

Article

Not peer-reviewed version

The Origin(s) of Luca: Computer Simulation of a New Theory

[Shiping Tang](#)^{*} and [Ming Gao](#)

Posted Date: 8 October 2024

doi: 10.20944/preprints202410.0599.v1

Keywords: LUCA; Computer simulation; Cellular evolution; Carl Woese



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

The Origin(s) of LUCA: Computer Simulation of a New Theory

Shiping Tang * and Ming Gao

Center for Complex Decision Analysis, Fudan University, Shanghai, China 200433; gaoming@fudan.edu.cn

* Correspondence: twukong@fudan.edu.cn; twukong@yahoo.com

Abstract: Carl Woese's thesis of cellular evolution emphasized that the last universal common/cellular ancestor (LUCA) must have evolved by drawing from "global inventions". Yet, existing theories regarding the origin(s) of LUCA have mostly centered upon scenarios that LUCA had evolved mostly independently. In an earlier paper, we have advanced a new thesis regarding the origin(s) of LUCA that extends Woese's original insights. Our thesis centers upon the possibility that different vesicles and protocells can merge with and acquire each other as a form of variation, selection, and retention, driven by wet-and-dry cycles and other similar cyclical processes. In this paper, we use computer simulation to show that under a variety of simulated conditions, LUCA can indeed be produced by our proposed processes. We hope that our study can stimulate laboratory testing of some key hypotheses that vesicles' absorption, acquisition, and merger has indeed been a central force in driving the evolution of LUCA.

Keywords: LUCA; Computer simulation; Cellular evolution; Carl Woese

1. Introduction

The coming of the Last Universal Common (Cellular) Ancestor (LUCA) was a major transition in the making of the biotic world (Woese and Fox 1977; Maynard Smith & Szathmáry 1995; Koonin 2014a; 2014b).¹ Recently, Tang (2021) advanced a new thesis regarding the origin of LUCA. His core thesis is that vesicles' absorption, acquisition, and fusion via breaking-and-repacking, proto-endocytosis, proto-endosymbiosis and other similar processes, driven by "wet-and-dry" cycles or other similar thermochemical cycles within Darwin's "warm little pond(s)" (Mulikidjanian et al. 2012; Damer and Deamer 2015; Higgs 2016; see also Zhu et al. 2012), has been a central mechanism that propelled the origin(s) of the First Universal Cellular Ancestors (FUCAs) and then the evolution of LUCA from FUCAs.

Tang (2021) further hypothesized that the evolution of LUCA came in two major stages: the evolution of FUCAs and then evolution of FUCAs to LUCA. What is common to both stages has been that both involve the same process of absorption, acquisition, and merger by these protocells (or vesicles), and this process had been the key behind Carl Woese's (1998; 2000; 2002) thesis that LUCA must have evolved by drawing from "global inventions". The whole evolutionary process from simple vesicles to FUCAs and then LUCA is summarized in Fig. 1 below (reproduced from Tang 2021).

Tang (2021) has provided some chemical and biological evidence that supports his hypothesis. Ultimately, however, chemical and biological evidence can only show that the hypothesis is possible but not necessarily viable for the making of LUCA because re-creating LUCA from scratch in a lab is highly unlikely.

We therefore turn to the next best thing: to "visualize" the viability of the proposed hypothesis with computer simulation, following similar efforts with different focuses (e.g., Klein et al. 2017;

¹ Although LUCA has been conventionally taken to be the "Last Universal *Common* Ancestor", it is now generally accepted that LUCA must have been a fairly complete cell (Koonin 2014a; 2014b).

Armstrong et al. 2018; Lancet et al. 2018; Takagi et al. 2020). More specifically, we aim to show that merger and acquisition by vesicles and other similar processes can indeed produce FUCAs, and then inevitably, LUCA.

Before proceeding further, two caveats are in order.

First, we are mostly interested in simulating the hypothesis that FUCAs and LUCA had most likely evolved by drawing from “global innovation” via vesicles’ merger and acquisition (Tang 2020; 2021; see also Woese 1998; 2000; 2002). As such, we do not simulate the chemical synthesis of polymers from nucleotides or amino acids (e.g., Ross & Deamer 2016; Hargrave et al. 2018), the evolution of autocatalytic cycles, the evolution of a RNA replication system (e.g., Ma et al. 2007; Ma & Hu 2012; Kim and Higgs 2017), the forming of vesicles (e.g., Klein et al. 2017; Armstrong et al. 2018; Lancet et al. 2018), the coupling of an RNA replicator system and vesicles (Yin et al. 2019), or the coupling of cellularity and metabolism (e.g., Takagi et al. 2020; Nunes Palmeira et al. 2022; see also Szathmáry 2007), partly because they have been performed earlier. Also, simulating the whole process from chemical evolution to the coming of LUCA will be a daunting, if not an impossible, task, without necessarily shedding much light on the central mechanism we highlight here.² We therefore focus on the late stage of the evolution of LUCA, somewhat similar to Takagi et al. (2020) and Nunes Palmeira et al. (2022).

Second, we concur with several earlier studies that LUCA has at least three hallmarks: 1) a full functioning membrane; 2) a fully evolved standard genetic code (SGC); 2) about one hundred proteins and several hundred genes (Charlebois & Doolittle 2004; Harris et al. 2003; Koonin 2003; Ranea et al. 2006; Wolf & Koonin 2007; Goldman et al. 2013).³ In our simulation, we focus on the coming together of one hundred peptides (plus several hundred genes) and the coming of the SGC, while assuming that LUCA has a full functioning membrane.

2. From FUCAs to LUCA: Summary of a New Thesis

This section summarizes key theses regarding the evolution from FUCAs to LUCA, advanced by Tang (2021), with Fig. 1 summarizing the thesis in sketches. To avoid repetition, most references have been left out (for extensive references, see Tang 2021).

2.1. Evolution before FUCAs

Abiotic synthesis of bioorganic molecules was the first step in the origin of life. Once bioorganic molecules came to exist, first as monomers (e.g., amino acids, nucleotides, fatty acids, and later on, phospholipids) and then as polymers (e.g., short peptides, small RNAs), they came under the force of natural selection even though replication did not operate back then. During this stage, there were two key selection yardsticks. The first is thermochemical stability or survivability within the system. The second is solvability and a minimum level of availability that allows a minimum level of concentration for monomers to be assembled into polymers and more complex hetero-biomolecules. Both stability and availability partly depend on the relative easiness of synthesis from simple precursors and protection from UV light.

² For instance, Yin et al. (2019) merely simulated how a RNA replication system can evolve and then spread. Their “cellular” evolution is really about vesicles engulfing the RNA replication system. Thus, their simulation does not deal with the evolution of LUCA, or even the evolution of FUCAs. Note also, our simulation (based on our theory) subsumes vesicles engulfing ingredients such as amino acids, nucleotides, peptides, and RNAs. Most critically, Yin et al. (2019) based their simulation on the “hydrothermal vent” scenario, which many others and we believe to be rather unrealistic (Mulikidjanian et al. 2012; Damer and Deamer 2015; Higgs 2016; Tang 2021).

³ Most likely, these genes would have been linked into a chromosome (Maynard Smith and Szathmáry 1993). Here, we simply assume this fact without modelling it.

Amphiphiles form vesicles in certain conditions. A vesicle succeeds in persisting in the system if it can retain its basic structure, float within a solution, absorb ingredients from its environment (e.g., via proto-endocytosis), and merge with other vesicles. Most likely, such vesicles also has the capacity of “dividing” without either strict reproduction or genetic replication. Rather, they divide via pinching or budding due to enlargement of size by 1) absorbing more lipids, peptides, (poly-)nucleotides, and other bioorganic molecules, 2) merging with and engulfing other vesicles [hence, each vesicle is also a target of (proto-) endosymbiosis by other vesicles], and 3) synthesizing new polymers within (Mansy et al. 2008; Zhu and Szostak 2009; Budin and Szostak 2011; Budin et al. 2012; Kurihara et al. 2015; Armstrong et al. 2018). During this stage, persistence and division have not been coupled with active reproduction, genetic replication or even sophisticated metabolism.

Even if RNA alone is capable of both replication and metabolism, RNA might have come to interact with amino acids and peptides quite early on, and the primitive translation apparatus (and the genetic code) had originated from this interaction and then coevolution of amino acids/peptides with RNAs. During this stage of coevolution, precision in RNA replication (and proto-translation) is not necessarily an advantage. Rather, during this stage of coevolution, the key was to make more RNAs and peptides without too much precision so that the structural diversity and hence the functional diversity of RNAs and peptides could increase more rapidly. With more diverse structures and functions, RNAs can then support the production of more diverse peptides with different properties, and these peptides in turn interact with RNAs more diversely to generate more emergent properties. This mutually reinforcing increase in structure and function of both RNAs and peptides, subject to natural selection, laid a key foundation for the evolution of more complex ribonucleoprotein (RNP) world, the standard genetic code, and eventually a more versatile metabolism system.

