Review

Not peer-reviewed version

The 'LUCAN' Concept, the Latest Universally Common Ancestral Network Stage of Early Life: From Molecular Cooperation and Horizontal Symbiosis to the Escape of Vertically Stabilized Lineages of Particulate Viral Propagules and Cellular Organisms

Richard Egel *

Posted Date: 9 May 2024

doi: 10.20944/preprints202405.0555.v1

Keywords: origin of evolution; surface metabolism; progenote state; collective optimization; genome nucleation; coevolution; Darwinian threshold; Darwinian speciation; tree of life

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2024 doi:10.20944/preprints202405.0555.v1

Review

The 'LUCAN' Concept, the Latest Universally Common Ancestral Network Stage of Early Life: From Molecular Cooperation and Horizontal Symbiosis to the Escape of Vertically Stabilized Lineages of Particulate Viral Propagules and Cellular Organisms

Richard Egel

Department of Biology, University of Copenhagen Biocenter, Ole Maaløes Vej 5, DK-2200 Copenhagen N, Denmark; regel@bio.ku.dk

Abstract: The main objective is to offer a functionally coherent counter-narrative to the polarized views prevailing now about the complex 'Eukaryogenesis problem' versus the "primitive" nature ascribed to so-called "Prokaryotes" (termed Akaryotes herein). The paradigm shift presented aims at bridging the conceptual gap between the rudimentary beginnings of biomolecular cooperativity on a pristine Earth (the Origins of Life or OoL) and the beginnings of lineage-wise diversification toward Darwinean speciation in the Tree of Life (or ToL). Although the research field as a whole is not aware of a corresponding 'Prokaryogenesis problem', the evolutionary process resulting in the partly independent archaeal and bacterial cell types is not fully understood. The present paper provides an explanation for these problems based on coevolutionary complementarity at multiple levels. The unconventional model connects various stages from "neighborhood selection" in film-like layers of unbounded "surface protoplasm" based on short "statistical proteins", via a closely knit "peptide/RNA partnership" in the "making of genes", toward the "making of genomes" in several RNA-to-DNA transitions. Three virus-related plasmids carrying different DNA replicases may have initiated the vertical stability of hereditable lineages that led to the three "domains" of organismal life, with a fourth lineage ending in eukaryotic organelles. These partly independent transitions established DNA as a late-comer of fundamental biomolecules.

Keywords: origin of evolution; surface metabolism; progenote state; collective optimization; genome nucleation; coevolution; Darwinian threshold; Darwinian speciation; tree of life

---

1. Preamble: — "What is Life?" — in Schrödinger's Legacy

Life as such cannot be defined in mathematical or quantum physical terms to my mind, [1]. but it can humbly be described in more or less poetic phrasings, such as "*Life is a self-sustaining chemical system capable of Darwinian evolution*" according to NASA's concerns '*About Life Detection*' elsewhere in the universe [2], or as '*peculiar pockets of space and soft organic matter, organized and universally connected as quasi-stable organismal lineages in the tangled web of deep evolutionary time, potentially ever-lasting by repeatedly adapting to newly opened ecological niches and inadvertently causing local and global changes in worldwide connected environmental networks*' (my personal musings), which in hindsight have successfully walked a precarious tightrope between diffusive dispersal into open voids and solidified stagnation by irreversible crystallization. My major concerns are rather how physical/chemical matter can possibly have achieved the '*living state*' on a pristine, pre-biotic Earth about four billion years ago when *Darwinian evolution by natural selection* began to make a difference, eventually also resulting in *Darwinian speciation*.

Notably, two distinguished heroes of the field have drawn on the *crystals* metaphor to characterize particular aspects of the enigmatic living state: (1) theoretical physicist Erwin Schrödinger in his highly influential essay "What is Life?" [3], where he conceptually considers the "*chromosome fibre*" as a special kind of "*aperiodic crystal*" (in the unidimensional sequence space of macromolecular order in RNA and DNA, as we know now) – in contrast to periodic crystals in "*three-fold periodic lattices*", as traditionally dealt with in physics and chemistry; and (2) biophysicist (turned evolutionary microbiologist) Carl Woese in his discussion of tentative ancestral states before the consolidation of genuine biological organisms when he uses the '*crystallizing*' metaphor in the context of accretionary consolidation of modular subsystems of biological functionality within a cell or organism. In analogy to *crystallization* by physical annealing, and similar to ribosomes, the composite replisomes represent an important functional module amongst "*new cellular subsystems that are refractory to major evolutionary change*" [4]. Just as Woese used ribosomal RNA as a superior proxy for lineage-wise diversification in the organismal *Tree of Life* (ToL), I herein suggest using three to four different modes of replisome consolidation as a differential proxy for projecting the formally dichotomous pattern of the canonical ToL on to potential networking interactions of systemwide functional significance at the evolutionary turning point of genomic RNA-to-DNA transitions.

Throughout this potentially integrative Feature Paper I agree with the central argument that "*Life arises when lineage-forming entities collaborate in metabolism*" [5] — at least in the *digital sophistication* of the living state. The latter caveat then leaves a conceptual loophole for *analog* life-like functionality to have paved the "*Path to the Digital Cell*" [6]. On second thought, the conception of lineage-forming entities is a two-sided issue of vertical descent: At organismal and/or cellular levels there are physically structured lineages of self-similar (not necessarily fully identical) reproduction; at the genomic level, however, there are the macromolecular lineages of digital replication nucleotide by nucleotide for hundreds or thousands of repetitive biochemical reaction steps in a row. This conceptual dichotomy is correlated with the existence of two fundamentally different classes of lineage-forming principles: one-dimensional chain-like molecules, which can be replicated step by step, vs. two-dimensional amphiphilic membranes, which in watery surroundings are stable only in a topologically closed configuration. The latter constraint — now based on more advanced, longer-chain lipids — has effectively locked the present state of membrane growth and formative reorganization into a general dependency on preexisting membrane-bounded entities of topologically closed configuration.

As for life as we know it, no living entity can persist for long without a supportive environment. Moreover, rather few if any living beings can now exist without supportive interactions with other creatures living near or having lived before. Accordingly, all life unfolds within the context of a conceptual "*Ecology–Evolution Continuum*" [7-9]. It is then reassuring to learn that also the conceivable scenarios for the primordial *Origins of Life* (OoL) on Earth must be judged within the formal framework of an *Ecology–Evolution Continuum* as generalized in terms of "*localized chemical reaction systems as autocatalytic chemical ecosystems (ACEs): food-driven (open) systems that can grow due to the action of autocatalytic cycles (ACs)*" [10]. This superior mandate does not only extend the existential continuity of Life into a transitional scenario of biogenic evolution, but it also begs the historical question of which local conditions on a differentiating pristine Earth could possibly have formed the most suitable environment for Life's natural emergence.

It also appears more generally that present life "*can act as a guide to the full origin-of-life continuum, although some significant gaps remain*" [11]. Whilst the high-profile Harrison et al. review just cited is very detailed and convincing at the metabolic level, it just represents one particular set of speculations at the structural and organizational levels. It thus remains for the future to develop additional overall scenarios — likewise contiguous from prebiotic chemistry to genuine cells but with different organizational input — for drawing meaningful conclusions from critical pair-wise comparisons as regards the historical conundrums concerning the evolutionary origins of fundamentally different cell types along with the tentative rooting of the organismal *Tree of Life*. In the present Feature Paper I cast a net of multiply interconnected systemic considerations — with potentially paradigm-shifting implications and significance for how just two or three cell types in the '*primary lines*' of organismal descent have consolidated from a collective mesh of polyphenotypic populations of ancestral '*Progenote*' entities.

As inspired by the Third Way of Evolution Group [12,13] in striving for a grander evolutionary synthesis, this Feature Paper gives a provisional answer to a rhetorical question. "A Third Way" of connecting organismal lineages to a collective 'Progenote Era' — How could that be?

2. Biogenic Environment

The long-favored assumption that a *deep-sea vents scenario* for 'Origins of Life' (OoL) [14,15] is being replaced with the more recent notion that alternative scenarios associated with a variety of *terrestrial geochemical reactors* at air-exposed (anoxic) hydrothermal fields appear more likely and advantageous to OoL for a variety of reasons. For one thing, this is more compatible with ionic composition of modern cells [16], for another, the variable contact of solid earth with both atmosphere and hydrosphere could also expose a newly emerging life-like system to recurring dehydration–rehydration cycles [17,18].

The cyclic recurrence of environmental fluctuations observed in terrestrial surface settings is in contrast to the quasi-steady state of hydrothermal gradients and flow patterns deep under the ocean. This was a great advantage from a theoretical point of view, which posits a requirement of high-frequency and multilevel oscillations in hydrothermal environments to actively drive prebiotic organic microsystems away from thermodynamic equilibration [19,20]. Such cyclic fluctuations were also appropriate to overcome the '*water paradox*' in prebiotic chemistry between the requirement of water for life's existence and hydrolytic breakdown of macromolecules in free solution [21].

3. The Hidden Path from OoL to ToL

Where do we all come from? — And how did we get here?

Two existential questions regarding the presence of organismal Life on Earth —humankind included — are well worth asking yet difficult to resolve definitively in a scientifically rigorous way. One problem concerns the *Origins of Life* (OoL) from prebiotic geochemical beginnings, the other means uniting the nested similarities of evolutionary relationships in the branching patterns of a canonical *Tree of Life* (ToL). Importantly, the conventional approach of calculating the '*best fit*' for models of the universal ToL from extant *phylogenomic* comparisons alone cannot find a plausible, robust solution for the second problem at the *common root* without having a robust basic theory from OoL research in the first place. For the time being, expert discussions hover between two opposing views, as to whether a basic split into two or three fundamentally different '*Urkingdoms*' or (*Phylo*)-*Domains* — 2D-ToL vs. 3D-ToL — is a better model of our deepest historical past, putting *Archaea* into a key position between *Bacteria* and *Eukarya*.

- A personal note on basic terms: Regarding the "*Irreducible Nature of Eukaryote Cells*" [22] and questioning the irrationally founded assertion that all of *Eukaryote* complexity 'must' have descended from the conceptionally more simplistic modalities of bacterial/archaeal cell organization, which is intuitively associated with the conventional '*Prokaryote*'/*Eukaryote* distinction, I herein follow the suggestion to use a neutral, less teleologically loaded term — *Akaryote* — for both bacterial and archaeal cells [23,24].

Whilst the *Phylogenomics* community has argued for a 2D-ToL version deriving eukaryotes from Asgard archaea [25,26], the complementary *Phyloproteomics* approach is supporting Woese's original 3D-ToL proposal by concluding that "*diversification of eukaryotes and akaryotes from* [a common root] *is a better supported hypothesis rather than a prokaryote-to-eukaryote transition being assumed to interpret poorly resolved trees*" [27]. The latter view upholds long-held concerns from the viral community that eukaryotes have merely shared a common ancestor with archaea but are not direct descendants from within a primary Archaea Domain — not even from Asgard archaea, despite the fact that this group in particular has more protein domains in common with eukaryotic cells than other archaea in general [28].

Accordingly, there is still a "need for more accurate construction of the topology of the phylogenetic trees, better focus on the **archaeal host** or the **common archaeal/eukaryotic ancestor …**

But the question of which method delivers trees with the best likelihoods is not identical to the question of which method delivers the trees that best reflect **biological history**" [29]. The latter part of this quote addresses the 'Achilles heel' of calculating the "trees with the best likelihoods" — which merely delivers some "Statistical Tree of Life" [30] no matter how sophisticated a filtering scheme of systematic readjustments one might apply to the ever-growing body of empirical data. The weakest spot is and remains to be a conceptual schism between the very narrow basis in an emaciated "tree of 1 %" [31], and the more relevant "Process Pluralism and Pattern Pluralism" of biological history, which is potentially hiding out in the residual majority of 99 % discarded by default [32].

In this paper I will provisionally refer to the '*Tree vs. History Schism*' as a latent '*Phylogenomics paradox*' (taken up further below), which deserves to be embraced rather than merely ignored. To this end I aim to develop a non-conventional, more plausible narrative to connect some novel trends in OoL research with the still controversial rooting problem of the ToL to provide a potentially more robust synthesis than what the received consensus is offering in this regard. In particular, *to better reflect biological history*, I will also draw attention to the very basic functional trait of chromosomal replisome organization to distinguish bacterial cells and genomes from both archaea and eukaryotes, which commonly is set aside in formal reconstructions at the domain level of the organismal ToL.

Scouting the '*Hidden Path from OoL to ToL*' (to trace the rooting and emergence of *biological history*) must conceive of a plausible link from a strong foundation in Boltzmann's stochastic principle of *monotonous entropy increase in statistical thermodynamics* (favoring spontaneous equilibration amongst innumerous molecular interactions) toward the upper reaches in Darwin's biological principle of *evolution by natural selection* (relentlessly avoiding indiscriminate thermodynamic equilibration at population-wide scales). Viewed from my background in molecular genetics, Carl Woese (a theoretical biophysicist who turned to evolutionary microbiology) [33] provided the most perceptive analysis of evolutionary conditions at a tentative pre-organismal state of early life, but he has not systematically considered all the interactive exit options that his unconventional inferences might have offered for principal cell-type evolution later on. The current article specifically discusses supplementary assumptions in support of Woese's original 3D-ToL proposal, including the putative existence of a "*nuclear-cytoplasmic lineage*" [34] to link eukaryotic cell organization directly to a universally common ancestral population at the very base of the canonical ToL.

4. Ready for a Superior Paradigm Shift in OoL Research?

The present Earth is teeming with life for about four billion years, the best part of the Globe's physical existence — in mutual co-existence with the present Moon. But how on Earth this life came into being in the first place has long been a matter of debate in *Origins of Life (OoL) research* [35]. This existential query draws scientific thoughts head-on into a profound epistemological perspective, probing at the outer limits of human knowledge. Like others [36], I am convinced that a major change of paradigm is overdue to simultaneously contest several widely held views on the geochemical emergence of life-like evolution, but trying to challenge the overall repository of a received consensus is not an easy matter.

There is a conceptual hierarchy of mutual relations among different scientific disciplines, not without ambivalent undertones: "*According to the current mainstream hierarchy, biology idolizes physics, the cognitive sciences idolize biology, the philosophy of mind idolizes the cognitive sciences, and the academic study of religion idolizes all above*" [37]. The '*idolizing*' of physics by biology, in particular, is problematic for OoL research in that current mainstream thinking has long been dominated by the mechanistic linearity of Newtonian physics, whereas modern physics itself has been advanced into the quantum world of holistic interconnectedness — with countless, unprecedented, paradigm-shifting consequences.

In general [38] scientific hypotheses should be testable for inferential consequences but Karl Popper's most stringent falsifiability concept — a demarcation criterion to differentiate scientific theories from nonscientific ones [39] — may not directly apply to the ancient evolutionary and historical aspects of OoL scenarios. To overcome this dilemma, it would be advisable to begin anew in developing viable, more inclusive theories, which subsequently could be evaluated in direct comparison with competing hypotheses [40]. I herein suggest a more embracing synthesis to accomplish such a goal, giving front-line priority to potential *networking capacity* across various levels of complexity. To connect the functionality of a living organism to physical laws in a modern sense, it is necessary to understand its organizational complexity in terms of atomic statistics, stochastic fluctuations, and regular flow patterns of matter coupled to influx and partial degradation of environmental energy [41]. This is the domain of statistical thermodynamics. As for the eventual 'testing' of the various hypothetical suggestions proposed herein, I only find it appropriate to draw on the modeling experience of *Evolutionary Game Theory* [42,43] and not to experimental approaches in any realistic type of bench-top setting.

Two prominent OoL conceptions have focused on protometabolic beginnings of a chemotrophic kind, assuming deep water at submarine hydrothermal settings [12,13]. and suggesting a general requirement for membrane-bounded encapsulation of vesicle-like protocells [44]. These attributions, however, were arguably insufficient for rapid protolife emergence.

Much of the received consensus is based on chemical considerations appropriate for study in bench top modeling experiments, but overly simplifying approaches have consistently failed to resolve several fundamental '*controversies*' [45,46]. Instead, just as physics-inspired reasoning has led to the founding of molecular biology — and of classical genetics before — a major impetus for the upcoming paradigm shift for OoL research is expected to come from system-building physical-chemical and biophysical considerations, backed up by computer-based simulation analyses instead of traditional bench-top chemistry of particular reactions pathways [36,37]. Experimental *prebiotic* chemistry, too, is beginning to make a system-building difference by mimicking metabolic reaction networks with prebiotic cofactor-like components [48].

Notably — with the irony of hindsight — two bearing biophysical elements of the forthcoming revision had already been considered in the earliest scientific proposals for realistic OoL scenarios: *photon absorption* from sunlight as a dependable source of energy utilization [49] and *liquid-liquid phase separation* in 'coacervates' as the principal driver of accretionary growth and borderless molecular compartmentation [50]. These basic concepts, though, have meanwhile been neglected or dismissed for supposedly "better" consensus propositions in currently held views.

To overcome the apparent lack of fundamental progress in much of traditional OoL research, it has been suggested that the "*Open Questions*" framework should be set aside in favor of alternative approaches to focus on "*paradoxes*" derived from different models of OoL theory [51]. Paradoxes generally arise by deliberately over-specifying a set of credibility assumptions, which can be resolved by allowing for one or more additional dimensions or degrees of freedom. Moreover, networks of independent models in a tenable synthesis of opposing views should primarily be judged on their paradox-resolving potential [52]. Likewise, the *networking potential* of different models as such should contribute substantially to the overall plausibility of the prospective meta-synthesis from a selected set of promising proposals.

To turn the point of view around, it is worth asking to what extent the various OoL conceptions themselves are least prone to evolve into increasing network complexity from within. Arguably, the so-called "*selfish-gene paradigm*" is no longer popular as an OoL-driving principle alone, and no single gene or other '*replicators*' have ever come about for strictly selfish reasons in the first place [53-56]. Next in line, various single-component OoL hypotheses gained widespread popularity in the past, such as the so-called *RNA World* or *Lipid World* scenarios, but the *RNA World* model itself has been subjected to critique, together with other '*privileged function*' proposals of this kind [57]. The notion of *privileged functions* applies to tentative OoL scenarios that postulate the evolutionary precedence of a particular, relatively simple yet universally prevalent facet of present life, implying that other universal features had to emerge as secondary innovations later on.

Of note to the latter reference, the authors deliberately subsumed *metabolism* under the '*privileged function*' category in *Metabolism First World* scenarios — mainly owing to the preconceived conviction that '*genetic takeovers*' from any non-genetic OoL scenario "*appear implausible*" (though not impossible). *Metabolism* was thereby put on equal footing with *energy harvesting* from chemical gradients, molecular *replication*, vesicular (or mineral) *compartmentalization*, and the like. In contrast to the overall set, however, *metabolism* concepts are categorically different in their capacity to cooperate in composite *networks* of geo/biochemical reactions and corresponding catalytic agents. From a system-building perspective, too — considering the living state as striking a precarious balance between structural persistence and processive change — metabolism can have played an important mediator role quite early on, not the least by way of the game-changing concept of "*surface metabolism*" at appropriate mineral substrates [58]. Gradual changes from relatively simple beginnings could first have led to a *protometabolic network* based on the uncoded *(stochastically assembled)* synthesis of '*statistical peptides*', and to more composite modularity at various levels later on. Especially I think the latter stage included the nucleation and progressive optimization of secondary feedback loops for the catalyzed storage and utilization of genetic information as a metabolism-stabilizing RNA-based back-up archive to reproducibly regenerate functional proteinaceous sets of catalytic, structural, and regulatory agents — to a lesser extent also functional RNA regulatory elements and certain ribozymes, with focus on a narrow range of RNA-modifying catalytic activity.

A conceptual paradox of particular relevance to this article arises from a lively dispute within the *Phyloproteomics* community in the pursuit of comparative evolutionary analyses complementary to the more conventional *Phylogenomics* approach. This points to an instructive problem, which further below is addressed as a '*Phyloproteomics paradox*' — related to the '*Phylogenomics paradox*' pointed out in the preceding Section. The two methodologically different yet complementary '*omics*'-approaches are going for a common goal: (i) searching (amongst the immense variety of living organisms) for ancestral phyletic relationships around a tentative common base (allowing to 'root' the universal ToL), and (ii) conceptually relating their respective results to the underlying, still unresolved '*Tree-vs.-History Schism*'. Of note, the *Phyloproteomics* approach is potentially more powerful in this regard by using a more comprehensive empirical data base of structural folding in characteristic protein domains, as compared to the overly narrow window of homology-based sequence alignments available to conventional *Phylogenomics* for deriving their emaciated "*tree of 1 %*". Phyletic inferences drawn from *Proteomics* analyses are thus more representative on a system-wide scale, and the biological relevance can be further refined by conceptually reducing biological complexity to fewer dimensions by '*Principal Component Analyses*' for graphical display as illustrative clustering effects [59]. — These tangled concerns are taken up again further below, regarding their potential import on the '*Superior Paradigm Shift*' advocated herein as a paradox-resolving proposition.

5. There is Room for Progress to be Made

There is widespread disquiet in the backyard of OoL research, no longer taking for granted all the various models and limited conceptions that had virtually congealed into a received consensus on what to popularize as authoritative presuppositions in this intrinsically speculative field [35,60,61]. A forthcoming *major paradigm shift* toward unified and integrated conceptions about the "*Hidden Path from OoL to ToL*" is picking up momentum on many partial aspects as follows:

5.1. Cohesive Coalescence

The prebiotic formation of structural compartments has temporarily been considered quintessential for preventing biogenic chemical components from outward diffusion against the ocean void — be it by amphiphilic soap or lipid molecules for membrane-like, vesicular encasement of tiny and simple '*protocells*' [62,63], or within small mineralized, environmental cavities [64], but this long-held presupposition is now also being left behind. As recognized decades ago (see above), a principal alternative exists as *liquid-liquid phase separation* in so-called '*coacervates*' for concentration-dependent, border-less compartmentation, which since has been revitalized in a modernized context [65]. In particular, the composite concepts of "*protobiofilms*" [66] and primordial "*surface protoplasm*" [67] have little to no need for simple *protocells* at the beginning but are based on the early accretion of much larger molecular conglomerates, especially if facilitated by recurring dehydration–rehydration cycles. To nucleate the physical *phase separation process* in a surface-dependent manner, the characteristic *Fourth Phase* of surface-structured water [68], with different dielectric properties [69], is also worth consideration.

As a matter of note, the notable cohesiveness of proteinaceous or RNP-rich *membrane-less organelles* appears crucially dependent on diverse populations of *intrinsically unstructured/disordered proteins* [70], which adds particular relevance to potential roles of *statistical peptides* in OoL scenarios based on primordial liquid-liquid *phase separation* [71]. Of particular importance for consistency and functionality in modern cells is the internal condition of '*molecular crowding*' [72]. This characteristic condition depends on partial dehydration [73], which could be periodically enforced by wet/dry cycling in appropriate prebiotic environments [17,18].

The downplay of membrane formation of prebiotic vesicular structures does not, however, exclude that the earliest membranes could have emerged and coevolved together with other macromolecular aggregations in larger layers of phase-separated hydrogels. The long-chain phospholipids of modern cells are certainly of secondary origin and evolution as witnessed by two different kinds in different cell-types [74,75]. More primordial beginnings of potentially membrane forming amphiphiles are thought to have multiple roots such as based on the catenation of (1) terpenoids for steroids in eukaryotes, biohopanoids in bacteria and archaeal phospholipids [76,77], (2) acetic acid for long-chain fatty acids in bacterial and eukaryotic phospholipids [51,52,78], and (3) amino acids with lipophilic side chains for membrane-associated amphiphilic peptides and proteins [79,80].

In the context of self-amplifying surface-coating hydrogels, it is worth noting that single-stranded RNA and cationic peptides or proteins can readily coalesce into phase-separated liquid droplets of membrane-less '*coacervates*' [81,82]. The rapidly expanding experimental field of *biological phase separation* adds an important dimension of structural significance to the OoL model of a reciprocal *peptide/RNA partnership* [83] mentioned further below. Moreover, the membrane-less state implied a natural tendency to coalesce by stochastic fission fusion cycles as a primordial trait of *surface protoplasm* [75], and the evolvable traits associated with *protoplasmic miscibility* may in turn have gathered a momentum of their own, reaching far beyond the limitations of received conceptions about relatively simple '*protocells*'.

