 doi:10.20944/preprints201710.0153.v1

Concept Paper

'Parabiotic Evolution': From Stochasticity in Geochemical and Subsequent Processes to Genes, Genomes and Modular Cells

Richard Egel

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

Abstract: This article reevaluates the Woesean concept of crossing a 'Darwinian threshold' from pre-genomic communality, as prevailing in an ancestral 'progenote' state, to vertically stable lineages of autonomous and self-similar cells. This transition from collective trunk-line evolution to Darwinian speciation is dependent on the generation of modular organismal genomes. The same general principle should be valid at subcellular levels, allowing the emergence of semi-autonomous genomic agents, such as viruses and plasmid-carrying endogenous vesicles with organelle-like properties. As compartmentalized agents of endogenous nature could start with smaller genomes than those required for fully autonomous cells, it is conjectured that stable subcellular lineages emerged earlier than their cellular counterparts. Referring to the recent 'pre-endosymbiont hypothesis', it is proposed that free-living bacteria (the first 'prokaryote' cells) arose by 'lineage escape' from plasmid-bearing organelle-like compartments, evolving inside the internally complexifying 'paracells' of the progenote community. The double-membrane envelopes of diderm bacteria may have resulted from cell-biological processes facilitating cellular lineage escape. The later emergence of archaeal cells (resembling bacteria in 'prokaryote' appearance with unichromosomal genomes) and eukaryotic organisms (with compartmented cells and multichromosal genomes) can also be interpreted in terms of this modified progenote hypothesis. Conceivably, the multichromosomal genomes of eukaryotes were bundled in endogenous nuclear compartments to organize a 'nuclear-cytoplasmic lineage', which became vertically stable by perfecting mitosis/meiosis-like divisions and yet retained some intra-species population confluence by sexual division-fusion cycles.

Keywords: Tree of Life; origins of species; cellular lineage escape; endogenous compartmentation; proto-organelles; eukaryogenesis; origins of sex; syntrophic biofilms; endosymbiosis

1. Introduction

OoL Theory in quest for adjoining biological evolution to physicochemical beginnings

This essay mainly addresses the formation of genomes from many individual genes and at various levels of infrastructural relationships. This is a relatively advanced problem of '*OoL*' research [the science concerning '*Origin(s) of Life*'], which had to be resolved one way or the other toward the upper end of the conceptual gap that still exists between the enormous variety of organisms living here on Earth versus the more general validity of physical and chemical regularities pervading the entire universe. In a very general sense, it is evident that chemistry plays a decisive role in forming a bridge from physics to biology, and modern biochemistry has indeed come a long way in rationalizing the functionality of living organisms in molecular terms. On the other hand, it is also fair to say that physicochemical considerations are not yet sufficiently advanced to seamlessly explain the emergence of living matter from inorganic sources on the pristine Earth some 3–4 billion years ago.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2017 doi:10.20944/preprints201710.0153.v1

One way of looking at the task is to appreciate the power of large numbers. This approach has laid the foundations of statistical mechanics and near-equilibrium methods in classical thermodynamics – one of the strongest links of chemistry to physics. It also paved the way to non-equilibrium thermodynamics as an open-ended enterprise. In these terms it has long been recognized that life on Earth is a very peculiar dynamic system, which has effectively avoided wholesale collapse into thermodynamic equilibrium for billions of years. Yet how could living matter achieve this long-term robustness in the first place? Knowing the answer to this question would also mean holding the key to the OoL conundrum. The highly non-random relationships in living cells must evolutionarily have been connected, in one way or the other, to the stochastic reactions prevailing in the distant past. Overall, however, the relevance of primordial stochasticity is underrated and rarely discussed at length in the context of OoL hypotheses, or when it is [1], the potential cooperativity of emergent pre-replicative catalysts is not really taken into account.

To be sure, various issues of potential relevance for an eventually successful OoL theory are subject of innumerable scientific publications, and I do not intend to survey their entire range (for references see [2–4]). I will here briefly mention two separate but potentially related conceptual problems in the most widely accepted OoL hypotheses and thereafter elaborate further on one of these. Serious concerns about commonly held assumptions have been raised before, and unconventional thinking is called for to potentially resolve the most persistent puzzles. By and large, the leading models are dominated by how chemical key experiments may be designed most rationally in the lab. Only sporadic attention, however, has been directed to conceivable system objectives robust enough to lend themselves to evolutionary optimization under more realistic field conditions, considering presumptive environments on the pristine Earth. This operationally restricted bias has left chemical bottom-up approaches to OoL in a frustrating bind: what appears most relevant to the emergence of life on Earth from a biological systems perspective might remain outside the reach of experimental chemistry for a long time to come, whereas the favorite theorems of chemical OoL research, such as the *RNA World model* or the postulation of a self-copying 'first replicator' molecule, are falling subject to sustained critique. Alternatively, some self-perpetuating organic coalescence processes commenced at geochemically suitable surfaces, from which, in turn, a stochastic mix of many prebiotic peptides and oligonucleotides may have fostered the coevolution of these two types of nascent organic macromolecules toward better and better mutual affinity and cooperation.

By arguing from the other side of the evolutionary gap, extrapolating backward from full-fledged life as we know it, the top-down approaches of comparative phylogenomics are complementary to the chemical perspective on the OoL transition. Yet these approaches, too, are riddled by lingering controversies, which may be covered up from time to time but have not rigorously been resolved by now. In fact, not even the widely held belief that eukaryotic cell organization must have descended from preexisting prokaryotic cells has been uncontested, and many subsidiary disputes actually hinge on whether or not the principle of *prokaryote-to-eukaryote succession* holds true. Alternatively, the characteristic nucleo-cytoplasmic relations in eukaryotes may, at least in part, have originated at a common ancestral stage when typical prokaryotic cells did not yet exist.

In very general terms, the superior umbrella spanning over the enigmatic *OoL transition* has been an intermediate kind of *evolution*, which had to change characteristics from no longer just being *physical* in the beginning to becoming fully *biological* of the Darwinian type at the upper end. In order to distinguish this gradual and long-lasting transition by a distinctive and unifying name, I suggest calling it '*Parabiotic Evolution*' [close to life]. This measure deliberately avoids the futile quest for a clear-cut, general and binding definition of life as such. It rather stresses the conviction that the origins of life cannot be pinpointed to any particular event or borderline but were embedded in an extended phase of optimizing evolution for system-maintaining consistency at a global scale. The overall optimization occurring in this transition period primarily concerned the operational functionality and robustness of the living system as such – in terms of molecular reaction mechanisms and their networking potential in multiple interactions – whereas the lineage continuity

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2017 doi:10.20944/preprints201710.0153.v1

of different modular organisms, as commonly observed throughout the modern biosphere, was only coming into existence at the upper end of transitionally *parabiotic* evolution. The latter way of looking at the OoL transition has been advanced primarily by Carl Woese in his later work, and I have more to say on this below.

For the purpose of the present paper, as argued elsewhere [5–7], I consider stochastic sequences of short oligopeptides and oligonucleotides as prime candidates for emergent cooperation relatively early on. This bilateral relation became fundamental in two ways: first of all, it initiated a very basic kind of *protoplasmic continuity*, which essentially has lasted ever since, and thereafter it opened up to the vastness of sequence spaces for chain-like macromolecules – unfolding explosively with increasing chain lengths of the participating constituents. The enormous multitude and variety of possibilities provided the raw material for evolutionary changes according to slight but reinforcing variations in terms of relative survival rates. This may sound like retrodicting *neo-Darwinian* evolution of biological organisms deep into the era of stochastic processes directed by geochemical relationships, but that is not exactly my intention. A crucial difference still concerns the virtual absence of modularity at the beginning. Whilst the material and organizational continuity of living matter is 'all or none' in its existential non-dimensionality, the modularity observed in modern life is hierarchical at various scales and levels, and structural modularity, in particular, has probably had a complex evolutionary history.

In the present essay, I will merge two concepts originally put forth by Carl Woese into one coherent narrative: the hypothetical '*progenote*' state of collective pregenomic optimization – together with the gradual consolidation of proto-genomes – and the later concept of '*Darwinian thresholds*' at the exit line to the first appearance of vertically stable organismal lineages at the first branch points of Darwinian *speciation*. As this is a programmatic paper to introduce the '*generalized (extended) progenote hypothesis*' as such, the early generation of bacterial genomes and cells is given the most detailed attention herein. The later generation of archaeal and eukaryotic lineages, however, is only discussed in brief thereafter, awaiting a more detailed follow-up.

In the next sections, I will first summarize the basic concepts as hitherto considered in the literature and then extend their conceptual range to also consider additional possibilities of subsystems formed by endogenous compartmentation. To present this generalization of Woese's insights in a conceptual overview, I'd prefer to use a partly rephrased vocabulary, as listed in the accompanying Glossary (Box 1).

Box 1. Glossary

***Archgenote*:** residual *progenote trunk line* after the branching of bacterial side lines and still before the generation of archaeal branches.

***Cellular (lineage) escape*:** the 'birth' of a free-living lineage of *prokaryote microcells*, no longer being directly dependent on the progenitor population of composite *progenote paracells*.

***Chromosome*:** a covalently linked string of genes that is duplicated by the common (or major) replication machinery of the organism.

***Darwinian threshold*:** the minimum level of genome complexity and transmissibility required to grant the vertical stability and persistence of clonal lineages, applicable not only to organismal clades but also to partly autonomous genomes in subcellular compartments.

***Domain (organisms)*:** high-ranking subdivision of biological organisms, close to the root of the canonical *Tree of Life* (ToL).

***Domain (proteins)*:** characteristically folded building block of protein, made up by a single contiguous amino acid sequence.

***Eukaryogenesis*:** the evolutionary process giving rise to complex eukaryote cell organization.

***Genome*:** a vertically stable, faithfully transmissible assembly of several to many genes cooperating in cells, organisms or other (acellular) biological entities.

***Genophores*:** gene-bearing molecular units, such as plasmids or chromosomes.

***Holoplasm*:** in composite paracells, the protoplasmic bulk outside the paraorganelles, including the proto-nuclear genome.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2017 doi:10.20944/preprints201710.0153.v1

***LUCAS*:** a formal designation to the *last universal common ancestral state*, from where the generally dichotomous tree of speciated organismal lineages emerged at a common root.

***Macrocells*:** much larger, *eukaryote*-type cells with a complex intracellular substructure, such as genome-carrying *organelles* and genuine nuclei with multiple chromosomes.

***Microcells*:** *prokaryote*-type microbial cells, interpreted here as direct descendants of genome-carrying *paraorganelle* compartments.

***Organelles*:** (herein restricted to semi-autonomous, genome-carrying compartments): secondarily reduced remnants of formerly free-living cells that have been internalized as endosymbionts by other cells.

***Pan-genome*:** the totality of genes occurring in the *progenote* community, here borrowed from modern usage of comprising all the genes present in a (bacterial) species.

***Paracells*:** the supposedly variable, ephemeral, irregular and composite embodiments of the progenote community, occurring in *confluent, promiscuous, polyphenotypic populations*.

***Paragenome*:** the totality of genes occurring in individual (temporary) *progenote paracells*, apart from the semi-autonomous *paraorganelles*.

***Paraorganelles*:** primordial subcellular plasmid-carrying vesicles with *organelle*-like properties, assumed to have formed endogenously in composite *progenote paracells*.

***Paraplasmodial trunk-line*:** the evolutionary continuity of cytoplasm and *paragenomes*, outside the *paraorganelles*.

***Phylogram*:** tree-like diagram of deep-rooted evolutionary relationships among biological taxa, now primarily deduced by comparison of informative genomic sequences.

***Plasmid*:** a covalently linked string of accessory genes that specifies a replication machinery of its own and is not generally essential for survival.

***Progenote*:** hypothetical state of communal sharing, before the manifestation of organismal genomes and modular cells.

***Prokaryotes*:** the simple cells of Archaea and Bacteria, with principally unichromosomal genomes.

***Protoplasm*:** here used as a shorter synonym of a primordial '*proto-cytoplasm*', before any effective compartmentation of genomic matter.

***Protoplasmic confluence*:** the likely primordial trait of frequent *membrane fusion* to represent the most pervasive mode of "*horizontal gene transfer*".

***Protoplasmic continuity*:** the most basic systems property of living matter, which under present Earth conditions cannot be re-established anew.

RNP World: evolutionary stage of *ribonucleoprotein* complexes, before the emergence of DNA as the predominant genetic material.

***Supercells*:** multigenomic eukaryotic cells of undivided cytoplasm, such as syncytia or coenocytes, here used as a general model for non-modular (acellular) organization in *progenote paracells*.

***Trunk-line evolution*:** communal evolution without effective speciation, such as it is assumed for the *progenote* era, formally equivalent to a single *Darwinian species* sharing a common gene pool.

***Woesean asymmetry*:** a peculiar phylogenic disparateness at the earliest branch points into *Darwinian speciation*, where only one of two branches begins to diversify whilst the other branch is still remaining at the communal stage of *trunk-line evolution*.

***Woesean (Darwinian) transition*:** crossing the *Darwinian threshold* at the organismal level. Only thereafter would different descendants of a common ancestor be capable of founding separate species, which in turn engaged in inter-species competition and resource partitioning.

***Virosphere*:** the halo of acellular, infectious genomic agents (viruses or phages) surrounding the domains of organisms and depending on the ribosome-containing cytoplasm of a host cell.

127

128

 doi:10.20944/preprints201710.0153.v1

2. Woese's Progenote Concept Revisited, with 'Darwinian Thresholds' for Escape

2.1. An Overview

This paper concerns the critical period of early evolution when the most fundamental activities of biological life were being established, such as gene-encoded and ribosome-mediated protein synthesis, but vertically stable lineages of modular cells could not yet exist. It also considers the peculiar transition(s) toward modular lineage stability and the earliest bifurcations into a diversifying organismic biosphere. The actual occurrence of such a non-modular and pre-genomic phase was postulated by Carl Woese, who also reasoned that sufficiently many genes had to be gathered in faithfully transmissible genomes first, before different and vertically stable lineages of modular cells or organisms could come into existence and successfully engage in competition with other genomic/organismal lines of slightly different characteristics. Accordingly, Woese assumed the eventual exit from this precursory state to constitute a very significant transition, the crossing of a so-called *Darwinian threshold*, which signified the beginning of speciation in the Darwinian sense [8,9]. This conceptual coupling of the emergence of modular cells to the establishment of self-sufficient genomes is a ground-breaking insight indeed, and it properly tops off a remarkable row of integrative essays [8–11], heralded as Woese's "*millennial series*" [12]. Thereby, the author confirmed his leading role as a progressive innovator of potentially integrating views in the long-established field of evolutionary biology, the deep rooted causes of which are still partly unknown. Not all the implications of Woese's visionary propositions, as presciently envisioned from a *nonclassical* perspective, have yet been fully explored.

Well before already, in pushing comparative sequence analysis across different species up to a universal scale [13–16], Woese's experimental research had caused an upturn of basic phylogenies and microbial classification by recognizing Archaea and Bacteria as distinctly different *domains*, in addition to the more complex Eukarya, at the deepest organismal branches on the universal *Tree of Life* (ToL). Adding to the puzzle, the superficially bacteria-like archaea were shown to carry a peculiar affinity to eukaryotes in many of their informational genes and proteins, rather than to the morphologically similar bacteria.

To be sure, Woese's revolutionary analysis was 'only' founded on a single-gene comparison, but the gene selected for this ambitious study was not an arbitrary one. The sampled sequence did not even code for any particular protein but represented a structural RNA inside the ribosome, which is responsible for assembling every gene-encoded protein in all the living organisms by similar means. In fact, together with specifically amino-acylated tRNAs, the ribosomes are the most highly conserved structural entities occurring in all the living organisms.

Notably, to rationalize the radiation of three disparate organismal branches from a common root, a primitive and not distinctly individualizable "*progenote*" state was conceived by Woese as a poly-disperse population of *acellular*, rather intangible entities precursory to modern life forms. As presciently conjectured about this ancestral state early on [14,17], its gradual evolution from a hodgepodge of many stochastic interactions toward template-directed polymerization mechanisms must have occurred under the relational umbrella of *communal sharing* across the entire population to represent the pre-genomic *progenote* state.

More recently [18], the reasonable working assumption of primordial communality has been substantiated by dynamic computer simulations to model "*Competition between Innovation Pools*" in the presence of genetic exchange between diverging populations. In particular, the study was targeted at the gradual optimization of the genetic coding system from marginally biased *statistical proteins* toward reducing ambiguity in the codon assignments for related amino acids first and leading to fully deterministic codon specification later on. By inference, however, other composite subsystems of many interacting parts should positively respond to gradual optimization by communal *trunk-line* evolution at the *progenote* stage as well, which may have led to synergistic cooperativity at various other levels than what is commonly presumed.