For a period of time, the evolution of peptide-lipid membrane and the evolution of RNA-peptide (as the proto-translation machinery) might have proceeded independently from each other. The two processes might even have operated in different locations such as different terrestrial hydrothermal ponds or fields (Mulkidjanian et al. 2009; Mulkidjanian et al. 2012; Damer & Deamer 2015; 2020; Koonin 2014b, 35-36). *Eventually, however, these two processes had to come together, and the moment in which these two processes merged was the first decisive step from replicators to reproducers that paved the way toward the first protocells or FUCAs* (e.g., Yin et al. 2019). The fusing of the two processes was perhaps achieved by a peptide-lipid vesicle absorbing several RNA-peptide complexes (as proto-endocytosis), mediated via the interaction between RNA and lipids or peptide on the vesicle’s surface.

2.2. From FUCAs to LUCA

Once FUCAs came to possess both a proto-machinery of survival (roughly, metabolism supported by peptides/proteins within a membrane) and a proto-machinery of replication (now supported by both peptides/proteins and RNAs), survival and replication began to co-evolve with each other. Because both machineries require some kind of metabolism machinery (including bioenergetics), metabolism came to join survival and replication within the coevolution process (Takagi et al. 2020). This coevolutionary process laid the foundation for all subsequent evolutionary processes, and became only possible within protocells with a regulated membrane rather than directly from the “naked” RNA world.

Different FUCAs not only competed against each other for various ingredients but also divided and survived differently within the system, as a form of pre-Darwinian selection process (Tang 2020). Along the way, FUCAs continued to absorb useful ingredients and integrate them into more complex, versatile, and effective macromolecules, including more complex proteins and RNAs. During this phase, FUCAs might have also continued to absorb other (sub-)cellular components from other vesicles and integrate them into more tightly regulated cellular components. Hence, rampant extinction of (proto-)cellular lineages occurred during this period. In this phase, a tight coupling of survival and replication might not have held any selective advantage. Indeed, the opposite might have been true: being more promiscuous and having more flexibility provides a protocell with significant advantage for survival.

Within the commune of FUCAs, each FUCA protocell competed against each other. After a period during which survival, division, and replication co-evolved with each other, some of the FUCAs eventually became protocells in which division and replication are more tightly coupled and smoothly regulated. Eventually, a few lucky FUCAs with the right and tight coupling of metabolism, translation machinery, division with genetic replication, and energy efficiency will dominate the system, and these lucky few FUCAs merged into a single lineage or only one lineage of FUCAs survived: this lone surviving lineage became LUCA. *Because LUCA possessed a tight coupling of cell division with genetic replication, it was a genote that had crossed “the Darwinian Threshold”* (Woese 1998; 2000; 2002). A mostly fully functioning translation system with the full standard genetic code had also “crystallized” in LUCA.

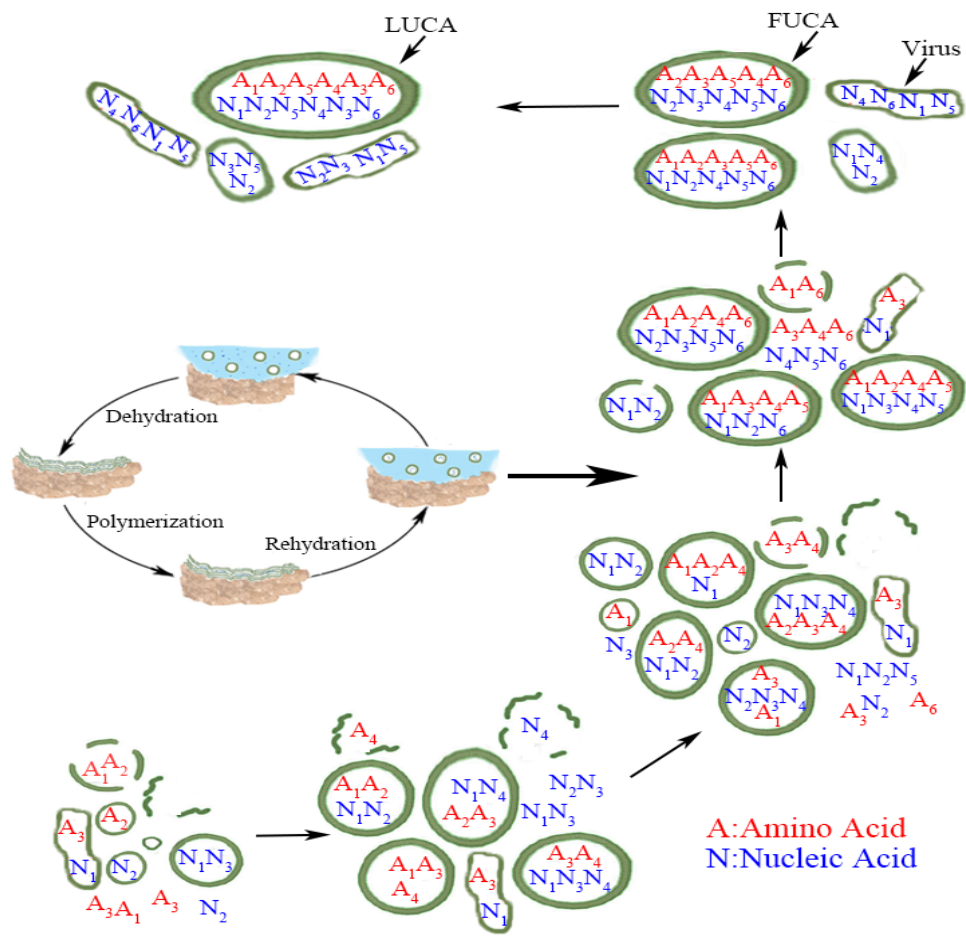


Figure 1. From Vesicles to LUCA (Reproduced from Tang 2021 with permission.) Numbers in subscript denote different amino acids and nucleic acids. The exact matching between amino acid and nucleic acid within LUCA, in a metaphorical sense, implies that the standard genetic code (SGC) had evolved most completely by the time of LUCA. The less than exact matching amino acid and nucleic acid within FUCA and vesicles before FUCA denotes the evolutionary path of SGC from a rudimentary form to a mature form in LCUA. Protocells or vesicles are in closed circles whereas broken vesicles are in broken circles. Viruses are in elongated or other non-circular shapes. The three to one ratio of virus versus cell at the stage of LUCA is to imply the fact that virus may be the most abundant biological entity in the biosphere. The wet-and-dry cycle (on the left of the diagram) might have played a key role in driving the process of breaking-and-re-encapsulation cycle that facilitates the merger and acquisition by vesicles. The wet-and-dry cycle part within the figure is adapted from Bruce Damer and David Deamer (2015) with permission.

3. Materials and Methods

Computer simulations are performed with the GAMA simulation platform (<https://gama-platform.org/>), on local servers hosted by our center or commercial cloud servers. All computer codes and experimental data are available in the online appendix.⁴

All simulations described below (i.e., whether the two stages of evolution from simple vesicles to FUCAs and then LUCA are simulated separately or together) start with one single “warm little pond” (see below for details). All simulations follow a similar flow (see Figure 2 below for the flowchart). Table 1 summarizes the parameters in the simulation.

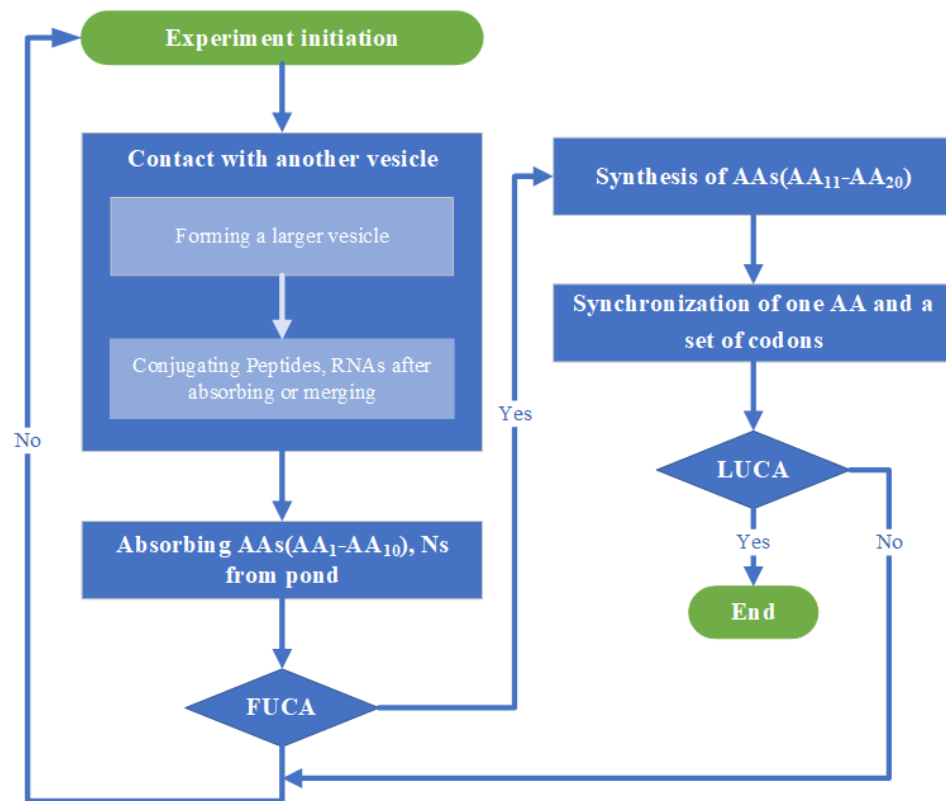


Figure 2. Flowchart of Simulation.

