International Journal of Astrobiology 15 (1): 27–31 (2016)
doi:10.1017/S1473550415000464 © Cambridge University Press 2015

A proposal of the proteome before the last universal common ancestor (LUCA)

Sávio Torres de Farias[1], Thais Gaudêncio Rêgo[2] and Marco V José[3]

[1] *Campus I, Departamento de Biologia Molecular, Laboratório de Genética Evolutiva Paulo Leminsk, Universidade Federal da Paraíba, João Pessoa, Paraíba, Brazil e-mail: stfarias@yahoo.com.br*
[2] *Campus I, Departamento de Informática, Universidade Federal da Paraíba, João Pessoa, Paraíba, Brazil*
[3] *Theoretical Biology Group, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, México D.F. 04510, México*

Abstract: The search for understanding the biological nature of the last universal common ancestor (LUCA) has been a theoretical challenge and has sparked intense debate in the scientific community. We reconstructed the ancestral sequences of tRNAs in order to test the hypothesis that these molecules originated the first genes. The results showed that the proteome before LUCA may have been composed of basal energy metabolism, namely, compounds with three carbons in the glycolytic pathway, which operated as a distribution centre of substrates for the development of metabolic pathways of nucleotides, lipids and amino acids. Thus, we present a proposal for metabolism in organisms before LUCA that was the initial core for the assembly of further metabolic pathways.

Received 6 August 2015, accepted 20 October 2015

Key words: last universal common ancestor, proteome, tRNA.

Introduction

The notion of a last universal common ancestor (LUCA) of all living forms is one of the focal points of biology involved with questions about the origin of life. Since the famous publication of Charles Darwin, On the Origin of Species by Means of Natural Selection, or The Preservation of Favoured Races in the Struggle for Life, in 1859, the concept of common descent to all forms of life triggered discussions about the components of this biological entity (Penny & Poole 1999; Forterre *et al.* 2005; Mushegian 2008; Glansdorff *et al.* 2008, 2009; Kim & Caetano-Anollés 2011; Morange 2011; Goldman *et al.* 2013). However, when we think about LUCA, we are already referring to a complex organism with a DNA-based genome with developed processing information pathways, thus, as a system of complex informational flows and efficient energetic metabolism (Forterre & Philippe 1999; Di Giulio 2003). However, the biological evolutionary process had already begun long before the LUCA. In this scenario, this organism or biological entity (sometimes called progenotes or ribocytes), had as informational molecule the RNA and its proto-metabolism and the compartmentalization by a membrane had begun. Woese (1998) developed the concept of progenotes, primitive cell before LUCA, with limited functions. They possessed an imprecise translation process generating statistical proteins, which upon further refinements may have fixed the system as we know it today. Thus, we cannot think of a single progenote that originated LUCA, but a community of sub-systems that developed separately and at a certain time of development began to cooperate, which culminated in the union of these sub-systems forming what we now know as LUCA. One of the central properties of this progenote was the ability to translate certain types of information (Eigen & Schuster 1978), followed by the development of basal metabolic pathways. Thus, we contend that not all steps of the modern metabolic pathways were present in a progenote, and that different steps of the metabolism appeared in different progenotes. The increasingly interdependence of different progenotes gave origin to more complex metabolic pathways.

Another important question is the nature of the molecules that constituted the progenotes. Large informational molecules did not have sufficient stability at this stage (Eigen & Schuster 1978; Woese 1998), and the roles of the synthesized proteins were the binding and stabilization than that of complex enzymatic functions. They acted more like binding domains, that later, evolved to modern proteins. Eigen & Winkler-Oswatitsch (1981) proposed that the initial molecular events were carried out by tRNAs, which have compatible size with the prebiotic polymerization conditions and these molecules could also have had a messenger function, that gave rise to the process of primitive translation and, in this way, the first modules of proteins could be synthesized. If biological systems started with tRNAs we pose the following questions: Which of the tRNAs would have started the system? Which amino acids would have been involved in the initial modules? In order to answer these questions, Eigen & Schuster (1978), proposed that the initial set of tRNAs that composed the progenote, comprised anticodons made up of repeating RNY triplets (where R stands for purine, Y pyrimidine and N either purine or pyrimidine). Thus, the first modules were rich in the amino acids alanine, serine, theonine, asparagine, aspartic acid, valine, isoleucine and glycine. We base our approach on the concept of progenote communities as proposed by Woese

https://doi.org/10.1017/S1473550415000464 Published online by Cambridge University Press