It is pertinent in general to note that the overall system performance resulting from multi-component interactivity can, in fact, be gradually optimized by the probabilistic means of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2017 doi:10.20944/preprints201710.0153.v1

communal evolution. This is an important conclusion applicable to the tentative progenote state as such and other evolutionary stages. At the very beginning, it potentially defuses the oft-repeated argument [2,19,20] that functionally adaptive evolution of self-organizing entities with life-like characteristics should require the faithful self-replication of some genetic substance, as being fed from an environmental pool of biochemical precursors. Instead, template-directed replication of genetic material becomes one of many catalytic abilities that are subject to optimizing bit by bit, in parallel with many other system-supportive activities in the '*protoplasm*', (here used in short as a contraction of '*proto-cytoplasm*'). Also quite early on, the general principle of *collective optimization of many interactive components* may have paved the way for endogenous membrane formation inside the protoplasmic mass when certain semi-soluble peptides began to aggregate just after leaving the early proto-ribosomes. Some of these served as hydrophobic scaffolds in coordinating catalytically active loops and crevasses, whilst others arranged themselves in planar sheets with membrane-like configurations [5]. This arguably happened well before the generation of long-chane lipids, which eventually optimized the advantageous properties of biomembranes in the long run. At the eventual exit of the progenote era later on, the *collective multipart optimization* principle became particularly relevant for the emergence of eukaryote cell complexity as well.

Overall, Woese's ground-breaking insight into the feasibility of collective cooperativity and communal evolution opens a new perspective for re-emphasizing the coherence and continuity of a communal *protoplasm* as very basic and important aspects of living matter [7,21], intrinsically selecting for macromolecular cohesiveness in space, functional connectivity in catalytic networks and continual persistence in time. Conceivably, this emergent system approached its functional complexity and informational connectivity by allowing the *coevolution of nucleic acids and proteins* to have an early start [5,6,24–26], commencing from more or less stochastic oligonucleotides and prebiotic, uncoded peptides.

Moreover, as further discussed in the present paper, re-focusing attention on the collective evolvability of a communal *protoplasm* can also shed new light on other issues of long-standing controversy concerning the early stages of biological evolution. Among the contentious issues, most notable is the commonly held opinion [27–30] that so-called *prokaryotic* cells once ruled the early biosphere exclusively but more complex and intricately organized *eukaryote-like* cells only appeared much later on and somehow, therefore, must have derived from 'prokaryotic' ancestors. On the other hand, various common aspects of eukaryotic cell organization appear more primitive in direct comparison with the streamlined features of even the simplest bacterial or archaeal cells. This might be explained more easily by a different model, assuming that the common precursory state ancestral to all present life forms resembled the basic organization of eukaryotes more profoundly than that of modern bacteria or archaea as such, or any combination of so-called prokaryotic traits. Whilst proponents of this unconventional view, implying some 'eukaryotes first' (EF) scenario [31], were more outspoken in a not so distant past [32–37], their voices are no longer counted much, even though their rational line of argument has not convincingly been proved to be erroneous.

Clearly, additional insights are needed for keeping this debate alive against a self-assertive consensus in the current literature. Such novel input may come from recent data concerning the mosaic nature of mitochondrial proteomes [38–40], of relevance to the enigmatic prehistory of mitochondrial endosymbiosis in eukaryotes. On this expanded basis, I will combine Woese's main theoretical insights into pre-genomic communality with the commonly dismissed, or just unnoticed, possibility of endogenous compartmentation in the primordial *protoplasm*, well before the establishment of modular genome-controlled cells in any modern sense. To motivate this conceptual expansion, I briefly summarize the significance of semi-autonomous eukaryotic organelles in general and the recent *pre-endosymbiont hypothesis* [41] as a novel twist to the conventional story.

All modern eukaryotes are composite, chimeric organisms in a very intrinsic way. Most of them have mitochondria as respiratory organelles, with the exception of certain anaerobic lineages, which nonetheless carry mitochondrial remnants, such as hydrogenosomes or mitosomes, or did so in their ancestral past [42–45]. It has long been established that mitochondrial genomes are specifically related to α-proteobacteria [46–48], in favor of *endosymbiotic theories for organelle origins* [49–53]. If this

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2017 doi:10.20944/preprints201710.0153.v1

were the whole story, the majority of mitochondrial proteins and the corresponding genes should have their common origin at the α-proteobacterial clade. Yet, when this particular prediction was tested systematically, it could only be confirmed for a widely conserved core, forming a minor part (~15 %) of all the mitochondrial proteins [44]. Each one of three other components appears more prevalent than the distinctly α-*proteobacterial* heritage: '*prokaryotic*' (~20 %), '*unique*' (~25 %) and '*eukaryotic*' (~60 %), showing considerable numerical variability in different organisms.

Where are all those other genes supposed to come from? Approaching this question rationally with an open mind, Michael Gray presented his *pre-endosymbiont hypothesis* [41] and made an important new assumption about the still enigmatic "host cell" that first took up an internally viable α-*proteobacterial* cell, which subsequently evolved into a clade of permanently endosymbiontic organelles. As archaeal cells in general would not likely accomplish such an incredible feat, alternative 'nonclassical' scenarios are worth taking seriously enough to argue over. The novel key assumption is as follows.

Making ends meet, Gray suggested that the host cell receiving the bacterial seed to initiate the current mitochondrial lineage already carried other organelles beforehand, tentatively called *premitochondria*, which presumably were of endogenous origin, but not necessarily so. In many ways, a complex cell can replace one endosymbiont by another more easily than any simple cell might acquire a novel endosymbiont from the outside. Having an effective protein import system in place already has arguably been advantageous in this regard.

Historically, the antagonism between endogenous/autogenous and exogenous/endosymbiotic aspects of eukaryogenic theories has shifted back and forth repeatedly (as summarized in [50–54]), with current emphasis on purely endosymbiotic explanations. Proponents of endogenous compartmentation, however, have never been completely silenced. In fact, quite complex compartments ensheathed by single membranes occur in both bacteria and eukaryotes [55–59] and are very likely of endogenous origin. A more general significance of endogenous processes in the earliest phases of eukaryogenesis has by no means been ruled out.

2.2. The Generalized (Extended) Progenote Hypothesis: A Synthesis Rephrased in Part

In the present essay I argue for a substantial reappraisal of subcellular compartmentation and the possible endogenous origins thereof. This is done so from a genome-focused Woesean retrospective, which looks backward into the past beyond the evolutionary split into Darwinian species of either eukaryote or so-called 'prokaryote' cell organization. To start with, I try to cast a generalized model of the Woesean progenote concept into a coherent narrative. This '*generalized (extended) progenote hypothesis*' assumes an early phase of '*compartmented trunk-line evolution*'. Selected aspects of mechanistic interest will be discussed in more detail further below.

The traditional designation of Archaea and Bacteria as *prokaryotes* ("before eukaryotes") has tentatively been replaced by "*akaryotes*" as a non-phylogenetic and more neutral term, in order not to imply a particular direction of organizational descendancy [37,60]. Although personally being in favor with this distinction, I will herein continue to use the firmly established *prokaryote* expression, hoping that the potentially provocative undertones connected to that term might be obviated by the suggestion of a population-wide transition from syntrophic symbiosis to endosymbiosis (as discussed further below). On the other hand, what is clearly needed is a separate identifier for the residual progenote-like trunk-line stage, as inferred by the *generalized progenote* hypothesis, after the bacteria have branched off as independent lineages. As this conceptual stage is still ancestral to both archaea and eukaryotes but does not fully resemble either one of the disparate organismal domains to follow, I seriously suggest *Archgenote* to designate this important transitional phase in our common ancestry.

This integrative narrative uses partly novel terms (Box 1). It begins at an evolutionary stage when a long period of coevolution of RNA and proteins, had reached its peak and the composite ribonucloprotein (RNP) machinery of a tentative "*RNP World*" [61,62] had culminated in the perfection of ribosomal protein synthesis. This was the time when DNA was about to take over as a more stable and reliable repository of genetic information. Thereafter, *genomes* in the modern sense

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2017 doi:10.20944/preprints201710.0153.v1

were being nucleated and became organized as informational back-up centers for operationally functional biological entities. In parallel with advancements on the genetic front, the gene-encoded proteins could become more specific in their affinity to binding partners and more complex in terms of domain size and number of multiple domains. However, there were no *modular cells* with well-defined *organismal genomes* yet, nor could they even exist as vertically stable lineages before propagable modular genomes were in place as well.

The intrinsic coupling between integral genomes and modular cells was first recognized in its full significance by Carl Woese. Accordingly, the initial establishment of a reliably transmissible genome marks a decisive point in evolutionary time for each monophyletic group of Darwinian species among the numerous organisms still living today. A peculiar before-and-after difference is noticeable at this turning point. This is when evolutionary dynamics changed from collective *trunk-line evolution* before the shift to competitive *Darwinian speciation* thereafter. Woese referred to this passage as the crossing of a *Darwinian threshold* [8], and he later suggested *Darwinian transition* for denoting this critical change of evolutionary dynamics [9]. The unsolved question still is: How many times did such a *Woesean/Darwinian transition* actually occur in the canonical *Tree of Life* (ToL)?

The hypothetical state of living matter before the manifestation of organismal genomes has been referred to as *Progenote* [14]. Initially, this stage was likely characterized by harboring a large number of coevolving yet separately distributed proto-genes, which were gradually changing into distinctive genes, but these were not yet bundled into integral genomes. This condition did not allow for the possibility of Darwinian speciation but called for a *high degree of communality* [8,18]. To comprehend the theoretical framework of this conception, it is mandatory to think in terms of '*fuzzy sets*', such as *probabilistic populations*, rather than of individually clonable cell-like entities to begin with. In essence, the evolutionary criterion for vertical stability can be rephrased as *long-term survival* (ideally spanning from an arbitrary period into the modern world), and the probabilistic risk assessment for survival versus sudden irrevocable extinction hinges on some critical population size, which in turn depends on various biological parameters to specify the momentary corpuscular composition of the momentary population members on the one hand, and the distributional characteristics of the heritability system on the other. Traditionally, the communal aspect of the ancestral *progenote* state has been ascribed to pervasive levels of "*lateral/horizontal gene transfer*" (LGT or HGT) [8,18], but since stable vertical lineages could not yet exist, this gene- and lineage-centric term appears somewhat inadequate. Also, as used for modern prokaryotes, the molecular mechanisms promoting LGT act in a unidirectional manner, from a donor source into the DNA genome of a particular recipient cell, and only a limited number of genes are being inserted at a time.

Preferring a more natural and intuitive explanation, I consider *protoplasmic confluence* and frequent *membrane fusion* to represent rather primitive, primordial traits that may have prevailed throughout the *progenote* era [6,63–66]. As to physical appearance, the material embodiments of the highly variable *progenote* state (herein termed *paracells*) were presumably larger and more primitive than modern bacterial cells, yet also became more complexified internally as time went by. The communality associated with fission–fusion equilibrium implies a long period of *trunk-line evolution* without effective speciation throughout the *progenote* stage. Formally, the population members of such a *trunk lineage* belonged to one and the same Darwinian species, all sharing a common gene pool [67]. Although there was no competitive coexistence between divergent clonal lineages, communality as such was particularly suited for optimizing large sets of interactive components simultaneously [18].

Originally, the *Woesean transition* of crossing the critical threshold was only considered for organismal lineages [8,9]. In general terms, however, this concept applies at any level of genomic propagation sufficient to grant vertical persistence, even to clonal lineages of a non-organismal kind. Its relevance also extends to plasmid-bearing subcellular compartments which, due to organelle-like properties, are herein referred to as *paraorganelles*. The assumption of such *paraorganelle compartments* is fully in line with the notion of '*premitochondria*' in a complex proto-eukaryotic host lineage, which is gaining support and credibility from comparative genomics [40,41]. In addition, the emergence of viral lineages (*virus* genomes and particles) is covered by this generalized principle as well.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2017 doi:10.20944/preprints201710.0153.v1

Notably, each *Woesean/Darwinian transition* implies phylogenic disparateness at the earliest branch points into Darwinian speciation; I call this peculiar aspect the *Woesean asymmetry*. This '*nonclassical* perspective' [8] is unavoidable and means that the earliest descendants crossing their Darwinian threshold would only thereafter begin to diversify into separate and vertically stable lineages. Together, therefore, these fortunate few just formed the first branch at the side of the collective *trunk-line* population; the large majority of other members, however, were still part of the collective *progenote* community. There is no reason to assume that this major lot disappeared at once thereafter or that it lost its evolutionary potential soon. Instead, the continued evolution in the communal trunk line might well have given rise to additional descendants later on, when they became able to accomplish another *Woesean transition* on their own, thereby founding more branches away from the collective trunk. The existence of such a persistent *trunk lineage* leading up to eukaryote ancestry has long been inferred [64,68] and is also indicated by comparative proteomics [69].

What about the likely physical appearance of the members supposed to represent the elusive *progenote* community? Woese himself did not directly deal with this speculative question, and his theoretical thinking was more about what they were not, genetically speaking, than how they were functionally organized or may have looked. Such intellectual sobriety is scientifically sound but can also delay progress in potential understanding at this scientifically challenging yet empirically intractable frontier.

Woese was fully aware of the "*nonclassical* perspective" in his insights [8], suggesting that Bacteria, Archaea and Eukarya had crossed their respective *Darwinian thresholds* in this order, at different times and more or less independently [8,16]. In detail, however, he was not yet prepared to modify his intuitive assumption that being "*simpler in structure*" also meant "*being closer to some ancestral form than are the eukaryotic ones*" [8]. Kandler, on the other hand, was more flexible in this regard, putting particular emphasis on the effects of independent *sampling* from a frequently exchanged common gene pool (his so-called '*pre-cells*', corresponding to Woese's *progenote* state) when the three organismal '*domains*' began to consolidate their domain-specific genomic repertoires [70]. Together, I think, Woese and Kandler provided us with a *nonclassical* legacy that has not yet been fully appreciated or explored. This is where I hope to make a further contribution here.

Perhaps it helps to suspend yet more of the preconceived notions that implicitly have shaped the currently most popular views about early evolution. Putting aside the prevalent connotations of symmetry at the earliest phylogenetic bifurcations, as Woese already did [3], was only a first beginning. Other widespread preconceptions concern the long-standing overemphasis on incidental fission (over spontaneous confluence), on protocells in free suspension (over sessile layers), on primordial simplicity as such (over self-organized internal complexification), or on submarine chemo-energetic origins of life (over surface-exposed, terrestrial, photo-activated biogenesis). Considering alternative possibilities, the nucleation of micro-genomes in endogenous compartments and the accompanying self-complexification of the *protoplasm* will be further discussed in this essay.

As of now, the ephemeral and irregular embodiments of the *progenote* community have no accepted common name; and they did not carry any definite genome either, just a variable number of more or less independent genes. These entities were previously referred to as "*multiphenotypical populations of pre-cells*" [70]. In a subtle shift of emphasis, I now prefer to use "*promiscuous, polyphenotypic populations of composite paracells*" instead. This is to draw special attention to *confluent protoplasmic fusion* as an essential factor for back-and-forth dynamics in the overall gene distribution of the communal pool, tentatively referred to as the '*pan-genome*', in analogy to all the genes observed by modern analyses in many strains of a (bacterial or archaeal) species [71].

Somewhat stochastic samples (the temporary '*paragenomes*' of individual *paracells*) were drawn at *irregular fission* events, and return flows could go back into the *pan-genome* pool by *protoplasmic fusion* at spontaneous encounters with other *paracells*. This notable feature is complementary to the *polyphenotypic* diversity of various subpopulations in different microenvironments. As seen from the perspective of collective *trunk-line evolution*, these cell-like non-modular *paracells* evolved together as a consortium of similar but not identical entities. In general terms, the overall degree of communal

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2017 doi:10.20944/preprints201710.0153.v1

sharing throughout the progenote population depended on the variable rates of residual gene flow between various subpopulations, which were temporarily scattered over a range of different microhabitats. The adaptive modulation of these flow rates may have gained particular importance during the ultimate transition toward the vertical stabilization of eukaryotic lineages.

As an aside, I strongly favor the idea that the *paracells* preferentially grew attached to mineral surfaces rather than freely in suspension. Such a sessile lifestyle would more readily conform to spontaneous confluence which here is assumed to constitute a primordial trait in early evolution. This trait would also link directly to yet earlier phases of emerging *surface metabolism* [72–74] and colloidally associated, sessile *protobiofilms* of *organic-hydrogel* consistency [7], and it would likewise link to a photo-activated early start [74,75], especially under *terrestrial surface conditions* [76–79]. The nonconventional assumption of placing the cradle of emergent life into the very heterogeneous environment of terrestrial geothermal fields has also certain 'proto-ecological' implications for the evolving population of *progenote paracells*, which will be considered more specifically further below.

What then can we learn by looking for potential genomic cues at a subcellular level? What kind of *non-organismal* entities can possibly propagate inside some pregenomic, acellular protoplasm as *vertically stable lineages* and thereby become *genomic agents* in their own right? *Viruses* are simple examples of this category, but since they are of little help explaining the eventual appearance of genuine, metabolically self-sufficient cells, they are not directly relevant to the objective of this paper. Much more important in this regard are subcellular *plasmid*-carrying vesicles with organelle-like properties. Conceivably, energy-converting vesicles of this kind [66] were particularly significant for the *composite progenote paracells* in general, and providing protein-synthesizing machinery inside such compartments may have fueled incremental growth of the corresponding plasmid genomes.

