate parasitic (BOP). Eukaryotic parasitic (EP). Eukaryotic obligate parasitic (EOP). Highly conserved proteins (HCPs). Clusters of orthologous genes (COGs). Fold Super-Families (FSF). Fold Families (FF). Folds (Fs). Markov chain clustering (MCL)

conserved genes, which are part of the translation machinery proposed for the LCA. Therefore, the conclusions, suggestions and speculations that will be presented in the following section are based on the comparisons of these reconstructions and their conserved genes from LCA's translation machinery.

A Common Conclusion: The Translation Machinery

As these approaches accumulate, the idea of extrapolating the consensus genetic content of the LCA emerged. As far as we know the very first attempt was LUCApedia (Goldman et al. 2012) a database that was presented as "*a unified framework for simultaneously evaluating multiple datasets related to the LCA*." It represented a tool for a quick reference in determining if a gene or a set of genes could be considered ancient. A more refined and detail attempt was recently published by Crapitto et al. (2022), they developed a series of bioinformatic and statistical procedures to compare the prediction of eight reconstructions of the genetic content of the LCA. Therein, the authors discuss that although most of the studies show a strong agreement with the consensus predictions, no single study shows even a moderate degree of similarity with any other. Of special interest is the conclusion, which derives from the consensus set, saying

that the LCA possessed a protein synthesis machinery, amino acid metabolism, and used nucleotide-derived cofactors. The latter immediately implied that the consensus set could in principle reveal the most conserved elements within the genetic content of the LCA. This is an idea that we independently explore in detail, with a more bounded scope, limited to the elements of the translation machinery.

Despite the different methodological strategies from the reconstructions of the genetic content of the LCA, all of them independently inferred that some portions of the translation machinery are among the most conserved features and therefore likely to have already been active at the time of the LCA. Although the vast majority of the elements that integrate extant translation machinery are not equally conserved among these reconstructions, by comparing the lists of each reconstruction, it was found that several key components are indeed well conserved across all of them.

The ribosome is a ribonucleoprotein complex that is regarded as the core of the translation system and it is composed of a small subunit (SSU) and a large subunit

Table 2 Highly conserved rProteins that might be within the genetic content of the LCA

rProtein	Mushegian and Koonin (1996)	Kyrpides et al. (1999)	Harris et al. (2003)	Mirkin et al. (2003)	Delaye et al. (2005)	Yang et al. (2005)	Ouzounis et al. (2006)	Ranea et al. (2006)	Kim and Caetano-Anollés (2011)	Weiss et al. (2016b)
rProteins from the LSU										
L1	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
L11	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
L14	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
L5	Y	Y	Y	Y	Y	Y	Y	–	Y	Y
L6	Y	Y	Y	Y	Y	Y	Y	Y	Y	–
L2	Y	Y	Y	Y	Y	–	Y	Y	–	Y
L22	Y	Y	Y	Y	–	Y	Y	Y	Y	–
L23	Y	Y	Y	Y	–	–	Y	Y	Y	–
L10	Y	–	Y	Y	–	Y	–	Y	Y	–
L13	Y	Y	Y	Y	–	–	Y	–	Y	–
L18	Y	Y	Y	Y	–	–	–	Y	Y	–
L29	Y	Y	Y	Y	–	–	–	Y	Y	–
L3	Y	Y	Y	Y	–	–	–	Y	–	–
L4	Y	Y	Y	Y	–	–	–	–	Y	–
L15	Y	–	Y	Y	–	Y	–	–	Y	–
rProteins from the SSU										
S5	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
S8	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
S7	Y	Y	Y	Y	Y	Y	Y	Y	Y	–
S10	Y	Y	Y	Y	Y	–	Y	Y	Y	Y
S13	Y	Y	Y	Y	Y	Y	Y	Y	Y	–
S2	Y	Y	Y	Y	Y	Y	–	Y	Y	–
S11	Y	Y	Y	Y	Y	–	Y	Y	–	Y
S19	Y	Y	Y	Y	Y	–	Y	Y	Y	–
S3	Y	Y	Y	Y	Y	–	Y	–	Y	–
S9	Y	–	Y	Y	Y	–	Y	Y	–	Y
S4	–	Y	Y	Y	Y	–	Y	Y	–	–
S12	Y	Y	Y	Y	Y	–	Y	–	–	–
S15	Y	Y	Y	Y	–	Y	–	Y	–	–
S17	Y	Y	Y	Y	–	–	Y	–	–	–