(1998), and in the propositions of Eigen & Schuster (1978) and Eigen & Winkler-Oswatitsch (1981), about the nature and operation of a translation system based on primitive tRNAs. Hence we reconstructed the ancestral sequences of tRNAs with anticodons of RNY type and by combinations of the eight kinds of tRNAs ancestors, three in three separately without repetition, the possible first genes of the progenote were reconstructed. Thus, as proposed by Eigen & Winkler-Oswatitsch (1981), the first tRNAs formed the first genes and the proteins encoded by them, constituting in this way, the proteome of the progenote or the set of binding domains that we call the *bindome*. This work proposes a proteome for the progenote based in the translation of ancestral tRNAs with a RNY pattern.

Methodology

Strategy

Traditionally studies have attempted to reconstruct the genetic composition of LUCA using top-down approaches, in which if a metabolic pathway is highly conserved then it counts as ancestral by evolutionary principles (Delaye *et al.* 2005). However, conservation of metabolic pathways does not necessarily indicate their ancestrality, since that may have appeared later in the evolution, and by gene transfer mechanisms, which were a common place in the tree of life. In this study, we decided to adopt a bottom-up approach and reconstructed the ancestral sequences for tRNAs that have codons with RNY pattern (purine – any nucleotide – pyrimidine), that was suggested for Eigen & Schuster (1978) as being the constitution of initial genetic code. The choice of analysis of tRNA molecules was based on the suggestion of Eigen and Winkler-Oswatitsch (1981), that these molecules gave origin to the first genes. The strategy used in this study was the reconstruction of ancestral sequences, from which we tried to retrieve the ancestral signal which could have originated the first genes.

Ancestral sequence reconstruction

The tRNA sequences were obtained from the tRNA database (http://trnadb.bioinf.uni-leipzig.de), corresponding to 361 organisms distributed in the three domains of life. In order to reconstruct separately the ancestor sequences for each type of tRNA with RNY anticodons, the total of 600 sequences for ^{Ala}tRNA, 313 for ^{Asn}tRNA, 267 for ^{Asp}tRNA, 608 for ^{Gly}tRNA, 326 for ^{Ile}tRNA, 855 for ^{Ser}tRNA, 634 for ^{Thr}tRNA and 538 for ^{Val}tRNA were analysed. Model tests were performed for each type of tRNAs to choose the best evolutionary model, which turned out to be a Kimura 2 parameters for every case. A phylogenetic tree was constructed by maximum likelihood method and from the tree for each type of tRNAs the ancestral sequences were obtained. In order to achieve statistical significance, bootstrapping with 1000 replicates was carried out. The parameters used for phylogenetic analysis and reconstruction of ancestor sequence were: Model = Kimura 2 parameters; uniform rates; all sites included; ML heuristic method = Nearest-Neighbour-Interchange (NNI); Initial tree = Neighbor Joining; Branch Swap filter = very strong; number of threads = 1. These evaluations were done with the MEGA5 program (Tamura *et al.* 2011).

Analysis of homology

From the ancestor sequences for each tRNAs with anticodons built for standard RNY (Ile, Asn, Vau, Asp, Gly, Ser, Thr and Ala), proposed as an initial set of tRNAs that composed the progenote (Eigen & Schuster 1978), sequences were reconstructed resulting from all possible combinations of those elements that took place without the presence of the same tRNA in two positions in the final sequence. A search for similar proteins from the combined ancestral tRNA sequences in National Center for Biotechnology Information with the Basic Local Alignment Search Tool X algorithm (Altschul *et al.* 1990) in the database UniProtKB/Swiss-Prot (swissprot) was performed. All analysed sequences were similar regardless of the level of similarity with the query sequence and these were then ranked according to their functional category with the support of the tool phylogenetic classification of proteins encoded in complete genomes in the Clusters of Orthologous Groups of proteins (Tatusov *et al.* 2003).