If such functionally useful entities are assumed to have formed endogenously, they should rather not be lumped together with mitochondria and other *organelles* in the modern sense, reserving the latter term to secondarily reduced remnants of formerly free-living cells that were internalized as *endosymbionts* by other cells [80]. To distinguish between organelle-like entities of endogenous and exogenous origin, I suggest using '*paraorganelle*' as a generic term for the primary kind, as here proposed to arise at the *progenote* stage in composite *paracells*. It is possible that the *pre-mitochondria* postulated by Gray [41] were of endogenous origin and thus would represent such *paraorganelle*s.

The above considerations generalize the Woesean concept of crossing *Darwinian thresholds* in a peculiar way. As soon as the first endogenous *paraorganelles* were generated at an early stage of *progenote paracells*, the beam of evolution was split into two very different components. Thereafter, the few genes carried on the compartmented plasmids and the many separate genes somewhere outside the *paraorganelle* compartments became subject to different evolutionary dynamics, in which selective trends were heading in opposite directions. This is because the plasmid-carrying *paraorganelles* had crossed their *Darwinian threshold*, whereas the surrounding bulk of the *progenote protoplasm* and the corresponding genes (the fluctuating '*paragenomes*' of *paracells*) had not yet done so. In the following, it is useful to distinguish this residual paracell protoplasm (outside the semi-autonomous paraorganelles) by another denomination, the '*holoplasm*' of *paracells*, irrespective of whether or not the corresponding *paragenomes* were collected in proto-nuclear compartments.

Modern *organelles* are subject to *reductive evolution* [81], as imposed by the now prevailing *host cell versus organelle selection*, so that most genes of the original endosymbiont genome have since been lost or transferred to the nuclear genome in competitive and diversifying lines of *eukaryote 'macrocells'*. Presumably however, the *paraorganelles* in *progenote paracells* tended to evolve in a more cumulative mode. Since the power of Darwinian selection depended on the heritable stability in vertical lineages, *paraorganelle versus progenote* selection gave *paraorganelle lineages* the advantage of '*accretive evolution*', at the expense of the hosting non-modular *paracells*. The preferential gain of essential genes for vital functions on the unimolecular genomes of *paraorganelles* may eventually have led to successful *lineage escape* of rather small yet fully autonomous *prokaryote 'microcells'*.

According to canonical core phylogenies [11,82], bacterial genomic/cellular lineages were first to segregate from the common root. Whilst standard views implicitly assume that genomic and

 doi:10.20944/preprints201710.0153.v1

cellular speciation commenced together and were naturally coupled [8–11,82,83], I'd rather question the intrinsic nature of such a linkage, at least for primordially acellular stages, suggesting a more flexible model in the '*generalized progenote hypothesis*'. More likely, I think, the initial seeding, or 'nucleation', of functional genomes may well have occurred at subcellular levels considerably earlier than at the superior level of cellular and/or organismic autonomy. In other words, even a subcellular, endogenously compartmented 'mini-genome' could act just as a genuine one and thereby initiate a *vertically stable lineage*; it 'only' had to stay within the common boundaries of a communal *protoplasm* for quite some time. After all, it was the emergence of a *vertically stable lineage* of self-similar genomes that characterized the *Woesean transition*.

This conceptual uncoupling of genomic nucleation and diversification from organismal speciation has significant consequences for interpreting the familiar *phylogram* of the conventional ToL in terms of organizational levels at the early branching points. Somewhat mockingly, the general significance of the standard ToL has even been played down as "*The tree of one percent*" [84], implying that non-treelike processes appear more important, at least in microbial evolution. This narrowly focused assessment is still deeply wedded to the prevalent opinion that microbial evolution, understood in terms of bacteria-like organization, was all that counted when *Darwinian speciation* began to leave a trace. According to this view, *lateral gene transfer* (LGT) and *endosymbiosis* (from one bacteria-like lineage to the other, and from a newly acquired exogenous endosymbiont to its virgin host, respectively) were the only mechanisms of non-treelike networking processes to be considered in conceptual model building at that early transitional stage.

To be sure, a universally conserved genomic core of "1 %" may not count much when it is about managing a self-sufficient cell, save any organism of yet higher levels of complexity; but this particular core set should have been more than sufficient when it merely came to the endogenous establishment of vertically stable organelle-like lineages, which thereafter could progressively evolve according to the Darwinian principles of *divergence* by non-identical reproduction and selective '*survival of the fittest*'. Notably, a functional protein synthesizing system, including genes for ribosomal RNA, may have been part of the *paraorganelle* genomes from early onwards, which would considerably relocate the common root of the Woesean ToL [10,11,13] in the backward direction. In organismic terms, for that matter, the deepest branch points would be embedded well within the collective state of a persistent yet internally composite *progenote* community.

To start with, the '*paraorganelles*' were bound to remain within the confines of the surrounding *progenote holoplasm*, which was mainly governed by the other "99 %" or more of the communal yet highly scattered '*pan-genome*' outside the *paraorganelles* as such. Certain membrane-associated proteins with energy-converting potential were probably most useful for the *progenote paracells* to be synthesized from within vesicular compartments, just as the integration of such components into small mitochondria-related organelles appears vitally essential for the energy management of the large eukaryotic cells of today [27,68]. Thereafter, '*accretive evolution*' may have taken over by acquiring more and more genes from the surrounding *pan-genome* pool, preferentially those that were also advantageous for the paraorganelles as self-propagative genomic agents.

Formally speaking, this gene flow into paraorganelles was also a kind of LGT but it was preferentially pointing in the opposite direction than what is commonly discussed in the context of *endosymbiotic* gene transfer today. Owing to the import of additional host genes and the preferential retention of functions advantageous to the *paraorganelle lineages* themselves, the best adapted ones would become more and more autonomous over time. Some of them would even be fully self-sufficient in the end, whereafter they could leave the host for good. This evolutionary model of *endogenous differentiation* allowed the composite *paracells* of the Woesean *progenote* population to act as '*neonatal incubators*' for the gradual maturation of fully autonomous *prokaryote* cells, at first of the bacterial kind. It thus provided a '*safe heaven*' for mini-genome maturation in semi-autonomous subcellular vesicles, which were separated from the external environment by functionally complex yet more slowly evolving *protoplasm*. In formal terms, this *generalized progenote hypothesis* is a novel approach to superimpose the canonical tree of early stabilized genomic cores with mixed topologies for many additional genes, even before the first organismal lineages had come into existence.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2017 doi:10.20944/preprints201710.0153.v1

Leaving the *progenote* community successfully along this route can be described as '*cellular lineage escape*'. The metaphorical expression itself is borrowed from another hypothetical scenario [85], where '*cellular escape*' was used with reference to leaving behind the physical encasement of mineral compartments. As suggested herein, the release from the buffering constraints of a collective and conservative *progenote* population, in favor of the pioneering and highly effective individuality of modular cells, should be even more appropriate for applying the term. Besides, the origin of virus particles has been rationalized by similar *escape hypotheses* [86,87], although for reproduction, virus genomes must always return for metabolic support from inside other members of the biosphere. It is also apparent that viruses in general are of polyphyletic origins, foremost preceding the emergence of cellular organisms, with which they since engaged in tightly linked relationships [88].

This structured *progenote* model provides bacterial and viral lineages with minimalistic yet rather "egocentric" escape routes from a communal state of large-scale *protoplasm sharing*. This model also provides a structural basis for the peculiar postulate of *Woesean asymmetry,* prevailing at the earliest branch points of *Darwinian speciation* in the canonical ToL. This *asymmetry* theorem is perhaps least appreciated among Woese's ground-breaking insights. Apparently, no structurally convincing scenario has yet been presented to substantiate or illustrate this particular theoretical inference. As for the *endogenous compartmentation* model presented here, the major question remaining is what has happened to the residual population of composite and communal *paracells*, after the first bacterial lineages had left this community as free-living and competitively diversifying cells. For a meaningful discussion of this evolutionary aspect, it is also important to consider the potential influence of ecological and environmental factors at that stage, at least in general terms. In the following I briefly highlight the main inferences drawn from this model. The discussion of several additional aspects will be taken up in other sections below.

At the stage when the first *bacterial microcells* just managed to leave the *progenote community* for good, the "*promiscuous, polyphenotypic populations of composite paracells*" (as herein proposed) had already undergone a prolonged optimizing period of evolution and, arguably, accumulated some locally adaptive variation. It is here also assumed that virtually all of these *paracells* were carrying clonal lineages of bacteria-like *paraorganelles* internally. Owing to their relatively small yet vertically stable genomes, these *paraorganelles* were capable of adapting and specializing faster than the larger and still communally shared *pan-genome* of the vertically unstable *paracells*. Thus, the differently specialized *paraorganelles* began to play important roles in the local variation of the communal *paracell* population at large.

It is of general interest that the first free-living bacteria should probably relate to a lineage of *paraorganelles* that more than others engaged in direct interaction with a particularly favorable external environment and effectively converted this interaction into local increments of organic matter. It was clearly in the evolutionary interest of such a lineage to increase its share of net production, which finally outcompeted the communal *paracells* around them and, subsequently, throughout that particular environmental niche. This happened when the first specialist lineage of *bacterial microcells* was born by "*lineage escape*".

On the other hand, this cannot mean that the entire population of *progenote paracells* was being wiped out immediately by that occurrence. Nor does it mean that all the bacterial lineages living today must have originated and diversified from the single specialist *escapee* line of first appearance. Here it is generally important to emphasize that the coeval *progenote* population presumably was '*polyphenotypic*', having spread out into accessible environmental niches other than the one most favorable for fully independent growth and propagation. To give an example from real life, when the most convenient surfaces exposed to the sun are occupied, it pays off to find an alternative way of living in the shade or even in permanent darkness.

As I see it, the residual *progenote* community survived outside the energetically most favorable environmental niche, still having different options to compete effectively with the clonally diversifying bacterial newcomers: (i) the previously successful strategy of generating and shedding specialized *prokaryote microcells* might have worked several times again; (ii) a different strategy was to retain the intrinsic qualities of composite infrastructure and some communal sharing by

streamlining into a vertically stable propagation system by other means, eventually leading to diversifying Darwinian species of *eukaryote macrocells*; (iii) the largest *paracells* might have resorted to a novel mode of making a living by actively feeding on the ever greater abundance of bacterial *microcells*; and (iv) some *paracells*, or their stabilized descendants, may also have engaged in long-lasting syntrophic relationships with bacterial partner cells.

More likely than not, it was the *residual progenote* community, referred to as *archgenote* herein, that continued its collective trunk-line evolution and, in one way or another, gave rise to both archaeal cells and major parts of nuclear genomes in eukaryotic organisms. Yet, by and large, the many current efforts to further resolve the latter bifurcation are solely focused on comparative genomics, as based on steadily increasing data sets. This retrospective approach, however, has limitations when it comes to reconstructing ancient ancestral phenotypes across phenotypically disparate divides of such proportions. Considering the pivotal acquisition of bacterial cells as mitochondrial endosymbionts into the ancestral proto-eukaryotic trunk line, I will give particular attention to option (iv) above.

The population-wide transition from syntrophic symbiosis to endosymbiotic integration could serve as a novel alternative to phagocytosis, as hitherto assumed to this effect. It still seems to me that characteristic features of two contrasting strategies, which later on have led to genomically stabilized organismal lineages of prokaryote- and eukaryote-type cell organization, respectively, were inherent as potential predispositions already in the communal *progenote* state of Woese's unconventional conception. A broadly based and internally compartmentalized *trunk line* would naturally bridge a smooth and protracted transition from the *communal progenote state* to the composite organization of *eukaryotic cells*, as contrasted with the narrower subset of *paraorganelles*, which herein are assumed to have given rise to various free-living *bacterial cells* via singular '*lineage escape*' events.

How *archaeal lineages* fit in with this model at an intermediate position is still open to further discussion. Since archaeal protein synthesis is based on a system with characteristics somewhat between the bacterial and eukaryotic counterparts, the archaeal tool set may have been sampled in a secondary wave of endogenous *paraorganelle* formation and a later wave of '*lineage escape*'. Alternatively, one or more archaeal lineages may have originated more directly from the *progenote paracells* as such, if the sampling process was mediated by genomic compression on to a single circular chromosome and severely reductive evolution at very marginal habitats. Either way, the escape of archaeal cells was perhaps assisted by certain bacterial genes that facilitated the management of circular genomes and cell division in a 'prokaryotic' manner, thus underlining the intermediate placement of archaeal lineages between bacteria and eukaryotes,

Some key assumptions in this scenario are modeled after certain modern examples, such as multinucleate amoebae and syncytial slime mold plasmodia. Large amoebae are also known today to act as evolutionary 'melting pots', which facilitate the emergence of chimeric microorganisms, such as giant viruses [83,89,90]. 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 [91] or between larger 'plasmodial' masses [92]. The present model explores the posited descendancy of *prokaryotes* from *paraorganelle* compartments inside the composite, polymorphic and amoeba-like *paracells* of the conjectured *progenote trunk line* population. The eventual transformation into eukaryotic cells will also be discussed, and assuming some plasmodial-like organization already at this early state appears particularly pertinent.

In discussing endogenous compartmentation it is also important to distinguish what was inside or outside the posited *paraorganelles*. While the plasmid-like genomes inside the vesicles began to diversify as vertical lineages, the bulk the cytoplasm remaining outside (together with eventual proto-nuclei) continued to evolve as a communal trunk line for quite some time. To emphasize the collective continuity in this differently evolving part of the *paracell* population, I suggest using *paraplasmodial trunk-line* for the extra-organellar portion, which may have formed the basis for a distinguishable eukaryote-specific "*nuclear-cytoplasmic lineage*" [93], as discussed further below.

 doi:10.20944/preprints201710.0153.v1

The basic concept of starting a nascent minigenome in a small vesicular compartment within the larger mass of communal protoplasm has an intriguing consequence. It relates the emergence of genuine microbial cells to two different points of origin, removed in evolutionary time – *genomic nucleation* first and *cellular lineage escape* considerably later on. In principle, therefore, a single event of genome nucleation can result in multiple escape events. Accordingly, successive cellularization events could give rise to differently organized cells in genomically related lineages. Deep-branching bacterial and archaeal phyla may indicate that this, in fact, has been the case, but I will not go into detail in this general presentation.

In considering the current model with regard to the relative timing of when the first plasmid minigenomes became separate from the bulk of other genes by endogenous compartmentation, it is relevant to note that bacterial and archaeal/eukaryotic replication enzymes, DNA primase included, supposedly have independent origins [94,95]. This implies that subcellular compartmentation already began before the genomic take-over from the preceding RNP world scenario, in which the optimization of ribosomal protein synthesis took place and various other vital activities must have been consolidated and optimized from rudimentary beginnings as well, just to keep the Woesean progenote system alive and sturdy all along.

The overview narrative is meant to illuminate – and potentially bridge – the conceptual gap between modern organismal life and hypothetical earlier states of protolife, when modular, self-similar organisms could not yet exist as separate and independently reproductive lineages. This concerns the enigmatic rooting of the canonical ToL and the widely divergent systems properties in the three domains of living organisms. Within each of these distinctly different early branches, a characteristic set of domain-specific features (or rather their corresponding genes) can tentatively be traced back to a *last common ancestor* of the particular domain. In all the outer twigs and intermediate branches, this traceability appears readily justified by the fundamental principle of *Darwinian speciation*, as resulting from treelike descent with modification. More often than not, the same algorithm is also extrapolated backwards to a conceptual LUCA (*last universal common ancestor*), from which all three domains of extant organisms eventually derived. It is also commonly implied or taken for granted that the inferential LUCAnian beings more or less resembled cell-like creatures of bacterial appearance and complexity. Altogether, this kind of backward extrapolation leads to solely retrospective views, which may lose their predictive power around the 'incipient singularity' of deriving disparate complex systems from an overly narrow common source.

By arguing the other way around, prospectively, Carl Woese was early to question the validity of such commonly held assertions. In conceptually connecting the fundamentals of molecular biology to their initial emergence from the potentially unbounded combinatorial sequence space of *statistical proteins* [15,18,96], his general reasoning was heavily drawing on the methodology of theoretical physics rather than on the more practical significance of biochemical metabolism and cytoplasmic infrastructure for evolutionary biology. To the extent that such insights could be parameterized, some major conclusions were indeed confirmed by modeling for numerical simulation [18].

At any rate, Woese successfully idendified the eventual consolidation of modular genomes [8] – from many originally unconnected genes – as a particularly relevant crossing point in the transition from communally supported collective optimization dynamics toward Darwinian speciation and inter-lineage competition in biological evolution. Again, however, this conceptual achievement was presented in general terms of theoretical physics – in analogy to *thermodynamic phase transitions*, which are *discontinuous changes of "state"* from of one set of temporarily stable genetic systems properties to another. He even used '*crystallization*' as a metaphor to characterize genome formation from individual genes; coalescence of soft matter might have been more appropriate in this regard. In recognition that the *Woesean progenote* comprised both a *communal* and a *common ancestor state* [10,11,25], the LUCA acronym has since been modified to LUCAS [97], so as to designate the *last universal common ancestor state* at the first branchpoint of the canonical ToL (see [98] for a pertinent discussion of the many shades of the LUCA/S concept).

 doi:10.20944/preprints201710.0153.v1

As of yet, the ample literature discussing the rooting of the ToL has virtually neglected potentially interfering influences of cytological infrastructure in the transition phase from pregenomic to fully genome-dominated population dynamics. The following sections of this paper will discuss in more detail how the possibility of intracellular compartmentation at this critical stage of early evolution may help solving some of the most enigmatic questions concerning the emergence of Woese's three domains of organismal life.