Table 1. Parameters in the Simulations.

Parameters	Explanation	Value
tickCount	count of ticks	1 tick = 2 cycles
currentVolume	units of volume	dry phase: (50, 80), wet phase: (80, 100); initiation: 100
numPeptide	number of peptide	see Table 2
numRNA	number of RNA	see Table 2
numAA	number of AAs in a peptide	see Table 2
numNBase	number of nucleotides in an RNA	3*numAA
V1InitNum	initial number of V1	3000
V2InitNum	initial number of V2	2000
totalnum	external supply of V1 and V2 in every cycle	dry phase: (2000, 3000), wet phase: (4000, 5000)
pContact	probability of contact with another vesicle	$(10^{-5}, 10^{-6}) * (100/\text{currentVolume})^2$
mergeP	probability of absorbing another vesicle	See Table 3 for details

⁴ <https://github.com/xymbgm/2023-05-Tang-and-Gao-Simulation-of-from-FUCA-to-LUCA-codes>

conjugateP	probability of conjugating peptides or RNAs after absorbing or merging	See Table 4A and Table 4B for details
fitnessScore	jointly determined by the number of AAs assigned (NA) and the total number of peptides (NP) within the vesicle	$FS=NA*NP$
pSurvival	Vesicles' probability of survival	$pSurvival=\ln(fitnessScore)/10$
rndIndexAA	type of AA within a vesicle	V1-V5: AA ₁ -AA ₁₀ , V6: AA ₁ -AA ₂₀
pABS	probability of absorbing more AAs and NNs from the pond	$5*10^{-3}$
pSynchronize	probability of assigning one AA to one set of codons within a dry-and-wet cycle	0.999

Following Klein et al. (2017, 35), every entity (i.e., amino acid, lipid, polypeptide, nucleic acid, or vesicle) is normalized to a ball-like entity, with a radius of 1 and a mass of 1. Vesicles can absorb amino acids, nucleotides, lipids, and other useful ingredients or components, which are assumed to have unlimited supply within each pond. The ratio between peptides and RNAs within a vesicle is set to be 2.5-3.5 (e.g., if a vesicle has 2-4 peptides, then it has 5-10 RNAs). Moreover, a vesicle must contain at least 2-3 peptides and 5-10 RNAs in order to qualify as a vesicle (i.e., V1 is the minimum threshold of being a vesicle). Finally, different vesicles have different numbers of AAs being assigned a proper set of codon.

There are six types of vesicles in our simulation (table 2), and FUCAs are protocells with at least 41-50 peptides and 120-160 RNAs. In other words, only type V6 vesicles are counted as FUCAs. FUCAs, however, have yet to possess a fully evolved SGC (Tang 2021).

LUCA is a fully functional (proto-)cell with at least one hundred proteins (peptides) and several hundred RNAs (as genes). Moreover, LUCA has a fully evolved SGC, symbolized by the tight coupling of nucleotides (within RNAs) and the amino acids (within peptides). In other words, only FUCAs that have evolved the tight linkage of RNA and peptide can become LUCA. Most likely, RNA and proteins have co-evolved with each other (Kovacs et al. 2017).

Vesicles' encounters and interactions are driven by wet-and-dry or other similar cycles (e.g., Damer and Deamer 2015; 2020; Deamer 2019; Zhu et al. 2012), which is captured by the increased-and-decreased volume of the "warm little pond". The pond's changing volume in turn drives changes in the probability of vesicles coming into contact with each other. When in contact, a vesicle can acquire another vesicle according to the probabilities dictated in Table 2.

Table 2. Types of Vesicles (V1 to V6).

Types of vesicles	No. of peptides (NP)	No. of RNAs (NR)	No. of AAs assigned (NA)	Fitness score (F)
V1	2-4	5-10	5-6	10-24
V2	5-10	10-30	6-8	30-80
V3	11-20	25-70	8-10	88-200
V4	21-30	50-100	10-12	210-360
V5	31-40	75-140	12-14	272-560
V6 (as FUCAs)	41-50	120-160	14-16	576-800

Table 3. Probabilities of absorbing another vesicle vs. being absorbed (The first number for a vesicle in rows; the second for a vesicle in columns).

	V1	V2	V3	V4	V5	V6
V1	0.5; 0.5	0; 1	0; 1	0; 1	0; 1	0; 1
V2	1; 0	0.5; 0.5	0.4; 0.6	0.3; 0.7	0.2; 0.8	0.1; 0.9
V3	1; 0	0.6; 0.4	0.5; 0.5	0.4; 0.6	0.3; 0.7	0.2; 0.8

V4	1; 0	0.7; 0.3	0.6; 0.4	0.5; 0.5	0.4; 0.6	0.3; 0.7
V5	1; 0	0.8; 0.2	0.7; 0.3	0.6; 0.4	0.5; 0.5	0.4; 0.6
V6	1; 0	0.9; 0.1	0.8; 0.2	0.7; 0.3	0.6 0.4	0.5; 0.5

Note: When two identical vesicles encounter each other, they have an equal probability of acquiring (as merging) each other (hence, 0.5 vs. 0.5). The actual probability of an acquisition is drawn randomly around the two numbers within a margin of $\pm 5\sim 10\%$.

When two vesicles are merged or one absorbs the other, the new vesicle gains all the peptides and RNAs from its two ancestral vesicles. For simplicity, we assume that peptides and RNAs within the vesicles are fully functional. Moreover, the peptides and RNAs within the new vesicle can be conjugated. The probabilities of conjugating different peptides and RNAs are set in Table 4A and Table 4B respectively.

Table 4. A. Rules for Conjugating Peptides after Absorbing or Merging.

Length of peptide (AAs)	3-10	11-25	26-50	>51
3-10	$1*10^{-5}$	$0.6*10^{-5}$	$0.3*10^{-5}$	$0.1*10^{-5}$
11-25	$0.6*10^{-5}$	$0.3*10^{-5}$	$0.1 *10^{-5}$	$0.05*10^{-5}$
26-50	$0.3*10^{-5}$	$0.1*10^{-5}$	$0.05*10^{-5}$	$0.02 *10^{-5}$
>51	$0.1*10^{-5}$	$0.05 *10^{-5}$	$0.02 *10^{-5}$	$0.01 *10^{-5}$

Table 4. B: Rules for Conjugating RNAs after Absorbing or Merging.

Length of RNA (NBs)	6-30	31-60	61-100	>100
6-30	$1*10^{-5}$	$0.6*10^{-5}$	$0.3*10^{-5}$	$0.1*10^{-5}$
31-60	$0.6*10^{-5}$	$0.3*10^{-5}$	$0.1 *10^{-5}$	$0.05*10^{-5}$
61-120	$0.3*10^{-5}$	$0.1*10^{-5}$	$0.05*10^{-5}$	$0.02 *10^{-5}$
>120	$0.1*10^{-5}$	$0.05 *10^{-5}$	$0.02 *10^{-5}$	$0.01 *10^{-5}$

Consistent with the thesis that there were two phases of SGC evolution (Wong 1975; 2005; Wolf and Koonin 2007; Francis 2013; Sengupta and Higgs 2015; Koonin & Novozhilov 2017; Yarus 2017; 2021), we dictate that a vesicle or protocell must first get the first 10 AAs assigned with their respective codons, and only after can a vesicle or protocell get the next 10 AAs assigned with their codons.⁵ Consistent with the co-evolutionary thesis of codons, proteins, and cellular functions, the more AAs are assigned to their codons, the more functions a protocells will gain from more complex peptides or proteins. Accordingly, a vesicle or protocell should gain a bit of fitness with each “optimal assignment” of AA to its proper codons.

The net fitness score (FS) of a vesicle is jointly determined by the number of AAs assigned (N_A) and the total number of peptides (N_P) within the vesicle. That is, $FS= N_A*N_P$.

For vesicles that do not go through the processes of merger-and-acquisition during the wet-and-dry cycles, those vesicles with higher fitness scores will have a greater probability of surviving in the pond than those with lower fitness scores. The probability of survival (P_S) is determined by the simple equation below:

$$P_S = \frac{\ln(FS)}{10}$$

Finally, when a FUCA has assigned all the 20 AAs (i.e., $N_A = 20$) and has more than 100 peptides (i.e., $N_P \geq 100$) and 300 RNAs, it becomes a LUCA, and the simulation ends.

3.1. The First Stage: the origin(s) of FUCAs

⁵ The ten early AAs are: Gly, Ala, Ser, Asp, Glu, Val, Leu, Ile, Pro, Thr. The ten late AAs are Phe, Tyr, Arg, His, Trp, Asn, Gln, Lys, Cys, Met.

In the initial state, there were anywhere between 2000 and 3000 V1 and V2 vesicles within a pond of 100 units of volume. Within each tick (consisting of a wet phase and a dry phase), the pond also produces or gains more V1 and V2 vesicles. During the wet phase, anywhere between 4000 and 5000 new V1 and V2 vesicles will be added to the pond whereas during the dry phase anywhere between 2000 and 3000 new V1 and V2 vesicles will be added. Vesicles can also absorb other biochemical components and ingredients (e.g., amino acids, nucleotides), and thus become larger vesicles. Biochemical components and ingredients such as amino acids, nucleotides, and lipids are exogenously generated and added to the pond, with unlimited supply.