15 rProteins from LSU and 14 rProteins from SSU are most likely part of the genetic content of the LCA according to 10 studies. Proteins are named as they are included within each reconstruction and the databases used to identify them. A "Y" symbol indicates the presence, and the "–" symbol indicates its absence for each study. We grouped the rProteins according to the number of studies that identified them, going from 10/10 to 5/10 for the LSU and for the SSU. There are more rProteins that are reported as part of the genetic content of the LCA, but they are mentioned by less than five studies; therefore, we considered that they are more likely to be false positives. Nevertheless, we show them in the supplementary material for the benefit of the reader

(LSU) (Steitz 2008; Fox 2010; Wilson and Doudna Cate 2012). The structure and contents of these subunits include both conserved and variable features. In prokaryotes, the SSU contains one 16S ribosomal RNA (rRNA) and ~ 21 ribosomal Proteins (rProteins), while the LSU contains 5S and 23S rRNAs and ~ 33 rProteins (Steitz 2008; Wilson and Doudna Cate 2012). Several ribosomal proteins including L1, L2, L5, L6, L11, L14, L22, S2, S5, S7, S8, S10, S11, S13, and S19 are found essentially in all the reconstructions (Table 2). LSU and SSU rProteins that are listed above were found in 80–100% of the studies, nevertheless all rProteins detected by even a single reconstruction are included in the supplementary tables. Such degree of conservation immediately suggests that the ribosome of the LCA was already exploiting the coexistence with large globular peptides. Even further, such observation implies that the LCA´s ribosome has already gone through several stages of rProtein evolution (Kovacs et al. 2017).

These highly conserved rProteins associate in various functional places within the extant ribosome (Schuwirth et al. 2005; Lin et al. 2015). L2 and L14 are located between the two subunits, most likely assisting in the ribosome assembly. L22 is associated with the last part of the exit tunnel, likely assisting with the folding and expulsion of the nascent polypeptide. S7 and S5 touch the SSU in helix 28, while S11 touches helix 45. Both helixes are at the core of the decoding center whose dependence on rProteins for appropriate folding has been established (Schedlbauer et al. 2021). Ribosomal proteins L5, S13, and S19 establish a bridge between the 5S, the 16S, and the 23S rRNAs. Of special interest are those conserved rProteins that potentially contained posttranscriptional modifications, such as methylation, acetylation, and/or phosphorylation. As shown by Ilag et al. (2005) phosphorylated rProteins bind more tightly to the rRNA scaffold. Highly conserved rProteins L5, L11, L22, S5, S7, S11, and S13 are phosphorylated in extant ribosomes (Soung et al. 2009). Although, enzymes capable of such modifications were not reported by any of the reconstructions. Therefore, it is less probable that they were modified by the enzymatic mechanisms of the LCA instead such modifications must have evolved later.

Transfer RNAs (tRNAs) are crucial components of the translation machinery. tRNAs are the "adaptors" that establish the complementarity between the mRNA codons and the amino acid alphabet (Crick 1958). The tRNAs are charged with specific amino acids by highly specialized enzymes called aminoacyl tRNA synthetases (aaRS). Each tRNA is specific for one amino acid and each aaRS specifically recognizes both the tRNA and its cognate amino acid. Extant proteins are made from 20 canonical amino acids although some variations occur like pyrrolysine a non-canonical amino acid (Srinivasan et al. 2002; Nozawa et al. 2009). Hence, there are at least an equal number of aaRSs. In the charging reaction, one canonical amino acid is ester bonded to its cognate tRNA by one specific aaRS. Based on their primary sequences and their tertiary structure, two classes (I and II) of aaRS were identified and usually, there are 10 aaRSs in each class (Eriani et al. 1990; Ibba and Söll 2000). Their distinct protein fold domains (Rossman and ATP-grasp, respectively) suggest they have a separate evolutionary history. A plausible explanation for this observation could imply earlier *progenotes* may have only one of the two classes, which would likely have restricted the usable code before they meet each other. Both class I and class II aaRSs are detected as elements from the genetic content of the LCA by most reconstructions (Table 3). Class I and class II aaRSs were detected by at least 50% of the studies, but 8 out of 10 aaRSs within each class were detected from 70 to 100% of the studies. This high conservation pattern strongly suggests that the translation machinery of the LCA had an almost complete version of the extant genetic code, if not fully consolidated.