3. The Energy Connection

Life is a surface phenomenon in more than one way. Not that it has to be attached to solid surfaces, it often isn't; but life cannot be imagined to persist without an abundance of internalized reactive surfaces within each living cell. The important dynamic processes of energy transfer and biochemical reactivity are mediated and channeled by surface-exposed epitopes on molecular nanoscale biostructures, such as distinctly folded protein enzymes, composite ribonucleoprotein nanomachines and protein-loaded lipid biomembranes. Photoinduced charge separation and energy transfer reactions, in particular, occur at membrane-integrated protein complexes.

Before the first self-organizing biostructures came into being, presumably, certain adsorptive and reactive mineral surfaces [72–74] were instrumental in paving the way for the emergence and early evolution of polymolecular and/or polymeric organic matter with biogenic potential. In this emergent coevolution of structural and functional relationships, the internalization of reliable energy transfer was particularly important, and various sources of environmental energy flux to drive such reactions early on have been proposed or favored over the years [99–104]. The most general common aspect of these suggestions is that volatiles from the atmosphere are chemically converted into larger organic molecules, which in turn are physically kept in place by various adsorptive interactions or binding forces. A potentially cooperative combination of chemical disequilibrium of geothermal or volcanic origin on the one hand and sunlight-induced charge separation on the other appears particularly attractive in the context of this paper. Such a fortunate coincidence, at anoxic primordial conditions, supposedly occurred frequently enough to be considered seriously; it is still perceptible at terrestrial geothermal fields [76,78], were it not for the highly oxygenated state of the modern atmosphere which would have severely interfered with basic key reactions in primordial biogenesis.

Colloidal nanoparticles of metal sulfides (MeS) only occur in significant amounts close to volcanic sources. These tiny grains have semiconductor properties and appear suitable as mineral photo-catalysts to have started organic synthesis cascades in the vicinity of volcanic vents [74–76,105,106], owing to charge separation at their surface in response to photon absorption from sunlight in the ultra-violet (UV) region [107,108]. Notably, UV sunlight has yet other photo-chemical effects of potential biogenic significance [109–111], and FeS clusters in various configurations are still central to the many electron transfer reactions mediated by modern FeS proteins [112,113], many of which are highly conserved and have very ancient roots. – Besides, a biological 'proof of principle' example has demonstrated that photo-active MeS–organic coupling can quite easily be made to work, using cadmium sulfide nanoparticles to induce self-photosensitization and photosynthesis in otherwise nonphotosynthetic bacterial cells [114].

It is not a long leap to envision that inorganic MeS nanoparticles and FeS cluster proteins have evolutionarily been connected by a gradual optimization process early on, involving minerals and self-aggregating hydrophobic peptides to start with, together with yet other organic amphiphiles. Such peptide-rich patches attached to photo-active mineral grains may thus have initiated the collective growth of polymolecular *organic hydrogels* in general [7] and membrane-like enclosures for *internal vesicular compartments* in particular [66]. The initially partial coverage of MeS particles with patches of predominantly hydrophobic organic matter would influence the duration of photo-induced charge separation. Various aspects were potentially amenable to further evolutionary optimization, such as the insulating capacity against the watery surroundings, the closure and permeability of biomembranes, the incorporation of organic electron donors and acceptors to utilize

 doi:10.20944/preprints201710.0153.v1

the 'free' energy in subsequent reactions and, in the longer run, the incorporation of organic antenna pigments to utilize other regions of the sunlight spectrum as well.

Resulting from mineral-based photocatalysis at daylight, organic matter could thus begin to accumulate, but emergent life must have outlasted the night and other periods of darkness. This could be achieved by degrading a limited amount of organic matter but saving the most critical components as long as possible. Notably, certain membrane-integrated energy-transferring protein complexes play similar roles in biosynthesis (anabolism) and degradation pathways (catabolism). With this in mind, and in consideration of the *generalized progenote hypothesis*, the endogenous *paraorganelles* held a key position for both anabolic and catabolic activities in early life. This notion has important implications concerning the influence of size.

In fact, the view that 'size matters' has been brought up before in a related context [27,68], entertaining yet allegedly rejecting the possibility of *eukaryogenesis* by other means than by the exogenous acquisition route for mitochondrial endosymbionts. The fundamental difference, in biophysical terms, means that the narrow prokaryote dimensions allow bacterial cells, as well as eukaryotic organelles, to be managed by diffusion-controlled kinetics, whereas the bulky eukaryotic cytoplasm depends on directional transport mechanisms in addition. As duly pointed out by Lane and Martin at that occasion [27], basic energetic principles and mechanisms are integrated differently into prokaryote and eukaryote cell types. It is worth noting, however, that the Lane/Martin postulate mainly applies to the high energy flows under aerobic conditions in the modern biosphere; this would have been different under pristine anoxic conditions, as fully demonstrated by the severe reduction or even the complete loss of mitochondrial remnants in various anaerobic, albeit heterotrophic, eukaryotes [42–45].

While electron transfer chains and primary ATP production in both bacterial and archaeal cells are tightly coupled to the cytoplasmic membrane surrounding the entire cell, the plasma membrane of eukaryotic cells is not involved in these essential processes. Moreover, the observed scaling differences between prokaryotic and eukaryotic cells are tremendous and affect all the major characteristics of intracellular relationships. As pointed out in the current essay, this innate difference in size is quite naturally related to the different routes of exiting from the *generalized progenote scenario*.

Notably, both volume and mass of the cytoplasm controlled by a modular genome in eukaryotes are larger by several orders of magnitude [68], as compared to bacterial or archaeal cells. The correspondingly low surface to volume ratio in eukaryotic cells is compensated for by elaborate membrane trafficking internally, together with a variety of interactive cytoskeleton components. In particular, primary energy procurement is delegated to less spacious organellar membrane systems, owing to their more favorable surface to volume ratio. On this observational basis alone, Lane and Martin 'conclude' that large and complex cells, such as eukaryotes, could not have existed before the exogenous acquisition of the α-proteobacterial progenitor clade to mitochondrial endosymbionts [27]. While much of this reasoning seems sound in general, a particular inference drawn appears unwarranted: that a singular event of mitochondrial acquisition by a prokaryotic host cell not only was an early event in the evolution full-fledged eukaryotes but also became the major cause of triggering all the innovations characteristic of eukaryote complexity in general. This evolutionarily unlikely singularity is not the only choice and the possibility of *endogenous compartmentation* should be taken into consideration as a more realistic alternative as well.

With this in mind, it seems advisable to recognize the potential role of *endogenous compartmentation* in organizing the precellular metabolism at the communal *progenote* state in the first place, as well as to reconsider its relevance for primordial *eukaryogenesis* thereafter, as supported by previous notions that eukaryote complexity is deeply rooted in complex primordial conditions, which already have prevailed at the common ancestral state [33–36,60]. These views fit in with the complementary theory that the emergence of *prokaryote* lineages has been dominated by genomic streamlining for high efficiency, rapid proliferation and minimal accessories, due to reductive evolution from a more complex source [37,115,116].

 doi:10.20944/preprints201710.0153.v1

4. Stage I: from Energy-Converting Vesicles to Bacterial 'Microcell' Escape

4.1. Bacterial Escape: Microcells of First Appearance

How the precursors to bacterial cells might have been nurtured as paraorganelles in the confines of the composite progenote community has already been mentioned above. These minigenomes of '*quasi-embryonic cells in the making*' needed many additional functions and genes before any 'escape' from the composite paracells was successful, granting the emigrant colonists not only with individual viability but also with perpetual propagation as modular cells. – What were the selective advantages of gradually accumulating such genes before the final mark was reached, so as to attain cellular and clonal independence?

In terms of Woesean transitions, the scenario is composite and has not yet been fully recognized as a genuine possibility before. The complexity arose since self-replicative plasmids in separate compartments were enabled to cross the Darwinian threshold on their own, but they did so while still depending on substantial support from the surrounding bulk of the communal progenote paracells, which collectively were still below the Darwinian threshold. Selective competition between distinctly different lineages, therefore, was not yet established at the organismal level, but it was already in effect internally – between the genomes of divergent paraorganelle lineages. In other words, as genomic competition was increasing between different lineages, this should intrinsically select for higher levels of selfishness – primarily against competing lineages of other paraorganelles, but later also against the host.

Presumably therefore, the ambivalent state led to directional changes, progressively shifting the balance of economic costs and benefits – not only in favor of the more rapidly evolving and diversifying paraorganelles but also in favor of more self-sufficient entities, which ultimately became resilient enough to leave the host for good. Further below, this line of reasoning will be resumed at the example of bacterial outer membrane functions. Presumably, the progenote trunk line of the host could not yet fight back with similar weapons, since it had not yet crossed the Darwinian threshold for the entire pan-genome. Today, however, the situation is entirely different in that the vertically stable lineages of eukaryotes can effectively compete amongst one another for permanent containment of their endosymbiont organelles, which over time has led to a substantial reduction of genome complexity and size in both mitochondria and chloroplasts [80,81]. The pivotal turning point was reached when also the eukaryotic ancestors had crossed their Darwinian threshold and thereby 'domesticated' the somewhat elusive pan-genome into modular, integrally manageable nuclear genomes. In addition, the paraorganelle lineages still residing inside and all the exogeneous endosymbionts to be acquired later on would be effectively domesticated as well.

As presumed for this scenario, the paraorganelles emerged as relay stations for the coupling of energy exchange to organic synthesis reactions and, therefore, held a key position controlling the overall growth rate of the progenote consortium. By recruiting more and more essential genes to the resident plasmid genomes, the paraorganelles gradually lessened their dependence on sharing resources with the surrounding bulk of interactive biomatter. In return, they needed fewer and fewer products of other essential reactions, for which the respective genes still resided in the variably dispersed pan-genome of the hosting progenotes. Conceivably, in this scenario, the presumptive takeover from mineral-coupled photo-activation to membrane-directed and protein-coupled energy procurement occurred in the complex progenote state already. In a first wave of internal compartmentalization, paraorganelle vesicles of the proto-bacterial lineage may have specialized on electron transfer reactions related to photo-reduction in the light, as well as oxidation of organic matter in reverse and in the dark.

Where does all this leave us in a pending contest between two outpost positions – sternly defended for so long? – The contending parties were split between the once-favored non-symbiotic ('*endogenous*' or '*autogenic*') origin of mitochondria [117] and the precarious assumption of an unprecedented singularity – "*once in four billion years*" – whence endosymbiosis between prokaryotes alone gave rise to mitochondria [27,50,118,119]. Somewhat miraculously, supposedly, this singular fusion then exploded in a creative burst of generating the entire cell complexity now characteristic of

 doi:10.20944/preprints201710.0153.v1

eukaryotes. Of recent research on this matter, and soundly based on comprehensive, comparative genomics, the *pre-endosymbiont theory* [39–41] strongly argues for the pre-existence of many mitochondrial functions in the presumptive host lineage that eventually took up and assimilated free-living α-proteobacterial cells as 'genuine' endosymbionts (of exogenous origin), which subsequently evolved into extant mitochondria and further reduced derivatives thereof. It takes yet more constructive thought to congeal some partial truth from either side of this, perhaps insubstantial controversy into a coherent and comprehensive theory to bridge one of the most glaring gaps from pregenomic protolife into the modern biosphere. As of late, the lingering controversy has even come full circle, reviving the endogenous origin of all mitochondria, as based on a phylogenetic cluster analysis of protein domain superfamilies in mitochondrial, bacterial and archaeal proteomes [120]. The results are not in conflict with the general concepts advocated here.

4.2. Thermoreduction in Ancient Thermophilic Lineages

It has long been surmised that life arose in a hot environment [103], which might explain the deep rooting of thermophiles and hyperthermophiles among both Archaea and Bacteria [121]. The general theory of a truly thermophilic origin of life, however, has been seriously questioned [115,122], in line with additional analyses, inferring a mesophilic state for the *last universal ancestor* or *ancestral state* [123,124]. The most severe problem arises from the chemical instability of polymeric RNA, especially at higher temperatures. This strongly argues against the hypothetical possibility that the Woesean progenote community as such might have emerged in a hot environment, since optimizing the genetic coding system of ribosomal protein synthesis supposedly happened in the primordial era of an emergent pan-genome, as being dominated by RNA.

Besides, in phylogenetic trees displaying thermophilic lineages, eukaryotes used to be the 'odd guys', not containing any hyperthermophiles at all and strangely being disconnected from a tentative thermophilic root, yet commonly dismissed for being considered uninformative to early evolution [124,125]. Together with the *thermoreduction hypothesis* of early prokaryote diversification [37,115], and taking the ecological setting of terrestrial geothermic fields into consideration [76,77,126], the *generalized progenote hypothesis* developed here can reasonably account for the early generation of thermophililic lineages, as well as the inherently mesophilic nature of early eukaryotes. Forterre's suggestion implies independent origins of archaeal and bacterial hyperthermophiles from mesophilic ancestors by *reductive evolution* and gradual adaptation to a more extreme environment [81,116]. This is supported by the derived and composite nature of reverse gyrase, a universal constituent of the hyperthermophilic lineages among prokaryotes [127]. Moreover, this particular DNA topoisomerase is also the only hyperthermophile-specific protein [128], which supposedly originated just once in very ancient times. The somewhat erratic distribution among very few bacterial lineages and most of the Archaea, however, is still open to discussion [129]. A tentative gene transfer from Archaea to Bacteria [130] is certainly not the only interpretation possible.

The recent consideration of terrestrial geothermic fields as an appropriate arena for the early transition from photo-activated and mineral-catalyzed organic syntheses to increasingly self-organized and self-propelled protometabolic systems [76,77,126] places origins of life in a patchwork of diverse microenvironments, potentially supportive to different forms of early life in close proximity to one another. As pointed out before [7], the metabolic takeover of photosynthetic activities from essentially geochemical reactions may reasonably have occurred in damp and cooler organic mats at the periphery of geothermal vents, rather than in the hot bulk volume of hydrothermal pools. Later on, Forterre's *thermoreduction* principle allowed tiny specialized cells with minimal genomes to colonize and populate the more hostile volume of hot-spring pools in the vicinity of the subaerial mats of progenotic paracells. The evolutionary damage award to compensate for the physiological hardship in hot brines must certainly be seen in the various possibilities of chemoautotrophic growth, if only thermoresistence of other vital features could be obtained by adaptive optimization. As this chemoautotrophic alternative has its local optimum

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2017 doi:10.20944/preprints201710.0153.v1

closer to the source of geothermal flows, it is appealingly complementary to the local optimum for photoautotrophic growth at the damp and cooler subaerial mats.

In line with such reasoning in general, much of Kandler's earlier conjecture may still be valid "*that pre-cells based on H_2/O_2 chemolithoautotrophy, and thus metabolically resembling the extant 'Aquificiales'* [Aquificae], *evolved faster and underwent cellularization earlier than the metabolically more archaic pre-cells based on H_2/S^0 and H_2/CO_2 chemolithoautotrophy, which amalgamated into the domain Archaea*" [70]. The hallmark of hyperthermophilic microorganisms, reverse gyrase, has actually resulted from an ancient fusion event combining a topo-IA domain with a helicase-like extra domain [131]. It now appears that the corresponding phylogenies for topoisomerase IA and reverse gyrase converge in a lineage ancestral to all the hyperthermophilic bacteria (*Aquifex*, *Thermus*, *Thermotoga*; in this order) [132], where the peculiar fusion with a helicase domain can have occurred in the common ancestor. This is readily compatible with the hypothesis that the most ancient 'escape' of paraorganelles from progenote paracells as the first independently viable microcells was actively selected for by thermo-adaptation and subsequent expansion into the hot environment of hydrothermal pools.

According to this view, the *Aquificae* phylum represents a particularly ancient line of free-living bacteria, and other lineages may have 'escaped' from the residual, still mesophilic progenote population at later occasions under different circumstances. As a corollary to this, in a later wave of secondary thermoreduction toward the lineage escape of archaeal microcells, the bacterial 'invention' of reverse gyrase may have been acquired laterally by the common ancestor of Archaea. Being metabolically complementary to H_2/O_2 chemolithoautotrophy in thermophilic bacteria, the Archaea would have optimized the modes of H_2/S^0 and H_2/CO_2 chemolithoautotrophy from the progenote holoplasm outside the bacteria-type paraorganelles.

It is one thing to suppose that the first rounds of lineage escape were releasing thermophilic microcells into the hot environment where the mesophilic paracells of the progenote community could not follow suit. It is another to consider the ultimate fate of other paraorganelles remaining under temperate conditions. Deep-branching bacterial phyla indicate indeed that different bacterial lineages may have originated from such a source at later times, without having passed through any hyperthermophilic stage.