Within each tick, a vesicle has certain probability of coming in contact with another vesicle. The probability of being in contact is regulated by the total volume of the pond, dictated by the following function: $P_c = (10)^{-5 \sim -6} \left(\frac{100}{X}\right)^2$, with X being the unit of volume of the whole pond at a given tick. We further assume that in the dry phase, the total volume of the pond decreases from 100 units of volume to 50-80 units of volume, thus increases the probability of contact by vesicles. The dry phase also breaks some vesicles. Reversely, in the wet phase, the total volume of the pond increase back to 80-100 units of volume, thus decreases the rate of contact by vesicles but also allows new vesicles to form.

When two vesicles come in contact with each other, they can merge with or acquire the other with certain probabilities specified in Table 3. And when smaller vesicles (e.g., V1, V2, and V3) merge with or acquire each other, they form larger vesicles (e.g., V3, V4, and V5).

3.2. The Second Stage: from FUCAs to LUCA

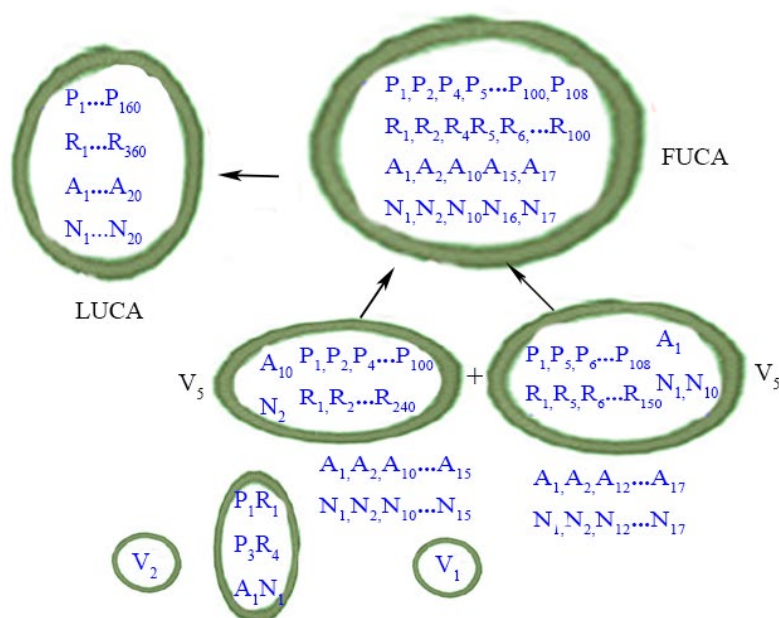
It is now widely accepted that LUCAs possessed about several hundred genes, about half of them were in RNA metabolism and translation, with the SGC fully in place. LUCA also possesses about 140 proteins or domains (e.g., Koonin 2003; Mirkin et al., 2003; Wolf and Koonin 2007; Ranea et al. 2006; Francis 2013; Koonin & Novozhilov 2017; Yarus 2021). We thus assume that LUCA must have about 100-120 proteins and 300-360 RNAs, with each protein being more than 50 AAs and each RNA being more than 150 nucleotide bases long. We also require that LUCA to have assigned all twenty amino acids to the full genetic code. For this second stage, we assume that only FUCAs (i.e., V6) can perform the task of assigning AAs to their proper codons. For this particular process, we assume the following:

- For simplicity, we skip the evolution of the stop codon. Hence, when all of the 20 AAs have been assigned to their codons, the full SGC has evolved, and we can consider that LUCA now exists. Accordingly, there may be a few LUCAs and they have different genomes, but they all have the same SGC. Very critically, according to our theory, SGC could have only evolved by drawing from “global inventions” via the merger and acquisition of vesicles (Tang 2020; 2021; see also Woese 1998; 2000; 2002).⁶
- Because there might have been 4 to 6 nucleotides available for making into the SGC before the SGC was finally fixed, the total combinations of 20 AAs with the possible sets of codons is anywhere between $20(4^3)$ and $20(6^3)$, or 1280 to 4320. Thus, for each simulation, the number of codons to be assigned is a random number anywhere from 1280 to 4320. For each tick as a wet-and-dry cycle, FUCAs can only assign one AA to one set of codons, with a fixed probability of 0.999.⁷

⁶ In this sense, we agree with Herron’s (2021) stand that SGC should not be classified as a distinct “major transition” (cf. Maynard Smith & Szathmáry 1995) because the evolution of SGC has been a (gradual) process. Notably, while Maynard Smith & Szathmáry (1995) identified SGC and protocell (FUCA) as two distinct transitions, Szathmáry (2015) argued that they might have co-evolved in prokaryotic cells. Our thesis holds that they had evolved mostly together, most likely by LUCA.

⁷ Biologically and mathematically, there is a positive feedback mechanism in the evolution of SGC: once one AA has been assigned a set of codons, the remaining AAs will have a smaller pool of codons to be assigned.

For simplicity, we assume that when two vesicles with different AAs already assigned to their specific codons merge with each other, the merged vesicle obtains all the codons and hence evolution of the universal codon accelerates. For example., vesicle-1 has A1, A2, A3, A4, A5 assigned whereas vesicle-2 has A1, A2, A3, A4, A6, then the merged vesicle of the two vesicles will have A1, A2, A3, A4, A5, A6 assigned. This is consistent with the dynamics underscored by Vetsigian et al. (2006), that SGC had most likely evolved via “collective evolution” by drawing from “global innovations” with “horizontal gene transfer” (HGT). Indeed, according to Tang (2021), absorption, acquisition, and merger by vesicles entail extensive “horizontal biomolecule transfer” (HBMT) rather than merely HGT: HBMT thus subsumes HGT because HBMT entails exchange and retention of other biological ingredients other than genetic materials. Of course, if two vesicles have the same set of AAs assigned (i.e., when both vesicle-1 and vesicle-2 have A1, A2, A3, A4, A5 assigned), the newly merged vesicle of the two vesicles does not gain a new AA assigned. But the new merged vesicle can still gain peptides and RNAs, thus also increases its fitness score according to $FS = N_A * N_P$.



Hence, an earlier assigning of one AA to a set of codons will accelerate the next cycle of assigning the remaining AAs and codons. As a result, if the probability of assigning the first AA with a set of codons is 10^{-3} , the probability of assigning the next AA with the remaining sets of codons becomes: $P(\text{codon} - \text{assign}) = 10^{-3} \left(\frac{20}{20-n} \right)^2$, with n denoting the number of AAs already assigned. So, for the first AA to be assigned, n is 0, and for the second AA was n is 1, and so on. $\left(\frac{20}{20-n} \right)^2$ is to magnify the cumulative impact of previous rounds of assigning upon the remaining synchronizations. We initially hope to implement such a dynamics in the simulation. However, due to the fact that many vesicles will die in each cycle (as predicted by our theory, Tang 2021), implementing such a dynamics requires significant computational resources. We therefore set the probability of successful codon assignment to a fixed probability of 0.999 to speed up the process.

Figure 3. Summarizes this evolutionary phase of FUCAs to LUCA. P: Peptide; R: RNAs. A: Amino acid; N: nucleotide (as codons). Protocells or vesicles are in closed circles. LUCA is depicted according to the definition in the main text. The exact matching between amino acids (A) and nucleic acids (N) within LUCA, in a metaphorical sense, implies that the standard genetic code (SGC) had evolved most completely by the time of LUCA. The less than exact matching amino acids and nucleic acids within the two FUCA denotes the evolutionary path of SGC from a rudimentary form to a mature form in LCUA. Contents within the two smaller vesicles (V1 and V2) are now shown. The wet-and-dry cycle part is omitted here.

3.3. *The Two Stages Together*

To further test our hypotheses, we also simulate the two stage together but still allow vesicles to merge and absorb each other. Moreover, we also allow for all vesicles (or protocells) to absorb and integrate amino acids, short peptides, nucleotides, and short RNA molecules into their existing peptides (or proteins) and RNA molecules according to the probabilities dictated in Tables 1, 2, and 3. We also simulate different number of ponds.

The purpose of these simulations is to show that LUCA can indeed emerge from simple vesicles (i.e., V1 and V2) as long as vesicles are allowed to merge and absorb each other, in addition to absorbing and integrating amino acids, short peptides, nucleotides, and short RNA molecules into their existing peptides (or proteins) and RNA molecules.

Again, because small vesicles inevitably perish or are merged and absorbed by other larger vesicles, if vesicles can only be generated de novo and in situ with each pond, simulation time will be extremely long within the constraint of computational resource. We therefore too add exogenously generated V1 and V2 vesicles to each pond to speed up the process. 2000 to 3000 V1 or V2 are added in the dry phase, while 4000-5000 V1 or V2 are added in the wet phase.

4. **Results**

We now present the simulation results.

4.1. *The First Stage: the origin(s) of FUCAs*

In various settings, including different volumes of the “warm little pond” and different starting numbers of V1 and V2 vesicles, with about 60-70 ticks, the first FUCA (i.e., V6) will be produced. Eventually, with about 100 to 200 ticks, anywhere from 70 to 300 FUCAs will be produced (Table 5 and Figure 4A-4D).

Table 5. Evolution of less than 100 FUCAs under different settings.