Results

Among the various metabolic pathways found in modern organisms, some already were present in LUCA and on the progenotes. In this way, we analysed the metabolic pathways, that might be early in the development of the sub-system or progenote. Thus, we analysed similarities of tRNAs that, when translated, had similarity related with pathways to amino acids pathways, glycolysis/gluconeogenesis, lipids pathways, nucleotides pathways, translation and transcription or RNA replication. We found similarities with the following proteins ordered by different categories.

Amino acids pathways

Diaminopimelato epimerase, L-asparaginase, ATP phosporibosyltransferase, histidinol-phosphate aminotransferase, 4-aminobutyrate aminotransferase, ornithine decarboxylase antizyme, N-acetyl-gama-glutamyl-phosphate reductase, homoserine kinase, aromatic-amino acid aminotransferase, ornithine carbomyltransferase and tryptophan synthase alpha chain.

Glycolysis/gluconeogenesis

Putative ribose/galactose/methylgalactose import, glycerate kinase, triose phosphate isomerase, Beta-glucosidase A, glucose 6-phosphate 1-dehydrogenase 2, glucose 6-phosphate isomerase, phosphoglycerate kinase, glycerol 3-phosphate dehydrogenase, transketolase and alpha-galactosidase.

Lipids pathways

Fatty acid synthase, CoA mutase, phosphate acyltransferase, lycopene cyclase and 3-beta-hydroxisteroid dehydrogenase.

https://doi.org/10.1017/S1473550415000464 Published online by Cambridge University Press

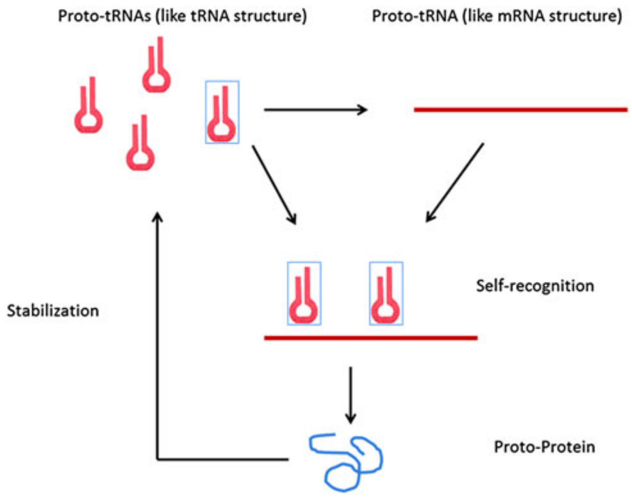
Proto-tRNAs (like tRNA structure)
Proto-tRNA (like mRNA structure)
Self-recognition
Stabilization
Proto-Protein

Fig. 1. A model proposal for self-translation in primordial tRNA.

Nucleotides pathways

Thymidylate kinase, cytidine deaminase, uridylate kinase, orotidine 5-phosphate decarboxylase, dihydroorotate dehydrogenase, phosporibosyl-formyl-glycinamidine cyclo-ligase and phosporibosyl-glycinamidine synthase.

Translation

Elongation factor-1 alpha, elongation factor 4, initiation factor 3, ribosomal protein L3, ribosomal protein L7a, ribosomal protein L27a-1, ribosomal protein L27a-3/4, ribosomal protein L27a-3, ribosomal protein L10, ribosomal protein S13, methyl-tRNA formyltransferase, RNA methyltransferase, tRNA uridine 5-carboximethylaminomethyl, glutamate–tRNA synthetase, lysine–tRNA synthetase, asparagine–tRNA synthetase, leucyl–tRNA synthetase, valine–tRNA synthetase, phenylalanine–tRNA synthetase and ribosomal RNA large subunit methyltransferase F.

Transcription or RNA replication

DNA-directed RNA polymerase and RNA-directed RNA polymerase.