4.3. One or Two Membranes Surrounding Bacterial Cells – a Deep-Branching Tale

Many bacterial cells are surrounded by two membranes (*diderms*), whereas others have only one (*monoderms*). The phylogenetic significance of this fundamental difference, however, is far from clear [133–138]. On different grounds, a deep split among primarily mesophilic bacterial phyla has been recognized as *terrabacteria* vs. *hydrobacteria* [139], with ancient adaptations to life on dry land or in aquatic environments. By and large, the *monoderms* and *diderms* are clustered in *terrabacteria* and *hydrobacteria*, respectively. Yet, this correlated dichotomy has exceptions, such as diderm *Cyanobacteria* amidst the other terrabacteria, and it is not representative for all the bacteria, leaving some additional diderms aside. Notably *Thermus*, *Thermotoga* and *Fusobacteria*, which are commonly placed at the base of the bacterial phylogram [134], are diderms. This indicates that the common bacterial ancestor already had two surrounding membrane systems. Accordingly, the single membrane around monoderms is considered a derived trait, viewing the loss of the outer membrane in many terrabacteria as one of their selective adaptations to long-term survival on dry land. This leaves the question of how the peculiar outer membrane system arose at all much earlier (or perhaps more often than just once).

At any rate, neither membrane system can be understood as merely consisting of a phospholipid boundary. Their respective membrane proteins are yet more important for biological function. The single monoderm membrane and the inner membrane of diderms have in common that all their transmembrane proteins are composed of tightly packed α-helix bundles [140]. On the other hand, the more spacious tunnels of β-barrel porins are exclusive to and highly characteristic of the outer membrane in diderm bacteria and related endosymbiotic organelles [139–142]. In light of *pre-endosymbiont theory* [39–41] and the *generalized progenote hypothesis* developed here, I conjecture

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2017 doi:10.20944/preprints201710.0153.v1

that the emergence of the peculiar outer membrane proteins (OMPs) could be derived quite naturally from the composite paracell organization in the progenote community.

Interestingly enough, similar tunnel-forming proteins are also produced by certain monoderm bacteria, such as staphylococcal α-hemolysin, which forms heptameric β-barrel pores [143]. Yet, this protein does not affect the cell's own membrane but is exported for punching transmembrane pores into other cells. By the same token, early precursors to present OMP tunnels may originally have been produced by '*pre-endosymbiont*' paraorganelles to be inserted into other, *extra-organellar*, membranes (Fig. 1). They could thereby have been very instrumental for nascent cells in paving the way into the outside world.

The pioneering *pre-endosymbiont theory* is based on the chimeric nature of the modern mitochondrial proteome in eukaryotic cells [39–41]. This is inferred from the somewhat surprising observation that the bulk of mitochondrial proteins appear to have arisen outside of the α-*Proteobacteria*, from which the endosymbiont lineage of all modern mitochondria is supposed to have originated [47]. In particular, the hypothetical premitochondria-carrying host to receive the first α-proteobacterial endosymbiont may have possessed a trans-membrane protein import system, which may have facilitated the acquisition of free-living α-Proteobacteria and their eventual transformation into genuine organelles [40]. Several core components of this import machinery have no bacterial homologs [144] and were already present in the common ancestor of present Eukarya. By inference, therefore, they may likewise have existed in the communal progenote paracells already.

In this perspective, the emergence of mitochondria no longer figures as a miraculous singularity but turns into a readily comprehensible step of a graded evolutionary series. I here propose to extend this mode of gradual evolution into the consolidation stage of microgenomes, well before the release of genuine microcells. It is also relevant in this context to look at the relative timing of mitochondrial acquisition in comparison to the emergence other sub-cellular compartments in eukaryotic cells [145]. This study has estimated proteomic diversion times from differential tree partitioning with regard to the functional networks of nucleolus, nucleus, endomembrane system and mitochondria, respectively. Notably, mitochondrial diversion was found to have begun latest in this ordered innovation series.

As already noted before, the assumption of compartmental microgenomes, embedded in a larger collective whole, has interesting kinetic implications in terms of evolutionary *organelle versus host* competition. That line of argument is here resumed at the example of emergent communication between evolving paraorganelles and the outer environment. To begin with, the small initial vesicles with just a few genes in their genomes were not yet in direct contact with the outer world (Fig. 1 a,b). Environmental fluctuations were filtered and attenuated through the progenote holoplasm lying in between. As time went by, the more the evolving paraorganelle lineages engaged in direct competition with one another for faster and more efficient response to environmental stimuli, the more advantageous became a direct window to the outer world.

Positioning in the paracells, for instance, became selectively important, since minimizing the distance to the outer boundary decreased the attenuating effect of the intervening protoplasm against environmental stimuli upon the paraorganelle vesicles. After coming close to the boundary repeatedly by chance (Fig. 1 c), it became selectively important to stay there firmly and more permanently (Fig. 1 d). Once being associated with the cytoplasmic membrane of the communal paracells, it further became advantageous to gain direct access to the outer world by punching sizable holes into the outer membrane within the area of contact (Fig. 1 e), even though that structure was not originally part of the paraorganelle's sphere of influence. – As an aside, according to this model, the functional import of bacterial-type lipids into proto-eukaryote membranes, as accompanied by the transfer of the corresponding genes from the paraorganelles into the proto-nuclear pan-genome, may have occurred during the extended period of inter-membrane association when the paraorganellar/bacterial outer-membrane system came into existence.

Approaching self-sufficiency by further enlarging their genome at such an exposed position, the most successful paraorganelles would eventually outgrow their host as biofilms and/or

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2017 doi:10.20944/preprints201710.0153.v1

suspensions of free living microcells (Fig. 1 f). This is a relatively straightforward way to understand how the characteristic double-membrane envelope of diderm bacteria emerged quite early in the bacterial phylogeny. A crucial step to integrate two membranes into a coupled modular system is indicated by the formation of structural connections between the membranes (Fig. 1 d), of which modern bacteria have several types. Some of these [146] became important at the time of cell division (Fig. 1 g). For so long as the early bacteria kept on growing in aqueous surroundings, there was no need to shed the newly acquired outer membrane, and more complex cell wall structures assembled in the intervening periplasmic space. In dry air, however, especially during extended periods of desiccation, the outer membrane system may have become a handicap rather than a blessing.

One way of losing outer membranes altogether might be seen in the germination process from draught-resistant spores or cysts. Notably, the formation of endospores and/or arthrospores is frequently observed in monoderm *Firmicutes* and *Actinobacteria*, the largest groups of terrabacteria, but spores are rare among diderm gram-negative bacteria. In eukaryotes, too, spores and cysts are frequently observed as particularly durable and resistant survival stages, and even a peculiar type giant virus particles – Pandoravirus – bears striking morphological resemblance to heavily encapsulated spores, which are designed to release their contents through a specialized 'germination pore' after ingestion by an unfortunate amoeba predator cell [147]. Yet more peculiarly, the remotely related Pithovirus was even discovered from Siberian permafrost, whence similar looking spore-like capsules managed to survive for some 30,000 years [148].

A certain note of caution is warranted, however, since the hierarchical clustering of insertion/deletion (indel) signatures separates monoderms and diderms in general as the deepest branching sister groups in eubacterial phylogeny [149], which seems to weaken the assumption that the diderm pattern as such is representative of the earliest free-living bacteria. On the other hand, the *generalized progenote hypothesis* proposed herein is very open to the possibility that lineage escape of bacterial cells has happened more than just once. In this relaxed scenario, monoderm bacteria may or may not have left the progenote consortium directly, and several deep-branching phyla of diderm bacteria may have resulted from multiple escape events along the general route depicted in Fig. 1.

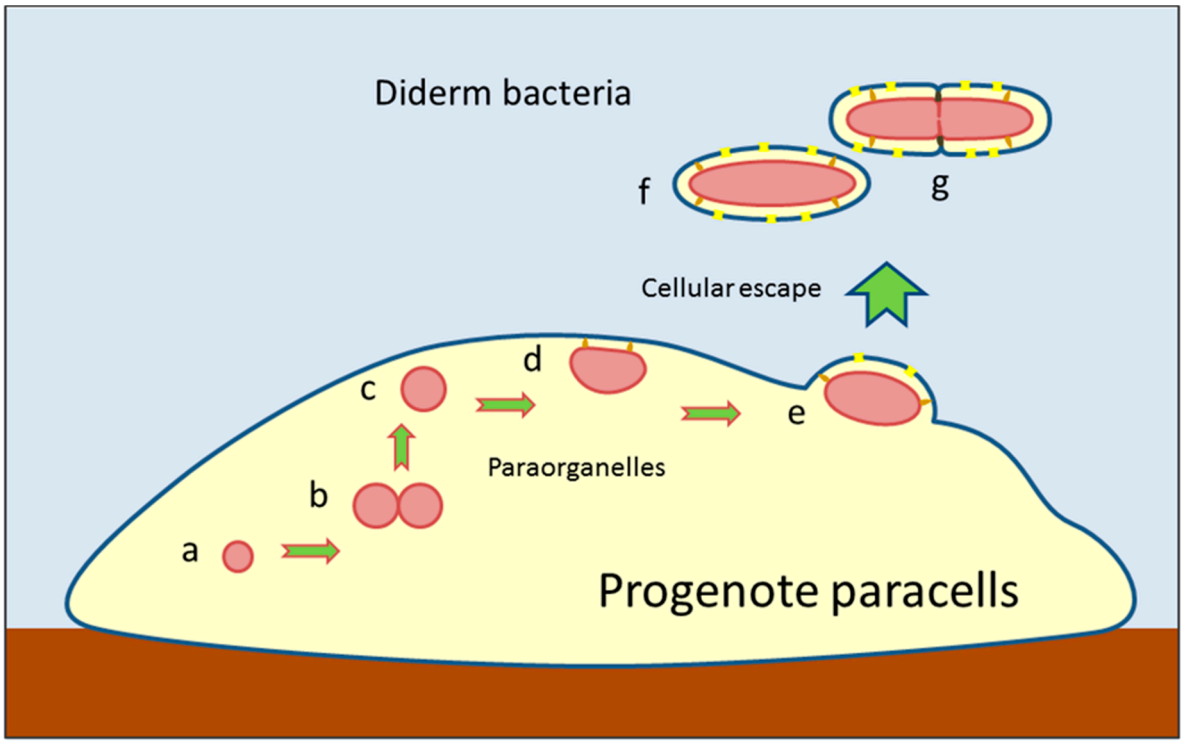

Figure 1. Emergence of diderm bacteria from preexisting paraorganelles of progenote paracells. See text for detail.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2017 doi:10.20944/preprints201710.0153.v1

5. Stage II: Residual Trunk-Line Evolution to Eukaryotes and Archaea

After the latest *lineage escape* of *bacterial microcells*, according to the scenario stated above, the residual *paracells* of the *progenote* trunk line did not cease to exist immediately. This is the essence of Woese's *Darwinian asymmetry* at such first-level bifurcations from the communal progenote state. Later on, this *archgenote* population of still largely communal paracells gave rise to both *archaeal* and *eukaryotic* lineages of *micro-* and *macro-cells*, respectively. This leaves the question of why and how these two cell types have become so incommensurably different.

As mentioned before, the problems of *eukaryogenesis* are far from being solved [29–35,50–53,118,119,150–156] and the present paper is not up to settling controversies once and for all. I want to emphasize, however, that the *generalized progenote hypothesis*, as here proposed, can offer a new perspective for further studies. The present mantra of "*Eukaryotes arose from prokaryotes – period!*" ows much of its appeal to the lack of a readily available alternative. The possibility of endogenous compartmentation opens up for a meaningful and reasonable negotiation between extreme and controversial views. As early denoted by Ford Doolittle, eukaryotes emerged from a peculiar "*nuclear-cytoplasmic lineage*" [93], in which the characteristic nuclear–cytoplasmic relationships evolved to levels of high sophistication before the last common ancestor of all present eukaryotes. Much of these intrinsic systems properties can now be rationalized as never having been part of tentative ancestor cells that were organizationally similar to prokaryotes living today.

5.1. The Puzzling Problem of Eukaryote Complexity

Generally speaking, "*bacteria simply have a fundamentally different strategy for cytoplasmic organization as compared to eukaryotes*" [157]. Very summarily, the most significant differences in system properties among the various evolutionary stages are compiled in Table 1. The evolutionary implications of this overall comparison become more relevant below, concerning the peculiar phylogenetic relationship between Archaea and Eukarya, which presumably unfolded in the intermediate 'Archgenote' stem line – before the diversion of archaeal and eukaryotic lineages, but after the separation from bacteria.

Considering the significance and magnitude of the organizational difference between prokaryotes and eukaryotes, together with profound resemblance in organizational characteristics between eukaryote cells and the pregenomic/precellular progenote stage (Table 1), it is astounding how little attention has been paid to rationalizing the posited transformation of one of these strategies into its complementary counterpart. It is a major issue of this paper to point out that the unlikely implications of this conundrum can be avoided altogether by considering the generalized progenote hypothesis, which only calls for the gradual optimization of the two contrasting organizational strategies in parallel, assuming this happened in different compartments of intrinsically composite progenote paracells. – Alternatively, one might begin to entertain the somewhat remote possibility of recreating progenote-like conditions at a later stage, reintroducing cytoplasmic confluence and communal sharing of all constituents throughout entire populations of symbiotic prokaryote cells and/or secondarily generated paracells with multi-partite, variable and highly redundant paragenomes of prokaryotic origins.

Table 1. System properties prevailing before and after organismal consolidation.

	Progenote I [1,2]	**Progenote II** [1,3]	**Bacteria**	**Archaea**	**Eukarya**
Noncoding RNA [4]	prevalent	abundant	spurious		high
RNA splicing [1]	?	abundant ?	spurious		abundant
Genophores [5]	>>> 1	>> 1	1		>> 1
Virosphere	RNA	RNA, DNA	DNA		RNA, DNA
Modular units	none	emergent	prokaryote cells		compartmental [6]
Subcompartments	?	emergent	spurious [7]	spurious	prolific
Body/cell size	emergent	substantial	small		large

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2017 doi:10.20944/preprints201710.0153.v1

Cohesiveness [8]	intrinsic	profound	spurious	?	profound
Mixability	unlimited	protoplasmal	DNA-LGT		sexual fusion
Line of descent	universal trunk lineage		clonal strains		trunk-line species

[1] conjectural, as presumed in the text; [2] early stage of RNP emergence; [3] later phase of emergent take-over by genomic DNA; [4] discounting rRNA and tRNAs; [5] different gene-bearing molecules defining the genome of a vertical lineage; [6] modular nuclei before modular cells (nuclear division often being uncoupled from overall cell division) and organellar (cytoplasmic) compartments with residual prokaryote-like genomes; [7] presumably beginning with vesicle-enclosed plasmid-like minigenomes; [8] internal connectivity among cytoplasmatic proteins.

The organizational schism between prokaryotes and eukaryotes calls for a choice between two very different evolutionary alternatives.

***Scenario 1*:** Eukaryotes emerged from prokaryote ancestors. This possibility must have gone through many intermediate stages, none of which has left extant descendants other than the monophyletic lineage of the last common ancestor of eukaryotes themselves, including the surviving eukaryotic progeny.

***Scenario 2*:** Eukaryotes and prokaryotes shared a common ancestor only in their distant past, and the organization of this common ancestral stage was neither fully eukaryotic nor prokaryotic in the present sense.

Hypotheses to resolve the impasse are many, but most of these are not focused on the central issue of critically discriminating between the two principle alternatives given above. Instead, the vast majority of current work is taking *Scenario 1* for granted, without asking further questions about its unsettled base of how the first prokaryote lineages of fully propagative and self-sufficient cells came into being in the first place. The nonconventional views presented here, however, are fully embracing *Scenario 2* on the grounds that functionally complementary subsystems of endogenous origin could early on be optimized in parallel, until eventual specialization resulted in self-propagative and fully competitive lineages of differently constituted organisms.

In stark contrast to the classical majority at large, Carl Woese put focus on the essentials of the problem, and he also gave valuable cues for further investigation. As I see it, his major theoretical insights are about three-fold. Most significantly, the emergence of vertically stable lineages of modular cells and cellular organisms is conceived as being dependent on the emergence of stably transmissible organismal genomes (demarking the threshold for a *Woesean transition*). Before that pivotal change, the collective evolutionary optimization of interactive multi-component systems proceeded steadily in a miscible population of non-modular, precellular entities (the communality of a progenote state). Each point of exit from the collective progenote community defined a branchpoint toward a stable lineage (or Darwinian species). Inevitably, this process implies an event of asymmetric bifurcation (the '*Woesean asymmetry*'), in that the new vertically branching lineage and the residual progenote trunk-line community continued to coexist side by side thereafter.

Above all else, Woese's theoretical framework has general validity and is independent of particular mechanistic detail as to how a stably transmissible genome could be generated or how an autonomously viable cell might be organized, as well as to how the miscibility at the population level required for the communal progenote state was actually brought about. For that matter, these superior insights are also independent of whatever illustrative detail Woese himself happened to have in mind when he put his abstract concepts into print for public presentation, and it is up to further investigation to identify eventually missing links or possible extensions.

Woese's notions have very significantly contributed to expand evolutionary theory into the primordial phase of Darwinian speciation, which Darwin was not yet capable of looking for himself. And just as Darwinian Theory did not cease to be developed after Darwin's death, the Woesean extension to Darwin's theorem is here to stay and be continued. Obviously, a critically missing part in a comprehensive theory of early evolution is a better comprehension of eukaryote origins as well. It is the objective of the present essay to develop generalizing options for closing the gap.