Settings	A	B	C	D	E	F	G	H
Volume of the pond (units)	60	72	66	69	52	93	85	62
No. of V1 vesicles	1863	1845	1888	1928	1784	1608	1624	1568
No. of V2 vesicles	1867	1849	1895	1891	1785	1610	1621	1552
No. of V3 vesicles	926	926	944	980	906	903	897	880
No. of V4 vesicles	265	280	288	317	374	439	438	471
No. of V5 vesicles	71	77	89	98	120	171	174	239
Total no. of vesicles (V1-V6)	5028	5019	5149	5266	5039	4865	4893	4994
No. of V6 as FUCAs produced	36	42	45	52	70	134	139	284
Ticks needed	112	115	118	125	128	140	142	170

Figure 4. Snapshots of the evolution of FUCAs (Stage 1): a single simulation. Note: The left window of the panels (A, B, C, D) records the evolutionary process of FUCAs (or type V6 vesicles) in simulation, with the red line highlights the appearance individual FUCAs. The right window plots the ticks (x-axis) with the number of different types of vesicles (y-axis) within the pond. The total number of FUCAs is in red in the right window.

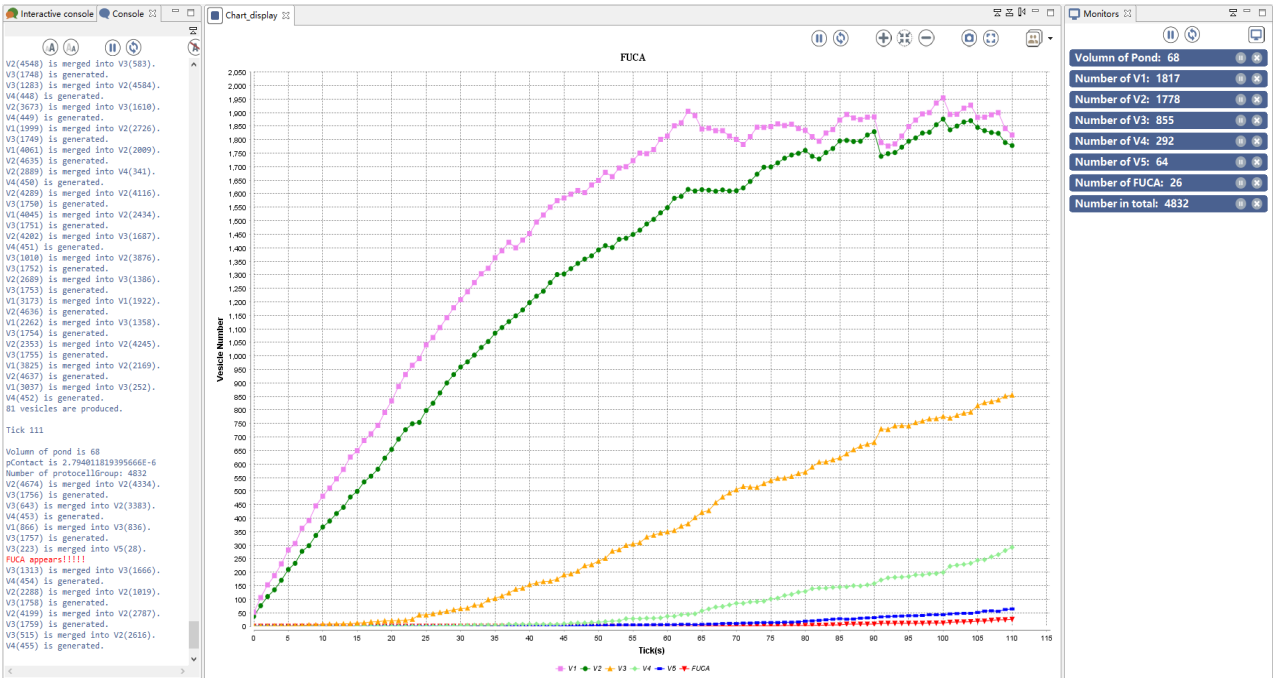


Figure 4. A. 26 FUCAs produced, 110 cycles.

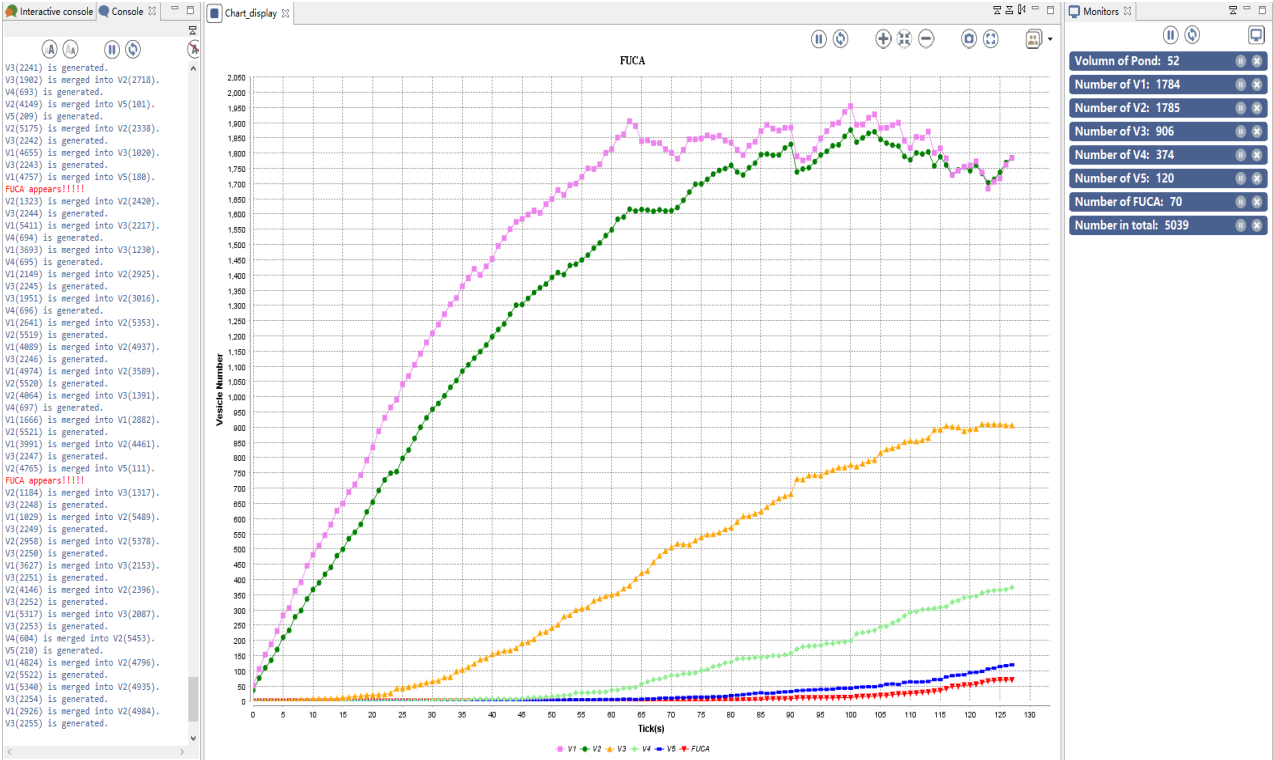


Figure 4. B. 70 FUCAs produced, 128 cycles.

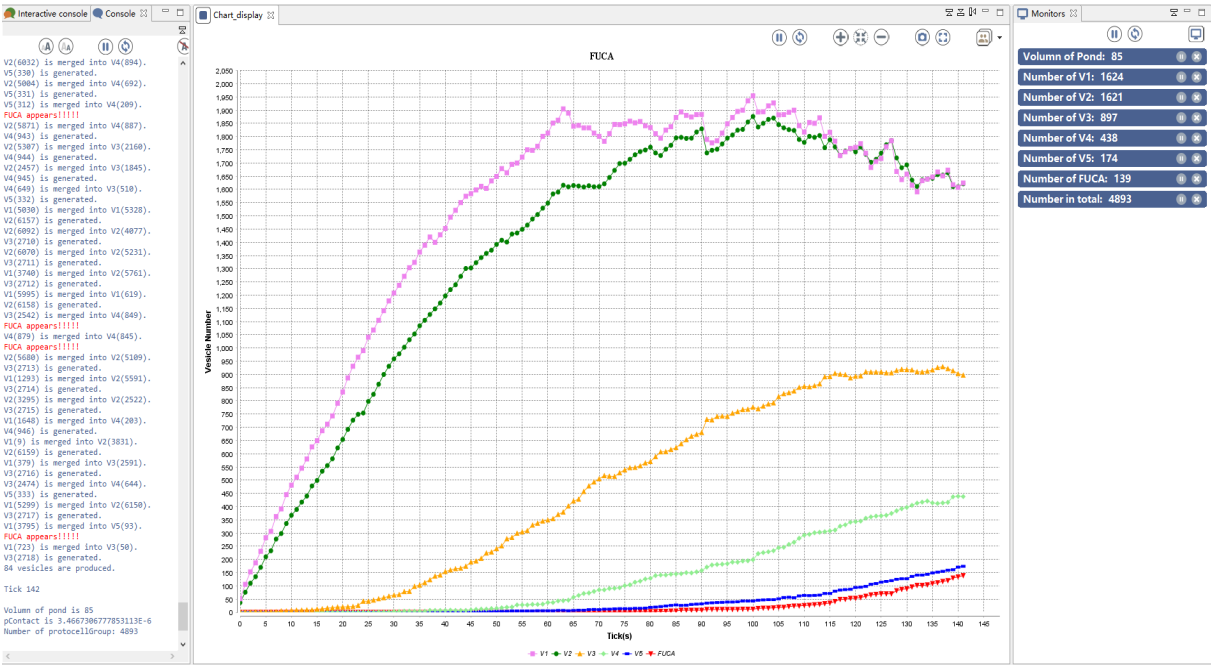


Figure 4. C. 139 FUCAs produced with 142 ticks.