Discussion

The evolutionary processes before LUCA must have occurred in a nonspecific way, but effectively to local needs. However, with the increasing complexity of these sub-systems (progenotes), the binding domains that at this time started to evolve to modern proteins. We cannot predict how many sub-systems were formed at this time and if all of them contributed to the emergence of LUCA. Our analysis focused on the prediction of a proteome of a progenote comprising a primitive translation system, based on the pattern of RNY codons, such as that suggested by Eigen & Schuster (1978). In our analysis, we consider that the initial tRNAs could be translated, and they were the source of the initial mRNA (Fig. 1).

Thus, the progenote began to take shape as a sub-system, which eventually led to the formation of LUCA. We can infer that the metabolic pathways were initially basal and that they have supplied demands of a formation system; in this case, the initial demands were related to the replacement of the constituents consumed and the preservation of its integrity.

Among the categories, the translation proved to be a system that began to get organized with proteins or modules that could promote the onset of a primitive protein synthesis. The factors responsible for the course of the proceedings arose at the beginning of the organization of progenote. This fact must have been very important for the development of the system as a whole, enabling that new proteins could be synthesized more accurately. This event may have led to the emergence of basal metabolic pathways, and thus, increasing the adaptability of the sub-system in formation. The presence of aminoacyl tRNA synthetase (aaRS) shows that the genetic code must have started their organization in the primary stages. An observation about the aaRSs is that only two correspond to the RNY pattern, which may indicate that this sub-system or progenote were in communication with other sub-systems in formation, where a progenote could supply other progenotes with molecules or products. Therefore, these processes induced the merging of sub-systems into a larger and more robust system, which may have led to the origin of the LUCA or pre-LUCA.

Among the ribosomal proteins found, the L3 and L10, in particular, show up as important in the organization of large subunit ribosomal being universally distributed. They are among the first proteins that bind to the ribosomal RNA. Other interesting fact to be noted is that the protein L3 is considered one of the oldest r-proteins, binding in the domain IV, V and VI of the rRNA. The domain V of the rRNA 23S is considered the most ancestral and it is where we find the peptidyl transferase center. This result strengthens the evidence of ancestral proteins in the system (Fox 2010; Korobeinikova *et al.* 2012). Another important ribosomal protein found was the S13 protein, which is important in the assembly of the ribosome, since it links the L5 protein and 38 helix of the 23S ribosomal. This protein also makes contact with tRNAs located in sites P and A of the ribosome. A mutation in the S3 protein reduces the affinity between the subunit 30S ribosomal and the tRNAs and increases the possibility of frameshift (Korobeinikova *et al.* 2012).

In progenote conditions, the appearance of a system, which would allow the synthesis of new RNA molecules, as well as the reproduction of existing RNA molecules with minimal fidelity, should have been necessary. We highlight the similarity with the enzyme RNA-dependent RNA polymerase, which uses an RNA strand as a template for the synthesis of another RNA strand. Currently these proteins are important in viruses with RNA genome, because they perform the duplication of the virus (Choi 2012). The appearance of these proteins in the progenote must have had a great importance, therefore promoting the maintenance of the molecules fixed in the beginning and by inserting errors enabled the emergence of new RNA molecules, increasing the range of proteins produced by translation of new RNA molecules.

Similarities were observed in proteins which are involved in the metabolism of sugars, lipids, nucleotides and pentoses. Then a proto-metabolism partially enclosed in progenotes

https://doi.org/10.1017/S1473550415000464 Published online by Cambridge University Press