Important additions advocated here concern protoplasmic coherence and confluence, as well as endogenous compartmentation. All these features are here considered to be intrinsic to and highly relevant for early evolution – already at the communal progenote state of the Woesean paradigm.

 doi:10.20944/preprints201710.0153.v1

Even though neither one of these features has been analyzed as such in Woese's papers, their consideration as potential factors under the basic assumptions would not affect the generality of Woese's theoretical framework as stated above. On the other hand, it is the evolutionary integration of just these properties as potentially primordial traits that draws the otherwise enigmatic emergence of eukaryotes into the limelight of the Woesean perspective as a rationally comprehensible transition. The possibility of endogenous compartmentation, in particular, would allow progenote paracells (as described above) to act as common precursory stages to both bacteria and eukaryotes, owing to their respective crossing of the Darwinian threshold by distinctly different strategies and means, leaving several intermediate options for archaea to derive from such composite paracells as well.

As stated before, the Woesean notion of asymmetric branching at the Darwinian threshold level implies the possibility of multiple escape events for the emergence of bacteria-like cell lines from the communal progenote population of internally compartmented paracells. The same inference is relevant for the eventual transformation of variable paracells into vertically stable lineages of eukaryotic organisms, as well as for the intermediate branching off of various archaeal lineages. Conceivably, approaching the critical Darwinian threshold level from below must inherently have been a very gradual transition process, involving the adaptive optimization of multiple, partly redundant, structural and functional parameters. It is also reasonable to assume that smaller and smaller subpopulations, in environmentally isolated conditions, sufficed to accomplish the Woesean transition on their own. This is equivalent to presuming that multiple crossings of the Darwinian threshold became more likely toward the end of the overall transition period for leaving the communal progenote state, which then led to the emergence of several vertically stable lineages with eukaryotic cell organization but significant differences in the founding genotypes.

Such multiple crossings would likely have resulted in the founding of a bushy tree with several branches from a not readily resolvable common origin, which actually is observable around the bases of all three Woesean domains of organisms. Eukaryote phylogeny, in particular, is deeply rooted among the wide range of protist lineages emerging around a common source [158–162]. The inferential tree diagrams have a rather bushy appearance at their root indeed (for a particularly bushy depiction of this relationship, see Fig. 7 of [163]). All the protist cells of this large variety [164] are fundamentally different from both bacteria and archaea, not to mention the few yet even more conspicuous clades of multicellular organisms that independently have arisen from some of the protist groups later on. Yet, several deep-branching protist lineages are also characteristically different from other protists, so some of their peculiar traits may already have been established as primitive poly-phenotypic possibilities in the trunk line of the Woesean/Darwinian transition period, whereafter they were further elaborated on in just one or few of the various protist phyla.

Of relevance to the largely unexplained phenological differences between eukaryotic and prokaryotic cells, the rise of modern genomics has opened new perspectives to this expanding field. A characteristic set of extra genes is present in virtually all the major subdivisions of eukaryotic organisms but, with rather few exceptions, is absent from both archaea and bacteria. These extra genes encode the so-called *eukaryotic signature proteins* [165,166] and contribute to a corresponding set of *cellular signature structures* [33], such as nuclei, nucleoli, spliceosomes, Golgi apparatus, endoplasmic reticulum, centrioles and – very conspicuously – molecular tracking motors [147]. Some functional aspects are further discussed below. By inference, therefore, many genes for this signature set of proteins and their functions can have separated from bacteria-specific gene phylogenies already at the *progenote* stage of composite, internally compartmentalized *paracells*.

As a corollary to the *eukaryotic signature* conception, all the genes encoding the *signature proteins* were supposedly present in the last common ancestor of all extant eukaryotes, and the *signature structures* were already expressed to some extent at that ancestral stage. Presumably, therefore, that enigmatic ancestor of *eukaryote macrocells* was a potentially dimorphic creature, capable of phagocytosing *prokaryote microcells* in both an amoeboid and a ciliated state; thus alternating between these phases in response to fluctuating conditions in the environment was a common ancestral option as well.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2017 doi:10.20944/preprints201710.0153.v1

As is indicated in Table 1, potential relics from an ancient RNA or RNP World are more profoundly represented in eukaryotes than in bacteria [167,168], which poses a severe challenge to the conventional *eukaryotes-from-prokaryotes* presupposition. In functional terms, this overabundance of eukaryotic noncoding RNAs involves a range of special guide RNAs for the modification or recombination of other RNA molecules, such as rRNAs, tRNAs and spliceosomal introns, as well as the composite spliceosome RNP machines themselves. The putative anciency of introns in eukaryotic mRNAs [169,170], however, has become subject to a highly contentious debate, which culminated in the claim that spliceosomal introns might rather have derived from rogue bacteria-born self-splicing introns [171,172], in the aftermath of the supposedly unique event of mitochondrial acquisition. On the other hand, the alternative position that "*bacterial group II introns share a common* [considerably earlier] *ancestor with spliceosomal introns*" can still be considered a "*more likely interpretation*" [173], and the generalized progenote hypothesis proposed herein is certainly in line with this conclusion.

In comparing general protein properties too, characteristic differences exist between bacteria and eukaryotes. It is common knowledge that permeabilized bacterial cells release much of their cytoplasmic contents quite freely, but this is not generally the case for eukaryotic cells [174]. A key factor in this difference is the predominance of soluble proteins in the bacterial cytoplasm, whereas more sticky proteins are predominantly associated with the cytoplasmic membrane [175]. This is in contrast to eukaryotic cells, in which most internal spaces are populated by fibrillar networks, such as the tubulin/actin-based cytoskeleton [176] and the less well characterized nuclear matrix [177]. At times, these internal meshworks are structurally quite robust but, now and then, rigidity alternates with periods of dynamic reorganization, involving various interaction partners and anchoring components. These differences point at antithetical strategies for organizing the cytoplasm of eukaryotic and bacterial cells, an important aspect to be taken up further below.

5.2. Propagation and Sexuality – Division and Mixis – Mitosis and Meiosis

The characteristic cytoplasmic infrastructure of eukaryotic cells is one thing, and much of this structural and organizational complexity may already have been established by collective optimization for growth as such and by endogenous compartmentation, as assumed and generalized by the *extended progenote hypothesis* herein. What really counted for eukaryotic lineages in crossing their Darwinian threshold(s) as modular organisms, was having a reliable propagation system with modular genomes as well. As a corollary of this hypothesis, I find it most natural to assume that the intrinsically multi-partite paragenomes of progenote paracells in general were subject to more stochastic fluctuation and, therefore, followed different evolutionary dynamics than the monomolecular (plasmid-like) genomes of vesicular paraorganelles. If (and what a big "if"!) there indeed was a direct connection from the primordial progenote state to intrinsic features of eukaryotic infrastructure, this evolutionary link must have mastered the effective modularization of multipartite nuclear genomes in the first place, and that was neither a trivial nor a simple matter.

With hindsight from a biocentric perspective, it appears that some of the most distinctive cyclic features of eukaryotic organisms have more in common with the stochasticity and communality inherent in the posited progenote state than conventional wisdom might expect. The complex and coordinated changes in eukaryotes recurring in cell-to-cell and generation-to-generation cycles have rather little in common with propagation of bacteria. In this introductory presentation, I will not go into detail with the many characteristics of cytoplasmic infrastructure but concentrate on the genome-related traits for developing a plausible eukaryote-specific alternative for crossing the Darwinian threshold from the collective community of progenote paracells.

In very general terms, the eukaryotic cytoplasm is dominated by the molecular '*swarm intelligence*' of multiple protein–protein interactions. This partly stochastic network greatly extends to the participation of noncoding RNAs in the nuclear–cytoplasmic relationships. To borrow a buzzword from evolutionary machine learning algorithms, I think that eukaryote complexity has much to do with '*Particle Swarm Optimization*' [178], exceeding bacteria-like organisms by far in these regards. The Woese school has long maintained that collective system optimization is deeply rooted

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2017 doi:10.20944/preprints201710.0153.v1

in the progenote state already [8,14], and this notion has been substantiated by modeling *in silico* [18]. Conceivably, the collective optimization of *molecular swarm intelligence* could have extended to the coordinated management of large and multipartite genomes in the eukaryote-specific contribution of Woese's progenote population to the modern biosphere. In eukaryotic organisms, for that matter, the exquisitely orchestrated ballets of chromosomes performed at each mitotic division – and even more so at meiotic prophase – represent outstanding examples of molecular *swarm intelligence*, which must have emerged and evolved over a long period of *Particle Swarm Optimization*.

Presumably, from a wide range of stochastic interactions in communal paracells, more time was required to reach the Darwinian threshold along this additional route, involving complex and non-linear relationships, than what it took plasmid-like monomolecular genomes to organize the first free-living bacterial cells by adding single genes one at a time to a preexisting replication unit. Indeed, the manageability of eukaryotic genomic material in general depends on numerous interactive components, many of which are related to cytoplasmic functionality as well; but only a fraction of the latter category is responsible for the coordinated propagation of multiple chromosomes as modular genomes. The following aspects of eukaryotic genome management are more easily understood in terms of collectively optimized multi-protein complexes and/or RNAs.

- The nucleosome-based chromatin structure, together with histone-coded variations is relevant for the regional control of gene expression [179], and bulky mediator complexes are responsible for the local activation of transcriptional RNA polymerases [180]. Heterochromatic variations of the histone code are also important for chromosome segregation at their centromeres [181].
- The nucleolar subcompartment of the nucleus and other 'nuclear bodies' of RNA reorganization [182] maintain their corpuscular integrity without being surrounded by membrane boundaries.
- The *nuclear envelope* separates one multipartite genome from others, but it also is related to and physically connected with the endoplasmic reticulum in the eukaryotic cytoplasm [183].
- The *nuclear pore complexes* (NPCs), which regulate macromolecular traffic into and out of the nucleus, are composed of some 30 different proteins, being arranged in 8-fold rotary symmetry around the central pores. Virtually all of the various protein domains present in these complexes are also involved in many other aspects of membrane trafficking in the cytoplasm [183–185].
- The microtubules, actin fibers and other components involved in genome segregation during nuclear divisions are also, in the long growing phase between divisions, engaged in motile activity throughout the cytoplasm [186].
- Various kinds of posttranslational protein modification, such as phosphorylation and dephosphorylation, are used in the temporal management of nuclear and cytoplasmic activities in similar ways.

To integrate these disparate features within a common frame of understanding, I presume that the most fundamental differences between bacterial and eukaryotic strategies, not the least their characteristic modes of chromosomal segregation, were optimized in parallel but to rather different ends. This divergent optimization may have continued for a long time after the initial separation into endogenous sub-compartments of a common protoplasmic whole, which herein is presented as an unconventional and hitherto unthought of variation of Woese's ancestral progenote state in general.

I strongly suspect that evolution has found a direct way to organize a multipartite genome by a gradual bundling process. This must have started from the many independent gene-bearing molecules expected to populate the early progenote state in the transition period from RNA- to DNA-based chromosomes. I also suspect that this gradual bundling process included the still enigmatic *Masterpiece of Nature* [187]: *The Evolution and Genetics of Sexuality*. The ancient system of sexual reproduction is part of the eukaryotic signature package of intrinsic functions. Generally speaking, it superimposes long-period generation cycles upon the stereotype pattern of short-period

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2017 doi:10.20944/preprints201710.0153.v1

mitotic nuclear division cycles. With much variation in detail, the composite life cycles involve the balanced alternation of sexual fusion of complementary partners and *meiotic* genome reduction.

Mechanistically, the additional features, particularly those characteristic of meiosis, comprise a variable range of modifications to vegetative processes ensuring growth and maintenance in eukaryotic organisms. Contrary to conventional wisdom, however, I am not convinced that these modifying features were just added onto a fully developed mitotic mechanism at some later stage. Alternatively, meiosis as such may have come into existence together with the mitotic apparatus of dividing vegetative nuclei. By complementary perfection of different principles, the progressive coevolution of mitosis and meiosis would naturally have linked the vertical lineage stability of eukaryotic species to the inherent stochasticity of multipartite probability distributions prevailing in the communal gene pool of the Woesean progenote state. Together with David Penney [188], I have discussed in more detail how the bimodal nature and cyclic alternation of mitotic and meiotic division mechanisms could gradually have evolved through intermediate stages, as driven by alternative needs at different times of the life cycle in a periodically changing environment.

This is to say that the best of confluent, pregenomic paracells was carried over into the Darwinian world of competitive speciation by the co-emergence of conservative genome segretation for multiple chromosomes in mitosis, periodic sexual fusion and controllable genome reshuffling in meiosis. By this token, the proneness to protoplasmic confluence, as here assumed for the progenote state, was channeled into the temporal and intra-species restrictions of programmed sexual fusion – by complementary sensing and cytoplasmic membrane fusion first and pairwise nuclear fusion later on. In addition, the "*rampant and pervasive*" occurrence of horizontal gene transfer (HGT) [18], as assumed by the Woese school for the progenote gene pool, was funneled into periodic intra-species exchange and the limited facilitation of homologous recombination in meiotic prophase. In principle, therefore, each Darwinian species on its own, comprising conspecific populations of sexual eukaryotic organisms [67], is still benefiting from similar gene pool dynamics as were at work throughout the entire progenote community.

5.3. Viewing Archaeal Descent as Being Intermediate between Bacteria and Eukaryotes

Traditionally – before the Woesean revolution of molecular phylogenetics – all the microbial cells without nuclei were pooled together as '*prokaryotes*', which is based on considerable resemblance of archaeal and bacterial cells (Table 1). This put a strong emphasis on the phenotypic properties they had in common. In general, bacteria-like cells are small in size and simple in terms of structural organization. Their unitary genomes are being contained on circular molecules of DNA, transcription and translation occur in the same compartment, functionally related genes are often clustered with operon-type coordination, and the surrounding cytoplasmic membrane is host to energy converting systems as well as protein secretion. On the other hand, the disparate branching pattern at the bottom of the canonical ToL – together with the *generalized progenote hypothesis* developed here – gives particular significance to the structural and organizational differences between bacterial cells on the one hand and archaeal cells on the other.

What differences could have been crucial enough to give proto-archaeal newcomers the competitive edge to leave the postulated polyphenotypic progenote consortium successfully at a later stage, when free-living populations of bacterial cells supposedly were well established? – This is one of those questions that cannot be answered by genomics alone. It also begs the more general question of how the residual *trunk-line* population of *progenote paracells* might have been capable of surviving and evolving further on, yet facing ever more stringent competition from the newly established cellular lineages of rapidly diversifying, specializing and further optimizing bacterial clades. Physiological and/or ecological possibilities are worth considering in addition to genomic data. At any rate, there is mounting evidence that the common ancestor of Archaea and Eukaria was considerably more complex than any archaeon living today [37,153,189]. On the one hand, the "*dispersed archaeal eukaryome*", as inferred in retrospect from comparative genomics [153], is indicative of considerably higher degrees of resemblance with the basic eukaryotic toolbox than hitherto expected; on the other hand, as viewed prospectively from the Woesean vantage point

adopted in this paper, the elusive ancestor of both Archaea and Eukaria was also very much closer to the communal *Progenote state* than hitherto appreciated – not only in absolute terms of evolutionary time but, presumably, in terms of subcellular diversification as well. To stress the latter point, considering the general impact of *Woesean asymmetry* on this issue, I herein refer to the residual trunk-line lineage leading up to the second dichotomy in the canonical ToL as the *Archgenote stem line* toward both Archaea and Eukarya.

Syntrophic cooperation is a powerful driver of microbial biofilm associations today [190–193]; it may also have been operational at the progenote stage already. Considering the anoxic state of the early Earth, a tightly coupled interaction between the subsystems that respectively produce and consume hydrogen and/or formate – "*the paradigm for anaerobic metabolic cooperation*" [194] – may have very ancient roots. Conceivably, the paracellular progenote community already made a living on the rudimentary beginnings of this cooperative principle. In particular, a first round of subcellular specialization may have gathered hydrogen-/formate-producing activities and FeS cluster formation in the paraorganelle vesicles proposed above, which subsequently developed further into the first free-living cells of bacterial lineages. Conspicuously, the formation of FeS clusters is the last recognizable activity remaining in severely reduced mitochondrial remnants before even that relict was lost in the anaerobic microbe *Monocercomonoides* [49], where the essential supply of FeS clusters has been rescued by lateral gene transfer of an alternative mechanism of cytosolic components.

Conversely, the complementary hydrogen-/formate-using activities would remain localized in the paracell holoplasm, outside the proto-bacterial vesicles, for quite some time. Eventually the latter may have become concentrated in a second generation of paraorganelle vesicles, together with a particular mini-chromosome from the otherwise fragmented and decentralized '*protokaryote*' pan-genome – still in the paracells of the residual *archgenote* trunk-line population. By that time, the replication mechanism of *paracell minichromosomes* would have diverged considerably from the plasmid-derived counterpart in the bacterial-type paraorganelle lineages [94,95,153,189]. Eventually, a few such minichromosomes – enclosed in their 'private' vesicles – may have diverged away from the bulk pan-genome, which remained outside of these vesicles. Certain key functions invented by bacteria-type plasmids to manage circular genomes in the preexisting paraorganelles may have been transferred laterally at this stage, as being facilitated by close proximity in the amoeba-like 'melting pot' of progenote paracells, which could explain the peculiar resemblance of archaea and bacteria in this respect.