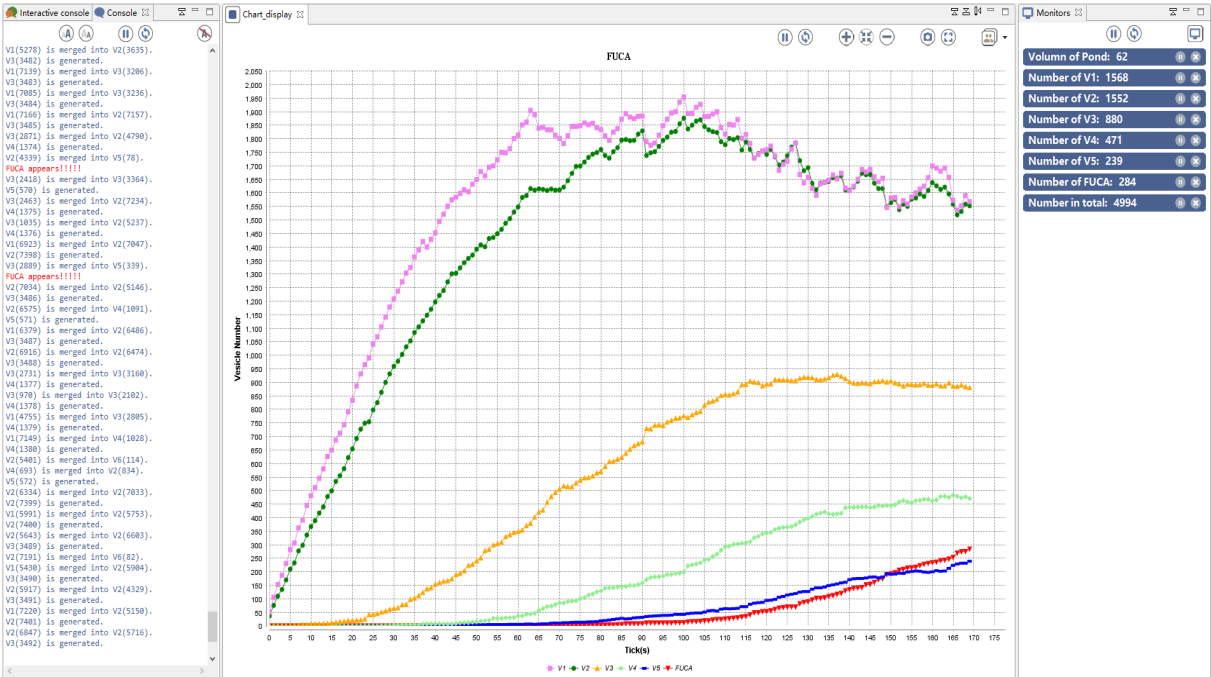


Figure 4D. 284 FUCAs produced with 170 ticks.

4.2. The Second Stage: the origin of LUCA

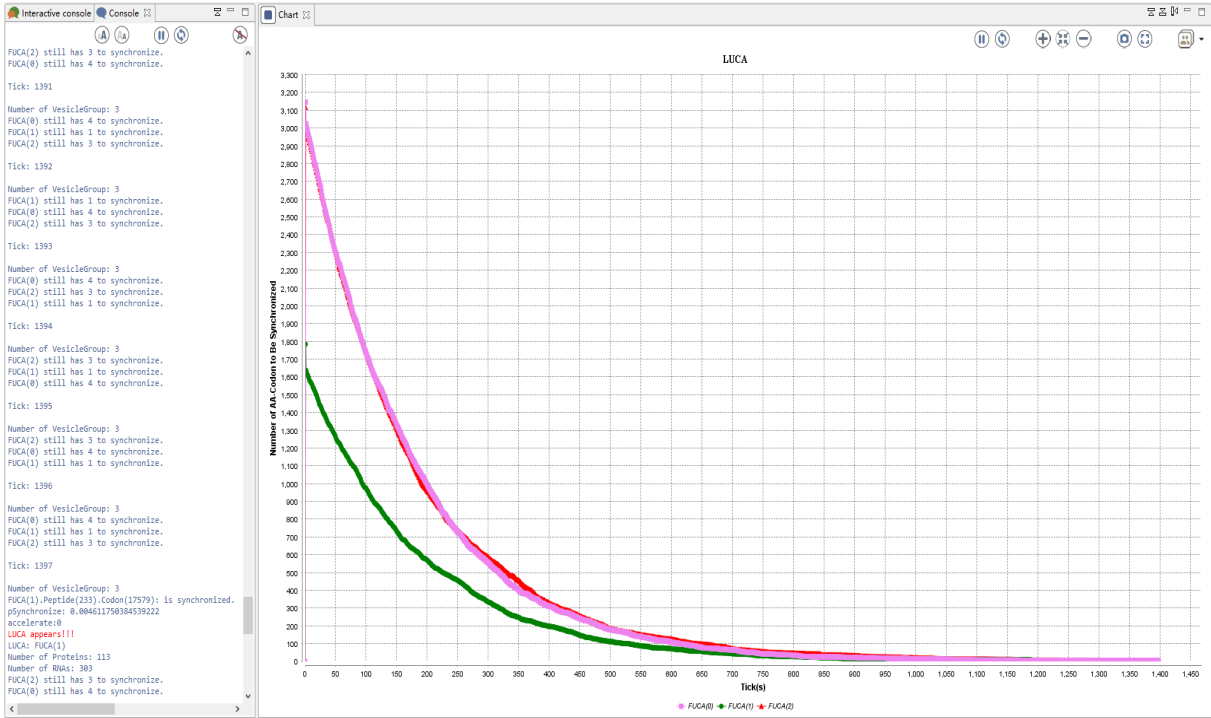
With different numbers of FUCAs (i.e., V6 vesicles), eventually a LUCA will appear, even though different FUCAs become the LUCA in different simulations in different time frames (Table 6 and Figure 5A-5B). Hence, as long as each FUCA is allowed to assign its amino acids and RNAs in each tick, LUCA is an inevitable evolutionary outcome.

Table 6. From FUCAs to LUCA (Stage 2): starting with 300 FUCAs.

Simulations	FUCAs						Which FUCA becomes LUCA?
	FUCA-(0)		FUCA-(1)		FUCA-(2)		
	Proteins	RNAs	Proteins	RNAs	Proteins	RNAs	

1	126	375	101	358	127	327	FUCA(0)
2	133	435	135	390	124	384	FUCA(2)
3	127	333	115	391	128	433	FUCA(2)
4	121	442	127	420	150	382	FUCA(2)
5	114	310	134	316	128	305	FUCA(2)
6	130	446	116	321	139	376	FUCA(1)
7	112	325	105	416	124	344	FUCA(1)
8	117	327	135	324	134	348	FUCA(0)
9	131	435	117	420	112	390	FUCA(2)
10	103	422	134	406	433	310	FUCA(0)
11	115	406	107	301	118	377	FUCA(1)
12	139	365	100	302	127	447	FUCA(1)
13	119	369	135	407	131	349	FUCA(2)
14	121	351	108	416	129	361	FUCA(1)
15	150	385	113	303	139	376	FUCA(1)
16	140	435	145	343	135	307	FUCA(2)
17	146	396	121	415	130	382	FUCA(2)
18	118	357	143	333	102	417	FUCA(2)
19	107	331	115	323	120	389	FUCA(1)
20	145	366	102	304	147	406	FUCA(2)

Note: FUCA(0), FUCA(1), and FUCA(2) denote the different lineages of FUCA. As shown in the table, eventually a LUCA will appear, even though a different FUCA may “win” in different simulations.