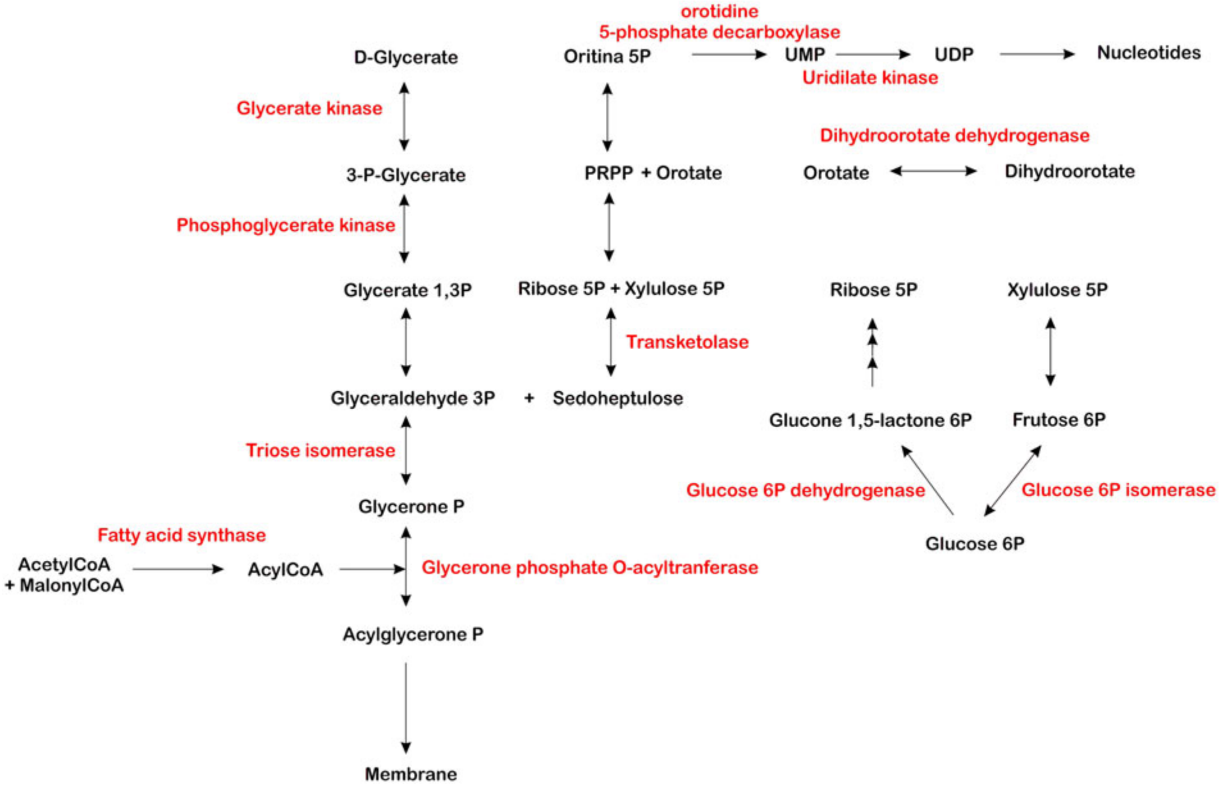

Fig. 2. A progenote proto-metabolism proposal based in tRNA with RNY anticodons. In red, the proteins that matched with the tRNAs translated like a mRNA.

emerged, and the core of this metabolism, are compounds of the three carbons that participate in glycolysis/gluconeogenesis (Fig. 2).

Thus, this part of the metabolism provided the compounds to be used in the synthesis of phospholipids, allowing a minimum of compartmentalization of the progenote, which could have led to an increase in the efficiency of other processes. It is also observed that these compounds were substrates for synthesis of pentoses, which now could serve as substrates for the synthesis of nucleotides that would be used by the RNA-dependent RNA polymerase to replicate sets of molecules already existing and also the synthesis of new RNA molecules. Therefore, the observed steps of the metabolism of glycolysis/gluconeogenesis should have functioned as the centre of distribution of organic compounds essential for structuring the progenote, as lipids, nucleotides and pentoses. Among the proteins that showed similarities with the synthesis of purines and pyrimidines, most of them are involved in the synthesis of uridine, which reinforces the idea of a biological information world based on RNA molecules (Schwartz 1995; Lazcano & Miller 1996; Kua & Bada 2011; Kawamura 2012; Higgs & Wu 2012).