In turn, by gradually building up an autonomously viable genome of their own, descendants of these hydrogen-/formate-processing vesicles became capable of leaving the progenotic paracells in a second round of '*microcellular escape*', giving rise to archaeal lineages. In line with this reasoning, the most highly developed mode of anoxic hydrogen consumption – *methanogenesis* – is a solely archaeal achievement of a supposedly unique and very ancient origin [195], most likely followed by multiple losses throughout the archaeal tree. In turn, free-living archaea became masters in making minute amounts of ATP on exceedingly shallow gradients of free energy available in both temperate and extreme environments [125].

To be sure, an endovesicular paraorganelle origin of Archaea as suggested above is not the only model to rationalize the emergence of archaeal microcells from a presumably complex progenote state of sufficient functionality in a Woesean layout. In any case, a single minichromosome (containing the so-called nucleolus organizer region and being subject to proto-eukaryotic replicative mechanisms) must ultimately have ended up with having gathered a minimalistic self-sufficient genome on itself, for the most part being sampled from the proto-nuclear *pan-genome* for the proto-eukaryotic protein synthesizing cytoplasm, as well as including some relevant genes from the proto-bacterial *paraorganelle* compartments in addition. Conceivably, the successive accumulation of essential genes on a single chromosome may even have been driven by selective pressures at the border of extreme environments, into which the composite paracells as such could not expand (see further above for *thermoreduction* as a typical example). Moreover, the separate encasement of archaeal microcells might also have been provided by a miniaturized and partly

 doi:10.20944/preprints201710.0153.v1

reorganized version of the external cytoplasmic membrane or – more remotely – even by some capside-coated megavirus.

In considering the endogenous formation of *paraorganelles*, as already mentioned, the *generalized progenote hypothesis* conceptionally distinguishes between two different bifurcation points in evolutionary time: the initiation of a monophyletic genomic lineage first and the corresponding emergence of free-living cellular lineages later on. Although a full discussion of these aspects concerning archaeal diversity is beyond the scope of this initial presentation, some striking observations are worth considering in this context. The initiation and processivity of DNA replication from chromosomal origin sites are of basic importance in all three domains of life but the proteins involved are rather different in bacteria on the one hand and in the archaeal/eukaryotic group on the other [196]. In the views presented here, the early onset of bacterial lineage verticality was facilitated by two isolating factors: a self-specific initiation mechanism of unimolecular plasmid replication and the physical separation of intravesicular gene expression from the extravesicular holoplasm. In contrast, the paracell paragenomes (particular samples from the progenote pan-genome) were being replicated from multiple similar origins, held in common on many different (mini-)chromosomes.

Characteristic of the archaeal/eukaryotic group, a composite assemblage of accessory factors is needed to activate the processive helicases at the replication fork [197,198], of which the GINS subcomplex is part of an external thrust bearing to stabilize the pumping motion at the internal motor ring domains of the hexameric helicase complex [199]. Comparative genomics of these composite initiation complexes may tell us something about the deep-branching relationships between archaeal lineages and the proto-eukaryotic trunk line [200–202]. Conventionally, the much higher number of different subcomponents throughout eukaryotes is interpreted as resulting from some "*cataclysmic event leading to a sharp drop in the population size of the proto-eukaryote and the ensuing weakening of purifying selection, which in turn led to an increase in the survival time of duplications*" [203], taking for granted that the last universal common ancestor (LUCA) consisted of modular organisms that were organized by a rather minimalistic genome, comprising little more than all those genes that still are common to all three domains of life.

The *generalized progenote hypothesis*, however, implies a more gradualistic and inclusive concept for early evolution, as based on rational, overarching and widely applicable principles, for which the LUCAS acronym appears more appropriate. By deriving not only Bacteria and Archaea directly from the large collective melting-pot of the Woesean progenote state but also the intrinsically complex eukaryotic cell type, this composite scenario has no need to invoke any catastrophic emergency rescue, such as the launching of an unprecedented genomic meltdown after the acquisition of mitochondrial endosymbionts [204] and a secondary phase of expansive-creative trunk-line evolution toward proto-eukaryote emergence later on.

As for the variety of deep-branching archaeal lineages, the general assumptions of this hypothesis give room for some unconventional reinterpretation of relational links between certain Archaea and eukaryotes at large. It has long been noted that Crenarchaeota in some respects appear to be closer related to eukaryotes than the larger and more diverse group of Euryarchaeota [205]. More recently, several newly discovered additional branches were shown to map even closer to eukaryotes in traditional proteomic phylograms [206,207]. Yet, should these archaeal lineages now represent putative ancestors to eukaryote complexity? – Not necessarily, I think. Alternatively, these deep-branching lineages might also be regarded as relative latecomers in a series of multiple cellular escape events, considering a longer phase of proto-archaeal evolution from compartmentalized single minichromosomes within the non-modular paracells of a confluent and still communal progenote population. Such particular archaeal cell types may also have engaged in ectosymbiont relationship with early eukaryotes for longer periods than other archaea.

More recently, large-scale sampling and screening of environmental DNA has increased our knowledge of still unculturable prokaryotes tremendously, and many of these can be mapped as archaeal side lines around the eukaryotic root, at least by taking concatenated protein-coding genes into account [208,209]. The troubling paradox still is that ribosomal rRNA genes keep telling a

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2017 doi:10.20944/preprints201710.0153.v1

different story by confirming the three-domain topology of the Woesean tree (Supplementary Fig. 2 in [208]), so that Woese's canonical tree and the challenge from an archaeal origin of eukaryotes remain competing concepts for further analyses. Additional analysis of the Lokiarchaeota data has also revealed chimaeric components in the proteomic reconstruction of the phylogenetic relationship between eukaryotes and Archaea [210], in that a core set of very basic protein functions significantly supports Woese's canonical tree of three monophyletic domains. In view of the broadened perspective offered here, the new data also fit to certain alternative interpretations, in which the chromosomal cluster of rRNA genes may have taken a leading role.

For example, when the *Archgenote* stem line to both Archaea and Eukarya was still at its residual progenote state, circular units of ribosomal rDNA and a proto-eukaryote-type replication origin may have become compartmented in proto-organellar vesicles, together with some genes for membrane-associated proteins that preferentially should be inserted from the inner side. Any such path would formally initiate a more or less stable lineage of proto-archaeal specificity, thus founding the archaeal domain for as long as mostly rDNA is being considered in the comparative phylogenetic analysis. To propagate such a minimalistic founder genome inside a vesicular compartment, it would be mandatory to import replication functions and other proteins from the outer holoplasm. If this import also involved most ribosomal proteins, the corresponding proto-nuclear genes may gradually have been added by bulk-to-organelle gene transfer at later times, thus leaving room for the deep-branching side lines of archaea-like phylogeny that were inferred from the concatenation of several conserved ribosomal-protein genes [207]. The special relationship in certain regards between bacterial and archaeal genomes, ascribed to highly asymmetric interdomain gene transfer – ocurring some fivefold more frequently from bacteria to archaea than *vice versa* [211] – may thus have started as transfer between different *para-organelles* in a common *holoplasm* of *progenote paracells* already.

At any rate, the simplistic reduction of overall-concatenated proteomic data to a single tree-like pattern may tell us rather little about the likely nature of symbiotic networks being active around the time of archaeal–protoeukaryotic divergence several billion years ago. Only system-oriented factorial analyses can give more valuable information for distinguishing between various conceptual possibilities, such as the recent "*inside-out hypothesis*" for eukaryotic karyogenesis from tight microbial ectobiont associations [212] and/or residual populations of the compartmented progenote-like archgenote trunk-line advocated here. The recently proposed nuclear compartment commonality (NuCom) hypothesis [213,214] is pointing in the same direction as the ideas presented here, so as to revitalize Doolittle's notion of a peculiar "*nuclear-cytoplasmic lineage*" [93] as an intrinsic constituent of Woese's canonical three-domain phylogeny.

6. Concluding Remarks

6.1. From Primordial Stochasticity to Ordering Constraints

Carl Woese challenged the scientific community in his serious quest for rethinking basic assumptions concerning the protogenomic origins of Darwinian speciation [8]. Follow-up studies to further substantiate his theoretical insights, however, have not really been done. The widespread reluctance to engage with this issue may relate to a rather low regard for critical bottom-up considerations among evolutionary biologists in general, who prefer to concentrate their efforts on top-down extrapolation from comparative genomics exclusively. This one-sided approach, however, loses its resolving power and predictability upon reaching a virtual '*event horizon*' at certain singularities: the odd primordial bifurcation points that for fundamentally different phenotypic ensembles are being claimed as their common evolutionary origins. In fact, the many uncertainties about the phenotypic organization throughout the pregenomic era reach even further back in time.

In the long run, it will not be possible to develop a comprehensive theory of life's emergence and consolidation without resorting to bottom-up approaches of physics-based models that still comply with the general framework of non-equilibrium thermodynamics. Furthermore, such physics-based abstractions must also be adapted to biologically relevant chemical and structural

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2017 doi:10.20944/preprints201710.0153.v1

particularities. Considerations of this kind, however, should not be confined by overly rigid preconceptions when other possibilities may have been available as well. My personal inclination to this challenge has previously been expressed as follows:

> "*The origins of life are founded on three major roots, in this order of temporal, functional and logical priorities: a lasting energetic gradient on the pristine Earth between the radiating solar source and the sink of outer space; self-accreting networks of prebiotic macromolecules that happened to work together slowly; and an emerging archive to let the consolidating network remember how it actually had worked in the preceding period. To begin with, both the consolidating catalytic network and the nascent working memory were rather 'fuzzy' systems. By resonating and falling into sync with one another, they became subject to coevolutionary optimization*" [6].

By and large, this inclination of mine is recognizing more life-like properties in cytoplasmic functionality alone than conventional wisdom is willing to concede. On the other hand, I am also prepared to go a long way to evade the idle quest for defining life as such in rigidly preconceived terms beforehand. My programmatic suggestion of using '*Parabiotic Evolution*' for the presumably gradual transition phase from non-life to life should be regarded in this context.

Although '*replication*' is not explicitly mentioned in the passage cited above, the basic concept is related to the "*replication-and-metabolism-together scenario*" [215], in that the '*nascent working memory*' involved some means of template-directed synthesis reactions, albeit of rather 'fuzzy' specificity in the beginning. In addition to many catalysts of house hold metabolic function, the intricate molecular machinery to accomplish genomic replication and mRNA transcription, as well as ribosomal protein synthesis, became subject to the principle of *coevolutionary optimization* prevailing throughout the era of a collective progenote state – the communal trunk-line *paracells* of the current presentation. It was this extensive feedback-adaptive optimization process that gradually confined the initial fuzziness of many intrinsically stochastic interactions into the narrow band width of very specific reaction steps, engaging in template-dependent replication/transcription and RNA-encoded translation.

These considerations touch upon the most central *controversy* that long has puzzled OoL research: "*Replicators or metabolism first*"? [19,216]. In order to rationalize the comparison of several alternatives of this kind, the specification of a particular '*privileged function*' has been recognized for conventional OoL models in general, and this label is not intended as a compliment: "*A privileged function is an extant biological function that is excised from its biological context, elevated in importance over other functions, and transported back in time to a primitive chemical or geological environment*" [4]. Notably, the most popular models excel by an appearance of simplicity, but "*the simplicity of these models is seen to be an illusion on the realization that the models require fluidity in principles of evolution*". In other words, most of these models become unreasonable and forbiddingly contrived at later stages when they require rather unlikely takeover maneuvers and activities. These grave concerns do not apply if "*RNA, DNA, protein and the Molecular Toolbox co-evolved in a cooperative and symbiotic process*" [4].

Addressing the early evolution of common ribosomal proteins, a very recent example happens to illustrate how deeply the '*critical function*' idol is still engrained in conventional thinking: "*Coded proteins originated as oligomers and polymers created by the ribosome, on the ribosome and for the ribosome … Protein catalysis appears to be a late byproduct of selection for sophisticated and finely controlled assembly*" [217]. These statements postulate a partisan selfishness for emerging ribosomes that seems strangely detached from their pivotal system-supporting role at later stages, when the vast majority of proteins made by ribosomes are NOT incorporated in the ribosomal particles themselves. This caveat, in fact, is calling for alternative interpretations about the transition from uncoded to coded protein synthesis when the stochastic variation of peptide chain elongation was strongly reduced. First of all, ribosomes are not acting alone in making proteins; they heavily depend on the presence of aminoacylated tRNAs.

Arguably, tRNAs are substantially older than ribosomes [218] and may originally have functioned as genuine peptidyl transferases for uncoded peptide/protein synthesis [219]. From a general *systems-continuity* perspective, therefore, a range of early uncoded peptides supposedly began participating in similar functions to those of coded proteins later on, and the progress made

 doi:10.20944/preprints201710.0153.v1

from the ribosomal assistance of tRNAs was a matter of vastly increased overall efficiency rather than a qualitative change in the most basic systems properties. Hence the pointed statement cited above could even be given a subtle twist: "*Uncoded proteins originated as oligomers created by tRNAs, on tRNAs and for tRNAs; ribosomal synthesis of coded protein appears to be a later result of selection for multiple repetition of the tRNA-mediated transfer reaction and its coupling to sequence-preserving protein assembly*".

Generally speaking, I find the deceptive or illusive simplicity of certain *privileged functions* very relevant to this discussion, but I should not like to put *(Self)-Replicators* (Genetics) and *Metabolism* into quite the same category [4] in this regard. There are marked differences between these features concerning their potential connections up or down: to start from purely stochastic reactions and to initiate a truly self-sustaining *material system* on the basis of robust, evolvable and long-lasting *functional principles*. From the perspective of systems biology, the *fully historical connectivity* and *variable iteration* of metabolic and morphogenetic processes are prominent principles of this kind [220]. *Iterative loops* in general are closely related to *autocatalytic amplification*, and the intrinsic fuzziness of quasi-stochastic iterative processes forms the very basis of biological evolution in the first place. As no amplification loop can keep on running for extended periods without being coupled to some energy conversion, the *retarded channeling of energy degradation* from cosmic or geochemical sources is another fundamental principle of life's existence [221]. Besides, the localized accumulation of (non-tarry) organic substances is yet another characteristic of the biosphere at large and, arguably, has accompanied the *historical continuity* of life-like systems from the very beginning.

Metabolism is the sum total of the chemical processes that occur in living organisms. This is a multiply interconnected network of many iterative loops, of which a central core is now almost invariant in many organisms whereas the outskirts can be extremely malleable, especially in prokaryotes. In such general terms, metabolism has always been at work since life's beginnings but the number of iterative loops and the nature of some may well have changed with time. Arguably one of the first self-amplifying loops for an emergent proto-metabolism is seen in the *'Reverse Krebs Cycle'* [222], which may have emerged by UV-activation at colloidal mineral grains [74,75,108].

At present the organizational connectivity of different loops can be ordered into three stratified levels. There are the chains and cycles of elementary reactions at the base, the various catalysts to channelize these reactions are being added next, and the genetic memory to repeatedly make the proper catalysts reigns at the uppermost level. This particular classification touches upon a critical difference between 'creating' a complex hierarchical system by evolution (historical and bottom up) or by design (rational and top down). It appears that certain idealizing models tend to mimic the latter approach rather than try to comprehend the evolutionary options and capabilities.

All this complexity of living systems is being integrated into structural embodiments of physical durability and strictly historical continuity. Conceptually, however, it must have been seamlessly connected to the virtual randomness of stochastic geochemical reactions early on. To emphasize the roles of active players at the initial transitions, the so-called '*Metabolism before Genetics*' paradigm can be rephrased in complementary terms as '*Rudimentary catalysts before digitally encoded, replicable templates*'. As argued in the present paper, the overarching Woesean principle of collective optimization in a communal proto-cytoplasm [18] has been applicable to both the perfection of digital catalysts and the catalyzed replication of the likewise digital memory archive.

When the first rudimentary organic catalysts (originally resulting from quasi-stochastic prebiotic reactions) began to interact in self-amplifying autocatalytic networks, this marked the presumptive beginning of *parabiotic evolution*. When Stuart Kauffman began to argue in this direction [223], deliberately avoiding the conventional presumption of digital molecular *replication* early on, he exemplified his abstract conceptual networking model by referring to oligopeptides making other oligopeptides. I find it obvious to add oligonucleotides to a primordial 'starting kit' of this kind. The initial set need not be limited to oligomers of the four now standard ribonucleotides but may have included a range of base-modified building blocks as well, as it can nowadays still be observed in many tRNAs – allowing a conjectural extrapolation into a prebiotic setting [224].