5.2. Kinetic Coupling

A revival of older paradigm-changing concepts can also extend to photochemical reactions in coupling to sunlight as the most dependable source of environmental energy on Earth, so as to drive a protometabolic core of geochemical reactions [84-86]. A coherent self-organizing and self-perpetuating system with progressively evolving life-like properties may have emerged on this basis. Conceivably, a marginally coherent network of geochemical reactions and a corresponding set of environmental catalytic agents already fulfilled the principle of autocatalytic closure in a prebiotic environment allowing autotrophic modes of photon-driven *surface metabolism*. In particular, the reductive tricarboxylic acid cycle (TCA, or reverse Krebs cycle) appeared appropriate to constitute the central amplifier loop that nourished an ever-expanding protometabolic network and material growth of protobiotic molecular conglomerates [84,87]. Seen in this context the natural transition — from mineral-catalyzed surface metabolism and photoelectrochemistry to mainly intra-cytoplasmatic catalysis — could be paraphrased as a selective domestication process whereby suitable inorganic compounds were sequestered by organic chelating agents instead of permanent crystallization in solid environmental minerals. The effective internalization of such cofactor-like compounds into the newly emergent biomatter, in turn, would have allowed the evolving surface protoplasm to colonize other light-exposed substrates irrespective of any surface-exposed catalytic activity.

To be sure, a seemingly serious objection against a UV-based contextualization of OoL to terrestrial settings holds that "*UV light kills cells and no cells can harness energy or live from UV light*" [88]. However, modern cells are far removed by evolution from any conceivable OoL conditions, and UV-light is mainly killing cells by irreparable damage to their DNA genomes. This is in stark contrast to the higher resilience of protein-based protoplasmic functionality as such, the evolutionary continuity of which may have reached farther back than the origins of genetic coding first, and of vertically stable genomes later on [79,89].

A central issue for the superior shift of paradigm advocated here is moving the geophysical environment for OoL from non-cyclic conditions deep beneath the ocean to solar cycling at surface-exposed terrestrial settings, associated with repetitive reinforcement of crucial biochemical or biophysical effects. The overall framework of non-equilibrium thermodynamics, in fact, suggests that the pumping effect of temporal oscillations in the environment is very important if not essential for life's emergence [19,20]. The *potentially biogenic oscillations* come at a range of different frequencies, spanning from vibrant sunlight and photochemical molecular reactions to daily or seasonal dry/wet cycling and the structural consequences this has for soft organic matter in terms of shrinking/swelling or congealing/dispersing of hydrogel-like organic matter. Of note, the vital pumping effects associated with environmental cycling are based on partial irreversibility in characteristic hysteresis loops. In energetic terms the incoming beam of energy — by inelastic absorption — is split into a working part (the functional effect) and a lower-energy residue dissipated as heat.

Characteristically, the life-maintaining feature of such hysteresis loops — the 'upstroke' of cyclic pumping effects — couples the working aspect of inelastic energy absorption to accretionary growth of biomatter at various levels. The accretion mode in turn is cumulative for as long as reverse reactions during cyclic 'downstroke' periods — in terms of molecular degradation or structural disintegration, such as hydrogel dissolution upon excessive rehydration — occur to lesser extents compared to upstroke-dependent accretionary growth. This means in a terrestrial, atmosphere-exposed setting that inorganic volatiles were effectively immobilized by nonvolatile clustering in affinity-bound complexes of intermediate-size organic molecules first and incorporation into various macromolecules in the long run. Furthermore, such settings can also constructively revive Oparin's classical suggestion of liquid-liquid phase separation in 'coacervates' as a very natural means of localized accretion in bulky aggregations of soft organic matter — considerably larger than conventional 'protocell' concepts, and not necessarily surrounded by topologically closed membrane-like boundaries [50,67,90,91] — especially favored by dehydration and regular wet/dry cycling [73].

5.3. Midway Limbo

A transitional '*Limbo state*' (not truly living yet not fully abiotic either) presumably connected the prebiotic phase of potentially biogenic geochemical reactions at mineral surfaces with self-contained, self-organized metabolic networks in real life — supporting the emergence of *adaptive evolution*. The conceptional link assumed herein represents one of the most significant "*major evolutionary transitions*" [92], but it differs substantially from the hitherto received consensus by considering the kinetic and dynamic implications of the novel concepts of '*surface protoplasm*' and '*neighborhood selection*' — "*a kind of group selection that can, in principle, result in improvements in collective multiplication and enhanced complexity, even though neighborhoods are not discretely bounded and do not, in any simple sense, self-replicate*" [67]. These game-changing conceptions are based on more general modeling studies on the reinforcement of cooperative and communicative interactions among neighboring partners in two-dimensional arrays of localized agents by intergroup selection [93], even though there are no discrete, nonoverlapping groups. Approaching *chemical ecosystem selection* experimentally [91] and in silico modeling of quasi-stochastic neighboring effects in two-dimensional arrays [94] can potentially corroborate these notions.

In biology at least, the *evolutionary principle* as such is closely linked with Darwin's notions of '*natural selection*' and '*descent with variation*' in serial generations of variable and prolific *reproduction* could naturally account for selective/adaptive changes, which then led to cumulative effects and divergent *speciation* over deep geological time [95]. In the context of OoL scenarios, however, life-like *evolution by natural selection* had much deeper roots than Darwin's personal interest in the *Origin of Species*, whereas the speculative origins of life as such lay far beyond scientific inquiry at his time. It now becomes relevant to analyze to what extent the likely multi-factorial causes contributing to *biological descent, variation and selection* could be generalized and still resulted in evolutionary changes over geological time [96,97] — not the least concerning short-term material *stability* and long-term *persistence* in life-like evolution [98]. Arguably, the more basic processes of *accretionary growth* and *structural survival* had to reach a certain level of complexity before other functions could be added to enable self-similar *reproduction* of increasingly life-like material entities.

To my mind it appears that the advanced genetic mechanisms of *biological descent* are primarily serving protective stabilizing functions, which were to some extent supported by more basic means of sheer survival and cumulative growth already at the '*Limbo state*' when life emerged. To serve as a more elementary *stability/persistence function*, the principle of "*differential molecular survival*" — in a mutually reinforcing fashion — was indeed suggested long ago for selective and adaptive peptide evolution from quasi-stochastic origins [99]. Presumably, affinity binding to medium-sized molecules of multifunctional potential, such as metal-sulfide complexes, phosphorylated organic compounds, metabolic cofactors, prebiotic peptides and the like [100,101], had a particularly high selective value by *differentially* increasing their *molecular survival* in affinity-assembled clusters early on. Intrinsically disordered protein regions in particular are prone to hydrolytic cleavage unless protected by the shielding effect of affinity-bound interaction partners [102]. The *differential survival* of cooperatively interacting peptides has in fact been modeled [103].

Importantly, de Duve's differential survival principle had the potential for autocatalytic reinforcement as follows. As I see it, the central engine of accretionary growth kept spinning already at that early stage, mainly by the use of certain peptides with marginal catalytic effect on metabolically relevant reactions. This potential was highest for the lucky few that happened to be stabilized by structural molding into affinity-selected complexes, whereas the majority of random-sequence peptides were intrinsically unstructured and devoid of metabolic activity. Viewed from an evolutionary cost–benefit perspective, such peptides would appear as useless by-products, the recycling of which should add selective advantage to peptide complexes with increasing proteolytic activity against unstructured sequences quite early on.

The multifunctional potential of useful non-peptide compounds could include catalytic cofactor activity, the storage of chemical potential energy in metastable bonds, a buffering capacity against environmental fluctuations, and not the least some structural utility as versatile *affinity handles* to temporarily immobilize small freely diffusible metabolites. There is a particularly instructive case for the important *handle* function in the Beta-Alpha-Beta motif believed to be ancestral to the ancient, versatile and most abundant group of related P-loop and Rossmann superfamilies. Such proteins commonly bind phosphorylated ribonucleoside ligands as substrates, co-substrates or cofactors [104]. More often than not the phosphate and/or ribose moieties do not partake in a catalytic reaction mechanism as such but serve an accessory carrier function as sturdy, reusable *affinity handles* to position and coordinate some smaller reactive group close to the active site of a P-loop or Rossmann-type enzyme [105].

5.4. Internal Compartmentation — The Proto-Coenocyte Scenario

The *Limbo* Section about geochemistry turning into cellular life should not be left without referring to spatial separation. How could diffusive dispersion of collectively cooperative constituents be progressively constrained in the initial absence of structural boundaries as here assumed? — Somewhat belittling and implicitly including *liquid-liquid phase separation*, a category of '*passive localization*' (at solid surfaces) and a '*trait-group type lifecycle*' in general have been dubbed "*a poor man's form of compartmentation*" [92], so subtle parallels are not impossible to realize. To better understand the evolutionary kinetics of '*neighborhood selection*' as a leading concept (see above), it is not only relevant to ask which prebiotic constituents most likely formed a physically coherent '*neighborhood group*' of cooperative molecules but also to identify a functional set of malleable traits that potentially could respond to uniform selective pressures for further adaptive evolution of structural parameters.

In the context of Baum's original proposal of initially unbounded "*surface protoplasm*" the newly emerging outer membranes were thought to serve the "*cells-as-propagules*" conception [67]. However, with the added assumption of a photon-energized metabolic core (see above) I prefer to consider a rather different driving mechanism for gradual membrane emergence and the potential for multiple evolutionary transitions from surface-bound protoplasm to genuine cells. As suggested earlier [106] this alternative approach gives primary significance to lipid membranes as topological insulators prolonging photon-induced charge separation and subsequent channeling through structurally coordinated electron transfer chains.

The insulator function alone may have given temporary amphiphilic clusters with micellar, *lipid raft*-like properties a selective advantage as internal nucleation sites for laterally extended vesicular or reticulate membrane structures. The early unconventional notion that the first mitochondria-like organelles may have resulted from the coalescence of intracellular, energy-converting, "*plasmid-associated thylakoids*" [107] was pointing in the same direction. The advanced gated membranes of topologically closed outer boundaries could then arise from vesicle fission–fusion cycles and internal membrane trafficking — in lock-step with fusion events at outer surfaces — and the coevolutionary optimization of long-chain phospholipids and membrane-spanning proteins. It is furthermore intrinsic to this model that both nuclear envelopes and proto-organellar boundaries were formed endogenously as gated double-membrane structures. Such internal membranes could also have served as anchoring platforms for membrane-associated catalysts and submembrane networks of fibrillar cytoskeleton components. Altogether, the boundless extent of primordially border-less *surface protoplasm* would have provided an early *coevolutionary potential* very similar to what has been inferred with the traditional assumption that the larger volumes of eukaryotic cell type only developed at a considerably later period [108,109].

Seen from this viewpoint, the early onset of directional membrane trafficking and internal compartmentation in larger bodies of phase-separated protoplasm can shed light on several open issues:

1. On this provision, plasmids and viruses can be incorporated in the more general concept of a lineage-based yet not necessarily organismal Tree of Life (ToL) [110,111].
2. Allowing plasmid-like self-replicating genetic elements to localize in internal vesicular compartments may have paved the way to endogenous proto-organelles, such as the circumstantially inferred 'premitochondria' [112], some of which may have evolved further and eventually 'escaped' from the surrounding protoplasmic bulk as lineages of bacteria-like quasi-autonomous cells. These conjectures have led to the '*Karyogenic Proto-Coenocyte Hypothesis*' [106].
3. Internal membrane trafficking systems within the common protoplasm should also have favored eukaryote-like karyogenesis from within — as opposed to tentative endosymbiotic origins from outside [113].
4. Somewhere between these two extremes, a single chromosome from the collectively coordinated, communally shared (non-plasmid) gene pool may eventually have gathered a subset of essential genes just large enough and sufficiently diverse to allow additional lineages of archaea-like cells to 'escape' from the residual protoplasmic bulk as well. [114].
5. It has long been recognized that eukaryotic cell organization still has many characteristics that may have been generally advantageous at much earlier stages of pre-organismal evolution [115,116] and therefore may represent a more direct lineage of vertical descent from the Woesean 'Progenote State' of population-wide collective, RNA-directed gene pool sharing. This notion is central to Woese's 3-domain (3D) canonical Tree of Life (ToL) [117], which is now seriously challenged by the 'Eocyte' version of a competing 2-domain (2D) model, nesting all the eukaryotes within the ancestral archaeal domain [25,26], but disagreement among experts is not ending there.
6. Also eukaryote-like cells — or rather nuclei — may finally have emerged as modular genomic units of vertically stable inheritance out of the residual Progenote-like population. However, establishing a reliable system for the coordinate co-segregation of multiple chromosomes in the course of nuclear division may have taken considerably longer time than gathering a minimum number of essential genes on a single plasmid molecule (as suggested here for the generation of Akaryote-like cells).

A classical phylogenetic notion holds that spatial separation can initiate lineage-wise diversification and branching speciation in the canonical ToL. This assumption should also be valid for the hypothetical scenario of "*spatially separated pre-cell populations*" [74] — a tentative stage

doi:10.20944/preprints202405.0555.v1

equivalent to the better-known concept of collectively evolving pre-organismal populations at a communal *Progenote State* (sensu Woese). On this basis it is conceivable that the apparent anomaly of finding a scattered array of otherwise eukaryote-specific proteins in *Asgard archaea* [26,118] can also have resulted from preferential gene pool sharing in a persistent symbiotic relationship between a small proto-archaeal partner ancestral to present *Asgard* lineages and a larger one ancestral to the *nuclear-cytoplasmic lineage component* of present Eukaryotes [28,34]. Coexisting in a particular spatial pocket of geographically dispersed populations, in the aftermath of a complexifying *Progenote State*, the two partnering proto-organisms would carry different samples of proto-chromosomes drawn from the originally collective '*lottery drum*' arising from '*pre-cell*' aka '*progenote*' communities in Kandler's and/or Woese's views. Some of the open issues numbered above are taken up in more detail further below.

As there is substantial ambivalence in the literature regarding the appropriate range of Woese's prescient *Progenote* conception, a note on terminology is added here which will gain further significance throughout this article. The original term was rather vaguely introduced to signify a primitive stage in cellular evolution when "*the link between genotype and phenotype*" had not yet been firmly established [119]. As I see it, it took two different steps evolving this important link at two different levels: (1) the making of *genes and gene products* to define the temporary *genotype* of any given cell, together with an expression system feeding into the cellular phenotype in structural/functional terms; and (2) the making of *lineage-defining genomes* of any organismal species with its particular collective phenotype. Importantly, consolidating the large genomes of cellular organisms took considerably longer time than establishing the genetic coding/decoding system as such. Accordingly, I find it natural to use the *Progenote* term for both successive stages of collective evolution — *sensu stricto* for generating RNA-directed genes and gene expression, and *sensu lato* for consolidating the DNA-directed genomes that eventually allowed different organismal lineages to leave the overall collective *Progenote State*. Evidently there are wider limits of the collective *Progenote* concept, so I explicitly include eventual pre-genomic stages of *residual trunk-line evolution* (see further below) in the conceptions of a '*Generalized Progenote Hypothesis*' [114].

In conclusion of this Section I should like to propose a smooth transition from Baum's '*Surface Protoplasm Concept*' to a '*Proto-Coenocyte Scenario*' (see '*Modular Cellularization*' further below), which allows for more integrative suggestions — also regarding the Woesean conception of collective *innovation sharing* before the advent of lineage-based genomics — than the traditional assumption of rather simple, vesicle-like '*protocells*' at the beginning. The important difference is that the *proto-coenocyte scenario* is based on close interactions between two complementary subsystems: (i) a profound, virtually unlimited variety of *unlinked protogenes* for quasi-statistic peptide sequences in communal support of cytoplasmic functionality, and (ii) newly emerging *self-directed genome lineages* with plasmid-like or viral properties. The corresponding gene sets should have followed different evolutionary dynamics, but gene transfer could readily occur between the two sets in either direction — facilitated by their local proximity in a common protoplasmic medium. To have a name for tentative entities supporting such macromolecular cooperativity, I suggest using '*ribo-coenocytes*' — instead of tentative '*ribocytes*' [120], which conceptually were too firmly connected to the now outmoded RNA-World model in its strictest sense.

5.5. Coemergence

The controversial question of "*Genetics First*" or "*Metabolism First*" [45], which too long has been overshadowed by the tentative *RNA World* model of primordial RNA molecules having both replicative and catalytic functions on their own [121], can now be resolved by a more inclusive, network-expanding alternative as follows: The co-emergence of and further coevolution in functional partnerships between different types of macromolecular agents is receiving more attention, not the least in terms of the closely knit peptide/RNA partnership of a *Peptide/RNA World* scenario [83], which conceivably began by structurally complementary affinity already at the oligomeric level [122]. Circumstantial evidence suggests that the acceptor stem region of tRNA-like oligomers played a critical role for peptide growth as such and initial codon–anticodon interactions for the emergence of the genetic coding/decoding system as well [123-127].

The co-emergent RNA–peptide alliance is part of an *RNP World* scenario as a primordial model without requiring a preceding *RNA World* stage to start with [128,129]. The bi-partite coevolution then consolidated the functional utility of some prebiotic peptides — in ribonucleoprotein (RNP) complexes with RNA — as "*molecular hubs in the origin of life*" [79]. An informational archive retrieval system materialized thereby for future use to remember what had worked best in the operational past. The coevolutionary partnership of catalytic peptides and peptide-assisted stabilization, extension, and replication of RNA can readily concur with the kinetic selection advantage of metabolic over non-metabolic replicators [130], even though the *metabolic replicator* concept itself has mainly been proposed to save the long-promoted *RNA world* scenario in its purest sense [131]. It is worth noting though that many operative conjectures made about the *RNA World* model as such remain valid regarding the putative absence of DNA as genomic material at the beginning, but they cannot strictly reject the network theory assumption that primordial RNA precursors and prebiotic peptides began their coevolutionary '*pas de deux*' quite early on.

5.6. Accretionary Growth towards Hierarchical Modularity

Structural compartmentation is vital for life as we know it, but as mentioned above, confining membranes may not have been required for life-like network evolution to emerge. Several theoretical considerations can substantiate this point of view. "*The evolution of structure in biology is driven by accretion and diversification. Accretion brings together disparate parts to form bigger wholes* [and subsequent] *diversification provides opportunities for growth and innovation*" [132] (Caetano-Anollés et al. 2018). Seen from this angle, *accretionary growth* in bulk of marginally protoplasm-like components comprised the most primitive and basic characteristics of the evolvable and persistent physical state of life-like soft organic matter, whereas the more stabilizing features of membrane-based compartmentation, heredity and gene expression were secondary matters of gradual optimization and step-wise *innovation* later on.

The empirical studies for this structural approach to evolutionary inferences concerned the structural folding patterns of RNA and protein molecules from an evolutionary perspective. Of particular interest in the present context, the authors could roughly estimate the evolutionary age of conserved protein folds and other structural features on a timescale reaching back to before the consolidation of the genetic coding system as such [89]. This still unconventional approach is highly complementary to mainstream comparison of sequence alignments since characteristic folding topologies — as stabilized by polar and non-polar interactions with watery surroundings — are conserved more strongly over evolutionary time than the corresponding '*fuzzy sets*' of many polymeric sequences, all supporting the same folding topology [133]. The most distinctive folding patterns are topologically separable *protein domains*, not to be confused with the *phylodomains* of cellular organisms.

The hierarchical modularity of characteristic protein folding patterns forms the empirical basis for comprehensive and standardized structural data sets. The self-contained stability of such well-defined stereo-topological configurations is based on the balancing of forces between steady-state affinity fitting of internal hydrogen bonding and water exclusion from hydrophobic clustering among the various amino acid side chains concatenated in a linear string. The ordering principle for *tensional* domain *integrity* is seen in unifying *tensegrity* conceptions [134,135]. The phylogenetic conservation of such structural modules in protein families and superfamilies is highly suggestive of common ancestry — sharing the topological characteristics of a particular protein domain from a common root, with only minor to modest variation of a basic geometric shape. Moreover, the structural framework of any given protein domain can be supported by a wide variety of amino acid sequences, due to the structural redundancy of the genetic code — relative to side-chain ordering by local hydrophobicity effects. Therefore, the structural homology of protein domains can be ascertained over considerably longer time spans than what is possible with more rapidly diverging amino acid sequences alone [136,137]. Unfortunately, however, this great advantage of *phyloproteomic*

doi:10.20944/preprints202405.0555.v1

analyses is not yet duly appreciated by the prevailing community of conventional *phylogenetics* based on patch-wise sequence alignment over significantly conserved regional segments.

The underlying *phyloproteomic* information can be 'data-mined' from structural data bases for phylogenetic analyses at the deepest level as a valuable alternative to the conventional sequence-based approach. In a proof-of-principle comparison, the mere presence or absence of protein folds (at the superfamily level) in fully sequenced, annotated genomes gave more similar results to conventional phylogenies than those gathered from domain abundance per genome [138,139]. This is a system-covering, more global alternative to the selective vagaries of functionally constrained sequence matching in local alignments of conventional molecular phylogenies. Although not fully acknowledged yet on equal ranks by the established sequence alignment community, the upcoming field of *Phyloproteomics* is potentially more useful when it comes to comprehending the not necessarily unique transition from collective networking, as inferred for Woese's ancestral *Progenote State*, to *Darwinian speciation* of free-living organisms in the modern biosphere — arguably by way of intermediate stages allowing emergent vertically stable lineages to undergo tree-like bifurcation already at sub-organismal levels.

There are still conceptional problems with the *Phyloproteomics* approach in general in that the theoretical foundations are not yet understood nearly as well as for conventional *Phylogenomics*, which has led to a critical and potentially constructive exchange of opinion regarding the availability and use of "*realistic evolutionary models*" [140-143]. These concerns (implying a latent '*Phyloproteomics paradox*') were concentrated on the rooting problem of the organismal ToL to begin with; they were subsequently extended to include the *Virosphere* as well (see further below, where the organismal aspects of the formal dispute will be resumed and set into a new perspective).

5.7. Integrative Networking

For *networking theory* in general, the innovative '*bigger wholes*' (resulting from stochastic accretion of '*disparate parts*' and being integrated for synergistic benefits) are referred to as *modules* in the sense that interactive intra-module connections between their '*parts*' are clustered in minor *subnetworks* more densely than pairwise interactions connecting the *module* to other system constituents outside the *module* itself. Subsequently certain *modules* with higher potential to cooperate within a group of modules associate as superior *subnetworks* and "*become new parts for a new generative cycle of higher-level organization*" [144]. Generally speaking, the network-based accretionary scaling properties in the origin and evolution of composite modular structures are very similar over wide ranges in space and time — spanning all the way from macromolecules to urban cities, sun-like stars with all their planets, and galaxies with all their stars [145].

At the lower end of prebiotic evolution, considering the molecular characteristics of polypeptides, phospholipids and polymeric monosaccharide derivatives (especially with RNA-like properties), several main parameters appear influential when it comes to the emergence of biological information. The key features comprise *viable cores* (aka *nucleation centers*) and *connectivity kinetics*, resulting in some form of *spatial separation*, together with *resource availability*, *scalability* and system-wide *information control* [146]. The cohesive soft matter properties of RNA, peptides, lipid chains and other biopolymers can be understood in terms of paracrystalline order or liquid crystal-like aggregation in watery conditions [147,148], and this type of molecular interactions may have been particularly important already at the initial level of forming effective nucleation centers for superior modules of biological functionality, such as hydrophobic cores of globular protein domains, composite protein complexes, extended protein fibers, and lipid membranes filled with membrane-spanning proteins.

These considerations link accretionary growth to prebiotic network evolution and the emergence of *collectively autocatalytic sets* — with an open potential for ever-increasing modular complexity, as exemplified by the collective optimization of the universal genetic coding/decoding system from a *primordial peptide–RNA partnership* [83]. More likely than not the *peptide–RNA alliance* of a primordial *RNP World* scenario evolved via two distinctly different substages — *uncoded peptide synthesis* first and *RNA-encoded ribosomal protein synthesis* secondarily. These modular subnetworks of two different molecular part sets coevolved adaptively on this time course, transforming oligopeptide populations into full-length proteins and oligomeric ribonucleotides into functional RNAs. By taking these steps, early life "*began in analog mode*", which has remained the structural basis for protoplasmic functionality ever since. Yet, the maturing *peptide–RNA alliance* then managed to buttress this analog functionality by paving "*The Path to the Digital Cell*" [6] — thereby facilitating the coevolutionary optimization of both "*template-directed replication of RNA molecules*" and "*a digital genetic encoding system*" (together with corresponding decoding facilities) to complement the analog functionalities of homeostasis and growth with regular system maintenance first, and faithful reproduction in the long run. The putative succession of early, RNA-dominated steps was eventually transcended by another significant transition, resulting in the consolidation of DNA-dominated organismal genomes:

1. The heroes of the initial phase are seen in the first emergent RNP tandem complexes of tRNA-like hairpinned oligonucleotides and '*urzyme*' peptides ancestral to tRNA-charging enzymes — aka *codases* or *aminoacyl-tRNA synthetases* /*AARSs* [149-151]. Whilst the sensible, functionally indicative *codase* term [152] preceded the characterization of *aa-tRNA synthetase* mechanisms, the latter expression is in wider use.