Contrary to what has been documented to occur with the rest of the translation machinery, several horizontal gene transfer (HGT) events appear to have dominated the history of the aaRSs. Using sequence reconstruction and phylogenetic analyses, Fournier et al. (2011) recognized the role of several HGT events prior and after the divergence of the LCA, revealing a complex and intricate evolution of the aaRSs. Thus, explaining why their phylogeny does not always match to that of other highly conserved phylogenetic markers, like rProteins or the rRNA, nor between aaRSs themselves. Nevertheless, as complex as it appears, most of its evolution seems to have happened before the time of the LCA. Analysis of atypical forms of aaRSs suggests that ancient HGT have occurred within sister groups of a diverse community that inhabited different niches at the same time the LCA existed and even before. Further, the paralog sequence reconstruction of isoleucyl- and valyl-RSs suggest that they did not co-evolve with the genetic code, and these amino acids were already part of it before the cognate aaRSs diverged from their common ancestor prior to the LCA (Fournier et al. 2011).

Protein synthesis is promoted by several translation factors, which bind transiently to the ribosome during the phases of the translation process (Lipmann 1969; Kaziro 1978). Although translation can be carried out without translation factors (Spirin 1978), the rates are many orders of magnitude below the ones of the modern system. Recently, several observations supported the spontaneous translation view (Shoji et al. 2006; Konevega et al. 2007); however, without the translation factors protein synthesis is very slow and error prone. Table 4 shows that several initiation factors and elongation factors are among the proposed genetic content of the LCA by several reconstructions. Initiation Factors 1 and 2 were reported by 60 and 80%

Table 3 Highly conserved aminoacyl tRNA synthetases that might be among the genetic content of the LCA

Aminoacyl tRNA synthetase	Mushegian and Koonin (1996)	Kyrpides et al. (1999)	Harris et al. (2003)	Mirkin et al. (2003)	Delaye et al. (2005)	Yang et al. (2005)	Ouzounis et al. (2006)	Ranea et al. (2006)	Kim and Caetano-Anollés (2011)	Weiss et al. (2016b)
Class I Aminoacyl tRNA synthetases										
aaRS-Ile (I)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
aaRS-Arg (R)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
aaRS-Leu (L)	Y	Y	Y	Y	Y	Y	Y	Y	Y	–
aaRS-Val (V)	Y	Y	Y	Y	Y	Y	Y	Y	Y	–
aaRS-Met (M)	Y	Y	Y	Y	Y	Y	Y	Y	Y+	–
aaRS-Glu (E)	Y	–	Y	Y	Y	Y	Y	Y	Y+	Y
aaRS-Gln (Q)	–	Y	Y	Y	Y	Y	Y	Y	Y+	Y
aaRS-Trp (W)	Y	Y	Y	Y	Y	Y	Y	Y	Y+	–
aaRS-Tyr (Y)	Y	Y	Y	Y	–	Y	–	Y	Y+	–
aaRS-Cys (C)	Y	–	–	Y	Y	Y	Y	–	Y+	–
aaRS-Lys (K)	Y*	–	–	Y*	–	Y	Y*	–	Y+	–
Class II Aminoacyl tRNA synthetases										
aaRS-Pro (P)	Y	Y	Y	Y	Y	Y	Y	Y	Y+	Y
aaRS-Phe (F)	Y	Y	Y	Y	Y	Y	Y	Y	Y+	Y
aaRS-Ser (S)	Y	Y	Y	Y	Y	Y	Y	Y	Y+	–
aaRS-Thr (T)	Y	Y	Y	Y	Y	Y	Y	Y	Y	–
aaRS-Ala (A)	Y	Y	Y	Y	Y	Y	Y	–	Y	Y
aaRS-His (H)	Y	Y	Y	Y	Y	Y	Y	Y	Y+	–
aaRS-Asp (D)	Y	Y	–	Y	Y	Y	Y	–	Y+	Y
aaRS-Gly (G)	Y	Y	–	–	–	Y	Y	Y	Y+	Y
aaRS-Lys (K)	Y*	–	–	Y*	Y	Y	Y*	–	Y+	–
aaRS-Asn (N)	Y	–	–	–	Y	Y	–	–	Y+	Y

Class I and class II aminoacyl tRNA synthetases that are reported as part of the genetic content of the LCA. A "Y" symbol indicates the presence, and the "–" symbol indicates its absence for each study. We grouped aaRS according to the number of studies that identified them for each class, going from 10/10, 9/10, 8/10, and so on. *There was no clear distinction if the identified LysRS belong to Class I or Class II within these studies. (+) These aaRSs were not presented as separate gene hits by Kim and Caetano-Anollés (2011)

Table 4 Highly conserved translation factors listed within the reconstructions of the genetic content of the LCA