(A)

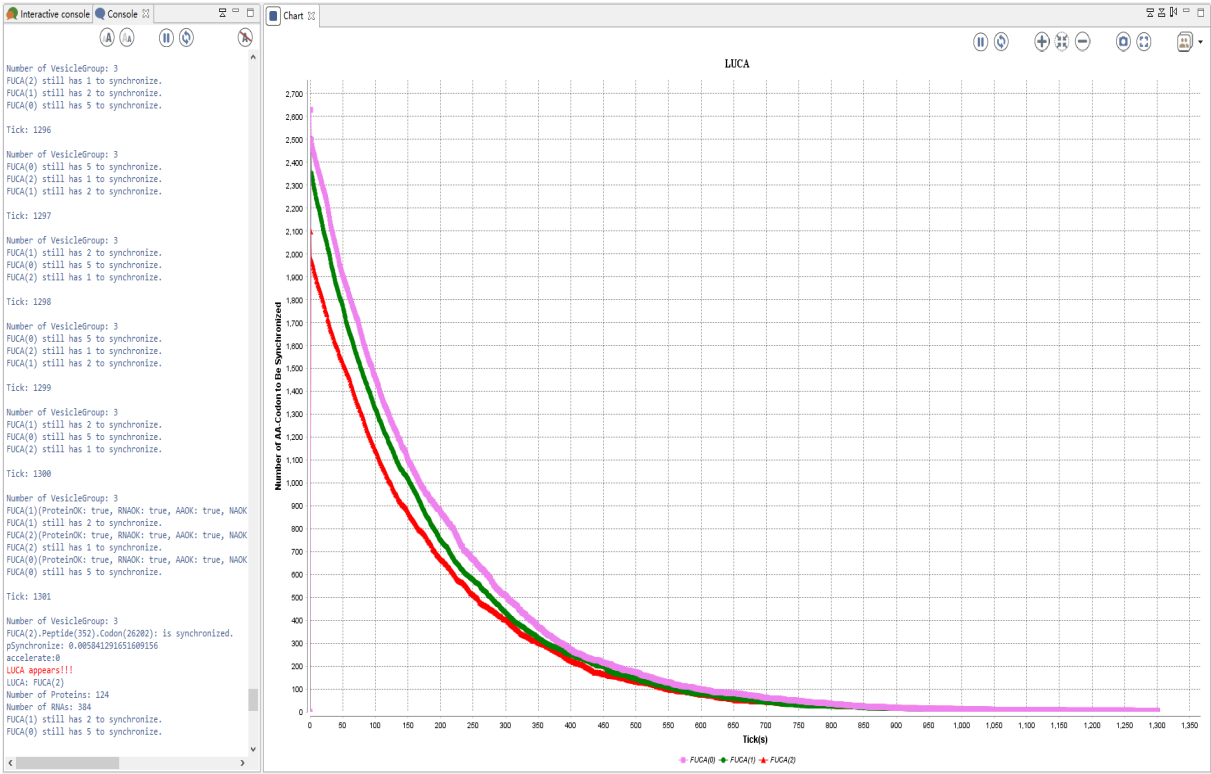


Figure 5. Number of Ticks for Synchronizing All the Codons (Stage 2), two simulations. Note: The left window of both panels records the evolutionary process of LUCA in simulation, with the red line highlights the appearance of the first LUCA. The right window plots the ticks for completing the synchronizing process, with its y-axis denoting the number of the sets of codons to be assigned.

4.3. The Two Stages Together

Again, with different numbers of V1 and V2 initially, LUCA can emerge with different number of cycles, with different number of other protocells or vesicles within a pond. Table 7 summarizes four such simulations while Figures 6A and 6B show results of two different simulations.

Table 7. Summary of Four Simulations with the Two Stages Pooled Together.

Parameters\Simulations	1	2	3	4
No. of ponds	58	76	55	57
Initial number of V1s and V2s	1169; 2036	1014; 1812	985;1839	993;1885
Percentage of peptides or proteins longer than 50 AAs	80%	80%	80%	80%
Percentage of RNAs longer than 150 NBs	70%	75%	78%	80%
Number of protocells perished	2946377	3712377	3083724	4085174
No. of alive protocells when the first LUCA emerged	3510	3125	3212	3337
No. of cycles until the first LUCA emerged	800	1000	840	1120

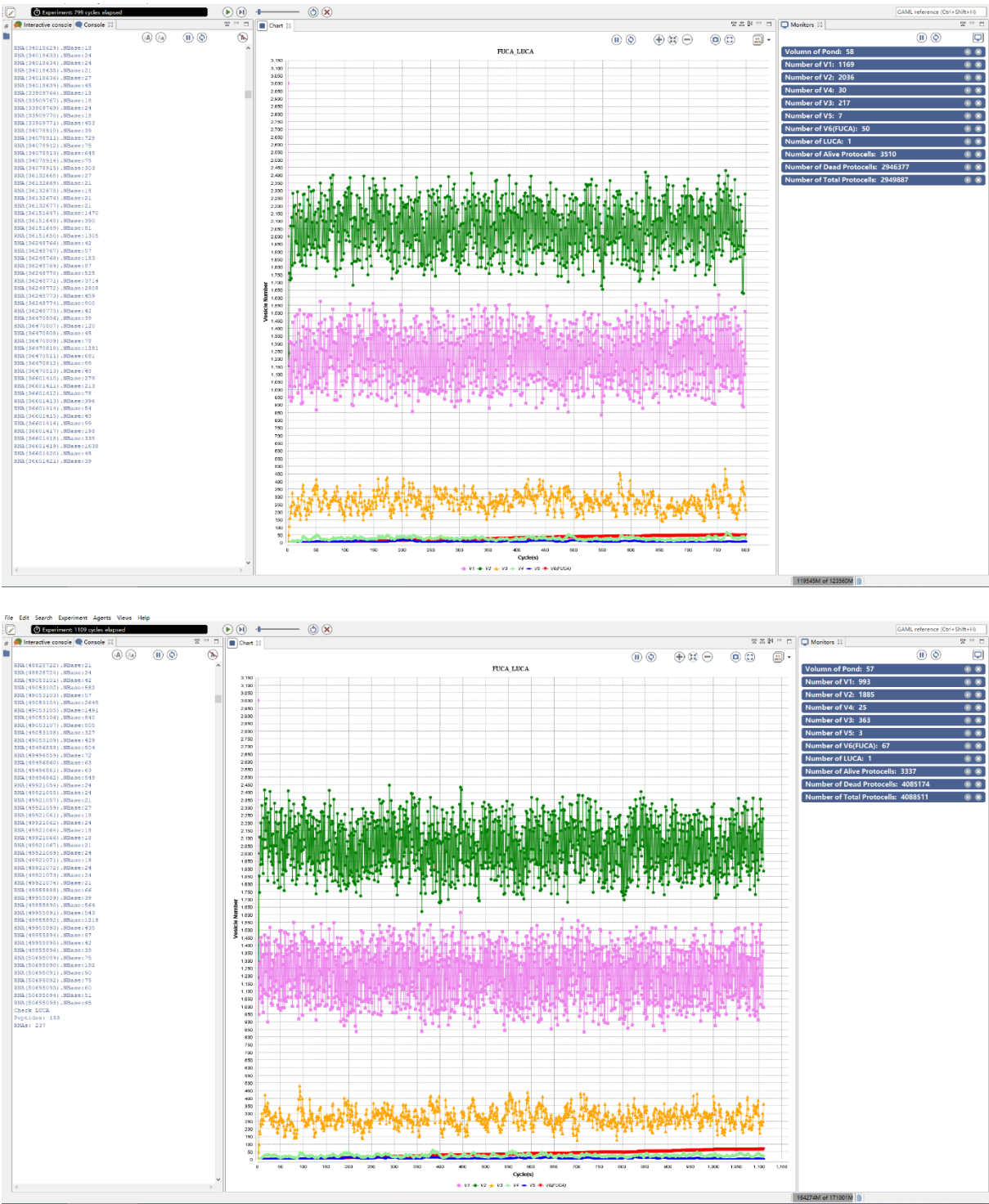


Figure 6. Results of Two Specific Simulations with the Two Stages Pooled Together.

5. Discussion

According to Tang (2021), vesicles’ absorption, acquisition, and fusion via breaking-and-repacking, proto-endocytosis, proto-endosymbiosis and other similar processes had been a central and powerful force in the pre-Darwinian evolution before LUCA, long before eukaryogenesis (e.g., Sagan 1967; Margulis 1981; 1991; O’Malley 2014), because these processes are processes of variation, selection, and retention.

Absorption, acquisition, and merger are processes of variation because they produce different compartmentalization and hence different crowding, combination, and coevolution of biomolecules within vesicles. Absorption, acquisition, and merger are also processes of selection and retention

because via these processes, some molecules will be retained and integrated within vesicles while some will be excluded from vesicles, and some vesicles will no longer exist. Moreover, absorption, acquisition, and merger entail extensive HBMT rather than merely HGT: HBMT thus subsumes HGT. Indeed, only with HBMT, could have pre-Darwinian evolution drawn from “global inventions” and overcome the seemingly insurmountable hurdle of bringing “the overwhelming amount of novelty needed to bring modern cells into existence” (Woese 2004, 182). This process of producing LUCA via vesicles’ absorption, acquisition, and merger is far more plausible than the scenario that LUCA has to evolve from a single FUCA *de novo* and in situ (Tang 2021).

In this paper, we have shown that Tang’s (2021) hypothesis is indeed viable, at least in computer simulation. If so, we can expect that some of the chemical and biological hypotheses advanced by Tang (2021) are also plausible. We therefore hope that our study can stimulate laboratory testing of these hypotheses to show that vesicles’ absorption, acquisition, and merger has indeed been a central force in driving the evolution of LUCA.

Finally, our computer simulation lends more support to the thesis that FUCAs and LUCA had most likely evolved from “fluctuating volcanic hot spring pools” (or Darwin’s “warm little ponds”) on land rather than from alkaline hydrothermal vents in the ocean (for overviews, see Damer and Deamer 2015; 2020; for evidence, see Mulikidjanian et al. 2012; Damer 2016; Milshteyn et al. 2018; Deamer et al. 2019; cf. Koonin and Martin 2005; Martin and Russell 2007; Lane and Martin 2012). Most critically, terrestrial hot spring pools allow the wet-and-dry cycles and hence can drive the breakup, repackaging, absorption, merger, and acquisition of vesicles whereas hydrothermal vents do not.