To rationalize this important step, the so-called *Rodin-Ohno hypothesis* that *two enzyme superfamilies descended from one ancestral codase gene* — by means of '*Sense/Antisense Coding*' in opposite directions — is gaining evolutionary significance from remarkable experimental progress [153,154]. This notion can neatly explain the structural divide between two classes of nonhomologous yet functionally equivalent *codase proteins*, which bind to the aa-acceptor stem regions of tRNAs from within the major or minor grooves of double-stranded RNA at opposite sides. Nowadays, the different tRNAs to cover the canonical codon table are served by members of one or the other class of *codase protein*. However, the nonoverlapping topologies of catalytic domains in certain mixed-class pairings is suggesting that this either/or distinction did not apply from the beginning, when the fledgling code of acceptor stem precursors was less diverse than the "*Operational RNA Code*" intrinsic to present tRNAs, when emergent ribosomes had not yet gained their full potential and many protein products consisted of quasi-stochastic, 'fuzzy' sequences. Instead, the single-hairpin tRNA precursors may have been sandwiched amidst a protective complex with a mixed-class protein dimer [155].

Conceptually, the peptide-RNA-peptide sandwich phenomenon may have emerged in a transition phase between small peptides being assembled according to "*structural complementarity between RNA and polypeptides*", as postulated by the Carter/Kraut model [122] and an '*urzyme*' stage of marginally effective *codase proteins* [151]. Detailed structural analyses suggest an evolutionary connection between these intermediate endpoints [156,157]. Such peptide-sandwiched RNA hairpins — on their way to forming compact little RNP machines — may not only have coordinated peptide-related processes (such as amino acid activation by phosphorylation and transphosphorylation, or the formation and interconversion of peptide bonding) but also served as primers for RNA extension and template replication [158,159].

Interestingly, the key events of amino acid activation — by aa-phosphorylation first, and subsequent transfer to the 3'-terminus of a tRNA — can be facilitated as ribozyme-catalyzed reactions by selectively enriched RNA oligonucleotides [160,161]. Although the presentation of these experiments was framed in the concept of a hypothetical *RNA world* before proteins, the results are perhaps even more compatible with the 'fuzzier' concept advocated here that quasi-random sets of short peptides and oligoribonucleotides emerged together and started to coevolve into a gradually perfecting *peptide–RNA partnership*.

In fact, Woese's long-held assumption that protein evolution begun with functional selection amongst quasi-random peptide populations as "*statistical proteins*" [162,163] (Woese, 1971, 2004) has been substantiated by two conclusions from the complementary *Phyloproteomics* approach: (i) "*ancient protein domains did not harbor translation functions* [but] *helped fulfill metabolic roles*" [164], and (ii) "*ancient forms of aminoacyl-tRNA synthetase (aaRS) catalytic domains and ancient non-ribosomal protein synthetase (NRPS) modules gave rise to primordial protein synthesis*" [165]. These conclusions imply that the emergent 'urzyme'–tRNA alliance mentioned above and various non-coded peptide ligases — by *analog means* — helped build marginally functional metabolic networks of primitive proteinaceous catalysts well before the emergence of ribosomes and mRNA translation for *digitally* gene-encoded protein synthesis.

2. To complete the dual molecular *alliance of RNA with proteins* in a second round of innovation, the central hub is still found in and around the *ribosome*: a processive, programmable, and tool-moving *RNP nanomachine*, producing virtually all the proteins a cellular organism can make. Together with its closest interaction partners, — such as charged, amino-acylated tRNAs, aa-specific tRNA charging enzymes, elongation factor proteins, and various accessory components for protein insertion into membranes or trans-membrane protein secretion — the ribosome interaction network amounts to about 100 different parts, many of which are universally conserved. This represents one of the largest macromolecular modules at the subcellular level, and Woese's decision to derive the rooting of the canonical ToL from just a single rRNA gene product — strictly required to keep the ribosomal module united as a compact structural machine — was not an arbitrary choice.

3. Conceptionally however, the processive importance of the ribosome should not unduly overshadow the more fundamental functional significance of tRNAs in catalyzing the iterative molecular reactions occurring at the *peptidyl transferase center* (PTC) of the ribosome. A partial bias and 'ribosome-domineering' attitude is culminating in the over-generalized assertion that "*Protopeptides, then polypeptides and proteins, were created by the ribosome, on the ribosome, and for the ribosome*" [166], but this alone is not the way evolutionary change is coming about. To be sure, the ribosome is vitally important to the current performance and early evolution of long-chain protein synthesis, yet it is too complex to have attained all its functionality for its own good or totally from scratch. First of all, the system-stabilizing and system-perpetuating roles of the surrounding protoplasm should also be considered, including the essential partner molecules without which the ribosome alone would be of little use.

   Alternatively, in the light of *hierarchical modularity of networks* [132], the transition from 'tinkering' with peptide bonding in 'fuzzy' sequences in a tRNA-dominated, pre-ribosomal era to fully processive ribosomal protein synthesis was marked by ribosome emergence as a case of integrative module accretion at a higher level. Whilst the *peptidyl transferase* reaction as such was primarily left to the participating pair of tRNAs alone — using their vicinal OH groups as a case of *substrate-assisted catalysis* [167], ribosomal RNP functionality contributed to the more overarching integration of structural and processive innovations: (i) a toroidal nozzle to colocalize two charged tRNAs around the reactive phosphate bonds as a nucleation center for the PTC (*peptidyl transferase center*) of the large-rRNP L-subunit; (ii) excluding excessive water molecules from the nozzle cavity to protect the nascent peptide bond from hydrolysis in the reverse reaction; (iii) affinity binding of a small-rRNP S-subunit around the distal tRNA loops and a ratchet function with flexible, non-covalent inter-subunit connections; (iv) extending the nozzle at the PTC into a funneled exit channel across the spacious L-subunit; (v) supporting affinity binding to the additional RNP *Signal Recognition Particle* (SRP) [168] to inject hydrophobic peptide products into membranes or lipid rafts. Further studies from the *Phyloproteomics* approach have led to the conclusion that the entire *ribosomal protein synthesis* module has only emerged relatively late on the evolutionary trajectory toward cellular complexity [169].

   Although the factual information given on the structural achievements of the Yonath group are firmly established [170], the explanatory narrative favored herein is at odds with the conventional mantra of declaring proto-ribosomes as primordial *ribozymes* for peptide bond formation [171]. If any type of RNA would deserve the distinction of being a '*ribozyme*' for catalyzing peptide bonding as a reaction mechanism, this should be tRNAs with their 3'-terminal, vicinal OH groups, whereas the superior significance of ribosomes concerns the overall robustness and processivity of long-chain protein synthesis at an upper level of functional networking modularity.

4. As I see it, "*The Path to the Digital Cell*" [6] would not be complete with the '*making of genes*' as such but should also extend to the '*making of lineage-defining genomes*' at the organismal level. A major innovation to make this happen during a period of genomic RNA-to-DNA transitions was the emergence of replisomes — several times over — as composite molecular machines for highly processive DNA synthesis on very long template molecules. Considering modular connectivity in general, replisomes are largely comparable to ribosomes. Just as ribosomes have made protein synthesis processive at the peptide chain level, not leaving their template strand before the job is

done, the various replisomes have become masters of processive DNA synthesis by also taking many repetitive reaction steps in a row.

To further comprehend the organized complexity of life on Earth in terms of *hierarchical modularity*, the ample in-between from macromolecular cooperativity to the infrastructural organization of urban cities [145] is to be filled with additional interactive modules, such as *unitary genomes* and *modular control* over the *cell division cycle*. It is of particular note that the accretion kinetics to initiate sufficient connectivity within the first *viable cores* or *nucleation centers* were not strictly deterministic but intrinsically probabilistic in both space and time. And seen in this perspective, the emergence and organizational diversity of genome-coordinated cells may well have had more in common with the overall formation of the solar system with planets and their satellites than one might think.

As tentative points of comparison, I suggest drawing an analogy between the probabilistic accretion kinetics at gravitational centers in star-forming molecular clouds in outer space and the tentative patches of primordial *surface protoplasm* [67]. In either case several minor accretionary cores may have emerged within the common catchment area of a larger attractor field. Keeping this analogy in mind, I presume that lineage-stabilizing proto-genomic units, such as viruses and plasmids, may have formed within a common structural and functional assemblage of pre-genomic protoplasm — just as galaxies have acted as orbit-stabilizing centers for their stars, the stars for planets, and planets for satellites. This far-flung analogy may provide a roundabout solution to the unsettled question of how to rationalize the organizational *eukaryote–akaryote dichotomy* as a natural outcome from an earlier Woesean state of *Progenote innovation sharing* from a universally communal tool bank [172]. So, adding viruses and plasmids to the collective *Progenote toolbox*, or proto-organelles resulting from internal compartmentation for that matter, may shed new light on possible answers to the wavering question "*Eukaryotes first: how could that be?*" — revitalizing former, now unconventional views "*that Bacteria and Archaea are convergently prokaryotic*" [173,174].

Comparing the origins of biotic evolution to the formation of the solar system in terms of hierarchical modularity points at more than just a fortuitous coincidence. Our local triad of cosmological bodies consists of the central Sun surrounded by the binary Earth–Moon planetary system of the rocky type. In simple terms of celestial mechanics the formation of both Earth and Moon is much easier to understand than the star-like formation of the central Sun, which as such became dominated by nuclear reactions at a massive scale. Astronomers are now convinced that all three bodies began their existence from three separate nucleation centers within a spinning and gradually contracting pre-solar accretion disc. In this article I have tried to collect sufficient circumstantial evidence that bacteria, archaea and eukaryotes originated from different genomic nucleation centers in the metaphorical concept of a pre-organismal 'accretion and complexification disc'.

5.8. Innovative Transitions

Some infrastructural *modules*, especially those relating to the integration of *genetic information*, are long known as *evolutionary units* [175], but the original proposal was overly reliant on the questionable *RNA world* scenario in that primordial, quasi-autonomous '*RNA replicators*' were thought to represent the first *evolutionary units* ever formed in early life. In a more recent update, the author then added a remarkable caveat: "*The emergence of novel evolutionary units belonging to a previously non-existing higher level is, however, a difficult problem, since if this emergence is to proceed through evolution requiring the presence of evolutionary units, the lack of these units seems to preclude pathways to their emergence. The clue lies, obviously, in the fact that there is a selective hierarchy involved in the process. The higher-level units must appear as the result of coevolution of units at the lower level*" [92]. This caveat in fact applies to any *major evolutionary transition* that rests upon the *coevolutionary* emergence of complex *innovations* composed of many interacting parts, as prominently represented by the emergence and subsequent optimization of the genetic coding and translation systems, which presumably occurred under *RNP World* conditions [129].

Moreover, seen from my personal perspective the latter general insight points at an intrinsic correspondence between *coevolution* and *symbiosis*: The *coevolving submodules* of a superior system have to coexist in close proximity — quasi '*symbiotically*', so to speak — which appears highly relevant for the co-emergence of different proto-genomic lineages before the eventual rise of genuine organisms. In other words, the collective networking required for Woese's ancestral *Progenote State* to persist corresponds to a kind of *symbiotic* relationships between structural submodules of a superior system.

This concerns the next most puzzling transitional stage, fictitiously termed LUCA in a pure *black-box* approach to represent the '*Latest Universal Common* (or *Cellular*) *Ancestor*' — aka LUCAS, with emphasis on its pre- or supra-cellular complexity as a composite '*Ancestral State*' [176]. All genuine organisms of the DNA-directed modern biosphere have subsequently descended from the legendary LUCA(S) in the ever-branching *Tree of Life* (ToL), diversifying by way of Darwin's principles of *evolution by natural selection and eventual speciation*. It took one transition from randomized sequences in smaller peptides to longer, gene-encoded proteins to establish a collective '*metagenome*' composed of numerous independent *genes*, which were still subject to stochastic redistribution throughout a communal population of highly variable pre-organismal biotic entities. It took another major transition to establish fully individualizable *organismal genomes* by keeping sufficiently many genes together in *vertically stable lineages* of different organisms. To emphasize the *symbiotic networking* required for the ancestral *Progenote State* to persist I have herein given the play on LUCA-derived acronyms yet another twist to 'LUCAN' — the Latest Universally Common Ancestral Network.

As the conjectural pre-LUCAS entities "*had not yet completed evolving the link between genotype and phenotype*" [119], their organizational status would not readily compare to modern *Akaryote* cells, nor to *Eukaryote* organization either. To signify the categorical difference, the primitive ancestral entities were called '*Progenotes*' to begin with. In summary, the major innovation to keep the variable progenote entities persistently alive was the *making of genes* — a "*progenote, stricto sensu*" [177] — but the *making of organismal genomes* was further required to leave behind the ancestral state of collective evolution, a '*Progenote State, sensu lato*' to my opinion. Correspondingly, Woese's somewhat nebulous notion of "*evolving the link between genotype and phenotype*" to completion needs qualifying by extending the phenotype concept as such from gene-dependent functionality within individual physical entities to temporal invariance of a particular pattern of many phenotypic traits at the '*pangenome*' level — encompassing the entire population of a given species but distinct from other species in the biosphere. While the *making of genes* initially gave a boost to momentary functionality in individual structural entities, the *making of well-coordinated genomes* was also required to warrant accurate performance of vertical descent and sustained lineage stability of *conspecific populations* for numerous generations in a row.

In hindsight now it appears that a transition from RNA to DNA was required to collect sufficiently many genes in discrete genomes for cellular organisms to persist as discrete populations in competitive coexistence with those of diverse other species. Yet, what had originally driven emergent innovation in this direction may well have been a different story, such as giving certain RNA-derived viruses or plasmids a distinct advantage over their by then conventional competitors (see further below).

Mainly from Carl Woese's break-through discoveries [119,178]. we can roughly imagine that the *LUCAS* stage was preceded by collective or communal populations of highly variable entities, and we have learned that three *superkingdom 'domains'* (aka *phylodomains*) of genuine organisms diversified and coexisted later on: *Bacteria*, *Archaea* and *Eukarya*. — Or were there only two of them [26,179] as posited by modern variations on the so-called *Eocyte Hypothesis*? — But what could possibly have gone on inside the *black box* of this crucial transitional period is not yet fully understood, nor have the evolutionary implications of a noticeable and deeply rooted mechanistic hierarchy been assessed for possibly innovative structural consequences. Furthermore, the potential survival of organizational remnants derived directly from molecular *protobiofilm* communities attached to mineral surfaces, such as Baum's '*surface protoplasm*', has not been evaluated for possible consequences at the *LUCAS black-box* transition or even beyond this conjectural stage.

To create the breeding ground for innovation as Woese put it, "*Vertically generated and horizontally acquired variation could be viewed as the Yin and the Yang of the evolutionary process*" [117]. As a subtle caveat, though, I should prefer putting Woese's more primitive *Yang* of cross-linked networks before the later emerging *Yin* of linear bifurcations in tree-like branching patterns. Evolutionary Ping Pong is appealing as a general concept. Its possible structural ramifications and kinetic channeling deserve further scrutiny — not the least as being applicable well before the tentative LUCAS transition when the overwhelming, population-wide influence of horizontal admixture gradually diminished in favor of vertically stable lineage identity at different levels.

Carl Woese had long fancied the notion that the evolutionary dynamics prevailing at the *Progenote* level were universally communal or collective, rather than being competitive and selective between separate and vertically stable organismal lineages — different species in the conventional Darwinian sense, that is. Computer-assisted modeling analyses have converted such theoretical conjectures about "*collective innovation sharing*" from prescient hunches into a respectable scientific theory [172]. The modeling was parameterized to simulate conditions for optimizing the genetic code as such, and mixed heritage by gene pool sharing throughout the population at large was found to be especially important, not the least for the extreme optimization observed for polar amino acid–codon interactions in the canonical genetic code [180]. Tentatively, the *Progenote* feature of population-wide communality was attributed to exceedingly high levels of lateral or horizontal gene transfer (LGT/HGT) but considering the tendency to coalesce by fission fusion cycles as a primordial trait of surface protoplasm allows for a more convincing narrative [75].

Subsequent modeling studies were likewise centered on how to optimize the coding system itself [181-183]. By further inference from nonlinear, coevolutionary dynamics, many other important subsystems should likewise be subject to the collective optimization principle facilitated by communal gene pool sharing [178]. The coevolutionary coupling of template-dependent replication, transcription and translation systems has since been supported by more detailed modeling [184,185]. Moreover, the collective optimization principle may also have forged the gradual coevolution of lipid precursors and amphiphilic peptides into functional biomembranes [78,186,187]. As this Feature Paper puts focus on the unconventional possibilities connected to more massive *protoplasmic bulk* than commonly considered in traditional *protocell* models, such conceptual extension should prompt further research on how finally to transit out of Woese's *acellular Progenote State* by an effective *cellularization* process.

In short, before modern organismal life forms could emerge from the hypothetical, more primitive *Progenote State*, virtually everything must have been different from what we know about cellular organisms today [188]. The sequence-determining processes of gene replication and protein synthesis were necessarily rather inaccurate to begin with, resulting in many different, partly randomized and relatively short product chains. With many different genes but no integral genome or assorting mechanism in the beginning, there was no regular cell division cycle either and vertically stable organismal lineages could not possibly exist. Instead, early evolutionary change was mainly focused on the collective optimization of many system-stabilizing interactions, which very much depended on communal innovation sharing and the frequent exchange of interactive components on a system-wide scale.

To emphasize the unbranching character of Woese's conceptual *Progenote State* — not yet allowing vertically stable sub-lineages to coexist in Darwinian speciation at the organismal level — I suggest the term *trunk-line evolution* for its gradual, collective maturation over time, when it took a well rooted trunk to bear the tree of Darwinian crown group diversification [189]. Of special note, the *trunk-line* concept is also relevant for modern organisms, especially with sexual reproduction in eukaryotes supporting population-wide gene pool sharing within a species by default. Darwin himself was well aware that his linearized model of a simple tree was a conceptual abstraction to gather larger populations at the species level and beyond, whereas the potential for identical reduplication of individually clonable cells or organisms was not part of his general evolutionary theory, which was solely based on experience with animals and plants when neither bacteria nor archaea were known to science or humanity at large.

Moreover, Mayr's *Biological Species Concept* as gene pool sharing in *groups of interbreeding natural populations* [190] is, strictly speaking, only valid for sexual eukaryotes, whereas a meaningful adaptation to asexual akaryote phylogenies requires additional practical considerations [191]. Modern eukaryotes are fulfilling this condition by the alternation of periodic clonality (mitosis) and episodic mixis (fertilization and meiosis) in a variety of organismal life cycles. Various barriers exist to keep effective mixis in the limits of a biological species and, therefore, eukaryotes in general are subject to an isolated kind of *trunk-line evolution* within each species: "*One Trunk Line – One Species*". Turning the argument around, the ancestral *Progenote* population as such could be considered the first — and only — *biological species* existing at its time. Although genetic barriers to keep two coexisting species apart had not yet been established, spatial separation by geographical scattering over suitable, increasingly fragmented environments could have led to early diversification of phylogenetic clustering effects, potentially still traceable in the modern biosphere [74]. For the following Sections it is worth keeping in mind that much of the initial phase of *Progenote trunk-line evolution* was dominated by RNA as the exclusive genetic material, whereas the transitional phase toward organismal speciation was highly influenced by the modified version of RNA that now is known as DNA — presumably introduced by certain viruses beforehand.

5.9. From Tangled Networks to Tree-like Lineage Evolution

The metaphorical *Tree of Life* is a powerful icon that has pervaded legendary myths about mankind's origins from pre-literal times in many cultures — long before Darwin and others drew it into the limelight of evolutionary science. More recently phylogenetic reconstructions from multi-genomic datasets have come to dominate the literature pertaining to the subject. The focus on sequence homology-based alignments alone, however, is being subjected to substantial critique [31,32]. In fact, the conceptual clarity of tree-like phylogenies is being blurred by the counteracting principle of multiple reticulation — different genes telling different historical tales, all the way up to the level of closely related species [192,193]. Even more so, the very rooting of the organismal ToL is virtually hidden by intrinsic communality throughout the pre-genomic reign of Woese's ancestral *Progenote State*. — Of note to this paper, the alternative approach of analyzing comparable sequence data sets as *sequence similarity networks* (SSN) [194,195] can assist in overcoming some shortcomings of homology-based alignments alone as substantiated further below.

In terms of partly intertwined enzymology [196,197], the lineage-tracking, tree-like component is dependent on the accuracy of semi-conservative RNA or DNA *replication*, together with structural functionality for distributive chromosome assortment by cytoskeleton constituents under periodic cell cycle control. Inherent inaccuracy and accidental damage, however, require the assistance of molecular *repair* mechanisms, such as intra-lineage *recombinational repair* to reestablish functional replication forks after stalling at double-strand breaks [198,199]. On the other hand, phylogenetic network reticulation is brought about by stochastic sampling among unconnected genes and inter-lineage sequence exchange by (non-homologous) molecular *recombination* from different sources, in the aftermath of protoplasmic fusion (sexually programmed in eukaryotes) or by the classically '*akaryotic*' processes of lateral/horizontal gene transfer (LGT/HGT), such as conjugation, transformation and bacteriophage transduction. To what extent Woese's ancestral *Progenote State* resembled one or the other principal mode of inter-lineage reticulation is still a matter of dispute. It is worth emphasis in this connection that a life-supporting network of protometabolic reactions and corresponding proto-catalysts supposedly preceded the emergence, nucleation and further evolution of lineage-stabilizing proto-genomic agents. Moreover, as faithful genome replication is likewise catalyzed by protein enzymes, this essential mediator of vertical lineage stability has in turn been fully integrated into the overall metabolic networking system as well.

The canonical *Tree of Life* has gained iconic power because it combines conceptual simplicity and familiar experience with phenomenological acuity and communicable illustration. Sadly, however, the more realistic concept of a '*tangled tree*' is also more cumbersome to discern from comparative data sets, more challenging to compute in model building, and less easily depictable in graphical form. Only a few examples have been presented yet to follow up on this still unconventional approach but deserve particular attention in the present context and for future work.

doi:10.20944/preprints202405.0555.v1

After wondering about the sparsity of universally comparable genetic sequence elements to reconstruct the basic ToL [31], a group of researchers coordinated by W. F. Martin pioneered to represent a more realistic ToL by network-related graphical displays [200-203]. Unfortunately, from my perspective, these authors took it for granted that "*Eukaryotes arose from prokaryotes, hence the root in the tree of life resides among the prokaryotic (phylo)domains*", so an additional '*Third way*' from progenotes toward *eukaryogenesis* would be entirely out of the question.

Meanwhile the SplitsTree algorithm [204] and illustrative Data Display Network (DDN) graphics, which "*are good for generating biological hypotheses but not for testing them*" [205], have also been applied to phylogenetic data sets derived from large-scale proteome comparison across the entire ToL [206]. Together with several preceding articles this important *phyloproteomic* paper strongly supports Woese's fundamental tenet that not only the two *Akaryote* (aka Prokaryote) phylodomains but also *Eukarya* with their very different organismal cell types are quasi-independent descendants of a "*Universal Common Ancestor*" (UCA) community [140,142,143]. Thus, a tentative '*Third way' from progenotes toward eukaryogenesis* is no longer just a weird idea. When looking at the phyloproteomic DDN displays [206], it is apparent that (i) present organisms divide into three well separated clusters, (ii) the ancestral triple junction to connect the outer clusters resides within the denser tangles of the DDN graphics which summarize the data conflicts with the conventional assumption of a uniquely definable *ancestral lineages* at the deepest branchpoints of a formally dichotomous *Tree of Life*), and (iii) including Asgard archaea in the recent data update does not drastically change the overall three-foil topology of the formal network plotted in these diagrams.

At this part of the Section I suggest using the clear-cut triple division of the global *phyloproteomic* data sets [143]. for a provisional redefinition of '*Phylodomain*' (or '*Superkingdom*') as a primarily taxonomic term for each of three different '*organismal cell types*', none of which has been derived as such from any of the respective other two cell types. This practical measure is meant to evade the formal rooting problem of the canonical ToL and the "*delusive state*" resulting from a slew of "*delusive investigations*" by disregarding that several to many basal nodes actually "*might correspond to evolutionary stages of premature cells, namely, progenotes*" [207] — including the ancestors of bacteria and archaea or even of proto-eukaryote ancestral lineages themselves. Such disconcerting ambiguity was also anticipated by Woese already when he discussed the conceptual consequences of passing the so-called *Darwinian Threshold* asymmetrically at the deepest branch points, only affecting one branch every time [208]. The intrinsic "*Woesean asymmetry*" at early branchpoints of the ToL— to my opinion — is most evident at the still-subcoenocytic founding stage of prospective *'Bacteriogenesis'*.