Ribosomal Factors	Mushegian and Koonin (1996)	Kyrpides et al. (1999)	Harris et al. (2003)	Mirkin et al. (2003)	Delaye et al. (2005)	Yang et al. (2005)	Ouzounis et al. (2006)	Ranea et al. (2006)	Kim and Caetano-Anollés (2011)	Weiss et al. (2016b)
Initiation factors										
IF2	Y	Y	Y	Y	Y	Y	Y	–	Y	–
IF-1	Y	–	Y	Y	–	–	Y	Y	Y	–
Elongation Factors										
EF-G	Y	Y	Y	–	Y	Y	Y	Y	Y	Y
EF-Tu	Y	–	Y	–	Y	Y	Y	Y	Y	Y

Translation factors that are mentioned by several reconstructions of the genetic content of the LCA. A "Y" symbol indicates the presence, and the "–" symbol indicates its absence for each study. We grouped the translation factors according to the number of studies that identified them. There are more translation factors that are reported as part of the genetic content of the LCA, but some of them are exclusive to certain phylogenetic groups; therefore, they are more likely to be false positives. Nevertheless, we show them in the supplementary material for the benefit of the reader

of the studies, respectively, while Elongation Factors Tu and G were detected in 80 and 90% of the cases. The high conservation of these factors suggests that the LCA's translation machinery was already fine-tuned and dependent on translation factors that enhance its translation rates and fidelity.

Furthermore, it is well known that several translation factors hydrolyze GTP. They belong to a family of GTP-hydrolyzing enzymes that is related to a larger family of ATP-hydrolyzing enzymes (Leipe et al. 2002). IF2, EF-G, and EF-Tu are among these factors which can be called translation GTPases which are indispensable for the extant translation machinery, as can be clearly seen by the fact that nowadays Archaea and Bacteria possess several backup copies in their genomes (Margus et al. 2007). IF2, EF-G, and EF-Tu are listed as part of the genetic content of the LCA by most reconstructions (Table 4). The latter immediately suggests that the LCA must have possessed an efficient energy production system able to meet the ribosome's extensive GTP demand.

It is widespread knowledge that several modifications to the RNA nucleobases are common features of ribosomal, messenger, transfer, and other noncoding RNAs. Such modifications are believed to play key roles in regulation, molecular recognition, and structural stabilization. Methylation, acetylation, and the chemical transformation of uridine into pseudouridine are examples of the most common. Some of these modifications even occur in the ribonucleobases that form the peptidyl transferase center, the very core of the translation machinery (Tirumalai et al. 2021). Mature tRNAs are also extensively modified (McCloskey and Crain 1998; Byrne et al. 2010). Such modifications are so typical that they have influenced the name of the tRNA structures. For instance, the D loop is named after the 5,6-dihydrouridines and the T loop after the thymine preceding a pseudouridine (Ψ). tRNA pseudouridine synthase catalyze the conversion of uridine to Ψ at several positions in the tRNA. tRNA pseudouridine synthase is also listed by 90% of the reconstructions among the genetic content of the LCA (Table S5) suggesting that fine tune regulation such as nucleotide modification with structural and functional implications was operational but still evolving within the translation machinery of the LCA.

Final Remarks

As reviewed here, there have been at least ten attempts to reconstruct the genetic content of the LCA since the release of the first completely sequenced organisms. Some tried to derive LCA's physiology and others even extrapolated into the possible environment that the LCA inhabited. They have implemented different methodological strategies and used a variety of completely sequenced organisms from the three domains of life. This was a deliberate effort to compensate for the methodological *biases* inherent to previous strategies. Although their specific conclusions are not always compatible, all ten studies have been successful in noticing, to some extent, that multiple aspects of the translation system are highly conserved.

Key elements of the translation machinery are found in essentially every reconstruction.

Many ribosomal proteins from the SSU and the LSU, representing over half of the total number of rProteins of extant prokaryotic ribosomes, are included within the genetic content of the LCA by the majority of reconstructions (Table 2). Almost every aaRS from both classes are listed by most reconstructions as present in the genome of the LCA (Table 3). GTP-dependent translation factors like IF2, EF-G, and EF-Tu are also regarded as elements of the

LCA translation system by most reconstructions (Table 4). Even the tRNA pseudouridine synthase is included among the genetic content of the LCA by many reconstructions (Table S5). All these key features indicate that LCA's translation machinery closely resembled a contemporary prokaryotic system. It contained many rProteins, a full set of aaRSs which directly imply a modern genetic code, several energy-dependent elongation factors and even specific nucleotide modification enzyme that most likely enhanced structure and may have influenced the overall translation rate alongside the translation factors.