Author Contributions: S.T. conceptualized and designed the study. M. G. performed the experiment. S. T. and M. G. analyzed the results together. The authors co-wrote the paper.

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

Acknowledgments: To Carl Woese (1928-2012), for his conceptual breakthrough of cellular evolution.

References

1. Armstrong DL, Lancet D, Zidovetzki R (2018) Replication of simulated prebiotic amphiphilic vesicles in a finite environment exhibits complex behavior that includes high progeny variability and competition. *Astrobiology* 18:419–430.
2. Charlebois RL, Doolittle WF (2004) Computing prokaryotic gene ubiquity: rescuing the core from extinction. *Genome Res* 14:2469–2477.
3. Damer B (2016) A field trip to the Archaean in search of Darwin’s warm little pond. *Life* 6:21. <https://doi.org/10.3390/life6020021>
4. Damer B, Deamer D (2015) Coupled phases and combinatorial selection in fluctuating hydrothermal pools: a scenario to guide experimental approaches to the origin of cellular life. *Life* 5:872–887.
5. Damer B, Deamer D (2020) The hot spring hypothesis for an origin of life. *Astrobiol* 20:429–452.
6. Deamer D (2019) *Assembling life*. Oxford University Press, Oxford.
7. Deamer, DW, Damer, B, Kompanichenko V (2019) Hydrothermal chemistry and the origin of cellular life. *Astrobiology* 19:1523–1537.
8. Francis BR (2013) Evolution of the genetic code by incorporation of amino acids that improved or changed protein function. *J Mol Evol* 77:134–158.
9. Goldman AD, Bernhard TM, Dolzhenko E, Landweber LF (2013) LUCApedia: a database for the study of ancient life. *Nucleic Acids Res* 41:D1079–D1082.
10. Hargrave M., Spencer S.K., and Deamer D.W. (2018) Computational models of polymer synthesis driven by dehydration/rehydration cycles: repurination in simulated hydrothermal fields. *J Mol Evol* 86:501–510
11. Harris JK, Kelley ST, Spiegelman GB, Pace NR (2003) The genetic core of the universal ancestor. *Genome Res* 13:407–412.
12. Herron MD (2021) What are the major transitions? *Biol. & Philosophy* 36:2 <https://doi.org/10.1007/s10539-020-09773-z>
13. Higgs PG (2016) The effect of limited diffusion and wet-dry cycling on reversible polymerization reactions: implications for prebiotic synthesis of nucleic acids. *Life* 6:24.
14. Kim YE, Higgs PG. (2016) Co-operation between polymerases and nucleotide synthetases in the RNA world. *PLoS Comput Biol*. 2016;12:e1005161
15. Klein A, Bock M, Alt W (2017) Simple mechanisms of early life-simulation model on the origin of semi-cells. *Biosystems* 151:34–42.

16. Koonin EV (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat Rev Microbiol* 1:127–136.
17. Koonin EV (2014a) Carl Woese's vision of cellular evolution and the domains of life. *RNA Biol* 11:197–204.
18. Koonin EV (2014b) The origin of cellular life. *Antonie Van Leeuwenhoek* 106:27–41.
19. Koonin EV, Martin W (2005) On the origin of genomes and cells within inorganic compartments. *Trends Genet* 21:647–654.
20. Koonin EV, Novozhilov AS (2017) Origin and evolution of the universal genetic code. *Annu Rev Genet* 51:45–62.
21. Kovacs NA, Petrov AS, Lanier KA, Williams LD (2017) Frozen in time: the history of proteins. *Mol Biol Evol* 34:1252–1260.
22. Lancet D, Zidovetzki R, Markovitch O (2018) Systems protobiology: origin of life in lipid catalytic networks. *J R Soc Interface* 15:20180159. <https://doi.org/10.1098/rsif.2018.0159>
23. Lane N, Martin WF (2012) The origin of membrane bioenergetics. *Cell* 151:1406–1416.
24. Ma W, Yu C, Zhang W, Hu J (2007) Nucleotide synthetase ribozymes may have emerged first in the RNA world. *RNA* 13:2012–2019
25. Ma WT, Hu JM. (2012) Computer simulation on the cooperation of functional molecules during the early stages of evolution. *PLoS One*. 2012;7: e35454.37
26. Margulis L (1981) *Symbiosis in cell evolution: life and its environment on the early earth*. WH Freeman, San Francisco.
27. Margulis L (1991) Symbiogenesis and symbiogenesis. In: Margulis L, Fester R (eds) *Symbiosis as a source of evolutionary innovation: speciation and morphogenesis*. MIT Press, Cambridge, pp 1–14.
28. Martin W, Russell MJ (2007) On the origin of biochemistry at an alkaline hydrothermal vent. *Philos Trans R Soc Lond B Biol Sci* 362:1887–1925.
29. Maynard Smith J, Szathmáry E (1993). The evolution of chromosome I. Selection for Linkage. *J. Theor Biol* 164:437–446.
30. Maynard Smith J, Szathmáry E (1997) *The Major Transitions in Evolution*. Oxford.
31. Milshteyn D, Damer B, Havig J, Deamer D (2018) Amphiphilic compounds assemble into membranous vesicles in hydrothermal hot spring water but not in seawater. *Life* 8:11.
32. 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* 3:2.
33. Mulikidjanian AY, Bychkov AY, Dibrova DV, Galperin MY, Koonin EV (2012) Origin of first cells at terrestrial, anoxic geothermal fields. *Proc Natl Acad Sci USA* 109: E821–E830.
34. Nunes Palmeira R, Colnaghi M, Harrison SA, Pomiankowski A, Lane N. (2022) The limits of metabolic heredity in protocells. *Proc. R. Soc. B* 289: 20221469. <https://doi.org/10.1098/rspb.2022.1469>.
35. O'Malley MA (2014) Endosymbiosis and its implications for evolutionary theory. *Proc Natl Acad Sci USA* 112:10270–10277.
36. Ranea JA, Sillero A, Thornton JM, Orengo CA (2006) Protein superfamily evolution and the last universal common ancestor. *J Mol Evol* 63:513–525.
37. Ross, DS, Deamer D (2016). Dry/Wet Cycling and the Thermodynamics and Kinetics of Prebiotic Polymer Synthesis. *Life* 2016, 6, 28; doi:10.3390/life6030028
38. Sagan L (1967) On the origin of mitosing cells. *J Theoret Biol* 14:225–274.
39. Sengupta S, Higgs PG (2015) Pathways of genetic code evolution in ancient and modern organisms. *J Mol Evol* 80:229–243.
40. Szathmáry E (2007) Coevolution of metabolic networks and membranes: the scenario of progressive sequestration. *Phil Trans R Soc B* 362:1781–1787.
41. Szathmáry E (2015) Toward major evolutionary transitions theory 2.0. *Proc Natl Acad Sci USA* 112: 10104–10111.
42. Takagi YA, Nguyen DH, Wexler TB, Goldman AD (2020) The Coevolution of cellularity and metabolism following the origin of life. *J Mol Evol* 88: 598–617.
43. Tang S (2020) Pre-Darwinian evolution before LUCA. *Biol Theory* 15:175–179.
44. Tang S (2021) The origin(s) of cell(s): pre-Darwinian evolution from FUCAs to LUCA. *J Mol Evol* 89: 427–447.
45. Vetsigian K, Woese C, Goldenfeld N (2006) Collective evolution and the genetic code. *Proc Natl Acad Sci USA* 103:10696–10701.
46. Woese CR (1998) The universal ancestor. *Proc Natl Acad Sci USA* 95:6854–6859.
47. Woese CR (2000) Interpreting the universal phylogenetic tree. *Proc Natl Acad Sci USA* 97:8392–8396.
48. Woese CR (2002) On the evolution of cells. *Proc Natl Acad Sci USA* 99:8742–8747.
49. Woese CR (2004) A new biology for a new century. *Microbiol Mol Biol Rev* 68:173–186.
50. Woese CR, Fox GE (1977) The concept of cellular evolution. *J Mol Evol* 10:1–6.

51. Wolf YI, Koonin EV (2007) On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization. *Biol Direct* 2:14. <https://doi.org/10.1186/1745-6150-2-14>
52. Wong JT (1975) A co-evolution theory of the genetic code. *Proc Natl Acad Sci USA* 72:1909–1912.
53. Wong JT (2005) Coevolution theory of the genetic code at age thirty. *Bioessays* 27:416–425.
54. Yarus M (2017) The genetic code and RNA-amino acid affinities. *Life* 7:13.
55. Yarus M (2021) Evolution of the Standard Genetic Code. *J Mol Evol* 89:19–44.
56. Yin S, Chen Y Yu C Ma W (2019). From molecular to cellular form: modeling the first major transition during the arising of life. *BMC Evolutionary Biology* (2019) 19:84 <https://doi.org/10.1186/s12862-019-1412-5>
57. Zhu TF, Adamala K, Zhang N, Szostak JW (2012) Photochemically driven redox chemistry induces protocell membrane pearling and division. *Proc Natl Acad Sci USA* 109:9828–9832.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.