As a practical measure a certain redefinition of '*Phylodomain*' should be helpful in redirecting priority resources toward "*the trees that best reflect biological history*" (see further above). It is symptomatic that the major *historical* problem has conventionally been viewed as affecting the unknown details of *Eukaryogenesis* alone, whereas *the related terms 'Prokaryogenesis', 'Bacteriogenesis' and 'Archaeogenesis'* were virtually non-existent in PubMed returns by 2015 [209]. Accordingly, the integrative paradigm shift called for in the introductory Sections should also properly account for the successful generation of modular *Akaryote* cells in stable lineages of vertical descent — thus leaving behind the acellular, tentatively coenocytic '*Progenote State*' of population-wide collective gene pool sharing.

Besides, as demonstrated in a global perspective for three particular protein families, large-scale data-mining for comprehensive analyses as *sequence similarity networks* (SSN) can be extended to (super)family-wide *phylogenetic analyses*: *Serine protease inhibitor (serpin)* proteins [210], *Ribonucleotide reductase* [211], and *7-Transmembrane Domain Ion Channels and Chemoreceptors* [212]. This concentric focusing on a single clan of fold-related proteins is also factoring structural information into tree-building algorithms from a recognized consensus folding pattern, which thus can lead to informative conclusions by identifying a nodal cluster at the supposedly ancestral center of a hub-like structural web. Accordingly, the more peripheral, spoke-like diversification into various functionally different subclusters is then inferred to be derived by divergent evolution. A prospective extension of such single-fold phylogenies to other universal protein families should potentially shed more light on the enigmatic origins of the organizational split between Akaryote and Eukaryote emergence.

doi:10.20944/preprints202405.0555.v1

6. A Web-like Trunk to Bear a Tangled Tree

6.1. The Virosphere — Yet What About It?

There is a fundamental flaw in the traditional ToL in that the large variety of viral agents are not usually considered relevant in phylogenetic reconstructions at the level of *organismal Phylodomain*s, but changing the historical neglect has not been easy [213-215]. As the conceptual ToL represents a tree of genomic lineages rather than a tree of cellular organization as such — and viral genomes follow lineage-wise patterns of vertical descent and variation just the same — there are no logical reasons for barring viruses from the universal ToL of ever-diversifying genomic lineages. Reaching even deeper yet, mobile genetic elements (MGEs) follow lineage-wise descent on their own, and viruses are suggested to represent a hybrid of two sources: replication-related functions from ancient MGEs, and capsid shells from ancestral cellular structural proteins [111]. It has been pointed out before that single genes are never '*selfish*' in the strictest sense, but becoming part of '*selfish modules*' is a different story. As soon as reliable replication mechanisms had been established — to serve many genes in parallel as a '*common good*', in support of system-level continuance — the emergence of temporally and/or locally over-replicating modules several collectively self-perpetuating components became virtually inevitable, as exemplified by mobile transposable elements, plasmids and viral agents [216].

Full recognition of viral evolutionary relevance has been unduly limited by academic debates about whether virus particles actually are alive or not [217,218]. Yet, viruses have system-sustaining qualities and self-directed modularity — thus dwarfing the potential of single genes in general [219]. But not having a metabolism of their own, viruses depend on susceptible hosting systems for particle propagation and recurring infectivity. Viruses in general and RNA viruses in particular are thought to have arisen well before the generation of modular cells in lineages of vertical descent [111,220], and it is worth asking how the primordial hosting system may have been organized in the first place. Arguably, the '*surface protoplasm to proto-coenocyte scenario*' put forth in the present article is a suitable framework to look at the origin of viruses from a more general perspective by also considering the so-called '*Virocell Concept*': The focus then is primarily on the *intracellular phase* of viral reproduction instead of *virus particles* as simple spore-like *propagules* [221-223], and the *ToL of organisms* becomes a more comprehensive *ToL of virocells*. The modern biosphere abounds with a variety of different viral lineages and other mobile elements, now coevolving with their hosts in every phylodomain of organismal life. *Nota bene*: bacterial viruses are also known as *bacteriophages* or *phages* for historical reasons.

Emerging viral genome lineages have undoubtedly become involved in to-and-fro gene transfer (LGT/HGT) with the host system — facilitated by intimate proximity at the *Virocell* stage. This concept represents an instructive case of '*symbiotic parasitism*', which has to strike a balance between lytic production of virus particles and the hosting system in terms of '*biological fitness*' of either partner. It is clear that any overly effective viral lineage would be on a suicide mission if all the potential host cells were wiped out upon acute virus infection. Accordingly many long-established viruses have found a way of entering a persistent '*latent state*', such as turning into a temporary plasmid stage or transposing reversibly into a host cell chromosome. Such virus genomes persisting in a latent phase can also be selectively advantageous for the hosting system [224].

Of note, different evolutionary dynamics prevailed on either side of the symbiotic virus–host relationship, with alternative or complementary selective advantages for the disparate partners: (i) A virtually unlimited variety of short, unlinked genes for quasi-statistic peptide sequences presumably resulted from a stochastic trial-and-error sifting through mutational changes in the communal cytoplasm and gradual enrichment for functional improvements and collective optimization of many enzymatic reactions. Yet, viral lineages could also pick and choose from this plentiful source of potential innovation. (ii) Once transferred to a virus genome, a candidate gene — potentially useful for viral functionality — could then selectively be specialized to perfection, in direct competition with other viral lineages. (iii) Some of these virus-encoded functions in turn — after transfer back into the communal gene pool of the protoplasmic hosting system — could become of particular use to the

doi:10.20944/preprints202405.0555.v1

emergence and diversification of organismal lineages as well. This is a superior category of constructive feedback by gene transfer from a youthful virosphere to a still embryonic, pre-organismal biosphere, which has been instrumental, I suppose, in shaping the conceptual connection from tentative OoL scenarios to unsolved problems about the rooting of the canonical ToL. (iv) Somewhat between viral and cellular genomes, plasmids were likewise forming self-directed, vertically stable lineages embedded in the progenote protoplasm. Of note, there is a floating conceptual border line between the categories since infectious viruses can turn latent as integrated MGEs or in a self-replicative plasmid stage not shedding virus particles. (v) Like viral lineages, plasmids could also have accelerated evolutionary perfection of the relatively few genes they carried along, and some of these genes may have become critically important for system persistence at large.

Instructive suggestions for virus-mediated organismal innovations are as follows:

1. *Recombinational hotspots*: Even previral agents, such as transposable MGEs, have found use in organismal genomes as entry points for horizontal gene transfer and shuffling of exons or protein domains [225].
2. *RNA to DNA transitions*: The modes of genomic DNA replication are generally equivalent in the organismal phylodomains. Their basic enzymology however is partly non-homologous — especially between *Bacteria* on the one hand versus *Archaea* and *Eukarya* on the other [226,227]. This means to me that the full transition to DNA has occurred stepwise and that the molecular diversity at different steps was potentially of critical relevance for connecting the core pattern in a residual "*ToL of 1 %*" to its historical roots at the organismal and/or pre-organismal levels. With tentative origins in different ancestral RNA polymerase genes of the preceding RNA-dominated era, the phylogenetically irregular distribution of replicative DNA polymerase subunits across the deepest branches of the organismal ToL [228] does not uniquely specify a single evolutionary paradigm regarding the biological history of cellular diversity. I'll return to this aspect more specifically in the following Sections. In contrast, the phylogenetic history of replication helicases in the archaeal branch appears more regular in following the rDNA-based standard tree [229].
3. *Heterochromatin-like clustering*: The emergence of the eukaryotic nucleus is still a mystery, which has been interpreted by two basic models: *endosymbiotic theories* from outside a prospective akaryote host cell versus *autogenous* (or *endogenous*) *hypotheses* suggesting diversifying membrane trafficking from within larger pre- or proto-eukaryotic cells [113,230,231]. The recent advances on '*viral replication factories*' have led to the notion "*that uncoupling of transcription from translation is a feature of giant viruses* [and] *the ability to uncouple transcription from translation potentially has a very long evolutionary history*" [232] — in support of the hypothesis that the nucleus is derived from a characteristic '*viral factory*' [233,234]. This model entertains the additional hypothesis that eukaryotic histones likewise derived from viral origins, allowing differentially compacted chromatinization for giant viruses first and for their host cells secondarily [235]. Molecular chaperones for histone assembly, too, may be related to viral proteins [236]. The intimate coevolution of viral and cellular membrane fusion proteins may likewise be relevant in this context [237], and bacterial viruses have perfected the translocation of DNA across membranes by molecular motors [238].
4. *Host line evolution and persistence*: The interaction of bacteriophages with chromosome-borne MGEs can be reciprocal and intense at characteristic '*Phage-inducible chromosomal islands*' scattered in the host cell genome. These gene-bearing clusters allow viral genomes to integrate in a '*lysogenic state*' and benefit the host cell by promoting genetic variability, protecting from the lytic stage, and shielding against super-infection by other viruses from outside [239,240].

To conclude this Section, structural conservation in virus evolution is also studied by the comparative *Phyloproteomics* approach [241], which has strengthened the belief that viral lineages began to emerge at the very onset of genomic evolution [242]. So, "*viruses should be considered drivers of cellular evolution rather than minimalistic genetic parasites*" [237]. Furthermore, the organized release from membrane-bound but not yet fully cellularized compartments — as '*Viral Escape*' of the earliest RNA viruses to begin with — has been integrated into an "*Extrusion model of viral Panspermia: from vesicles to viruses*" [238].

Incidentally, the conceptual inclusion of viruses in *phyloproteomic* analyses has refueled the critical dispute about "*realistic evolutionary models*" referred to as a latent '*Phyloproteomics paradox*' further above [244,245]. Yet, the Kurland group has kept to rather discordant views about Woese's concept of a collective *Progenote State* and its unconventional implications (herein quoted verbatim; my emphasis in bold): "*Second is the discovery that the most recent universal common ancestor (MRUCA) of the modern crown is not a bacterium (or an archaeon or a eukaryote). Rather, MRUCA has extensive phylogenetic affinities with eukaryotes as well as both bacteria and archaea, which* ***could mean that MRUCA has matured beyond the progenote stage****. /// It is* ***difficult to identify MRUCA with the progenote*** *postulated by Woese* [4] *because there is nothing elementary or simple about its proteome, but then there is* ***nothing simple in Woese's sketches of the progenote****. /// If there were* ***a progenote*** *in the early evolution of organisms, it* ***would have appeared much before the debut of MRUCA****. It is anyone's guess* ***how long the progenote mode of gene exchange persisted****. /// However, it is* ***inconceivable*** *that a cell as* ***complex as MRUCA could have been a progenote*** *– or so* ***it seems now***" [136].

To be sure, the conventional *Phylogenetics* research community is no longer fond of Woese's *ab initio* notions about collective sharing from the earliest *Progenote* stages (in the strictest sense of perfecting the genetic code) up to the rooting problem of the organismal ToL. To illustrate this point, Eugene Koonin has seriously dealt with Woese's concepts in 2014 at the latest [246], but in his otherwise highly informative review on the replication machinery of a tentative LUCA [228] there is no mention of Woese's considerations whatsoever — very much in contrast to Forterre's conceptions [177].

I will argue in the following Sections that Woese's collective *Progenote State*, in fact, has more potential bearing on the open issues with the formal ToL than the Kurland group and many others have been ready to admit. The potential role of plasmids in early evolution will play a key role in my reasoning in this regard. As for DNA lineages, viruses and plasmids have engaged in multiple interactions early on [247]. Correspondingly, similar RNA-based interactions may have developed even earlier. Evidence from modern life confirms the actual existence of double-stranded RNA plasmids [248]. I will argue for the possibility that an *RNA-based plasmid associated to 'protothylakoids'* may have founded the deeply rooted genomic lineage that ultimately has led to the emergence of free-living bacteria-like cells, and other plasmid lineages may have comparable effects.

- NB — For want of a better word, I deliberatively 'borrow' the *thylakoid* term for application in a more general ancestral sense than its well established meaning for highly advanced photosynthesis as represented in 'purple bacteria', cyanobacteria and plastids of eukaryotic plants.

6.2. Modular Cellularization — Progenote and Lineage Aspects Reconsidered

More generally speaking about the '*Phyloproteomics paradox*' mentioned above, the issues raised may even go deeper than some arguably unrealistic assumptions set up by the opponents. At the bottom line of the main problem, apparent anomalies may have arisen under the kinetic influence of several opposing trends — at various scales with regard to organizational levels or temporally speaking. To my understanding, the most generally relevant trend reversal (not yet fully understood or systematically analyzed, though) concerns the transition from system-wide (or module-wide) *accretionary evolution* at the collective *Progenote* stage to begin with, vs. partly *reductive evolution* later on when organismal lineages began to selectively adapt to narrowly defined environmental niches, such as marginal survival in extreme environments, cf. the '*Thermoreduction Hypothesis*' [249,250] or in predatory and parasitic lineages in upper sectors of the ToL [251].

In other words, the now prevailing views expect that one and the same model be applicable '*from top to toe*' — or '*root to branches*' — in the tree to be computed on a graph. It was Carl Woese's momentous and far-reaching insights to realize (i) that evolutionary dynamics must have been very different before and after the onset of organismal Darwinian speciation (as separated by "*Darwinian thresholds*") [188,172,208], and (ii) that this overall "*Woesean–Darwinian transition process*" was intrinsically composed of several, more locally defined principal components. This means that as soon as "*only one of the major evolving cell designs were to cross its Darwinian threshold, tree representation* *would appear to be appropriate because that one lineage (only) would be distinguishable from all the rest* [as a '*primary line of descent*'], *despite the fact that the others did not yet exist as discrete stable lineages*" [163]. It also means that all three Woesean "*Urkingdoms*" (aka '*Phylodomains*') may well represent "*primary lines of descent*" resulting from the "*Woesean–Darwinian transition process*" by different and partly independent means.

Woese himself has already compiled a string of arguments (herein quoted verbatim; my emphasis in bold) that the universal ancestor to start the ToL has most likely been a *Progenote*:

"In principle ***the universal ancestor*** *could have* ***[1] resembled any one of the three major type****s of extant organisms. It also could have in essence been* ***[2] a collage of all three****, or have been* ***[3] very unlike any of them****. I will argue that the last alternative is the correct one and that the universal ancestor was a progenote." /// The alterations "required to* ***change one*** *of the three phenotypes* ***into either of the others*** *are* ***too drastic and disruptive*** *to have actually occurred." /// The "only solution to the problem is for the universal* ***ancestor to have been a progenote****." /// In "the* ***transition*** *from the universal* ***ancestor*** *to its* ***descendants*** *we are witnessing the* ***evolution of biological specificity itself****." /// Since "the progenote is far simpler and more rudimentary than extant organisms, the* ***significant differences*** *in basic molecular* ***structures and processes*** *that* ***distinguish the three major types*** *of organisms would be* ***attributes that the universal ancestor never possessed****. In other words, the more rudimentary versions of a function present in the progenote would become* ***refined and augmented independently****, and so* ***uniquely****, in each of its progeny lineages. This* ***independent refinement*** *(and* ***augmentation****) of a more rudimentary function,* ***not the replacement*** *of one complex function by a different complex version thereof (the beginning stages of which would be strongly selected against), is why* ***remarkable differences in detail*** *have evolved for the* ***basic functions in each of the urkingdoms****" [252] … "If modern large proteins could not be produced by progenotes, then a modern type of* ***genome replication/repair mechanism did not exist****. As with translation, a* ***rudimentary mechanism implies a less accurate one****, and the resulting high mutation rates necessitated small genomes. The structure of these genomes must reflect the primitive evolutionary dynamic in general. Therefore, I see the progenote genome as* ***organized rather like the macronucleus*** *of some ciliates today []: it comprised* ***many small*** *linear chromosomes (****minichromosomes****), each* ***present in multiple copies****. /// Small primitive genomes with low genetic capacity and* ***imprecision in both translation and genome replication*** *imply a primitive cell that was rudimentary in every respect /// It was [] a* ***community of progenotes****, not any specific organism, any single lineage, that was our universal ancestor — a genetically rich, distributed, communal ancestor. It was also* ***this loose-knit biological unit*** *that ultimately* ***evolved to a stage in which it somehow pulled apart*** *into two, then three communities, isolated by the fact that they could no longer communicate laterally with one another in an unrestricted way.* ***Each had become sufficiently complex and idiosyncratic that only some genes, some subsystems, could be usefully transferred laterally****. Each of these* ***three self-defining communities*** *then further congealed, giving rise to what we perceive as the* ***three primary lines of descent****" [4].*

These passages support several major points to motivate the reasoning put forth in the present paper: (1) The high levels of *genetic redundancy* expected for macronucleus-like clustering in progenote entities were easier to attain from pre-genetic stages that already consisted of relatively large bodies of functionally interactive protoplasmic matter. (2) The initial *inaccuracy of processive mechanisms* affected *replication* as well as *transcription* and *translation* but not necessarily to the same extent at every particular substage. (3) A natural series of temporally ordered evolutionary perfection can be inferred from additional input as follows (not yet explicitly considered in Woese's presentation): (i) initial perfection of tRNA charging at a genuinely pregenetic, *analog* stage of pre-progenote life-like molecular networking and functionality; (ii) intermediate perfection of RNA replication with RNA-to-protein coding and decoding mechanisms for individual genes at the early *Progenote* stage (*sensu stricto*); (iii) final perfection of genome-wide accuracy and processivity of DNA replication at the late *Progenote* stage (*sensu lato*), so as to warrant the conservation of chromosomal synteny in a species-wide population over considerable evolutionary time; and (iv) whilst the direct impact of chromosomal synteny was strongest for monomolecular plasmids and akaryotic genomes, it was less prevalent for the multi-molecular gene pool for cytoplasmic functionality in the progenote population at large, which may have been the major reason why it took longest to attain vertical

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2024 doi:10.20944/preprints202405.0555.v1

stability for multi-chromosomal genomes from an ancestral *nuclear-cytoplasmic lineage* to eukaryotic organisms.

The high degree of polyphenotypic variability intrinsic to the communal *Progenote state* is generally underappreciated in current studies of the rooting problem for the universal ToL This putative indistinctness has led Kandler to suggesting "*allopatric speciation of a multiphenotypical pre-cellular population*" [74], as pointed out by Wächtershäuser: "*These precells are seen as 'multi-phenotypical', having distinctly different metabolic phenotypes. Some sub-populations may be autotrophic, others heterotrophic; some anaerobic, others micro-aerophilic; some H2-consumers, others H2-producers, etc.*" [252].

It is the purpose of this Feature Paper to devise a plausible model to further support the prescient insights of Woese and Kandler — without falling prey to cladistic orthodoxy alone, which seem to imply that the composite eukaryote cell-type has been derived secondarily from a primary archaeal lineage [25,118,253]. The subcellular strategy of self-directed lineage stabilization — pursued by viruses and plasmids alike — should also pave the way to understanding how Woese's organismal 3-D ToL could have originated from a somewhat indistinct, internally complexifying '*progenote collage*' of protoplasmic masses, combining rudimentary properties of all three major cell types during a collective phase of *trunk-line evolution* at a formally common root.

It is important in the present context to be critical about what is meant by a 'cell' as a morphological and organizational unit of life. Is it the shell-like container or modularity of its functional contents that is more important in comprehending the tentative origins and early evolution of these complex biological entities? Almost two decades ago, Juli Peretó considered the question of "*Early or late cellularization*" as one of several controversies still waiting to be resolved, and he expressed a personal bias that "*life would have been cellular 'ab initio'* [from the very start]" because he found it difficult to imagine how the necessary bioenergetics could be managed in '*acellular*' systems [45]. This preconception appears related to received consensus views that '*protocells*' had to be small, vesicle-like to begin with for subsequently to evolve into *Akaryote* (aka '*Prokaryote*') cells before additional features resembling eukaryote complexity emerged at considerably later stages.

As applied to eukaryotes however, the basic concepts of unitary (mononuclear) *cells* and the corollary of *cellularization* as a process of generating preferentially mononuclear cells from larger *acellular* (*coenocytic, polynuclear*) systems had originally been introduced when bacteria-like *Akaryote* cells were not even known to science. On second thought, the organismal modularity of eukaryotes — the ancestral *nuclear-cytoplasmic lineage* included — may not even be of a cellular nature primarily but rest on the modularity of equational nuclear division. The classical concept of *cellularization from acellular syncytia* in eukaryotes has led me to conjecture that *Progenote entities* systemically adopted a composite, "*plasmodial-like organization*" as '*proto-coenocytes*' [106,114], which is not equivalent to the conventional conception of simpler, vesicle-like configuration in various '*protocell*' models.

Some modern *coenocyte* examples comprise of multinucleate amoebae, siphonal green algae, or syncytial slime mold plasmodia. Large amoebae are also known today to act as evolutionary '*melting pots*', which facilitate the proliferation of chimeric microorganisms, such as giant viruses [254-256]. Foraminifera and plasmodial slime molds are of particular interest in this context because of their tendency to coalesce by cytoplasmic fusion, respectively occurring within an extensive '*reticulopodial*' network [257], or between larger '*plasmodial*' masses [258].

In analogy to the concept of '*Viral Escape*' (above), the compound model suggested here assumes that bacterial cells were the first *Akaryote* lineages to descend from the communal *Progenote State* by '*Cellular Escape*' after a considerable period of accretionary evolution as plasmid genomes enclosed in endogenous *proto-organelle* compartments — inside the composite, polymorphic and amoeba-like *proto-coenocytes* of a collective, polyphenotypic '*proto-plasmodial trunk-line population*'. Similar ideas were also being developed in the 1970ies suggesting that a *protein synthesis system* was implanted into the *respiratory organelle* by incorporating a *stable plasmid* with genes for *ribosomal components* [259,260] — equivalent to combining a '*ribosomal DNA episome*' with '*plasmid-associated thylakoids*' [107].

These early ideas have been effectively shunned by proponents of the now prevalent doctrine that mitochondrial lineages solely originated from once free-living *α-proteobacteria* [261,262]. On the other hand, *mitochondrial phyloproteomics* can also tell a different story of most mitochondrial proteins not showing any particular relationship to *α-proteobacteria*, which led to alternative views implying the pre-existence of mitochondria-like modules of eukaryotic ancestry — as potentially endogenous *premitochondria*, well before eventual genomic interactions with free-living *α-proteobacteria* [112,263,264]. Comparative reviews of mitochondrial origins with regard to "*symbiogenic-chimeric vs autogenic-incremental*" conceptions have since been taken up in favor of multiple symbiotic interactions very early on [265-267]. However, the '*early on*' in this debate has not yet explicitly included the collective *Progenote* stage at the common root of Woese's 3D-ToL.

As briefly mentioned earlier, I refer to the *Progenote* concept in both its narrow and its broader meaning — corresponding to an early and a later phase of collective system optimizing — which initially concerned '*the making of genes and gene products*' for immediate usage on the spot, and '*the making of genomes*' for faithful inheritance in vertically stable cells and organismal lineages later on. Evolutionary optimization came about via differential survival of the better-fit performance within two to three coupled modes of molecular catenation in a processive manner: mRNA-directed protein synthesis by composite *ribosomes*, and template-dependent nucleic acid synthesis by *transcriptase* action and/or composite *replisomes*. Yet, why should it require longer periods to optimize processive replisomes than what it took for the considerably larger ribosomes? — The most reasonable answer may be a complex one of matching the chemical reactivity and instability of RNA against the stochastic limits of tolerable genome size, as imposed by the large intrinsic error rates to start with [268,269]. In consequence, the relatively large genomes needed for cellular organisms to stably coexist with others could only have come into existence in the aftermath of one or more RNA-to-DNA transitions — presumably after DNA viruses and plasmid interactions had paved the way [247,270].

It is my present opinion that the notion of *plasmid-associated thylakoids* [107] offers the best potential for rooting the entire bacterial domain deeply in the predominantly collective *Progenote State* of Woese's early theorizing, and key to this notion is the conceptual separation of *modular cellularization* from the molecular nucleation steps that established the earliest *vertically stable lineages* of different minimalistic '*genomic agents*' as such. This is where certain plasmid lineages could have made a difference well before genuine cell lines had been established. There is a general understanding that much of the universally communal *Progenote State* unfolded under the influence of RNA as the predominant informational molecule, but the lineage-wise establishment of organismal cell types very much depended on the adoption of DNA as genomic material [176,208,226,271]. It is thus reasonable to assume that the tentative *energy-harvesting genomic agent associated with 'protothylakoids'* began as an RNA plasmid relatively early in the *Progenote Era*.