These reconstructions focus on the distribution of the proteins rather than the rRNAs or tRNAs. Typically, RNA secondary structure is defined by the occurrence of helical regions. When comparing large RNAs one can monitor the presence or absence as well as the extent of conservation of each individual helical region. The history of the individual helical regions can also be correlated with ribosomal protein interaction sites. When this is done, some aspects of rRNA structure are essentially universal and could be useful to include them in future LCA's reconstructions as it has done for the rRNA (Petrov et al. 2014, 2015; Bernier et al. 2018). Independent comparisons of atomic-resolution ribosomal structures suggested that the size of the LCA's rRNA must be closer to extant prokaryotic ribosomes (Bernier et al. 2018). Efforts to include RNA structural features that were useful in reconstructing the history of rRNA, such as GNRA tetraloops (Hsiao et al. 2009), A-minor interactions (Bokov and Steinberg 2009), and insertion fingerprints (Petrov et al. 2014), await future studies focused on the translation machinery of the LCA.

A usual conclusion is that the genetic code is essentially universal and likely already established in the LCA. This should be determined by looking at tRNA populations, but instead it has been inferred from the conserved sequences of key enzymes like the aaRSs. As described above, the history of the aaRSs turned out to be intricate due to several HGT events that occurred after the divergence of the main cellular domains (Fournier et al. 2011). More important are those HGT events that occurred before that divergence since they molded the extant genetic code, whose origin and early evolution seem to be the consequence of multiple forces acting differentially throughout their history (Knight et al. 1999).

Different reconstructions produce different scenarios for the physiology and possible environment of the LCA. Whether it was an autotroph or a heterotroph is still unclear. What every reconstruction agrees on is that it possessed an almost fully functional translation machinery that closely resembles a modern prokaryotic one. Therefore, we propose that the prokaryotic nature of the LCA was largely established when the divergence of the three main cellular domains occurred.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00239-024-10199-4.

Acknowledgements This work was supported in part by a subcontract to the University of Houston from NASA Contract 80NSSC18K1139 under the Center for the Origin of Life, at the Georgia Institute of Technology.

Author Contributions MRM and GEF conceived the work. MRM conducted all comparative work. Results were discussed with GEF. Both authors contributed equally to the writing process and preparation of the manuscript.

Declarations

Conflicts of Interest The authors declare no conflict of interest.

References

Agmon I (2009) The dimeric proto-ribosome: structural details and possible implications on the origin of life. Int J Mol Sci 10(7):2921–2934. https://doi.org/10.3390/ijms10072921

Agmon I (2016) Could a proto-ribosome emerge spontaneously in the prebiotic world? Molecules 21(12):1701. https://doi.org/10.3390/molecules21121701

Agmon I, Bashan A, Zarivach R, Yonath A (2005) Symmetry at the active site of the ribosome: structural and functional implications. Biol Chem 386(9):833–844. https://doi.org/10.1515/BC.2005.098

Agmon I, Davidovich C, Bashan A, Yonath A (2009) Identification of the prebiotic translation apparatus within the contemporary ribosome. Nat Proceed. https://doi.org/10.1038/npre.2009.2921.1

Bashan A, Zarivach R, Schluenzen F, Agmon I, Harms J, Auerbach T, Baram D, Berisio R, Bartels H, Hansen HAS, Fucini P, Wilson D, Peretz M, Kessler M, Yonath A (2003) Ribosomal crystallography: Peptide bond formation and its inhibition. Biopolymers 70(1):19–41. https://doi.org/10.1002/bip.10412

Becerra A, Delaye L, Islas S, Lazcano A (2007) The very early stages of biological evolution and the nature of the last common ancestor of the three major cell domains. Annu Rev Ecol Evol Syst 38(1):361–379. https://doi.org/10.1146/annurev.ecolsys.38.091206.095825

Bernier CR, Petrov AS, Kovacs NA, Penev PI, Williams LD (2018) Translation: the universal structural core of life. Mol Biol Evol 35(8):2065–2076. https://doi.org/10.1093/molbev/msy101

Bokov K, Steinberg SV (2009) A hierarchical model for evolution of 23S ribosomal RNA. Nature 457(7232):977–980. https://doi.org/10.1038/nature07749

Bose T, Fridkin G, Bashan A, Yonath A (2021) Origin of life: chiral short RNA chains capable of non-enzymatic peptide bond formation. Isr J Chem 61(11–12):863–872. https://doi.org/10.1002/ijch.202100054

Bose T, Fridkin G, Davidovich C, Krupkin M, Dinger N, Falkovich AH, Peleg Y, Agmon I, Bashan A, Yonath A (2022) Origin of life: protoribosome forms peptide bonds and links RNA and protein dominated worlds. Nucl Acids Res 50(4):1815–1828. https://doi.org/10.1093/nar/gkac052

Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, Kerlavage

AR, Dougherty BA, Tomb JF, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM et al (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. Science 273(5278):1058–1073. https://doi.org/10.1126/science.273.5278.1058

Byrne RT, Konevega AL, Rodnina MV, Antson AA (2010) The crystal structure of unmodified tRNAPhe from *Escherichia coli*. Nucleic Acids Res 38(12):4154–4162. https://doi.org/10.1093/nar/gkq133

Crapitto AJ, Campbell A, Harris A, Goldman AD (2022) A consensus view of the proteome of the last universal common ancestor. Ecol Evol 12(6):e8930. https://doi.org/10.1002/ece3.8930

Crick FH (1958) On protein synthesis. Symp Soc Exp Biol 12:138–163

Davidovich C, Belousoff M, Bashan A, Yonath A (2009) The evolving ribosome: from non-coded peptide bond formation to sophisticated translation machinery. Res Microbiol 160(7):487–492. https://doi.org/10.1016/j.resmic.2009.07.004

Davidovich C, Belousoff M, Wekselman I, Shapira T, Krupkin M, Zimmerman E, Bashan A, Yonath A (2010) The proto-ribosome: an ancient nano-machine for peptide bond formation. Isr J Chem 50(1):29–35. https://doi.org/10.1002/ijch.201000012

Delaye L, Becerra A, Lazcano A (2005) The last common ancestor: What's in a name? Orig Life Evol Biosph 35(6):537–554. https://doi.org/10.1007/s11084-005-5760-3

Doolittle WF (1999) Lateral genomics. Trends Cell Biol 9(12):M5-8

Doolittle WF (2000) The nature of the universal ancestor and the evolution of the proteome. Curr Opin Struct Biol 10(3):355–358. https://doi.org/10.1016/s0959-440x(00)00096-8

Eriani G, Delarue M, Poch O, Gangloff J, Moras D (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature 347(6289):203–206. https://doi.org/10.1038/347203a0

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. https://doi.org/10.1101/sqb.1987.052.01.085

Fournier GP, Gogarten JP (2007) Signature of a primitive genetic code in ancient protein lineages. J Mol Evol 65(4):425–436. https://doi.org/10.1007/s00239-007-9024-x

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(6):621–632. https://doi.org/10.1007/s11084-011-9261-2

Fox GE (2010) Origin and evolution of the ribosome. Cold Spring Harb Perspect Biol 2(9):a003483–a003483. https://doi.org/10.1101/cshperspect.a003483

Fox GE, Stackebrandt E, Hespell R (1980) The phylogeny of prokaryotes. Science 209(4455):457–463

Freeland SJ, Hurst LD (1998) The genetic code is one in a million. J Mol Evol 47(3):238–248. https://doi.org/10.1007/PL00006381

Gogarten JP, Olendzenski L (1999) The progenote. In: Creighton T (ed) Encyclopedia of molecular biology. Wiley. ISBN 0471-15302-8

Gogarten JP, Deamer D (2016) Is LUCA a thermophilic progenote? Nat Microbiol 1:16229. https://doi.org/10.1038/nmicrobiol.2016.229

Goldman AD, Bernhard TM, Dolzhenko E, Landweber LF (2012) LUCApedia: a database for the study of ancient life. Nucl Acids Res 41(D1):D1079–D1082. https://doi.org/10.1093/nar/gks1217

Harris JK, Kelley ST, Spiegelman GB, Pace NR (2003) The genetic core of the universal ancestor. Genome Res 13(3):407–412. https://doi.org/10.1101/gr.652803

Hsiao C, Mohan S, Kalahar BK, Williams LD (2009) Peeling the onion: ribosomes are ancient molecular fossils. Mol Biol Evol 26(11):2415–2425. https://doi.org/10.1093/molbev/msp163

Huang L, Krupkin M, Bashan A, Yonath A, Massa L (2013) Protoribosome by quantum kernel energy method. Proc Natl Acad Sci 110(37):14900–14905. https://doi.org/10.1073/pnas.1314112110

Ibba M, Söll D (2000) Aminoacyl-tRNA synthesis. Annu Rev Biochem 69:617–650. https://doi.org/10.1146/annurev.biochem.69.1.617

Ilag LL, Videler H, McKay AR, Sobott F, Fucini P, Nierhaus KH, Robinson CV (2005) Heptameric (L12)6/L10 rather than canonical pentameric complexes are found by tandem MS of intact ribosomes from thermophilic bacteria. Proc Natl Acad Sci 102(23):8192–8197. https://doi.org/10.1073/pnas.0502193102