The similarities found in the amino acid biosynthesis pathways, showed only few individual steps. However, these steps are related to amino acids encoded by RNY codons in most cases, demonstrating that at this moment of the process, the selective forces must have been the replacement of the components that were being utilized by the progenote in formation. In experiments simulating the prebiotic atmosphere of the earth, seven out of the eight amino acids were encoded by the standard RNY (alanine, valine, aspartic acid, asparagine, glycine, isoleucine and serine), being asparagine the exception (Bada 2013). This may reflect the need for immediate replacement of the same, for showing up in relative abundance, may have diminished the selective pressure and its pathways could have been developed in a second phase, where the carbon molecules could be substrates for amino acids synthesis. Among these steps, we observed similarities with the homoserine kinase, a protein involved in the synthesis of threonine, an amino acid encoded by the RNY codons. In experiments simulating volcanic activity, which must have occurred frequently during the early stages of the formation of the primitive atmosphere, one of the non-proteic amino acids that appeared was homoserine (Parker *et al.* 2011; Bada 2013), which is a substrate for homoserine kinase enzyme in the pathway of threonine formation (Fig. 3). This shows that the consumption of some substrates increased the selective pressure for the emergence of the proteins or modules to synthesis, and thus, replacing the molecules that were consumed.

Our data show a scenario where a progenote or sub-system emerged with basal metabolic pathways. At this stage of development, information exchange must have occurred among progenotes where local needs could be supplied. Thus, the evolution by merging with other sub-system may have originated more complex organisms, which in turn led directly to appearance of LUCA. Delaye *et al.* 2005, through the analysis of universally

https://doi.org/10.1017/S1473550415000464 Published online by Cambridge University Press

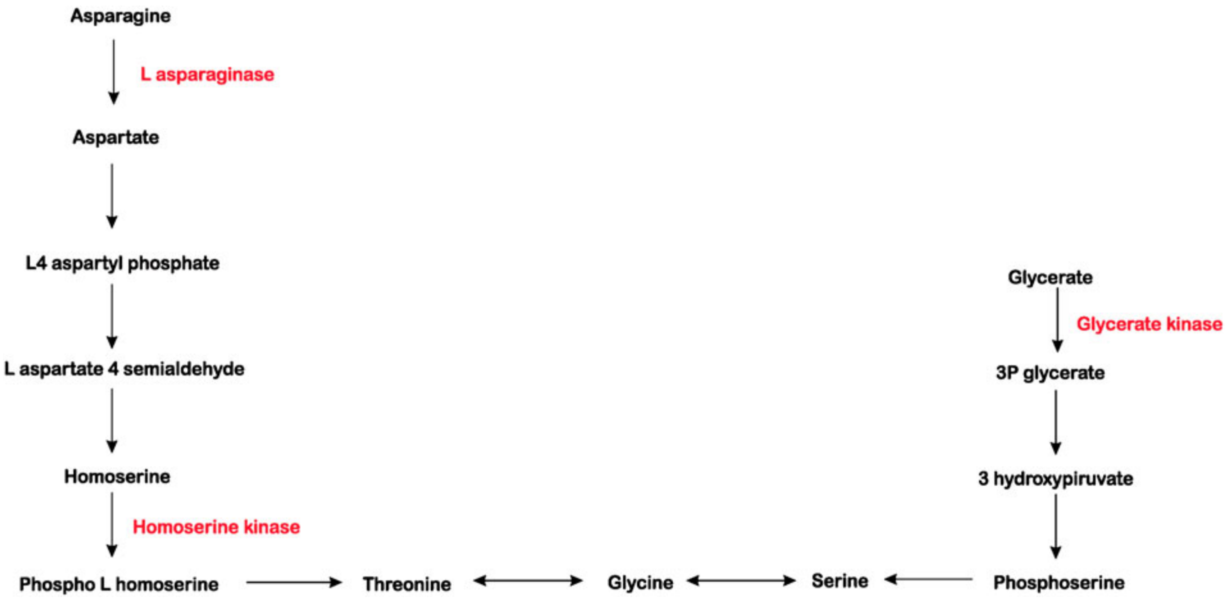

Fig. 3. Aminoacids pathways proposal based in the tRNA with RNY anticodons. In red, the proteins that matched with the tRNAs translated like a mRNA.

conserved pathways, proposed a group of proteins that could be composing the genome of LUCA. Analysing the results obtained by Delaye *et al*. (2005), we observed that several proteins proposed as participants in the composition of LUCA are also in our data set, which shows the consistency of our data with independent methodologies. These similarities could indicate that these proteins pertained to an era prior to LUCA and that these metabolic systems were gradually being mounted through cooperative processes among progenotes. These results suggest an origin for the system based on tRNA molecules, being the key to start the organization of progenote and with the function of the mRNAs and tRNAs and thus synthesizing the first functional modules.