The founding concept of *plasmid-associated protothylakoids* promoted here argues that a certain RNA plasmid made itself indispensable by giving prominence to a triad of system-sustaining innovations already at the RNA-dominated beginnings of the collective *Progenote State*: (1) self-reproductive capacities at a miniscule scale, including self-directed replication and self-serving ribosomes quite early on; (2) micellar to membrane-like association of amphiphilic peptides and lipid constituents; (3) directional channeling of photon-induced charge separation in coupling to inter-molecular transfer chains for electrons and/or protons, coordinated at lipid-raft-like nucleation centers. The early generation of endogenous vesicles in coupling to environmentally driven charge separation appears particularly attractive since this topology should allow for the simultaneous utilization of reactive electrons and protons for different redox reactions on either side of the emerging lipid rafts and early membranes. The plasmid-coordinated compartmented module of subcellular structure–function integration could thereby have provided the mechanistically organized "*Engine of Free Energy Conversion*" needed to get life-like metabolism under way [272], albeit augmented by different mechanisms and amplifying effects of repetitive environmental pumping than what arguably could not be expected from constant flow patterns under the long-favored OoL scenario at submarine hydrothermal vents.

On the sub-coenocytic basis of this model I find it natural to argue that the composite machinery of genomic DNA replication specific for bacteria began to consolidate first and did so by partly independent means as compared to functionally equivalent replication modules of other origins, which are partly shared by archaeal and eukaryotic organisms [220]. Plasmids are often mentioned together with transposable elements and viruses for their early lineage-wise emergence as self-serving reproductive units at subcellular scales [273,274]. Their evolutionary potential, however — affecting the hosting system at large — could have been distinctly different, as exemplified above by the posited coupling of membrane-based energy harvesting directed by an early-emerging, system-supportive plasmid lineage, presumably as a *compartmented RNA plasmid associated with protothylakoids* to start with.

This unconventional perspective can shed a new light on the puzzling complexity of RNA-to-DNA transitions ancestral to the major organismal cell types. A major uncertainty may point at the heart of the puzzle: Is it still reasonable to assume that there actually existed a cellular 'LUCA' with a uniquely definable replicative polymerase? [228] — Or should this questionable presupposition rather be dismissed in favor of a historically more plausible scenario? Woese's conception of a communal *Progenote Phase* (in a broader sense) has, arguably, the highest potential to unravel this conundrum.

There is a particular null hypothesis worth keeping in mind when thinking in terms of the *Progenote State*, in that all three processive reactions at the heart of *Progenote* networking complexity are *egalitarian* with regard to substrate spectrum and *communal* as regards their systemwide effects: 1. *Translation*: A certain kind of ribosome is responsible for producing all the gene-encoded proteins in the communal system. 2. *Transcription*: A certain kind of RNA- (or later DNA-) dependent RNA polymerase is responsible for making all (or most of) the potential mRNAs in the communal system. 3. *Replication*: A certain kind of RNA- (or later DNA-) dependent principal *replicase complex* is responsible for duplication of all the genomic ('chromosomal') molecules in the communal system, and all the potential *replication origin* sites respond to the same communal mechanism of organizing nascent *replication forks*. The early deviation of viruses and plasmids from this principle of general communality — as well as cellular lineage consolidation later on — would then require additional steps for explanation. These early semi-autonomous modular agents changed the rules by limiting replication mechanisms toward reproduction of their own genes preferentially.

By comparing replicon organization in bacteria and eukaryotes I find it remarkable that DNA replication in bacteria resembles plasmid replication in various aspects: Bacterial genomes are foremost contained on single molecules of circular DNA, including a single, bidirectional replication origin. The same is true for most bacterial plasmids, which are non-essential for cell growth in general [140,275]. Furthermore, some 10% of bacteria have large, essential '*extra-numerous chromosomes*' which actually are mega-plasmids with other — plasmid-specific — replication origins and partition machineries [175]. I am confident therefore that the founding principle of a *genomic plasmid* compartmentalized with *protothylakoids* is a viable model for gradual accretive evolution within a larger mass of potentially coenocytic protoplasm as early as the upcoming *peptide–RNA alliance* in a primordial *RNP World* scenario.

We can only guess what actually happened at that tentative RNA-dominated stage since there is no extant organism surviving to fully represent that era. Only RNA viruses, RNA plasmids and retrotransposons can give us some insight into its nature and tentative history, though experts are still divided on how to explain the spurious evidence [244,277,278]. There are good indications that *reverse transcriptase* (RT) of retro-transposons and RNA viruses had remarkable roles in RNA-to-DNA transitions at the organismal level, at least for the catalytic domains of major DNA replicase complexes in archaeal and eukaryotic cells [228]. The RT enzyme relates to the characteristic core domain of RNA polymerase, aka transcriptase, whereas the major bacterial DNA replicases are based on a different RNA-making protein family, which in eukaryotes is including polyA polymerase, aka terminal riboadenylate transferase. At the superior superfamily level, though, all the known *Nucleic Acid Polymerase* proteins are structurally related by resembling a right hand with *fingers, thumb, and palm regions* and thus may have evolved from a very ancient common ancestor. [279]. Moreover, the

doi:10.20944/preprints202405.0555.v1

only class of DNA topoisomerase occurring in all three domains of life (Type IA) is often associated as well with RNA topoisomerase activity [280,281].

From early, relatively open discussions of potential possibilities at the major RNA-to-DNA transitions [226,282], the expert community is now more categorically divided between opposing views: (1) Assuming beforehand that there was a distinctive — *prokaryote*-like — common cellular ancestor and that Woese's communal *Progenote State* had ended long before [228], vs (2) accepting that the tentative *Progenote Phase* ended stepwise by giving rise to different primary cell types one by one — as an implicit prerequisite for organizing the composite machineries of initiating and performing processive DNA replication at the organismal level, more than just once and by partly independent means [270,271,283]. It is my present aim to bridge this virtual gap of understanding by conceptually separating the early stage of genomic nucleation in viral and plasmid lineages from the later steps of gaining genome integrity and cellular autonomy at the organismal level.

This explorative conception leans on a nesting principle by allowing certain semi-autonomous minigenomes to emerge and be compartmentalized inside a larger systemic whole, which in turn kept nourishing some of these embryonic genome lineages for accretionary growth up to their later '*escape*' as quasi-autonomous cellular entities. To substantiate this notion I herein suggest reinterpreting the rooting problem of linking both Bacteria and the partly related mitochondrial lineage to the base of the canonical ToL according to the following considerations:

1. The founding core of intracellular genomes was a '*protothylakoid'-associated RNA plas*mid.
2. It carried an operational core for independent protein synthesis, perhaps assisted by a productive combination of recombinational bypass of replication-blocking lesions and/or the superior principle of '*rolling circle replication*' of ribosomal RNA sequences, — analogous to a commonly observed mode of differential gene amplification today [284]. Somewhat indirectly, *rolling circle replication* can be initiated by recombination between circularly permuted linear sequences and/or terminal redundancy [285-287]. The plasmid-based coding potential comprised one or more membrane-interacting amphiphilic proteins with directional charge transfer or other energy-converting capacity, which became vitally important for the surrounding protoplasmic system at large.
3. Micellar, vesicular or cisternal *protothylakoids* accumulated around the associated plasmid molecules and eventually fused to form internal, organelle-like compartments [107].
4. Topological closure of a surrounding envelope, however, could only be achieved in coevolution with appropriate transport systems into and out of the emerging compartments.
5. The RNA-to-DNA transition of the compartmentalized plasmid was partly independent from the larger systemic whole. The process began with plasmid-specific replication origins and ancestral primase–helicase complexes — presumably similar to the metazoan mitochondrial initiation system [288]. Accordingly, the peculiar resemblance between DNA replicases in mitochondria and T7-like bacteriophages [289] and the discovery of T7-like lysogenic prophage modules, which are inferred to better represent ancestral stages than the better known, strictly virulent T7-like phages themselves [290], fall neatly in line with the case study of a lineage-defining gene exchange equilibrium between viruses and plasmids with regard to certain host-related replication specificity factors [229]. The main point here is that minimal lineage-defining modules comprised of particular sequence elements in DNA to function as preferential internal replication origins and corresponding proteins to recognize the starting sites for processive template replication.
6. The nascent lineage-tracking genome modules in turn had to deal with increasingly multidimensional concerns for subsequent accretionary growth, not the least in bargaining overall genome length against the cumulative hazards of accidental damage and momentary replication infidelity. Inasmuch as the resolution of many such replication-blocking events required '*trans-lesion synthesis*' of DNA for *recombinational repair*, all organismal genomes — and larger viruses too — depended on more than just a single kind of DNA polymerase and also needed more effective processivity clamps for long-term lineage persistence and stability. Arguably the most significant modular innovation in this regard is the establishment of bidirectional replication by multi-enzyme replisomes [291]. A pair of sister replisomes is set in

motion at a common origin of replication — only to be dissolved after pairwise collision at certain replication-termination zones [292].

7. Each replisome is assembled at a nascent replication fork after ds-DNA has been opened at a replication origin by helicase/primase deposition. The overall gearing of these composite molecular machines appears comparable in all domains of life, but many individual components are structurally non-homologous in bacteria as compared to archaea and eukaryotes [293]. As composite replisomes too (similar to ribosomes) represent an important *functional module* amongst "*new cellular subsystems that are refractory to major evolutionary change*" [4], they should resist the replacement of single components by LGT/HGT. However, while proto-ribosomes were vitally important throughout the *RNA-directed early phase* of collective *Progenote Evolution*, typical replisomes became important only during the *later stage of RNA-to-DNA transitions* — with a potential for multiple emergences at different proto-organismal branchpoints of the formal ToL.

6.3. RNA-to-DNA Transitions at the Crossroads

Generally speaking it is not any replicative DNA polymerase as such that is central to replisome organization but a spider-like hub with tethers to coordinate several molecular shackles to the effect that the loop-assisted fork structure remains united and functional to duplicate the entire replicon with end-to-end reliability. In particular the discontinuous, looped-out synthesis of Okazaki pieces on the lagging strand – away from the advancing helicase — is an intrinsic challenge to persistent replisome integrity. Accordingly, up to three DNA polymerase complexes are being tethered to the pivotal helicase [294], each replicase being associated with a circular sliding clamp to warrant long-distance processivity of template-directed DNA synthesis [295,296]. Multi-subunit clamp loaders are needed nowadays for locking the ring-shaped sliding clamps in a full circle around ds-DNA, thus begging the question of how the intricate interdependence within composite replisome modules could possibly have derived from simpler structures with fewer capabilities.

Organismal replisomes come in just two different varieties, distinct in composition, structural topology and putative evolutionary origins: bacterial lineages on the one hand and archaeal/eukaryotic versions on the other. The mobilome and virosphere provide for further variation and may give us some insight on the evolutionary emergence of organismal replisomes [293,297]. Various viral DNA replication machineries are diverse enough to warrant meaningful statistical analyses, and numerical correlations appear relevant in two different contexts [298]. Above all, the complexity of virus-encoded DNA replication machineries is positively correlated with increasing genome size, and the non-random patterns of co-occurrence for several key components may reflect step-wise, coevolutionary emergences of structurally interactive sub-modules within the composite replisome. The preferential coupling of replicative helicases, primases and DNA polymerases — in this order — underlines the central role of helicase action in assembling the composite replisomes and in defining a self-directed hereditary lineage by choosing where to initiate replication amongst different molecules of DNA. Another functional relationship connects accessory clamp loaders to DNA polymerases and sliding clamps, which are predominantly found in the largest virus genomes of all three domains of organisms, e.g. T4 phage (~170 kbp) in bacteria and giant viruses of eukaryotes (300 to 1200 kbp). Whilst all toroidal polymerase-binding sliding clamps require clamp loaders to be installed, there are numerous viruses of intermediate size having DNA replicases not in need of DNA-encircling sliding clamps, e.g. T7 phage (~40 kbp) in bacteria and herpes-like viruses (120 to 250 kbp) in eukaryotes. This structural distinction may also be functionally relevant for the consolidation of organismal genome lineages.

In general, virus evolution is locked in an arms race with cellular hosts. There is no organismal species known to be free of viral attack altogether, but the different varieties of viral DNA replication proteins are not spread evenly across the three domains of life. Considering the structural correspondence of viral key components to either bacterial or archaeal/eukaryotic counterparts, there is a striking host range bias as follows. Viruses with primases and DNA polymerases of the latter category are numerous in all three domains of life, whereas the bacterial counterparts are confined to rather few bacteriophages. A similar trend is also observed for the accessory proteins of sliding clamps and clamp loaders [298]. These fundamental differences indicate that virus hallmark replication genes descended from primordial replicators well before the consolidation of cellular organismal lineages [299]. It also indicates to me that the origins of early viral lineages lead back to the polyphenotypic '*proto-plasmodial trunk-line*' population characteristic of the universally communal *Progenote State* — well before the emergence of free-living bacterial cells.

To me at least, the asymmetry in host range distribution also signifies a remarkable inequivalence regarding the ways and means of how the first lineages of bacterial and archaeal cells eventually emerged from a common ancestral *Progenote State*. An inherently asymmetric model appears necessary to explain the empirical observations in terms of natural evolution. The compartmented, plasmid-based conception proposed herein does actually provide for a potentially appropriate asymmetry condition in that a small internally compartmented plasmid genome lineage could draw on the large quasi-stochastic gene pool of the surrounding protoplasm for further accretionary growth and adaptive selection toward cellular independence. The two-phase model implies a distinct asymmetry with particular kinetic consequences significant for differential evolution on either side of the compartmented partnership. This intrinsic disparity appears relevant to contextualize the notion of "*Woesean asymmetry*" concept mentioned further above, which has been postulated to affect the earliest branches of the organismal ToL — as a "*nonclassical perspective* [that] *takes some* [time for] *getting used to*" [208], and the clock is still ticking for taking extra time.

The early duality may have set a path toward contrasting strategies in bacterial and eukaryotic cells to make a living. The fundamental complementarity of these strategies has long been recognized and discussed in terms of early evolutionary divergence [116,249,300]. These general considerations are still valid and deserve renewed attention as revised, more basic OoL conceptions are adding weight to assuming sunlight-exposed *terrestrial* settings, cohesive *coalescence* and environmental *wet/dry cycling* as major driving factors for early evolution.

The strategic differences between Akaryote and Eukaryote cell organization appear correlated with small and large cytoplasmic volume, respectively. Additional parameters derived therefrom are also relevant for differing modes of density-dependent population dynamics [300-302]. As the larger eukaryotic cells are generally limited to lower growth rates and final densities than what is possible for smaller and more specialized bacteria, eukaryotes tend to evolve in *equilibrium populations* by so-called "*K-selection*", as opposed to "*r-selection*" dominating in the more *opportunistic populations* of bacteria subjected to recurrent catastrophic mortalities. As '*r-selection*' generally results in high degrees of genomic streamlining, this evolutionary mode is probably of secondary origin [173]. whereas the ancestral pre-genomic state of low-fidelity copying systems with high error rates had to cope with high levels of genetic redundancy [303]. Accordingly, Woese's collective *Progenote State* primarily responded to '*K-selection*' — being more concerned with system maintenance, perseverance and homeostasis at moderate growth rates than with faithful reproduction at high growth rates right away. Furthermore the smaller population sizes attainable for larger systemic entities may have been subject to "*constructive neutral evolution*" resulting in directional, "*ratchet-like increase in complexity*", which now pervades much of eukaryotic cell organization [304].

The '*Karyogenic Proto-Coenocyte Hypothesis*' [106] is thought to match the conditional requirements listed above by considering a collective and persistent population of large systemic *K-selected* entities, which hosted compartmented, primarily *r-selected* proto-organellar lineages of smaller size and with limited gene numbers per circular (or circular permuted) genome — first based on RNA and later DNA. This notion of complementary dynamics and synergistic evolutionary potential carries endogenous *symbiotic relationships* to a new level of networking complexity. By reaching deeply back into the collective *Progenote Era*, a more superior and integrative *Symbiosis* principle should not just ascribe the emergence of mitochondria to a somewhat fortuitous singularity as "*a fateful symbiotic encounter*" (*sensu* Martin et al.) [261,305] between fully individualized 'prokaryotic' cells. Instead I find it more natural to conceptually link the arguably most significant symbiotic relationships in eukaryotic cell organization to the endogenous generation of subcellular genomic lineages based on plasmids, which in turn could adaptively respond to different environmental challenges. Various anaerobic modes of syntrophic cross-feeding are fundamental in many methanogenic microbial communities today, such as specific microbial biofilms [306,307]. The principle of *anaerobic syntrophy* between hydrogen-producing and hydrogen-consuming microorganisms may have been operative in a precursory scenario to energize the '*plasmid-associated thylakoids*' as suggested herein. More generally speaking, the nascent plasmid-based lineages could grow constructively by 'feeding' on the relatively large genetic redundancy accumulating in the collective hosting system.

This scenario calls for syntrophic cross-feeding in densely packed anaerobic biofilm communities as a collective mode to nourish variable progenote entities that were genetically based on short mini-chromosomes and plasmids, none of which had yet assembled sufficiently many different genes for granting individual cells a chance to colonize pristine environments completely on their own. How then could ancestral proto-bacterial entities push up their monomolecular genomes to more appropriate sizes? The burden of achieving this capacity, I suppose, was mainly placed on organizing more dependable replisomes with sliding clamps and clamp loaders in the course of evolutionary *RNA-to-DNA transitions* as pioneered by virus and/or plasmid lineages. A critical look from Woese's '*nonclassical perspective*' may shed new light on the ancestral relationship of mitochondrial and bacterial lineages.

6.4. An Unconventional Alternative View Concerning Eukaryotic Organelles

Traditional views have it now that all mitochondria in eukaryotic cells owe their existence to an ancient internalization event of free-living bacterial cells as endosymbionts, which since have undergone a long succession of genetic streamlining [308,309]. Also, the photosynthetic plastids of plant cells are thought to descend from a similar event at a later stage [310]. Potentially however the disparate replication modules are telling a different story of early divergence at the dawn of genomic *RNA-to-DNA transitions*. The changing conceptualization hinges upon the lineage-defining parameters that couple various types of replisome organization to corresponding initiation sites on genomic strands of DNA. The main cause of concern is that neither mitochondria nor plastids are still carrying the replisomes characteristic for their putative bacterial ancestors but resemble a heterologous viral type in this regard, and a similar discrepancy concerns the distribution of RNA polymerases, respectively comprised of multi- or single-subunit proteins in bacteria and eukaryotic organelles. These strange concerns are typically explained away by the ad-hoc hypothesis of modular replacement at a later stage of merger-based eukaryogenesis, as driven by a strong yet unexplained selection pressure in organelle evolution [311-313].

However, I herein suggest a different solution, which also considers the evolutionary potential of energy-converting '*protothylakoid*'*-associated RNA plasmids*. This now unconventional proposal conceptually re-attaches to the "*autogenous origin hypotheses*" held earlier as a principal alternative to the now prevalent assumption of "*xenogenous*" origins of endosymbiotic organelles [314]. Accordingly, a virus-related plasmid replication module would thus have been present in proto-mitochondrial ancestral lineages from very ancient origins. The energy-related coding part of this association may have started with lipid/protein-facilitated participation in iron–sulfur cluster assembly for ferredoxin-like components of electron transfer chains, which still contributes to essential aspects of mitochondrial functionality today [315]. Notably, iron–sulfur cluster assembly is also the last metabolic activity remaining in the mitosome-like, extremely reduced mitochondrial remnants of certain anaerobic protists [316]. Modern mitochondria, of course, serve many additional purposes and metabolic roles, which depend on mitochondrial genes derived from bacteria (α-proteobacterial ancestors, in particular). Under the unconventional new hypothesis these genes may then have been introduced by multiple rounds of LGT/HGT into the proto-eukaryotic stemline population. This is not a preposterous assumption, considering the accepted fact that the large group of halophilic archaea derived from an archaeal anaerobic ancestor that had acquired more than 1000 genes from a bacterial aerobic partner — perhaps in close cellular interaction by anaerobic syntrophy [317.318].

Furthermore, I also think that the ancestors of photosynthesizing plants did not necessarily originate by physical engulfment of entire cyanobacteria and subsequent reductional organelle domestication into plastids. Instead, the photosynthesizing gene set may have likewise been acquired by LGT/HGT into part of the preexisting mitochondria population, which already was fully organelle-adapted to cytoplasmic interactions beforehand. As the viral mobilome-related replication module of mitochondria was being retained by this hypothesis, there is no longer need for any late-acting "*highly efficient selection pressure*" [311], which has evaded explanation since the coincidental discovery of a rather archaic virus-related replication mode in two functionally different organelles of eukaryotic cells. — Although symbiosis with free-living bacteria has played a crucial role in organellar evolution, mitochondria and plastids are mosaic organelles, and the bacterial role in the origins of eukaryotic endosymbiogenesis may not have been as revolutionary, all-pervasive and all-encompassing as it was initially envisaged [319,320].

6.5. Two or three Superkingdom Phylodomains? — That's the Question

Where did Bacteria and Archaea come from; how did they gain their general status of clonable, potentially autonomous cells; and how did eukaryotic cells enter the stage in the first place — or rather a 'third way'? Before trying to answer any of these questions, we must first make plausible assumptions about historical detail at the preceding state(s) of biophysical/biochemical consistency and interactions, and then sift through various stochastic possibilities for further evolutionary change. One way or another: The three extant cell types appear to have some connections to the pre-cellular, pre-genomic *Progenote State*, which in turn is herein assumed to be influenced by yet earlier, pre-genetic stages derived from relatively bulky layers of phase-separated, surface-coating protoplasm — already supporting an emergent network of protometabolic interactions. Intriguingly, however, the overall validity of a three-fold exit pattern of modular cell types from communal collectivity in the ancestral Progenote State has been discounted by the discovery of Asgard archaea, which appeared to nest the entire eukaryotic branch within the archaeal domain [25,26]. Yet, the growing global data sets on this newly detected branch of microorganisms do not really substantiate the criticality of that serious challenge to the canonical 3-D model of the organismal ToL [27,28,321].

When the *Progenote* concept was first proposed, the reasoning was based on the notion that "*evolution of the cell is the evolution of the genotype-phenotype relationship*" and "*the link between genotype and* [cellular] *phenotype*" was not fully established at such primitive stages when catalytic functionality was just emerging at many different levels [119]. For as long as stochastic error rates were high, the corresponding signal-to-noise ratios were rather low, and sustainable chain lengths of macromolecular products were quite limited to begin with. To characterize the nature of *Progenote* entities in more positive terms, these are networking qualities foremost, resulting in the "*refinement and selection of innovation-sharing protocols, such as the genetic code*". The community then "*rapidly developed complexity through the frictionless exchange of novelty enabled by the genetic code*" [172], thus accelerating the collective optimization of other system-bearing traits as well — eventually resulting in genomic lineage stability by vertical descent of modular cells [208].

In formal terms Di Giulio has taken care of Woese's legacy by arguing in the top-down direction of the ToL, and a group around Francisco Prosdocimi & Sávio Torres de Farias is joining ranks with this approach. Regarding "*rapid and progressive evolution*" as being *typical of the Progenote State*, Di Giulio focuses on the fundamental differences between bacterial and archaeal cell organization, such as DNA polymerases [271], methyltransferases [322], cell division system [323], RNase P proteins [324], and membrane lipids [325]. The implication of many nonhomologous protein constituents in these and other basic system-bearing traits is taken as cumulative circumstantial evidence that the deep nodes of the canonical ToL — the founding ancestors of the *Superkingdom Phylodomains* included — were still part of the conceptual *Progenote State* [326]. Moreover, he extends similar arguments with regard to *Eukaryogenesis* from the collective *Progenote State* no matter whether *Eukarya* shared a common ancestor with *Archaea* in general or *Asgardarchaeota* in particular [327]. The Prosdocimi/Farias group, too, is taking seriously the possibility that cellularization from an acellular *Progenote State* occurred several times independently for different cell types [129,328].

Furthermore, the central tenet of Woese's 3D-ToL is expressed more distinctly as "Domain Cell Theory", which states that "the descendants of each of the three domains retained its [cell-type] identity throughout its own unique evolutionary pathway" [329]. Seen in this light, the rather fragmentary Asgard–Eukaryote affinity may well be considered an anomaly potentially explained by HGT events from sharing a particular environment at early ancestral stages [28,177]. Alternatively certain data quality problems with metagenomic Asgard genome reconstructions from environmental DNA sampling alone may even lead to artefactual conclusions [330,331].