Kaziro Y (1978) The role of guanosine 5'-triphosphate in polypeptide chain elongation. Biochem Biophys Acta 505(1):95–127. https://doi.org/10.1016/0304-4173(78)90009-5

Kim KM, Caetano-Anollés G (2011) The proteomic complexity and rise of the primordial ancestor of diversified life. BMC Evol Biol 11(1):140. https://doi.org/10.1186/1471-2148-11-140

Knight RD, Freeland SJ, Landweber LF (1999) Selection, history and chemistry: the three faces of the genetic code. Trends Biochem Sci 24(6):241–247. https://doi.org/10.1016/S0968-0004(99)01392-4

Konevega AL, Fischer N, Semenkov YP, Stark H, Wintermeyer W, Rodnina MV (2007) Spontaneous reverse movement of mRNA-bound tRNA through the ribosome. Nat Struct Mol Biol 14(4):318–324. https://doi.org/10.1038/nsmb1221

Kovacs NA, Petrov AS, Lanier KA, Williams LD (2017) Frozen in time: the history of proteins. Mol Biol Evol 34(5):1252–1260. https://doi.org/10.1093/molbev/msx086

Krupkin M, Matzov D, Tang H, Metz M, Kalaora R, Belousoff MJ, Zimmerman E, Bashan A, Yonath A (2011) A vestige of a prebiotic bonding machine is functioning within the contemporary ribosome. Philos Trans R Soc Lond Ser B, Biol Sci 366(1580):2972–2978. https://doi.org/10.1098/rstb.2011.0146

Kyrpides N, Overbeek R, Ouzounis C (1999) Universal protein families and the functional content of the last universal common ancestor. J Mol Evol 49(4):413–423. https://doi.org/10.1007/PL00006564

Lazcano, A., Fox, G. E., & Oro, J. (1992). Life before DNA: the origin and early evolution of early Archean cells. In: Mortlock RP (ed) The evolution of metabolic function. CRC Press. ISBN 0-8493-8863-5

Leipe DD, Wolf YI, Koonin EV, Aravind L (2002) Classification and evolution of P-loop GTPases and related ATPases. J Mol Biol 317(1):41–72. https://doi.org/10.1006/jmbi.2001.5378

Lin J, Gagnon MG, Bulkley D, Steitz TA (2015) Conformational changes of elongation factor G on the ribosome during tRNA translocation. Cell 160(1–2):219–227. https://doi.org/10.1016/j.cell.2014.11.049

Line MA (2002) The enigma of the origin of life and its timing. Microbiology 148(1):21–27. https://doi.org/10.1099/00221287-148-1-21

Lipmann F (1969) Polypeptide chain elongation in protein biosynthesis. Science 164(3883):1024–1031. https://doi.org/10.1126/science.164.3883.1024

Margus T, Remm M, Tenson T (2007) Phylogenetic distribution of translational GTPases in bacteria. BMC Genom 8:15. https://doi.org/10.1186/1471-2164-8-15

McCloskey JA, Crain PF (1998) The RNA modification database–1998. Nucl Acids Res 26(1):196–197. https://doi.org/10.1093/nar/26.1.196

Mirkin BG, Fenner TI, Galperin MY, Koonin EV (2003) Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. BMC Evol Biol 3(1):2. https://doi.org/10.1186/1471-2148-3-2

Mushegian AR, Koonin EV (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc Natl Acad Sci 93(19):10268–10273. https://doi.org/10.1073/pnas.93.19.10268

Nozawa K, O'Donoghue P, Gundllapalli S, Araiso Y, Ishitani R, Umehara T, Söll D, Nureki O (2009) Pyrrolysyl-tRNA synthetase-tRNA(Pyl) structure reveals the molecular basis of orthogonality. Nature 457(7233):1163–1167. https://doi.org/10.1038/nature07611

Ouzounis CA, Kunin V, Darzentas N, Goldovsky L (2006) A minimal estimate for the gene content of the last universal common ancestor—exobiology from a terrestrial perspective. Res Microbiol 157(1):57–68. https://doi.org/10.1016/j.resmic.2005.06.015

Peretó J, López-García P, Moreira D (2004) Ancestral lipid biosynthesis and early membrane evolution. Trends Biochem Sci 29(9):469–477. https://doi.org/10.1016/j.tibs.2004.07.002

Petrov AS, Bernier CR, Hsiao C, Norris AM, Kovacs NA, Waterbury CC, Stepanov VG, Harvey SC, Fox GE, Wartell RM, Hud NV, Williams LD (2014) Evolution of the ribosome at atomic resolution. Proc Natl Acad Sci 111(28):10251–10256. https://doi.org/10.1073/pnas.1407205111