Acknowledgments

M. V. J. was financially supported by PAPIIT-IN224015, UNAM, México.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). *J. Mol. Biol.* **215**, 403–410.

Bada, J.L. (2013). *Chem. Soc. Rev.* **42**(5), 2186–2196.

Choi, K.H. (2012). *Adv. Exp. Med. Biol.* **726**, 267–304.

Delaye, L., Becerra, A., Lazcano, A. (2005). *Orig. Life Evol. Biosph.* **35**(6), 537–554.

Di Giulio, M. (2003). *J. Mol. Evol.* **57**(6), 721–730.

Eigen, M. & Schuster, P. (1978). *Naturwissenschaften.* **65**, 341–369.

Eigen, M. & Winkler-Oswatitsch, R. (1981). *Naturwissenschaften.* **68**(6), 282–292.

Forterre, P. & Philippe, H. (1999). *Biol. Bull.* **196**(3), 373–375.

Forterre, P., Gribaldo, S. & Brochier, C. (2005). Med. Sci. *(*Paris*)*. **21**(10), 860–865.

Fox, G.E. (2010). *Cold Spring Harb. Perspect. Biol.* **2**(9), a003483.

Glansdorff, N., Xu, Y. & Labedan, B. (2008). *Biol. Direct.* **3**, 29, Doi: 10.1186/1745-6150-3-29.

Glansdorff, N., Xu, Y. & Labedan, B. (2009). *Res. Microbiol.* **160**(7), 522–528, Doi: 10.1016/j.resmic.2009.05.003.

Goldman, A.D., Bernhard, T.M., Dolzhenko, E. & Landweber, L.F. (2013). *Nucleic Acids Res.* **41**(Database issue) D1079–D1082, DOI: 10.1093/nar/gks1217.

Higgs, P.G. & Wu, M. (2012). *Orig. Life Evol. Biosph.* **42**(5), 453–457.

Kawamura, K. (2012). *Biochimie.* **94**(7), 1441–1450.

Kim, K.M. & Caetano-Anollés, G. (2011). *BioMed. Cent. Evol. Biol.* **11**, 140, Doi: 10.1186/1471-2148-11-140.

Korobeinikova, A.V., Garber, M.B. & Gongadze, G.M. (2012). *Biochemistry (Mosc).* **77**(6), 562–574.

Kua, J. & Bada, J.L. (2011). *Orig. Life Evol. Biosph.* **41**(6), 553–558.

Lazcano, A. & Miller, S.L. (1996). *Cell* **85**(6), 793–798.

Morange, M. (2011). *Res. Microbiol.* **162**(1), 5–9, Doi: 10.1016/j.resmic.2010.10.001.

Mushegian, A. (2008). *Front. Biosci.* **13**, 4657–4666.

Parker, E.T., Cleaves, H.J., Dworkin, J.P., Glavin, D.P., Callahan, M., Aubrey, A., Lazcano, A. & Bada, J.L. (2011). *Proc. Natl. Acad. Sci. U. S. A.* **108**(14), 5526–5531.

Penny, D. & Poole, A. (1999). *Curr. Opin. Genet. Dev.* **9**(6), 672–677.

Schwartz, A.W. (1995). *Planet. Space Sci.* **43**(1–2), 161–165.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). *Mol. Biol. Evol.* **28**(10), 2731–2739.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., Smirnov, S., Sverdlov, A.V., Vasudevan, S., Wolf, Y.I., Yin, J.J., Natale, D.A. (2003). *BioMed.Cent. Bioinf.* **4**(41), 1–14.

Woese, C. (1998). *Proc. Natl. Acad. Sci. U. S. A.* **95**(12), 6854–6859.

https://doi.org/10.1017/S1473550415000464 Published online by Cambridge University Press