Patrick Forterre is taking a somewhat ambivalent position on these issues from a semantic point of view. Even though he has long been supporting Woese's 3D-ToL, he only considers using Woese's *Progenote* term in the strictest sense, which corresponds to '*the making of genes*' early on — not the '*making of genomes*' in a maturing phase of still collective *trunk-line evolution* as suggested herein. What is clearly needed from my perspective is a separate identifier for the residual trunk-line stage (still progenote-like in a broader sense, as inferred by the generalized progenote hypothesis) after the bacteria had branched off as separate lineages. This conceptual stage was still ancestral to both Archaea and Eukarya but not equivalent to either one of two different descendants — and not to Archaea, in particular [250]. I therefore suggest using '*Arkarya stem line*' for the conjoint ancestral '*Supradomain*' interval specifically.

The neologism '*Arkarya*' was originally proposed to include the crown groups of Archaea and Eukarya *domains* as well [332]. which actually would have made the classical Woese tree, stricto sensu, a two domains tree: "*Bacteria and Arkarya*" [281]. But Woese himself, being openly reluctant to address the common stem line by a distinctive name according to standard *cladistic principles*, considered the formal branchpoints before the three '*primary lines*' not as conventional *cladistic* branchpoints. Instead he postulated multiple '*Darwinian thresholds*' as the pivotal evolutionary turning points "*when the transmission of genetic information moves from a predominantly horizontal mode based on lateral gene transfer (LGT) to a predominantly vertical mode*" [177]. I suggest addressing the common group of Archaea and Eukarya *domains* as an '*Arkarya Supradomain*' to mediate between the two controversial extremes. By this provision both Archaea and Eukarya would retain their status of legitimate domains, originating from a common '*Arkarya stem line*', which as such still could be considered part of the *LGT-dominated Progenote State (sensu lato)*.

To conclude this section, it is still maintainable that Woese's three-domain model of life best reflects the biological reality of three distinctly different basic cell types [333]. Moreover, the favorite loophole of modern 2-D ToL proponents deriving Eukarya from a fortuitous singularity — an ancient 'merger' between an archaeon and a bacterium — shows a peculiar disregard for mechanism [334]. The extraordinary affinity of Eukarya to Asgard archaea, however, can also be explained by preferential horizontal gene transfers before the crown-group radiation of extant eukaryotes [28,177] — for example, if the ancestral Asgard lineage had stayed in symbiotic contact with the ancestral "*nuclear-cytoplasmic lineage*" (*sensu* Doolittle [34]) for the longest time. The LUCAN networking model put forth herein substantiates the salient notion that "*Life was born complex*" [335], well before it started to diversify at the organismal level. These considerations do not strictly contradict the now prevailing views that eukaryogenesis (also) was accompanied by rapid "*bursts of gene gain*" before the corresponding crown-group radiation [336].

7. The "LUCAN" Stage of Early Life

7.1. The Latest Universally Common Ancestral Network — Concluding Bird View

Hierarchical modularity is a fundamental characteristic for the '*living state*' of soft organic matter throughout the global biosphere. Various aspects of this general statement have been discussed in a recent monograph [337]. In terms of *networking theory*, smaller parts are being connected into larger wholes at multiple levels. The part-connecting wholes — aka modules — are networks, which in turn can become parts in superior wholes, i.e. innovative modules at a higher level of complexity. Accordingly the global biosphere can be conceived as a network of networks of networks … in a nested, virtually bottomless series of modular interconnected subsets. Not all of these subsets are equally well defined, however, as to their individualizable identity: *Cells* and *Organisms* represent the most characteristic categories of bearing biological identity in this regard. This is why the evolutionary origins of cells and organisms deserve to be given paramount attention at the transition line between the bottom-up considerations of OoL research and the top-down approach to rooting the ToL from retrodictive phylogenetic comparisons alone. This is also why a generally important insight cited further above deserves special attention at this particularly crucial point: "*The higher-level units must appear as the result of coevolution of units at the lower level*" [92].

That insight seriously begs the question of which conceivable kinds of intra-protoplasmic modules — from within a still acellular progenote hosting system — had the highest chances of mutual coevolution and accretionary growth for gaining cellular or organismal autonomy and eventual independence in gradual response to natural selection. The general concept of "*lineage-forming entities*" [5] can shed new light on this pivotal evolutionary problem, especially if it is being applied to plasmids and proto-organelles at subcellular levels as emphasized herein.

Considering the phylogenetic origins and distribution of replicative DNA polymerases [270,338-341], the predominant catalytic replicase family may serve as a proxy for lineage stability of vertical descent in the corresponding group of organisms. On this basis, and also considering ecological drivers prevailing across the contrasting environments at terrestrial hydrothermal fields, I argue more specifically for a model that virus-related plasmids paved the grounds for three to four distinct DNA-directed lineages to be founded at the critical period of genomic *RNA-to-DNA transitions*:

- Three different types derived independently from sub-lineages of a '*protothylakoid-associated RNA plasmid*', which in turn gave rise to the surviving lineages of (1) mitochondria (in eukaryotic cells), and free-living cells of (2) bacteria (type C), and (3) archaea (type D), respectively.
- The co-occurrence of both pre-bacterial and pre-archaeal proto-organellar lineages in a common proto-coenocytic cytoplasm may have been a matter of *anaerobic syntrophy* between hydrogen-producing and hydrogen-consuming proto-organelles.
- In addition, the pre-mitochondrial kind of replicase (type A) diversified to also serve collective replication (type B) for (4) the entire (non-organellar) gene pool of the still communal '*Arkarya Stemline*' (defined further above).
- Furthermore, the scattered occurrence of DNA type B polymerases in certain lineages of bacterial and archaeal cells should then be ascribed to secondary acquisition from the non-organellar common gene pool, facilitated by intimate symbiosis in the aftermath of the founding stage.
- As a historical corollary to the two of four lineage-founding principles, the cell biological entities characterized by type A and type B polymerases happened never to have separated physically, in that both mitochondria and related plastid lineages (type A) have not as such ever left the common cytoplasm directed by type B.
- Finally the common cytoplasm directed by type B DNA polymerases could have formed the "*nuclear-cytoplasmic lineage*" (*sensu* Doolittle [34]) conserved in eukaryotic cells.

The "*LUCAN*" concept just summarized is changing the paradigm of how we can rationalize the formidable differences between eukaryotes and akaryotes in terms of cellular infrastructure (composite vs simple compartmentation) and genome modularity (multi-chromosomal or basically mono-molecular, respectively). Generally speaking, "*bacteria simply have a fundamentally different strategy for cytoplasmic organization as compared to eukaryotes*" [342]. The decisive question still is: Why is that so?

The main point here is that "*size matters*" inasmuch as eukaryotes are generally larger in total cell size and morphological complexity compared to akaryotic organisms, and this basic difference must have had direct consequences as to which '*evolutionarily stable strategies*' (ESS) [343] were being selected for in the different cell types from very early on. Whilst the tiny cells of akaryotes do not normally exceed the physical range of '*diffusion limitation*' for rapid metabolic interactions [344,345], the much larger and complex eukaryotes have complemented simple diffusion with more effective intracellular membrane trafficking and transport mechanisms based on cytoskeletal filaments and linear-stepping motor proteins — in cooperation with numerous signaling GTPases specialized in large-scale cellular organization [342]. Although some archaea have more such GTPases than bacteria, none of these is associated with membranes — in stark contrast to the majority of their eukaryotic counterparts [346].

It is now worth asking whether the two alternative cell type strategies necessarily had to appear in temporal succession, which actually is implied by conventional wisdom that eukaryotes originated from 'prokaryotes' — well after the divergence of Archaea and Bacteria from a legendary prokaryote-like common ancestor. In my opinion this received presumption is neither reasonable nor strictly necessary, and the LUCAN concept presented here is based on possible co-emergence and further coevolution of different cell types from rudimentary beginnings at a common source. It seems to me that film-like organic associations as proposed by David Baum [67], and also by myself [66,106], entail the coevolutionary potential required for virtually simultaneous origins of both small and large appearing "*lineage-forming entities*", which subsequently consolidated to the different cell types prevailing in the modern biosphere. Of note to physical concerns, the film-like morphology is supportive to rapid exchange of small metabolites in diffusion-limited reactions across the film and incentive to the emergence of regulated transport functionality to move less diffusible larger products and structures sideways in the film plane.

To conceive of such a solution with paradox-resolving potential in evolutionarily plausible ways I had to go back in time to very early stages in the innovative OoL scenario based on terrestrial hydrothermal fields instead of ocean depths [16-18]. The game-changing factor in this approach is replacing the naïve model of vesicular '*protocells*' (floating freely in a three-dimensional water body) by the alternative model of *hydroscopic films* — adsorbed to mineral surfaces below, and exposed to atmosphere and sunlight from above.

There will be considerably more to say about how cellular representatives of the three domains of modern organisms may eventually have descended from the overall communal LUCAN stage. They presumably did so by way of '*cellular escape*' — in analogy to the escape hypothesis proposed for infective viral particles [114,277]. By also considering the ambient *Ecology–Evolution Continuum* into which such an escape scenario should be conceptually embedded, Forterre's notion of *thermoreduction* [250] may have provided the selective gradient to favor the eventual 'escape' of compact akaryotic cells from the originally temperate birthplace of life into the more extreme environmental niches that also abound at terrestrial hydrothermal fields.

7.2. On the Desirability of Comparative and Quantifiable Model Testing

From the philosophical or epistemological point of view, the objective of my current considerations is not to replace any particular scenario for ancient historical events by just another hypothetical variety. What really matters is to formulate respectable alternatives for competitive testing by critical comparison in kinetic modeling analyses, as demonstrated in two relevant cases: (1.) The communal aspects of Woese's prescient Progenote conception [252] were formulated several decades before they became subject to critical testing by kinetic simulation studies with very positive support [172]. (2.) Another example pertaining to this context is Baum's notion of '*neighborhood selection*' for the spontaneous appearance of surface-associated life in the quasi-two-dimensional planar masses of '*surface protoplasm*' [67,347], which also is testable by experimental and *in silico* modeling approaches [91,93,94].

It is concerning that the Woese school did not seriously attend to structural or organizational issues regarding the polyphenotypic variability of interactive relationships that potentially prevailed at the communal Progenote stage of early life. The infrastructurally diverse LUCAN concept is herein proposed as a transitional stage of various collaborative genomic lineages and compartments, none of which could fully qualify as modular cells in the modern sense. As presented it is derivable by natural selection from an OoL scenario based on biogenic hydroscopic films adsorbed to mineral surfaces to begin with. It is my hope to motivate younger colleagues with expertise in kinetic modeling analyses to accept this proposal as a distinct alternative to the conventional conception of vesicular protocells. In future analyses it should be the pairwise competition of alternative models that matters most for the evaluation of potentially useful kinetic simulation protocols.

The highly intermeshed LUCAN concept is comprised of digitally driven lineage-forming genomic modules and structurally coherent entities with non-digital systemic functionality. This inclusive view would not only form a novel basis for rooting the organismal ToL — the transitional networking stage would also make room for including a lineage-forming yet non-organismal continuum of *MGEs, plasmids and viruses* — the so-called '*Mobilome*' — in natural ways. If set into a yet wider perspective of longer evolutionary time spans, the *LUCAN* mesh is just a particular focus on placing a systemic root on to the canonical ToL. There were other structural/functional networks of more or less symbiotic relationships — both before and after the tentative LUCAN stage — from affinity-driven adsorption at mineral surfaces of proto-metabolic and biogenic impact in a prebiotic environment to local ecosystems and the global biosphere today. With this in mind, and in the pursuit of Schrödinger's rhetorical quest "*What is Life?*", I basically share the views expressed by John Dupré and Maureen A. O'Malley [5]:

- "*… competitive relations are a transitional state, with multi-lineage metabolic wholes eventually outcompeting selfish competitors, and … this process sometimes leads to the emergence of new types or levels of wholes. Our view of life as a continuum of variably structured collaborative systems leaves open the possibility that a variety of forms of organized matter — from chemical systems to ecosystems — might be usefully understood as living entities.*"

Funding: This research received no external funding.

Acknowledgments: I wish to thank Tore Sivertsen and David Baum for critical and constructive comments on the manuscript.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Egel, R. (2012). On the misgivings of anthropomorphic consensus polling in defining the complexity of life. Journal of Biomolecular Structure and Dynamics, 29(4), 615-616.
2. Benner, S. A. (2010). Defining life. Astrobiology, 10(10), 1021-1030; cf. Astrobiology at NASA — Life in the Universe Internet] <https://astrobiology.nasa.gov/research/life-detection/about/>
3. Schrödinger, E. (1944). What is life? The physical aspect of the living cell and mind. Cambridge: Cambridge University Press.
4. Woese C (1998) The universal ancestor. Proc Natl Acad Sci USA 95:6854–6859.
5. Dupré, M., & O'Malley, M. (2009). Varieties of living things: life at the intersection of lineage and metabolism. Philos Theor Biol 1:e003.
6. Baum, D. A., & Lehman, N. (2017). Life's late digital revolution and why it matters for the study of the origins of life. Life, 7(3), 34.
7. Marske, K. A., Rahbek, C., & Nogués-Bravo, D. (2013). Phylogeography: spanning the ecology-evolution continuum. Ecography, 36(11), 1169-1181.
8. Gillespie, R. G. (2016). Island time and the interplay between ecology and evolution in species diversification. Evolutionary applications, 9(1), 53-73.
9. McGaughran, A., Liggins, L., Marske, K. A., Dawson, M. N., Schiebelhut, L. M., Lavery, S. D., ... & Riginos, C. (2022). Comparative phylogeography in the genomic age: Opportunities and challenges. Journal of Biogeography, 49(12), 2130-2144.
10. Baum, D. A., Peng, Z., Dolson, E., Smith, E., Plum, A. M., & Gagrani, P. (2023). The ecology–evolution continuum and the origin of life. Journal of the Royal Society Interface, 20(208), 20230346.
11. Harrison, S. A., Rammu, H., Liu, F., Halpern, A., Nunes Palmeira, R., & Lane, N. (2023). Life as a Guide to its Own Origins. Annual Review of Ecology, Evolution, and Systematics, 54, 327-350.
12. Shapiro, J., & Noble, D. (2021). What prevents mainstream evolutionists teaching the whole truth about how genomes evolve? Progress in Biophysics and Molecular Biology, 165, 140-152.
13. THE THIRD WAY - evolution in the era of genomics and epigenomics. Available online: https://www.thethirdwayofevolution.com/ (accessed on 3rd April 2024).
14. Russell, M. J., & Hall, A. J. (1997). The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. Journal of the Geological Society, 154(3), 377-402.
15. Martin, W., & Russell, M. J. (2007). On the origin of biochemistry at an alkaline hydrothermal vent. Philosophical Transactions of the Royal Society B: Biological Sciences, 362(1486), 1887-1926.
16. Mulkidjanian, A. Y., Bychkov, A. Y., Dibrova, D. V., Galperin, M. Y., & Koonin, E. V. (2012). Origin of first cells at terrestrial, anoxic geothermal fields. Proceedings of the National Academy of Sciences, 109(14), E821-E830.
17. 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(1), 872-887.

18. Damer, B., & Deamer, D. (2020). The hot spring hypothesis for an origin of life. Astrobiology, 20(4), 429-452.
19. Kompanichenko, V. N. (2012). Inversion concept of the origin of life. Origins of Life and Evolution of Biospheres, 42(2), 153-178.
20. Kompanichenko, V. (2019). The rise of a habitable planet: four required conditions for the origin of life in the Universe. Geosciences, 9(2), 92.
21. Deal, A. M., Rapf, R. J., & Vaida, V. (2021). Water–Air Interfaces as Environments to Address the Water Paradox in Prebiotic Chemistry: A Physical Chemistry Perspective. The Journal of Physical Chemistry A, 125(23), 4929-4942.
22. Kurland, C. G., Collins, L. J., & Penny, D. (2006). Genomics and the irreducible nature of eukaryote cells. Science, 312(5776), 1011-1014.
23. Forterre, P. (1992). Neutral terms. Nature, 355(6358), 305-305.
24. Forterre, P., & Gribaldo, S. (2010). Bacteria with a eukaryotic touch: a glimpse of ancient evolution? Proceedings of the National Academy of Sciences, 107(29), 12739-12740.
25. Spang, A., Saw, J. H., Jørgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., ... & Ettema, T. J. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature, 521(7551), 173-179.
26. Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Bäckström, D., Juzokaite, L., Vancaester, E., ... & Ettema, T. J. (2017.). Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature, 541(7637), 353-358.
27. Harish, A., & Morrison, D. (2020). The deep (er) roots of Eukaryotes and Akaryotes. [version 2; peer review: 2 approved, 1 approved with reservations] F1000Research 9:112.
28. Da Cunha, V., Gaïa, M., & Forterre, P. (2022). The expanding Asgard archaea and their elusive relationships with Eukarya. mLife, 1(1), 3-12.
29. Xiao, J., Fan, L., Wu, D., Xu, Y., Lai, D., Martin, W. F., ... & Zhang, C. (2019). Archaea, the tree of life, and cellular evolution in eukaryotes. Science China Earth Sciences, 62(3), 489-506.
30. Koonin, E. V., Makarova, K. S., & Wolf, Y. I. (2021). Evolution of microbial genomics: conceptual shifts over a quarter century. Trends in Microbiology, 29(7), 582-592.
31. Dagan, T., & Martin, W. (2006). The tree of one percent. Genome biology, 7(10), 1-7.
32. Doolittle, W. F., & Bapteste, E. (2007). Pattern pluralism and the Tree of Life hypothesis. Proceedings of the National Academy of Sciences, 104(7), 2043-2049.
33. Prakash, O., Jangid, K., & Shouche, Y. S. (2013). Carl Woese: from Biophysics to evolutionary microbiology. Indian J. Microbiol., 53(3), 247-252.
34. Doolittle, W. F. (1980). Revolutionary concepts in evolutionary cell biology. Trends in Biochemical Sciences, 5(6), 146-149.
35. Preiner, M., Asche, S., Becker, S., Betts, H. C., Boniface, A., Camprubi, E., ... & Xavier, J. C. (2020). The future of origin of life research: bridging decades-old divisions. Life, 10(3), 20.
36. Spitzer, J., Pielak, G. J., & Poolman, B. (2015). Emergence of life: Physical chemistry changes the paradigm. Biology direct, 10(1), 1-15.
37. Wallace, B. A. (20O7). Hidden dimensions: The unification of physics and consciousness (p.15) Columbia University Press, New York.
38. Popper, Karl R. 1968. The Logic of Scientific Discovery. London: Hutchinson & Co.
39. Olszewski, W., & Sandroni, A. (2011). Falsifiability. American Economic Review, 101(2), 788-818.
40. MacKay, D. G. (1992). Problems with Popper: The initial goal is to develop viable theories, not disconfirm them. Consciousness and Cognition, 1(3), 231-240.
41. Egel, R. (2012). Life's order, complexity, organization, and its thermodynamic–holistic imperatives. Life, 2(4), 323-363.
42. Lyon, P. (2011). To be or not to be: Where is self-preservation in evolutionary theory? In: Calcott, B., & Sterelny, K. (Eds.). (2011). The major transitions in evolution revisited. MIT Press, pp. 105-125.
43. Arnoldt, H., Strogatz, S. H., & Timme, M. (2015). Toward the Darwinian transition: switching between distributed and speciated states in a simple model of early life. Physical Review E, 92(5), 052909.
44. Deamer, D. (2016). Membranes and the origin of life: a century of conjecture. Journal of molecular evolution, 83(5), 159-168.
45. Peretó, J. (2005). Controversies on the origin of life. INTERNATIONAL MICROBIOLOGY, 8, 23-31.
46. Jowett, P., Rayne, R., & Tomas, S. (2017). Myth and fact in the origins of cellular life on Earth. Bioscience Horizons: The International Journal of Student Research, 10.
47. Walker, S. I. (2017). Origins of life: a problem for physics, a key issues review. Reports on Progress in Physics, 80(9), 092601.
48. Muchowska, K. B., Varma, S. J., & Moran, J. (2020). Nonenzymatic metabolic reactions and life's origins. Chemical Reviews, 120(15), 7708-7744.
49. Haldane, J. B. S. (1929). The origin of life. Rationalist Annual, 148, 3-10.

50. Oparin, A. I. (1938) The origin of life. Macmillan, New York.
51. Benner, S. A. (2014). Paradoxes in the origin of life. Origins of Life and Evolution of Biospheres, 44(4), 339-343.
52. Benner, S. A. (2018). Prebiotic plausibility and networks of paradox-resolving independent models. Nature communications, 9(1), 1-3.
53. Shapiro, R. (1984). The improbability of prebiotic nucleic acid synthesis. Origins of life, 14, 565-570.
54. Shapiro, R. (2000). A replicator was not involved in the origin of life. IUBMB life, 49(3), 173-176.
55. Wicken, J. S. (1985). An organismic critique of molecular Darwinism. Journal of theoretical biology, 117(4), 545-561.
56. Balázs, A. (2004). What does a molecule want? The myth of the self-replicating molecule (comments on the "selfish-gene" paradigm). Biosystems, 73(1), 1-11.
57. Lanier, K. A., & Williams, L. D. (2017). The origin of life: models and data. Journal of Molecular Evolution, 84(2-3), 85-92.
58. Wächtershäuser, G. (1988). Before enzymes and templates: theory of surface metabolism. Microbiological reviews, 52(4), 452-484.
59. Ma, S., & Dai, Y. (2011). Principal component analysis based methods in bioinformatics studies. Briefings in bioinformatics, 12(6), 714-722.
60. Rogers, K. L. (2021). Reimagining Origins of Life Research: Innovation and Synthesis via Experimentation, Instrumentation, and Data Analytics. Bull. Am. Astron. Soc., 43, 215.
61. Simons, M. (2021). Dreaming of a universal biology: synthetic biology and the origins of life. Hyle: International Journal for Philosophy of Chemistry, 27.
62. Segré, D., Ben-Eli, D., Deamer, D. W., & Lancet, D. (2001). The lipid world. Origins of Life and Evolution of the Biosphere, 31, 119-145.
63. Lancet, D., Segrè, D., & Kahana, A. (2019). Twenty years of "lipid world": a fertile partnership with David Deamer. Life, 9(4), 77.
64. Russell, M. J., & Martin, W. (2004). The rocky roots of the acetyl-CoA pathway. Trends in biochemical sciences, 29(7), 358-363.
65. Butch, C. J., Meringer, M., Gagnon, J. S., & Cleaves, H. J. (2021). Open questions in understanding life's origins. Communications Chemistry, 4(1), 1-4.
66. Egel, R. (2014). Origins and emergent evolution of life: The colloid microsphere hypothesis revisited. Origins of Life and Evolution of Biospheres, 44(2), 87-110.
67. Baum, D. A. (2015). Selection and the origin of cells. Bioscience, 65(7), 678-684.
68. Pollack, G. H. (2022). EZ Water and the Origin of Life. Annual Research & Review in Biology, 18-28.
69. Hwang, S. G., Hong, J. K., Sharma, A., Pollack, G. H., & Bahng, G. (2018). Exclusion zone and heterogeneous water structure at ambient temperature. PLoS One, 13(4), e0195057.
70. Kulkarni, P., & Uversky, V. N. (2018). Intrinsically disordered proteins: the dark horse of the dark proteome. Proteomics, 18(21-22), 1800061.
71. Vieregg, J. R., Lueckheide, M., Marciel, A. B., Leon, L., Bologna, A. J., Rivera, J. R., & Tirrell, M. V. (2018). Oligonucleotide–peptide complexes: phase control by hybridization. Journal of the American Chemical Society, 140(5), 1632-1638.
72. Ellis, R. J. (2001). Macromolecular crowding: obvious but underappreciated. Trends in biochemical sciences, 26(10), 597-604.
73. Park, S., Barnes, R., Lin, Y., Jeon, B. J., Najafi, S., Delaney, K. T., ... & Han, S. (2020). Dehydration entropy drives liquid-liquid phase separation by molecular crowding. Communications Chemistry, 3(1), 1-12.
74. Kandler, O. (1994). Cell wall biochemistry and three-domain concept of life. Systematic and applied microbiology, 16(4), 501-509.
75. Wächtershäuser, G. (2003). From pre-cells to Eukarya — a tale of two lipids. Molecular microbiology, 47(1), 13-22.
76. Ourisson, G., & Rohmer, M. (1992). Hopanoids. 2. Biohopanoids: a novel class of bacterial lipids. Accounts of Chemical Research, 25(9), 403-408.
77. Ourisson, G., & Nakatani, Y. (1994). The terpenoid theory of the origin of cellular life: the evolution of terpenoids to cholesterol. Chemistry & biology, 1(1), 11-23.
78. Egel, R. (2009). Peptide-dominated membranes preceding the genetic takeover by RNA: latest thinking on a classic controversy. Bioessays, 31(10), 1100-1109.
79. Frenkel-Pinter, M., Samanta, M., Ashkenasy, G., & Leman, L. J. (2020). Prebiotic peptides: Molecular hubs in the origin of life. Chemical Reviews, 120(11), 4707-4765.
80. Fiore, M., Chieffo, C., Lopez, A., Fayolle, D., ... & Buchet, R. (2022). Synthesis of Phospholipids Under Plausible Prebiotic Conditions and Analogies with Phospholipid Biochemistry for Origin of Life Studies. Astrobiology, 22(5), 598-627.
81. Brangwynne, C. P., Tompa, P., & Pappu, R. V. (2015). Polymer physics of intracellular phase transitions. Nature Physics, 11(11), 899-904.