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 112(50):15396–15401. https://doi.org/10.1073/pnas.1509761112

Philippe H, Forterre P (1999) The rooting of the universal tree of life is not reliable. J Mol Evol 49(4):509–523. https://doi.org/10.1007/pl00006573

Ranea JAG, Sillero A, Thornton JM, Orengo CA (2006) Protein superfamily evolution and the last universal common ancestor (LUCA). J Mol Evol 63(4):513–525. https://doi.org/10.1007/s00239-005-0289-7

Raymann K, Brochier-Armanet C, Gribaldo S (2015) The two-domain tree of life is linked to a new root for the Archaea. Proc Natl Acad Sci 112(21):6670–6675. https://doi.org/10.1073/pnas.1420858112

Rivas M, Fox GE (2023) How to build a protoribosome: structural insights from the first protoribosome constructs that have proven to be catalytically active. RNA 29(3):263–272. https://doi.org/10.1261/rna.079417.122

Schedlbauer A, Iturrioz I, Ochoa-Lizarralde B, Diercks T, López-Alonso JP, Lavin JL, Kaminishi T, Çapuni R, Dhimole N, de Astigarraga E, Gil-Carton D, Fucini, P, Connell SR (2021) A conserved rRNA switch is central to decoding site maturation on the small ribosomal subunit. Sci. Adv. https://doi.org/10.1126/sciadv.abf7547

Schuwirth BS, Borovinskaya MA, Hau CW, Zhang W, Vila-Sanjurjo A, Holton JM, Cate JHD (2005) Structures of the bacterial ribosome at 3.5 Å resolution. Science 310(5749):827–834. https://doi.org/10.1126/science.1117230

Shoji S, Walker SE, Fredrick K (2006) Reverse translocation of tRNA in the ribosome. Mol Cell 24(6):931–942. https://doi.org/10.1016/j.molcel.2006.11.025

Soung GY, Miller JL, Koc H, Koc EC (2009) Comprehensive analysis of phosphorylated proteins of *Escherichia coli* ribosomes. J Proteome Res 8(7):3390–3402. https://doi.org/10.1021/pr900042e

Spirin AS (1978) Energetics of the ribosome. Prog Nucl Acid Res Mol Biol 21:39–62. https://doi.org/10.1016/s0079-6603(08)60266-4

Srinivasan G, James CM, Krzycki JA (2002) Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. Science 296(5572):1459–1462. https://doi.org/10.1126/science.1069588

Steitz TA (2008) A structural understanding of the dynamic ribosome machine. Nat Rev Mol Cell Biol 9(3):242–253. https://doi.org/10.1038/nrm2352

Tirumalai MR, Rivas M, Tran Q, Fox GE (2021) The peptidyl transferase center: a window to the past. Microbiol Mol Biol Rev 85(4):e0010421. https://doi.org/10.1128/MMBR.00104-21

Wächtershäuser G (2003) From pre-cells to Eukarya–a tale of two lipids. Mol Microbiol 47(1):13–22. https://doi.org/10.1046/j.1365-2958.2003.03267.x

Weiss MC, Neukirchen S, Roettger M, Mrnjavac N, Nelson-Sathi S, Martin WF, Sousa FL (2016a) Reply to "Is LUCA a thermophilic progenote?" Nat Microbiol 1:16230. https://doi.org/10.1038/nmicrobiol.2016.230

Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S, Martin WF (2016b) The physiology and habitat of the last universal common ancestor. Nat Microbiol 1(9):16116. https://doi.org/10.1038/nmicrobiol.2016.116

Williams TA, Foster PG, Cox CJ, Embley TM (2013) An archaeal origin of eukaryotes supports only two primary domains of life. Nature 504(7479):231–236. https://doi.org/10.1038/nature12779

Wilson DN, Doudna Cate JH (2012) The structure and function of the eukaryotic ribosome. Cold Spring Harb Perspect Biol 4(5):a011536–a011536. https://doi.org/10.1101/cshperspect.a011536

Woese CR, Fox GE (1977a) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci 74(11):5088–5090

Woese CR, Fox GE (1977b) The concept of cellular evolution. J Mol Evol 10(1):1–6. https://doi.org/10.1007/BF01796132

Yang S, Doolittle RF, Bourne PE (2005) Phylogeny determined by protein domain content. Proc Natl Acad Sci 102(2):373–378. https://doi.org/10.1073/pnas.0408810102

Yonath A (2017) Quantum mechanic glimpse into peptide bond formation within the ribosome shed light on origin of life. Struct Chem 28(5):1285–1291. https://doi.org/10.1007/s11224-017-0980-5