82. Yoshizawa, T., Nozawa, R. S., Jia, T. Z., Saio, T., & Mori, E. (2020). Biological phase separation: cell biology meets biophysics. Biophysical Reviews, 12(2), 519-539.
83. Carter, C. W. (2015). What RNA world? Why a peptide–RNA partnership merits renewed experimental attention. Life, 5(1), 294-320.
84. Guzman, M. I., & Martin, S. T. (2009). Prebiotic metabolism: Production by mineral photoelectrochemistry of α-ketocarboxylic acids in the reductive tricarboxylic acid cycle. Astrobiology, 9(9), 833-842.
85. Mulkidjanian, A. Y. (2009). On the origin of life in the zinc world: 1. Photosynthesizing, porous edifices built of hydrothermally precipitated zinc sulfide as cradles of life on Earth. Biology Direct, 4(1), 1-39.
86. Green, N. J., Xu, J., & Sutherland, J. D. (2021). Illuminating life's origins: UV photochemistry in abiotic synthesis of biomolecules. Journal of the American Chemical Society, 143(19), 7219-7236.
87. Muchowska, K. B., Varma, S. J., Chevallot-Beroux, E., Lethuillier-Karl, L., Li, G., & Moran, J. (2017). Metals promote sequences of the reverse Krebs cycle. Nature ecology & evolution, 1(11), 1716-1721.
88. Martin, W. F., Nagies, F. S., & do Nascimento Vieira, A. (2021). To What Inanimate Matter Are We Most Closely Related and Does the Origin of Life Harbor Meaning? Philosophies, 6(2), 33.
89. Caetano-Anollés, G., Wang, M., & Caetano-Anollés, D. (2013). Structural phylogenomics retrodicts the origin of the genetic code and uncovers the evolutionary impact of protein flexibility. PLoS One, 8(8), e72225.
90. Baum, D. A., & Vetsigian, K. (2017). An experimental framework for generating evolvable chemical systems in the laboratory. Origins of Life and Evolution of Biospheres, 47(4), 481-497.
91. Vincent, L., Berg, M., Krismer, M., Saghafi, S. T., Cosby, J., Sankari, T., ... & Baum, D. A. (2019). Chemical ecosystem selection on mineral surfaces reveals long-term dynamics consistent with the spontaneous emergence of mutual catalysis. Life, 9(4), 80.
92. Szathmáry, E. (2015). Toward major evolutionary transitions theory 2.0. Proceedings of the National Academy of Sciences, 112(33), 10104-10111.
93. Grim, P., Wardach, S., & Beltrani, V. (2006). Location, location, location: The importance of spatialization in modeling cooperation and communication. Interaction Studies, 7(1), 43-78.
94. Ishida, T. (2022). Emergence Simulation of Biological Cell-like Shapes Satisfying the Conditions of Life Using a Lattice-Type Multiset Chemical Model. Life, 12(10), 1580.
95. Darwin, C. (1859). On the origin of species by means of natural selection. Murray, London.
96. Tessera, M. (2011). Origin of evolution versus origin of life: a shift of paradigm. International journal of molecular sciences, 12(6), 3445-3458.
97. Tessera, M. (2018). Is pre-Darwinian evolution plausible? Biology direct, 13(1), 1-18.
98. Pascal, R., & Pross, A. (2014). The nature and mathematical basis for material stability in the chemical and biological worlds. Journal of Systems Chemistry, 5(1), 1-8.
99. de Duve, C. (1987). Selection by differential molecular survival: a possible mechanism of early chemical evolution. Proceedings of the National Academy of Sciences, 84(23), 8253-8256.
100. Milner-White EJ, Russell MJ. 2011 Functional capabilities of the earliest peptides and the emergence of life. Genes 2, 671–688. (doi:10.3390/genes2040671)
101. Milner-White, E. J. (2019). Protein three-dimensional structures at the origin of life. Interface Focus, 9(6), 20190057. https://doi.org/10.1098/RSFS.2019.0057.
102. Suskiewicz, M. J., Sussman, J. L., Silman, I., & Shaul, Y. (2011). Context-dependent resistance to proteolysis of intrinsically disordered proteins. Protein Science, 20(8), 1285-1297.
103. Sibilska-Kaminski, I. K., & Yin, J. (2021). Toward Molecular Cooperation by De Novo Peptides. Origins of Life and Evolution of Biospheres, 51(1), 71-82.
104. Longo, L. M., Jabłońska, J., Vyas, P., Kanade, M., Kolodny, R., Ben-Tal, N., & Tawfik, D. S. (2020). On the emergence of P-Loop NTPase and Rossmann enzymes from a Beta-Alpha-Beta ancestral fragment. Elife, 9.
105. Raffaelli, N. (2011). Nicotinamide coenzyme synthesis: a case of ribonucleotide emergence or a byproduct of the RNA world? In Origins of life: the primal self-organization (pp. 185-208). Springer, Berlin, Heidelberg.
106. Egel, R. (2012). Primal eukaryogenesis: On the communal nature of precellular states, ancestral to modern life. Life, 2(1), 170-212.
107. Cavalier-Smith, T. (1975). The origin of nuclei and of eukaryotic cells. Nature, 256(5517), 463-468.
108. Cavalier-Smith, T. (2009). Predation and eukaryote cell origins: a coevolutionary perspective. The international journal of biochemistry & cell biology, 41(2), 307-322.
109. Cavalier-Smith, T. (2010). Origin of the cell nucleus, mitosis and sex: roles of intracellular coevolution. Biology direct, 5(1), 1-78.
110. Forterre, P. (2016). To be or not to be alive: How recent discoveries challenge the traditional definitions of viruses and life. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 59, 100-108.
111. Harris, H., & Hill, C. (2021). A place for viruses on the tree of life. Frontiers in Microbiology, 11, 3449.

112. Gray, M. W. (2014). The pre-endosymbiont hypothesis: a new perspective on the origin and evolution of mitochondria. Cold Spring Harbor Perspectives in Biology, 6(3), a016097.
113. Lake, J. A., & Rivera, M. C. (1994). Was the nucleus the first endosymbiont? Proceedings of the National Academy of Sciences, 91(8), 2880-2881.
114. Egel, R. (2017). 'Parabiotic Evolution': From Stochasticity in Geochemical and Subsequent Processes to Genes, Genomes and Modular Cells. Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2017 doi:10.20944/preprints201710.0153.v1.
115. Jeffares, D. C., Poole, A. M., & Penny, D. (1998). Relics from the RNA world. Journal of molecular evolution, 46(1), 18-36.
116. Poole, A., Jeffares, D., & Penny, D. (1999). Early evolution: prokaryotes, the new kids on the block. Bioessays, 21(10), 880-889.
117. Woese, C. R. (2000). Interpreting the universal phylogenetic tree. Proceedings of the National Academy of Sciences, 97(15), 8392-8396.
118. Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., ... & Banfield, J. F. (2016). A new view of the tree of life. Nature microbiology, 1(5), 1-6.
119. Woese, C. R., & Fox, G. E. (1977). The concept of cellular evolution. Journal of molecular evolution, 10(1), 1-6.
120. Yarus, M. (2002). Primordial genetics: phenotype of the ribocyte. Annual Review of Genetics, 36(1), 125-151.
121. Gilbert, W. (1986). Origin of life: The RNA world. nature, 319(6055), 618-618.
122. Carter, C. W., & Kraut, J. (1974). A proposed model for interaction of polypeptides with RNA. Proceedings of the National Academy of Sciences, 71(2), 283-287.
123. Schimmel, P, Giegé R, Moras D, Yokoyama S (1993) An operational RNA code for amino acids and possible relationship to genetic code. Proc Natl Acad Sci USA 90:8763–8768.
124. Rodin, S., Rodin, A., Ohno, S. (1996) The presence of codon-anticodon pairs in the acceptor stem of tRNAs. Proc Natl Acad Sci USA 93:4537–4542.
125. Di Giulio, M. (2004) The origin of the tRNA molecule: implications for the origin of protein synthesis. J Theor Biol 226:89–93.
126. Müller, F., Escobar, L., Xu, F. et al. (2022). A prebiotically plausible scenario of an RNA–peptide world. Nature 605, 279–284; https://doi.org/10.1038/s41586-022-04676-3.
127. Martínez Giménez, J. A., & Tabares Seisdedos, R. (2022). A Cofactor-Based Mechanism for the Origin of the Genetic Code. Origins of Life and Evolution of Biospheres, 1-15.
128. Francis, B. R. (2011). An alternative to the RNA world hypothesis. Trends in Evolutionary Biology, 3(1), e2-e2.
129. de Farias, S. T. D., & Prosdocimi, F. (2022). RNP-world: The ultimate essence of life is a ribonucleoprotein process. Genetics and Molecular Biology, 45.
130. Wagner, N., Pross, A., & Tannenbaum, E. (2010). Selection advantage of metabolic over non-metabolic replicators: a kinetic analysis. Biosystems, 99(2), 126-129.
131. Czárán, T., Könnyű, B., & Szathmáry, E. (2015). Metabolically coupled replicator systems: overview of an RNA-world model concept of prebiotic evolution on mineral surfaces. Journal of theoretical biology, 381, 39-54.
132. Caetano-Anollés, D., Caetano-Anollés, K., & Caetano-Anollés, G. (2018). Evolution of macromolecular structure: A 'double tale' of biological accretion and diversification. Science progress, 101(4), 360-383.
133. Wang, M., Jiang, Y. Y., Kim, K. M., Qu, G., Ji, H. F., Mittenthal, J. E., ... & Caetano-Anollés, G. (2011). A universal molecular clock of protein folds and its power in tracing the early history of aerobic metabo-lism and planet oxygenation. Molecular biology and evolution, 28(1), 567-582.
134. Zanotti, G., & Guerra, C. (2003). Is tensegrity a unifying concept of protein folds? FEBS letters, 534(1-3), 7-10.
135. Azulay, H., Lutaty, A., & Qvit, N. (2022). How Similar Are Proteins and Origami? Biomolecules, 12(5), 622.
136. Kurland, C. G., & Harish, A. (2015). The phylogenomics of protein structures: the backstory. Biochimie, 119, 284-302.
137. Harish, A. (2021, September 2). Peering into the Abyss with Evolutionary Telescopes: How protein structure-telescopes can look further back in time (and why sequence-telescopes cannot). https://doi.org/10.31219/osf.io/hyknj.
138. Yang, S., Doolittle, R. F., & Bourne, P. E. (2005). Phylogeny determined by protein domain content. Proceedings of the National Academy of Sciences, 102(2), 373-378.
139. Caetano-Anollés, G., Wang, M., Caetano-Anollés, D., & Mittenthal, J. E. (2009). The origin, evolution and structure of the protein world. Biochemical Journal, 417(3), 621-637.
140. Harish, A., Tunlid, A., & Kurland, C. G. (2013). Rooted phylogeny of the three superkingdoms. Biochimie, 95(8), 1593-1604.

141. Kim, K. M., Nasir, A., & Caetano-Anollés, G. (2014). The importance of using realistic evolutionary models for retrodicting proteomes. Biochimie, 99, 129-137.
142. Harish, A., & Kurland, C. G. (2017). Empirical genome evolution models root the tree of life. Biochimie, 138, 137-155.
143. Harish, A., & Kurland, C. G. (2017). Akaryotes and Eukaryotes are independent descendants of a uni-versal common ancestor. Biochimie, 138, 168-183.
144. Aziz, M. F., & Caetano-Anollés, G. (2021). Evolution of networks of protein domain organization. Scientific reports, 11(1), 1-18.
145. Caetano-Anollés, D., & Caetano-Anollés, G. (2021). Growth: Molecular accretion, growth and innovation. In Untangling Molecular Biodiversity: Explaining Unity and Diversity Principles of Organization with Molecular Structure and Evolutionary Genomics (pp. 347-393).
146. Nghe, P., Hordijk, W., Kauffman, S. A., Walker, S. I., Schmidt, F. J., Kemble, H., ... & Lehman, N. (2015). Prebiotic network evolution: six key parameters. Molecular BioSystems, 11(12), 3206-3217.
147. Hamley, I. W. (2010). Liquid crystal phase formation by biopolymers. Soft Matter, 6(9), 1863-1871.
148. Wilson, R. M. (2017). Biological tissue can behave like a liquid crystal. Physics Today, 70(6), 19-21.
149. Woese, C. R., Olsen, G. J., Ibba, M., & Söll, D. (2000). Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. Microbiology and molecular biology reviews, 64(1), 202-236.
150. Safro, M. G., & Moor, N. A. (2009). Codases: 50 years after. Molecular biology, 43(2), 211-222.
151. Carter, C. W. (2014). Urzymology: experimental access to a key transition in the appearance of enzymes. Journal of Biological Chemistry, 289(44), 30213-30220.
152. Engelhardt, W. A., Kisselev, L. L., & Nezlin, R. S. (1970). Das "Prinzip des fixierten Partners' als Werkzeug auf dem Gebiet der Molekularbiologie /The Principle of "Fixed Partner" as Experimental Tool in Molecular Biology Research. Monatshefte für Chemie /Chemical Monthly, 101(5), 1510-1517.
153. Rodin, S. N., & Ohno, S. (1995). Two types of aminoacyl-tRNA synthetases could be originally encoded by complementary strands of the same nucleic acid. Origins of Life and Evolution of the Biosphere, 25(6), 565-589.
154. Carter, C. W., Li, L., Weinreb, V., Collier, M., Gonzalez-Rivera, K., Jimenez-Rodriguez, M., ... & Chandrasekharan, S. N. (2014). The Rodin-Ohno hypothesis that two enzyme superfamilies descended from one ancestral gene: an unlikely scenario for the origins of translation that will not be dismissed. Biology Direct, 9(1), 1-23.
155. de Pouplana, L. R., & Schimmel, P. (2001). Two classes of tRNA synthetases suggested by sterically compatible dockings on tRNA acceptor stem. Cell, 104(2), 191-193.
156. Carter Jr, C. W., & Wills, P. R. (2018). Hierarchical groove discrimination by Class I and II amino-acyl-tRNA synthetases reveals a palimpsest of the operational RNA code in the tRNA acceptor-stem bases. Nucleic acids research, 46(18), 9667-9683.
157. Carter Jr, C. W., & Wills, P. R. (2019). Class I and II aminoacyl-tRNA synthetase tRNA groove discrimination created the first synthetase–tRNA cognate pairs and was therefore essential to the origin of genetic coding. IUBMB Life, 71(8), 1088-1098.
158. Maizels, N., & Weiner, A. M. (1994). Phylogeny from function: evidence from the molecular fossil record that tRNA originated in replication, not translation. Proceedings of the National Academy of Sciences, 91(15), 6729-6734.
159. Chen, Q., Zhang, X., Shi, J., Yan, M., & Zhou, T. (2021). Origins and evolving functionalities of tRNA-derived small RNAs. Trends in biochemical sciences, 46(10), 790-804.
160. Murakami, H., Saito, H., & Suga, H. (2003). A versatile tRNA aminoacylation catalyst based on RNA. Chemistry & biology, 10(7), 655-662.
161. Ishida, S., Terasaka, N., Katoh, T., & Suga, H. (2020). An aminoacylation ribozyme evolved from a natural tRNA-sensing T-box riboswitch. Nature Chemical Biology, 16(6), 702-709.
162. Woese, C. R. (1971). Evolution of macromolecular complexity. Journal of Theoretical Biology, 33(1), 29-34.
163. Woese, C. R. (2004). A new biology for a new century. Microbiology and molecular biology reviews, 68(2), 173-186.
164. Caetano-Anollés, D., Kim, K. M., Mittenthal, J. E., & Caetano-Anollés, G. (2011). Proteome evolution and the metabolic origins of translation and cellular life. Journal of molecular evolution, 72, 14-33.
165. Caetano-Anollés, G., Kim, K. M., & Caetano-Anollés, D. (2012). The phylogenomic roots of modern biochemistry: origins of proteins, cofactors and protein biosynthesis. Journal of Molecular Evolution, 74, 1-34.
166. Bowman, J. C., Petrov, A. S., Frenkel-Pinter, M., Penev, P. I., & Williams, L. D. (2020). Root of the tree: the significance, evolution, and origins of the ribosome. Chemical reviews, 120(11), 4848-4878.
167. Weinger, J. S., Parnell, K. M., Dorner, S., Green, R., & Strobel, S. A. (2004). Substrate-assisted catalysis of peptide bond formation by the ribosome. Nature structural & molecular biology, 11(11), 1101-1106.
168. Lütcke, H. (1995). Signal recognition particle (SRP), a ubiquitous initiator of protein translocation. European journal of biochemistry, 228(3), 531-550.

169. Caetano-Anollés, G., & Seufferheld, M. J. (2013). The coevolutionary roots of biochemistry and cellular organization challenge the RNA world paradigm. Microbial Physiology, 23(1-2), 152-177.
170. Bose, T., Fridkin, G., Davidovich, C., Krupkin, M., Dinger, N., Falkovich, A. H., ... & Yonath, A. (2022). Origin of life: protoribosome forms peptide bonds and links RNA and protein dominated worlds. Nu-cleic acids research, 50(4), 1815-1828.
171. Baram, D., & Yonath, A. (2005). From peptide-bond formation to cotranslational folding: dynamic, regulatory and evolutionary aspects. FEBS letters, 579(4), 948-954.
172. Vetsigian, K., Woese, C., & Goldenfeld, N. (2006). Collective evolution and the genetic code. Proceedings of the National Academy of Sciences, 103(28), 10696-10701.
173. Reanney, D. C. (1974). On the origin of prokaryotes. Journal of Theoretical Biology, 48(1), 243-251.
174. Mariscal, C., & Doolittle, W. F. (2015). Eukaryotes first: how could that be? Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1678), 20140322.
175. Szathmáry, E. (1989). The integration of the earliest genetic information. Trends in Ecology & Evolution, 4(7), 200-204.
176. Koonin, E. V. (2009). On the origin of cells and viruses: primordial virus world scenario. Annals of the New York Academy of Sciences, 1178(1), 47-64.
177. Forterre, P. (2022). Carl Woese: Still ahead of our time. mLife, 1(4), 359-367.
178. Vetsigian, K., & Goldenfeld, N. (2009). Genome rhetoric and the emergence of compositional bias. Proceedings of the National Academy of Sciences, 106(1), 215-220.
179. Doolittle, W. F. (2020). Evolution: two domains of life or three? Current Biology, 30(4), R177-R179.
180. Butler, T., Goldenfeld, N., Mathew, D., & Luthey-Schulten, Z. (2009). Extreme genetic code optimality from a molecular dynamics calculation of amino acid polar requirement. Physical Review E, 79(6), 060901.
181. Aggarwal, N., Bandhu, A. V., & Sengupta, S. (2016). Finite population analysis of the effect of horizontal gene transfer on the origin of an universal and optimal genetic code. Physical biology, 13(3), 036007.
182. Goldenfeld, N., Biancalani, T., & Jafarpour, F. (2017). Universal biology and the statistical mechanics of early life. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 375(2109), 20160341.
183. Froese, T., Campos, J. I., Fujishima, K., Kiga, D., & Virgo, N. (2018). Horizontal transfer of code fragments between protocells can explain the origins of the genetic code without vertical descent. Scientific re-ports, 8(1), 1-10.
184. Carter, C. W., & Wills, P. R. (2018). Interdependence, reflexivity, fidelity, impedance matching, and the evolution of genetic coding. Molecular biology and evolution, 35(2), 269-286.
185. Wills, P. R., & Carter Jr, C. W. (2018). Insuperable problems of the genetic code initially emerging in an RNA world. Biosystems, 164, 155-166.
186. Cemin, S. C., & Smolin, L. (1997). Coevolution of membranes and channels: A possible step in the origin of life. arXiv preprint adap-org/9709004.
187. Lingwood, D., & Simons, K. (2010). Lipid rafts as a membrane-organizing principle. science, 327(5961), 46-50.
188. Woese, C.R. (2005) Evolving biological organization. In: J. Sapp, ed. Microbial Phylogeny and Evolution: Concepts and Controversies. Oxford University Press, pp. 99–117.
189. Egel, R. (2021). Origins: It takes a well rooted trunk to bear a tree — reductionist models of divergent evolution and further aspects of reality. In Untangling Molecular Biodiversity: Explaining Unity and Diversity Principles of Organization with Molecular Structure and Evolutionary Genomics (pp. 223-282). World Scientific, Singapore.
190. Mayr, E. (2000). The biological species concept. Species concepts and phylogenetic theory: a debate. Columbia University Press, New York, 17-29.
191. Doolittle, W. F. (2018). Speciation without species: a final word. Philosophy, Theory, and Practice in Biology, 11(014).
192. Ragan, M. A., McInerney, J. O., & Lake, J. A. (2009). The network of life: genome beginnings and evolution. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1527), 2169-2175.
193. Mallet, J., Besansky, N., & Hahn, M. W. (2016). How reticulated are species? BioEssays, 38(2), 140-149
194. Atkinson, H. J., Morris, J. H., Ferrin, T. E., & Babbitt, P. C. (2009). Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. PloS one, 4(2), e4345.
195. Copp, J. N., Akiva, E., Babbitt, P. C., & Tokuriki, N. (2018). Revealing unexplored sequence-function space using sequence similarity networks. Biochemistry, 57(31), 4651-4662.
196. Hanawalt, P. C. (2007). Paradigms for the three Rs: DNA replication, recombination, and repair. Molecular cell, 28(5), 702-707.
197. Lisby, M., & Mortensen, U. H. (2017). 3Rs tightly intertwined to maintain genome stability. FEMS Yeast Research, 17(1).
198. Kuzminov, A. (1999). Recombinational repair of DNA damage in Escherichia coli and bacteriophage λ. Microbiology and molecular biology reviews, 63(4), 751-813.

199. Weterings, E., & Chen, D. J. (2008). The endless tale of non-homologous end-joining. Cell research, 18(1), 114-124.
200. Dagan, T., Artzy-Randrup, Y., & Martin, W. (2008). Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. Proceedings of the National Academy of Sciences, 105(29), 10039-10044.
201. Dagan, T., Roettger, M., Bryant, D., & Martin, W. (2010). Genome networks root the tree of life between prokaryotic domains. Genome Biology and Evolution, 2, 379-392.
202. Bapteste, E., O'Malley, M. A., Beiko R. G., Ereshefsky, M., ... & Martin, W. (2009). Prokaryotic evolution and the tree of life are two different things. Biology direct, 4(1), 1-20.
203. Thiergart, T., Landan, G., Schenk, M., Dagan, T., & Martin, W. F. (2012). An evolutionary network of genes present in the eukaryote common ancestor polls genomes on eukaryotic and mitochondrial origin. Genome Biology and Evolution, 4(4), 466-485.
204. Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. Molecular biology and evolution, 23(2), 254-267.
205. Morrison, D. A. (2010). Using data-display networks for exploratory data analysis in phylogenetic studies. Molecular Biology and Evolution, 27(5), 1044-1057.
206. Harish, A. (2018). What is an archaeon and are the Archaea really unique? Peer J, 6, e5770.
207. Di Giulio, M. (2018). On Earth, there would be a number of fundamental kinds of primary cells — cellular domains — greater than or equal to four. Journal of Theoretical Biology, 443, 10-17.
208. Woese, C. R. (2002). On the evolution of cells. Proceedings of the National Academy of Sciences, 99(13), 8742-8747.
209. Gouy, R., Baurain, D., & Philippe, H. (2015). Rooting the tree of life: the phylogenetic jury is still out. Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1678), 20140329.
210. Spence, M. A., Mortimer, M. D., Buckle, A. M., Minh, B. Q., & Jackson, C. J. (2021). A comprehensive phylogenetic analysis of the serpin superfamily. Molecular Biology and Evolution, 38(7), 2915-2929.
211. Burnim, A. A., Spence, M. A., Xu, D., Jackson, C. J., & Ando, N. (2022). Comprehensive phylogenetic analysis of the ribonucleotide reductase family reveals an ancestral clade. Elife, 11, e79790.
212. Himmel, N. J., Moi, D., & Benton, R. (2023). Remote homolog detection places insect chemoreceptors in a cryptic protein superfamily spanning the tree of life. Current Biology, 33(22), 5023-5033.
213. Koonin, E. V. (2009). Darwinian evolution in the light of genomics. Nucleic acids research, 37(4), 1011-1034
214. Moreira, D., & López-García, P. (2009). Ten reasons to exclude viruses from the tree of life. Nature Reviews Microbiology, 7(4), 306-311.
215. Hegde, N. R., Maddur, M. S., Kaveri, S. V., & Bayry, J. (2009). Reasons to include viruses in the tree of life. Nature Reviews Microbiology, 7(8), 615-615.
216. Doolittle, W. F., & Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. Nature, 284(5757), 601-603.
217. Brown, N., & Bhella, D. (2016). Are viruses alive? Microbiol. Today, 43, 58-61.
218. Van Regenmortel, M. H. (2016). The metaphor that viruses are living is alive and well, but it is no more than a metaphor. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 59, 117-124.
219. Yolles, M., & Frieden, R. (2022). Viruses as Living Systems — A Metacybernetic View. Systems, 10(3), 70.
220. Forterre, P. (2006). The origin of viruses and their possible roles in major evolutionary transitions. Virus research, 117(1), 5-16.
221. Forterre, P. (2010). Giant viruses: conflicts in revisiting the virus concept. Intervirology, 53(5), 362-378.
222. Forterre, P. (2013). The virocell concept and environmental microbiology. The ISME journal, 7(2), 233-236.
223. Forterre, P. (2017). The Origin, Nature and Definition of Viruses (and Life): New Concepts and Controversies. In: Evolutionary Biology of the Virome, and Impacts in Human Health and Disease;
224. Villarreal, L. P. (2009). Persistence pays: how viruses promote host group survival. Current opinion in microbiology, 12(4), 467-472.
225. Nicolau, M., Picault, N., & Moissiard, G. (2021). The evolutionary volte-face of transposable elements: From harmful jumping genes to major drivers of genetic innovation. Cells, 10(11), 2952.
226. Leipe, D. D., Aravind, L., & Koonin, E. V. (1999). Did DNA replication evolve twice independently? Nucleic acids research, 27(17), 3389-3401.
227. Forterre, P. (2006). Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. Proceedings of the National Academy of Sci-ences, 103(10), 3669-3674.
228. Koonin, E. V., Krupovič, M., Ishino, S., & Ishino, Y. (2020). The replication machinery of LUCA: common origin of DNA replication and transcription. BMC biology, 18(1), 1-8.
229. Krupovič, M., Gribaldo, S., Bamford, D. H., & Forterre, P. (2010). The evolutionary history of archaeal MCM helicases: a case study of vertical evolution combined with hitchhiking of mobile genetic elements. Molecular Biology and Evolution, 27(12), 2716-2732.

230. Martin, W. F., Garg, S., & Zimorski, V. (2015). Endosymbiotic theories for eukaryote origin. Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1678), 20140330.
231. López-García, P., & Moreira, D. (2015). Open questions on the origin of eukaryotes. Trends in ecology & evolution, 30(11), 697-708.
232. Bell, P. J. (2020). Evidence supporting a viral origin of the eukaryotic nucleus. Virus Research, 289, 198168.
233. Forterre, P., & Raoult, D. (2017). The transformation of a bacterium into a nucleated virocell reminds the viral eukaryogenesis hypothesis. Virologie, 21(4), 28-30.
234. Takemura, M. (2020). Medusavirus ancestor in a proto-eukaryotic cell: updating the hypothesis for the viral origin of the nucleus. Frontiers in Microbiology, 2169.
235. Liu, Y., & Krupovič, M. (2022). Genome chromatinization in giant double-stranded DNA viruses. Trends in Biochemical Sciences, 47(1), 3-5.
236. Nasir, A., & Caetano-Anollés, G. (2015). A phylogenomic data-driven exploration of viral origins and evolution. Science advances, 1(8), e1500527.
237. Livingstone Bell, P. J. (2001). Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? Journal of Molecular Evolution, 53, 251-256.
238. Rao, V. B., & Feiss, M. (2008). The bacteriophage DNA packaging motor. Annual review of genetics, 42, 647-681.
239. Chiang, Y. N., Penadés, J. R., & Chen, J. (2019). Genetic transduction by phages and chromosomal islands: The new and noncanonical. PLoS pathogens, 15(8), e1007878.
240. Ibarra-Chávez, R., Brady, A., Chen, J., Penadés, J. R., & Haag, A. F. (2022). Phage-inducible chromosomal islands promote genetic variability by blocking phage reproduction and protecting transductants from phage lysis. PLoS pathogens, 15(8), e1007878.
241. Abrescia, N. G., Bamford, D. H., Grimes, J. M., & Stuart, D. I. (2012). Structure unifies the viral universe. Annual review of biochemistry, 81, 795-822.
242. Mughal, F., Nasir, A., & Caetano-Anollés, G. (2020). The origin and evolution of viruses inferred from fold family structure. Archives of virology, 165(10), 2177-2191.
243. Nasir, A., & Caetano-Anollés, G. (2021). Panspermia: An early cellular origin of viruses. In Untangling Molecular Biodiversity: Explaining Unity and Diversity Principles of Organization with Molecular Structure and Evolutionary Genomics (pp. 421-452). World Scientific, Singapore.
244. Harish, A., Abroi, A., Gough, J., & Kurland, C. (2016). Did viruses evolve as a distinct supergroup from common ancestors of cells? Genome biology and evolution, 8(8), 2474-2481.
245. Nasir, A., Kim, K. M., & Caetano-Anollés, G. (2017). Phylogenetic tracings of proteome size support the gradual accretion of protein structural domains and the early origin of viruses from primordial cells. Frontiers in microbiology, 8, 1178.
246. Koonin, E. V. (2014). The origins of cellular life. Antonie Van Leeuwenhoek, 106, 27-41.
247. Kazlauskas, D., Varsani, A., Koonin, E. V., & Krupovič, M. (2019). Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. Nature communications, 10(1), 3425.
248. Ebbesen, K. K., Hansen, T. B., & Kjems, J. (2017). Insights into circular RNA biology. RNA biology, 14(8), 1035-1045.
249. Forterre, P. (1995). Thermoreduction, a hypothesis for the origin of prokaryotes. Comptes rendus de l'Academie des sciences. Serie III, Sciences de la vie, 318(4), 415-422.
250. Forterre, P. (2013). The common ancestor of archaea and eukarya was not an archaeon. Archaea, 2013.
251. Andersson, S. G., & Kurland, C. G. (1998). Reductive evolution of resident genomes. Trends in microbiology, 6(7), 263-268.
252. Wächtershäuser, G. (2006). From volcanic origins of chemoautotrophic life to Bacteria, Archaea and Eukarya. Phil. Trans. R. Soc. B, 361, 1787-1808.
253. Eme, L., Tamarit, D., Caceres, E. F., Stairs, C. W., De Anda, V., Schön, M. E., ... & Ettema, T. J. (2023). Inference and reconstruction of the heimdallarchaeial ancestry of eukaryotes. Nature, 1-8.
254. Boyer, M., Yutin, N., Pagnier, I., Barrassi, L., ... & Raoult, D. (2009). Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. Proceedings of the National Academy of Sciences, 106(51), 21848-21853.
255. Thomas, V., & Greub, G. (2010). Amoeba/amoebal symbiont genetic transfers: lessons from giant virus neighbours. Intervirology, 53(5), 254-267.
256. Maumus, F.; Blanc, G. Study of gene trafficking between *Acanthamoeba* and giant viruses suggests an undiscovered family of amoeba-infecting viruses. *Genome Biol. Evol.* **2016**, *8*, 3351–3363. [**Google Scholar**] [**CrossRef**] [**PubMed**]
257. Travis, J. L., Welnhofer, E. A., & Orokos, D. D. (2002). Autonomous reorganization of foraminiferan reticulopodia. Journal of Foraminiferal Research, 32(4), 425-433.

258. Guttes, S., & Guttes, E. (1968). Regulation of DNA replication in the nuclei of the slime mold Physarum polycephalum: Transplantation of nuclei by plasmodial coalescence. The Journal of Cell Biology, 37(3), 761-772.
259. Raff, R. A., & Mahler, H. R. (1972). The Nonsymbiotic Origin of Mitochondria: The question of the origin of the eucaryotic cell and its organelles is reexamined. Science, 177(4049), 575-582..
260. Reijnders, L. (1975). The origin of mitochondria. Journal of molecular evolution, 5(3), 167-176.
261. Martin, W. F., & Müller, M. (Eds.). (2007). Origin of mitochondria and hydrogenosomes. New York, NY, USA: Springer.
262. Lane, N., & Martin, W. (2010). The energetics of genome complexity. Nature, 467(7318), 929-934.
263. Gray, M. W. (2012). Mitochondrial evolution. Cold Spring Harbor perspectives in biology, 4(9), a011403
264. Gray, M. W. (2015). Mosaic nature of the mitochondrial proteome: Implications for the origin and evolution of mitochondria. Proceedings of the National Academy of Sciences, 112(33), 10133-10138.
265. Gabaldón, T. (2018). Relative timing of mitochondrial endosymbiosis and the "pre-mitochondrial symbioses" hypothesis. IUBMB life, 70(12), 1188-1196.
266. Speijer, D. (2020). Debating Eukaryogenesis — Part 1: Does Eukaryogenesis Presuppose Symbiosis Before Uptake? BioEssays, 42(4), 1900157.
267. Speijer, D. (2020). Debating Eukaryogenesis — Part 2: How Anachronistic Reasoning Can Lure Us into Inventing Intermediates. BioEssays, 42(5), 1900153.
268. Eigen, M., & Schuster, P. (1977). A principle of natural self-organization: Part A: Emergence of the hypercycle. Naturwissenschaften, 64, 541-565.
269. Eigen, M., McCaskill, J., & Schuster, P. (1988). Molecular quasi-species. The Journal of Physical Chemistry, 92(24), 6881-6891.
270. Kazlauskas, D., Krupovič, M., Guglielmini, J., Forterre, P., & Venclovas, Č. (2020). Diversity and evolution of B-family DNA polymerases. Nucleic acids research, 48(18), 10142-10156.
271. Di Giulio, M. (2021). The late appearance of DNA, the nature of the LUCA and ancestors of the domains of life. Biosystems, 202, 104330.
272. Branscomb, E., & Russell, M. J. (2013). Turnstiles and bifurcators: the disequilibrium converting engines that put metabolism on the road. Biochimica et Biophysica Acta (BBA)-Bioenergetics, 1827(2), 62-78.
273. Eberhard, W. G. (1990). Evolution in bacterial plasmids and levels of selection. The Quarterly review of biology, 65(1), 3-22.
274. Krupovič, M., & Koonin, E. V. (2016). Self-synthesizing transposons: unexpected key players in the evolution of viruses and defense systems. Current opinion in microbiology, 31, 25-33.
275. Kim, J. W., Bugata, V., Cortés-Cortés, G., Quevedo-Martínez, G., & Camps, M. (2020). Mechanisms of Theta plasmid replication in Enterobacteria and implications for adaptation to its host. EcoSal Plus, 9(1).
276. Espinosa, E., Barre, F. X., & Galli, E. (2017). Coordination between replication, segregation and cell division in multi-chromosomal bacteria: lessons from Vibrio cholerae. Int. Microbiol, 20, 121-129.
277. Forterre, P., & Krupovič, M. (2012). The origin of virions and virocells: the escape hypothesis revisited. Viruses: essential agents of life, 43-60.
278. Wolf, Y. I., Kazlauskas, D., Iranzo, J., Lucía-Sanz, A., Kuhn, J. H., Krupovič, M., ... & Koonin, E. V. (2018). Origins and evolution of the global RNA virome. MBio, 9(6), e02329-18.
279. Trakselis, M.A., Murakami, K.S. (2014). Introduction to Nucleic Acid Polymerases: Families, Themes, and Mechanisms. In: Murakami, K., Trakselis, M. (eds) Nucleic Acid Polymerases. Nucleic Acids and Molecular Biology, vol 30. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-39796-7_1.
280. Ahmad, M., Xu, D., & Wang, W. (2017). Type IA topoisomerases can be "magicians" for both DNA and RNA in all domains of life. RNA biology, 14(7), 854-864.
281. Forterre, Patrick; Da Cunha, Violette; and Gaïa, Morgan (June 2018) Universal Tree of Life. In: eLS. John Wiley & Sons, Ltd: Chichester. DOI: 10.1002/9780470015902.a0001525.pub3.
282. Forterre, P. (2002). The origin of DNA genomes and DNA replication proteins. Current opinion in microbiology, 5(5), 525-532.
283. de Farias, S. T., Furtado, A. N. M., dos Santos Junior, A. P., & José, M. V. (2023). Natural History of DNA-Dependent DNA Polymerases: Multiple Pathways to the Origins of DNA. Viruses, 15(3), 749.
284. Hourcade, D., Dressler, D., & Wolfson, J. (1973). The amplification of ribosomal RNA genes involves a rolling circle intermediate. Proceedings of the National Academy of Sciences, 70(10), 2926-2930.
285. Maniloff, J., & Ackermann, H. W. (1998). Taxonomy of bacterial viruses: establishment of tailed virus genera and the other Caudovirales. Archives of virology, 143, 2051-2063.
286. Oldenburg, D. J., & Bendich, A. J. (2001). Mitochondrial DNA from the liverwort Marchantia polymorpha: circularly permuted linear molecules, head-to-tail concatemers, and a 5′ protein. Journal of molecular biology, 310(3), 549-562.
287. Piano, A. L., Martínez-Jiménez, M. I., Zecchi, L., & Ayora, S. (2011). Recombination-dependent concatemeric viral DNA replication. Virus research, 160(1-2), 1-14.

288. Sarfallah, A., Zamudio-Ochoa, A., Anikin, M., & Temiakov, D. (2021). Mechanism of transcription initiation and primer generation at the mitochondrial replication origin OriL. The EMBO Journal, 40(19), e107988.
289. Shutt, T. E., & Gray, M. W. (2006). Bacteriophage origins of mitochondrial replication and transcription proteins. Trends in Genetics, 22(2), 90-95.
290. Boeckman, J., Korn, A., Yao, G., Ravindran, A., Gonzalez, C., & Gill, J. (2022). Sheep in wolves' clothing: Temperate T7-like bacteriophages and the origins of the Autographiviridae. Virology, 568, 86-100.
291. Yao, N. Y., & O'Donnell, M. (2010). SnapShot: the replisome. Cell, 141(6), 1088-1088.
292. Claussin, C., Vazquez, J., & Whitehouse, I. (2022). Single-molecule mapping of replisome progression. Molecular Cell, 82(7), 1372-1382.
293. Yao, N. Y., & O'Donnell, M. E. (2016). Evolution of replication machines. Critical reviews in biochemistry and molecular biology, 51(3), 135-149.
294. McInerney, P., Johnson, A., Katz, F., & O'Donnell, M. (2007). Characterization of a triple DNA polymerase replisome. Molecular cell, 27(4), 527-538.
295. Yao, N. Y., & O'Donnell, M. E. (2021). The DNA replication machine: structure and dynamic function. Macromolecular Protein Complexes III: Structure and Function, 233-258.
296. Mulye, M., Singh, M. I., & Jain, V. (2022). From Processivity to Genome Maintenance: The Many Roles of Sliding Clamps. Genes, 13(11), 2058.
297. Weigel, C., & Seitz, H. (2006). Bacteriophage replication modules. FEMS microbiology reviews, 30(3), 321-381.
298. Kazlauskas, D., Krupovič, M., & Venclovas, Č. (2016). The logic of DNA replication in double-stranded DNA viruses: insights from global analysis of viral genomes. Nucleic acids research, 44(10), 4551-4564.
299. Koonin, E. V., Dolja, V. V., & Krupovič, M. (2022). The logic of virus evolution. Cell Host & Microbe, 30(7), 917-929.
300. Carlile, M. (1982). Prokaryotes and eukaryotes: strategies and successes. Trends in Biochemical Sciences, 7(4), 128-130.
301. Pianka, E. R. (1970). On r-and K-selection. The American naturalist, 104(940), 592-597.
302. Stenseth, N. C., & Ugland, K. I. (1985). On the evolution of demographic strategies in populations with equilibrium and cyclic densities. Mathematical biosciences, 74(1), 89-109.
303. Reanney, D. C. (1987, January). Genetic error and genome design. In Cold Spring Harbor symposia on quantitative biology (Vol. 52, pp. 751-757). Cold Spring Harbor Laboratory Press.
304. Gray, M. W., Lukeš, J., Archibald, J. M., Keeling, P. J., & Doolittle, W. F. (2010). Irremediable complexity? Science, 330(6006), 920-921.
305. Yu, R. Y., & Martin, W. F. (2016). Symbiotic Associations: All About Chemistry. The Mechanistic Benefits of Microbial Symbionts, 3-11.
306. McInerney, M. J., Sieber, J. R., & Gunsalus, R. P. (2009). Syntrophy in anaerobic global carbon cycles. Current opinion in biotechnology, 20(6), 623-632.
307. Nozhevnikova, A. N., Russkova, Y. I., Litti, Y. V., Parshina, S. N., Zhuravleva, E. A., & Nikitina, A. A. (2020). Syntrophy and interspecies electron transfer in methanogenic microbial communities. Micro-biology, 89, 129-147.
308. Margulis, L. (1971). Symbiosis and evolution. Scientific American, 225(2), 48-61.
309. Gray, M. W., Burger, G., & Lang, B. F. (2001). The origin and early evolution of mitochondria. Genome biology, 2, 1-5.
310. Keeling, P. J. (2010). The endosymbiotic origin, diversification and fate of plastids. Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1541), 729-748.
311. Filée, J., & Forterre, P. (2005). Viral proteins functioning in organelles: a cryptic origin? Trends in microbiology, 13(11), 510-513.
312. Koonin, E. V. (2016). Viruses and mobile elements as drivers of evolutionary transitions. Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1701), 20150442.
313. Baum, B., & Baum, D. A. (2020). The merger that made us. BMC biology, 18(1), 1-4.
314. Gray, M. W., & Doolittle, W. F. (1982). Has the endosymbiont hypothesis been proven? Microbiological Reviews, 46(1), 1-42.
315. Braymer, J. J., & Lill, R. (2017). Iron–sulfur cluster biogenesis and trafficking in mitochondria. Journal of Biological Chemistry, 292(31), 12754-12763.
316. Tachezy, J., & Doležal, P. (2007). Iron–sulfur proteins and iron–sulfur cluster assembly in organisms with hydrogenosomes and mitosomes. In: Martin W, Müller M (eds) Origin of mitochondria and hydrogenosomes (pp. 105-133). Berlin, Heidelberg: Springer Berlin Heidelberg.
317. Nelson-Sathi, S., Dagan, T., Landan, G., Janssen, A., Steel, M., McInerney, J. O., ... & Martin, W. F. (2012). Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea. Proceedings of the National Academy of Sciences, 109(50), 20537-20542.

318. Méheust, R., Watson, A. K., Lapointe, F. J., Papke, R. T., Lopez, P., & Bapteste, E. (2018). Hundreds of novel composite genes and chimeric genes with bacterial origins contributed to haloarchaeal evolution. Ge-nome Biology, 19(1), 1-12.
319. Gray, M. W. (2017). Lynn Margulis and the endosymbiont hypothesis: 50 years later. Molecular biology of the cell, 28(10), 1285-1287.
320. Lazcano, A., & Peretó, J. (2021). Prokaryotic symbiotic consortia and the origin of nucleated cells: A critical review of Lynn Margulis hypothesis. Biosystems, 204, 104408.
321. Liu, Y., Makarova, K. S., Huang, W. C., Wolf, Y. I., Nikolskaya, A. N., Zhang, X., ... & Li, M. (2021). Expanded diversity of Asgard archaea and their relationships with eukaryotes. Nature, 593(7860), 553-557.
322. Di Giulio, M. (2021). Errors of the ancestral translation, LUCA, and nature of its direct descendants. Biosystems, 206, 104433.
323. Di Giulio, M. (2021). The phylogenetic distribution of the cell division system would not imply a cellular LUCA but a progenotic LUCA. Biosystems, 210, 104563.
324. Di Giulio, M. (2022). The RNase P, LUCA, the ancestors of the life domains, the progenote, and the tree of life. Biosystems, 212, 104604.
325. Di Giulio, M. (2022). The origins of the cell membrane, the progenote, and the universal ancestor (LUCA). Biosystems, 104799.
326. Di Giulio, M. (2020). LUCA as well as the ancestors of archaea, bacteria and eukaryotes were progenotes: Inference from the distribution and diversity of the reading mechanism of the AUA and AUG codons in the domains of life. Biosystems, 198, 104239.
327. Di Giulio, M. (2020). Common ancestry of eukaryotes and Asgardarchaeota: Three, two or more cellular domains of life? Journal of Theoretical Biology, 486, 110083.
328. Prosdocimi, F., & de Farias, S. T. (2023). Origin of life: Drawing the big picture. Progress in Biophysics and Molecular Biology, Volumes 180–181, July–August 2023, Pages 28-36.
329. Staley, J. T. (2017). Domain Cell Theory supports the independent evolution of the Eukarya, Bacteria and Archaea and the Nuclear Compartment Commonality hypothesis. Open Biology, 7(6), 170041.
330. Da Cunha, V., Gaïa, M., Gadelle, D., Nasir, A., & Forterre, P. (2017). Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. PLoS genetics, 13(6), e1006810.
331. Garg, S. G., Kapust, N., Lin, W., Knopp, M., Tria, F. D., Nelson-Sathi, S., ... & Martin, W. F. (2021). Anomalous phylogenetic behavior of ribosomal proteins in metagenome-assembled Asgard archaea. Genome biology and evolution, 13(1), evaa238.
332. Forterre, P. (2015). The universal tree of life: an update. Frontiers in microbiology, 6, 717.
333. van der Gulik, P. T. S., Hoff, W. D., & Speijer, D. (2017). In defence of the three-domains of life paradigm. BMC evolutionary biology, 17, 1-5.
334. Poole, A., & Penny, D. (2007). Engulfed by speculation. Nature, 447(7147), 913-913.
335. Glansdorff, N., Xu, Y., & Labedan, B. (2008). The last universal common ancestor: emergence, constitution and genetic legacy of an elusive forerunner. Biology direct, 3, 1-35.
336. Koonin, E. V. (2015). Origin of eukaryotes from within archaea, archaeal eukaryome and bursts of gene gain: eukaryogenesis just made easier? Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1678), 20140333.
337. Caetano-Anollés, G. (Ed.). (2020). Untangling molecular biodiversity: Explaining unity and diversity principles of organization with molecular structure and evolutionary genomics. World Scientific.
338. Makarova, K. S., Krupovič, M., & Koonin, E. V. (2014). Evolution of replicative DNA polymerases in archaea and their contributions to the eukaryotic replication machinery. Frontiers in Microbiology, 5, 354.
339. Trakselis, M. A., & Bauer, R. J. (2014). Archaeal DNA polymerases: Enzymatic abilities, coordination, and unique properties. Nucleic Acid Polymerases, 139-162.
340. Mönttinen, H. A., Ravantti, J. J., Stuart, D. I., & Poranen, M. M. (2014). Automated structural comparisons clarify the phylogeny of the right-hand-shaped polymerases. Molecular biology and evolution, 31(10), 2741-2752.
341. Mönttinen, H. A., Ravantti, J. J., & Poranen, M. M. (2016). Common structural core of three-dozen residues reveals intersuperfamily relationships. Molecular Biology and Evolution, 33(7), 1697-1710.
342. Theriot, J. A. (2013). Why are bacteria different from eukaryotes? BMC biology, 11(1), 1-17.
343. Thomas, B. (1984). Evolutionary stability: states and strategies. Theoretical Population Biology, 26(1), 49-67.
344. Koch AL. 1990. Diffusion: the crucial process in many stages of the biology of bacteria. Adv. Microbiol. Ecol. 11:37–70
345. Koch, A. L. (1996). What size should a bacterium be? A question of scale. Annual review of microbiology, 50(1), 317-348.
346. Dey, G., Thattai, M., & Baum, B. (2016). On the archaeal origins of eukaryotes and the challenges of inferring phenotype from genotype. Trends in Cell Biology, 26(7), 476-485.

347. Baum, D. A. (2018). The origin and early evolution of life in chemical composition space. Journal of theoretical biology, 456, 295-304.