The latter consideration points to a peculiar ambivalence in the mutual interactions between stochastic peptides and oligonucleotides in terms of their differential networking abilities. The

 doi:10.20944/preprints201710.0153.v1

dynamic functionality of Kauffman-type networks depends on three partly independent features of the participating building blocks. In being incorporated into composite molecules (the nodes of the network) the added components can contribute to chemical reactivity on the one hand and structural connectivity on the other. The latter aspect can be further subdivided into intra-compound scaffolding potential and external stickiness with other partners in the agglomerating organic matter. The resultant catalytic potential in terms of reaction specificity and speed, in turn, is influenced in complex ways by all three features contributed by the different building blocks. In the long run it is just the complexity of these interactions that allows the gradual optimization of overall functionality by small, incremental, compositional adjustments.

The arrangement of chemically important characteristics of amino acids and ribonucleotides affects the interactivity of proteins and RNA in strikingly different ways, taking repetitive backbones and variable side chains or nucleobases into consideration. Whilst much of RNA variability is hidden away inside the base-paired stem regions, the monotonous and strongly acidic backbones are turned outside. The first level of protein folding, however, is more dependent on direct backbone interaction, which tends to expose side-chain variability at the surface where hydrophobic residues, in particular, determine local stickiness for intra- and inter-domain cohesive interaction. On the other hand, even unfolded protein motifs can effectively bind to RNA, which on its own would be readily dissolvable in water. Hence uncoded proteins may, early on, have provided the glue letting tRNA-like molecules participate in gel-like phase separation, where rudimentary catalyst action began to diversify by RNP complexes in combination.

During the Kauffman-type phase of early evolution, presumably, both RNAs and uncoded proteins were produced by the tinkering or taylor-made approach of fitting together a range of quasi-stochastic sub-assemblies. In comparison to modern usage, the primordial variety of building blocks included considerably fewer than 20 amino acids but more than just four nucleobases. These numbers subsequently changed when the genetic code was optimized and expanded to its present range but the direct incorporation of non-standard (base-modified) nucleotides was abandoned by processive and sequence-preserving replication. Correspondingly, the catalytic activities of coded protein enzymes expanded enormously, at the expense of diminishing ribozyme activity in residual RNP complexes.

The views presented here give evolutionary precedence to a long phase of perfection in the *cytoplasmic activities* needed just to keep life going, before modular propagation of self-similar organisms became possible by the reliable replication of individual genes and, in turn, the faithful transmission of modular genomes, which eventually completed the perfection of *'nuclear' activities* at the *Woesean transition* [8,9]. Proposing a *generalized progenote hypothesis* – *extended* by allowing for endogenous compartmentation – the exit stage of the Woesean progenote state is here further subdivided into three different modes of making a modular genome from stochastic populations of many originally unconnected genes:

- The endogenous vesicularization of circular autonomously replicating plasmids, which began to gather more and more essential genes from the surrounding cytoplasm, before they eventually could evade by cellular escape as free-living bacterial cells.
- The collection of multiple chromosomes in proto-nuclei, which later on were stabilized as eukaryotic multichromosomal genomes by the periodic alternation of mitotic and meiotic nuclear divisions.
- Somewhere between these two extremes, a single chromosome of the proto-eukaryotic set, which happened to carry a nucleolus organizer region, began to prevail over others and eventually gained independence in modular archaeal cells

As I see it, this *generalized progenote hypothesis* is not just one of many 'just-so' stories to subsume a surge of data as they happen to accumulate. It is rather conceived to integrate a range of otherwise puzzling evolutionary conundrums by way of more general principles, acknowledging the need of starting from virtually stochastic processes and applying the Woesean principle of collective optimization at additional levels to those considered by the original author(s).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2017 doi:10.20944/preprints201710.0153.v1

6.2. Terrestrial Progenotes: a Novel Perspective, also for a Communal Protoeukaryote Trunk Line

As for the geological setting that several billion years ago may have been most favorable to initiate the kind of *coevolutionary optimization* scenario envisioned here, we are currently witnessing an important paradigm shift away from the long favored submarine hydrothermal vents toward surface-exposed terrestrial geothermal fields [76,78,225,226]. This shifting concept appears highly significant for serious reconsideration of a eukaryote-specific '*nuclear-cytoplasmic lineage*' and its potential for early origins as well. Notably, the Sci.-Am. essay presenting this opinion shift [226] gives room to transient periods of "*moist gels*", which briefly formed during the rehydration of dried organic films to the wet suspension of free-floating protocells in the water body of temporary pools.

It is a matter of debate how much emphasis should be given to the different phases during such iterative wet-dry cycles very early on. Personally I have long held the view that moist surface-attached, biofilm-like mats of potentially confluent molecular associations were key to emergent autotrophic evolution [5,7,66,188,227]. It appears to me that the general emphasis on free-floating protocells is still highly influenced by the now outdated model of heterotrophic biogenesis by extracting organic precursor molecules from a rich "*primary broth*" ([1], p. 378; [228]) or "*prebiotic soup*" [229]. Also, the simplified assumption of putting rather few and simple precursor molecules into tiny membrane-bounded vesicles [230,231] is building on similar grounds, and the tentative importance given to external membrane formation early on may well be overestimated. Instead, I presume that dynamic surface adsorption of newly arising oligomeric organics initiated the formation and growth of sessile films or hydrogels and much of early evolution was taking place in stagnant periods of such a moist and spontaneously phase-separated state [7].

Whilst stratified growth under such conditions selectively retained all components that contributed to gel-like phase separation, intermittent episodes of turbulent mixing provided a mechanistic basis for material exchange between otherwise separate "*Innovation Pools*" [18] throughout the Woesean progenote population. Based on such assumptions, rather large and lumpy paracells carried many separate genes or minichromosomes and may have populated much of the progenote era with an intrinsic tendency to compartmentalize internally. This evolutionary potential is in stark contrast to the more naïve assumption of an organics-rich primordial ocean, in which a few genes enclosed in tiny vesicles would have but little chance of scaling up into a robust population of veritable protocells. Being surrounded by the *plasmodial trunk-line* cytoplasm of bulky paracells, as presumably directed by many chromosomes in multiple proto-nuclei, the posited plasmid-like genomes in endogenous paraorganelles had much better odds of becoming fully self-sufficient as eventually free-living bacterial cells.

Assuming a terrestrial-vent scenario [76], sunlight exposure was available for primary organic syntheses, and the *plasmodial trunk-line* population considered here was born into scattered moist environments surrounded by mostly arid land. One obvious possibility is that most of the primordial trunk line population continued to be attached to similar fresh-water habitats, whereas more extreme environments, the saline anoxic oceans included, were only colonized later on when vertically stable lineages of metabolically more progressive bacterial and/or archaeal cells had successfully 'escaped' from the less adaptable communal source.

Accordingly, the emergence and early diversification of eukaryotic protist organisms may also have been limited to fresh-water environments on land, and adaptive optimization to periodically fluctuating terrestrial conditions had thus been integrated into their cytoplasmic tool box from very early on. Regular cycles, such as daily, monthly or seasonal repetition, had natural training effects on systemic physiological responses, such as the canonical nuclear division cycle or the alternation of mitotic and meiotic divisions – with sexual confluence of the cytoplasm and nuclear fusion (karyogamy) occurring in between. Less regularly recurring atmospheric perturbations affecting the terrestrial hydrosphere were bound to have lasting effects on early life as well. Not only did rain storms fill streams and pools, which resulted in the rehydration and mixing of sessile proto-biofilms already referred to above, but dust storms, too – as they still do today [232] – could have important effects by transporting dried-down bio-matter around the globe. Thereby the emergent terrestrial

plasmodial trunk line was capable of establishing a globally connected population for further communal evolution very early on, even though suitable micro-habitats were scattered far apart.

6.3. Future Outlook

When Charles Darwin put forward his well-known theory of evolution by natural selection, this was a giant leap for biology and its mechanistic understanding in terms of more general principles than what observational comparison can reveal as such. Considering that virtually nothing was known at his time about the molecular basis of inheritance and its relationship to subcellular processes in general, this was a prescient act of abstraction to unprecedented heights. With increasing experimental data, however, a series of major conceptual syntheses was needed to firmly entrench the Darwinian principles into modern biology [233,234]; but the need for further rounds of synthesis is not over yet. First and foremost, the Darwinian legacy must still be connected to a reasonable scenario for origins of life on Earth [22,234]. To this end, Carl Woese's elaborate "*millennial series*" [8–11] is a pioneering work, addressing the upper reaches of the open gap and revealing some general principles of pre-modular genome dynamics. Yet, also Woese's insights are still in need of being integrated into a more comprehensive theory.

In the present paper, I suggest unconventional assumptions to envisage such integration, as based on endogenous compartmentation in particular. How can a thus modified *progenote hypothesis* be tested or substantiated by other considerations? Experimental testing would traditionally resort to modeling by chemical methods, but chemistry is perhaps approaching practical limits in this field of potentially unlimited complexity. To be sure, more and more composite '*One-Pot*' reactions are being devised and can in fact give astonishing results [110,235,236], but the modeling of endogenous '*paraorganelle*' formation might present a not yet surmountable challenge to a chemical approach.

On the other hand, the basic assumption to motivate the *generalized progenote hypothesis* is placing a plasmid-like self-replicating genome with energy-related functionality inside a vesicular compartment, which in turn depended on being nourished from a surrounding cytoplasm that still was organized by many partly unconnected genes on more or less stochastically distributed multiple chromosomes. In other words, to use the modified Woesean terminology, the plasmid genomes of vesicular *paraorganelles* had already crossed their *Darwinian threshold* whilst the *paragenomes* of the surrounding *holoplasm* had not yet done so. On this basis I suspect that the dynamic balance between selfish interests and communal cooperation would slowly shift toward metabolic independence for the most successful lineages of gradually complexifying *paraorganelles*. This is just an educated guess for the time being, but it should be possible to parameterize the various components of the underlying hypothesis so as to simulate the evolution to be expected by *in silico* analysis.

This kind of exercise would expand an emergent discipline of "*Theoretical Genome Dynamics*" into the pregenomic transition phase of organismal genome consolidation, which should be complementary to the traditional dominance of prospective chemical approaches to the OoL enigma on the one hand and retrospective comparative phylogenomics on the other. A very significant first step in this direction has, in fact, been done already when Woese's hypothetical conception of collective optimization by *communal innovation sharing* was subjected to appropriate computer simulations [18]. This challenge calls for a more visionary collaboration between evolutionary biologists and informatics departments than what is presently dominated by conventional genomics. – Moreover, especially pertaining to the agenda of this paper, very recent models of genome evolution are found to intrinsically root the global *Tree of Life* (ToL) and suggest partly independent descendancy for Bacteria, Archaea and Eukaryotes from a common source [237,238], which is in line with the inferences presented here.

Acknowledgments: The invitation from Life to contribute to this Special Issue is gratefully acknowledged.

Conflicts of Interest: The author declares no conflict of interest

References

1. Koonin, E.V. *The logic of chance: the nature and origin of biological evolution*; FT Press: Upper Saddle River, NJ, 2012.
2. Pressman, A.; Blanco, C.; Chen, I.A. The RNA World as a model system to study the origin of life. *Curr. Biol.* **2015**, *25*, R953–R963.
3. Spitzer, J.; Pielak, G.J.; Poolman, B. Emergence of life: Physical chemistry changes the paradigm. *Biol. Direct* **2015**, *10*, 33.
4. Lanier, K.A.; Williams, L.D. The origin of life: models and data. *J. Mol. Evol.* **2017**. *84*, 85–92.
5. Egel, R. Peptide-dominated membranes preceding the genetic takeover by RNA: latest thinking on a classic controversy. *Bioessays* **2009**, *31*, 1100–1109.
6. Egel, R. Life's order, complexity, organization, and its thermodynamic–holistic imperatives. *Life* **2012**, *2*, 323–363.
7. Egel, R. Origins and emergent evolution of life: the Colloid Microsphere Hypothesis revisited. *Orig. Life Evol. Biosph.* **2014**, *44*, 87–110.
8. Woese, C.R. On the evolution of cells. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 8742–8747.
9. Woese, C.R. A new biology for a new century. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 173–186.
10. Woese, C. The universal ancestor. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6854–6859.
11. Woese, C.R. Interpreting the universal phylogenetic tree. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 8392–8396.
12. Koonin, E.V. Carl Woese's vision of cellular evolution and the domains of life. *RNA Biol* **2014**, *11*, 197–204.
13. Woese, C.R.; Fox, G.E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5088–5090.
14. Woese, C.R.; Fox, G.E. The concept of cellular evolution. *J. Mol. Evol.* **1977**, *10*, 1–6.
15. Woese, C.R. Bacterial evolution. *Microbiol. Rev.* **1987**, *51*, 221–271.
16. Woese, C.R.; Kandler, O.; Wheelis, M.L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 4576–4579.
17. Darnell, J.E.; Doolittle, W.F. Speculations on the early course of evolution. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 1271–1275.
18. Vetsigian, K.; Woese, C.; Goldenfeld, N. Collective evolution and the genetic code. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 10696–10701.
19. Pross, A. Causation and the origin of life. Metabolism or replication first? *Orig. Life Evol. Biosph.* **2004**, *34*, 307–321.
20. Joyce, G.F. Bit by bit: the Darwinian basis of life. *PLoS Biol* **2012**, *10*, e1001323.
21. Monnard, P.A.; Walde, P. Current ideas about prebiological compartmentalization. *Life* **2015**. *5*, 1239–1263.
22. Wicken, J.S. *Evolution, information and thermodynamics: Extending the Darwinian program*; Oxford University Press: New York, 1987.
23. Wicken, J.S. Evolution and thermodynamics: The new paradigm. *Systems Res.* **1989**, *6*, 181–186.
24. Fishkis, M. Steps towards the formation of a protocell: the possible role of short peptides. *Orig. Life Evol. Biosph.* **2007**, *37*, 537–553.
25. Wills, P.R. The generation of meaningful information in molecular systems. *Phil. Trans. R. Soc. A* **2016**, *374*, 20150066.
26. Caetano-Anollés, G.; Seufferheld, M. J. The coevolutionary roots of biochemistry and cellular organization challenge the RNA world paradigm. *J. Mol. Microbiol. Biotechnol.* **2013**, *23*, 152–177.
27. Lane, N.; Martin, W. The energetics of genome complexity. *Nature* **2010**, *467*, 929–934.
28. Williams, T.A.; Foster, P.G.; Cox, C.J.; Embley, T.M. An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* **2013**, *504*, 231–236.
29. Williams, T. A.; Embley, T.M. Changing ideas about eukaryotic origins. *Phil. Trans. R. Soc. B* **2015**, *370*, 20140318.
30. Koonin, E.V. Origin of eukaryotes from within archaea, archaeal eukaryome and bursts of gene gain: eukaryogenesis just made easier? *Phil. Trans. R. Soc. B* **2015**, *370*, 20140333.
31. Mariscal, C.; Doolittle, W.F. Eukaryotes first: how could that be? *Phil. Trans. R. Soc. B* **2015**, *370*, 20140322.
32. Poole, A.; Jeffares, D.; Penny, D. Early evolution: prokaryotes, the new kids on the block. *Bioessays* **1999**, *21*, 880–889.
33. Kurland, C.G.; Collins, L.J.; Penny, D. Genomics and the irreducible nature of eukaryote cells. *Science* **2006**, *312*, 1011–1014.

34. De Duve, C. The origin of eukaryotes: a reappraisal. *Nat. Rev. Genet.* **2007**, *8*, 395–403.
35. Poole, A.M.; Penny, D. Evaluating hypotheses for the origin of eukaryotes. *Bioessays* **2007**, *29*, 74–84.
36. Glansdorff, N.; Xu, Y.; Labedan, B. The last universal common ancestor: emergence, constitution and genetic legacy of an elusive forerunner. *Biol. Direct* **2008**, *3*, 29.
37. Forterre, P. The common ancestor of Archaea and Eukarya was not an archaeon. *Archaea* **2013**, *2013*, 372396.
38. Kurland, C.G.; Andersson, S.G.E. Origin and evolution of the mitochondrial proteome. *Microbiol. Mol. Biol. Rev.* **2000**, *64*, 786–820.
39. Gray, M. W. Mitochondrial evolution. *Cold Spring Harb. Persp. Biol.* **2012**, *4*, a011403.
40. Gray, M.W. Mosaic nature of the mitochondrial proteome: Implications for the origin and evolution of mitochondria. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 10133–10138.
41. Gray, M.W. The pre-endosymbiont hypothesis: a new perspective on the origin and evolution of mitochondria. *Cold Spring Harb. Persp. Biol.* **2014**, *6*, a016097.
42. Tovar, J.; Fischer, A.; Clark, C.G. 1999, The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*. *Mol. Microbiol 32*, 1013–1021.
43. Embley, M.; Van Der Giezen, M.; Horner, D.S.; Dyal, P.L.; Foster, P. Mitochondria and hydrogenosomes are two forms of the same fundamental organelle. *Phil. Trans. R. Soc. B* **2003**, *358*, 191–203.
44. Schneider, R.E.; Brown, M.T.; Shiflett, A.M.; Dyall, S.D.; Hayes, R.D.; Xie, Y.; ... Johnson, P.J. The *Trichomonas vaginalis* hydrogenosome proteome is highly reduced relative to mitochondria, yet complex compared with mitosomes. *Int. J. Parasitol.* **2011**, *41*, 1421–1434.
45. Burki, F. Mitochondrial evolution: going, going, gone. *Curr. Biol.* **2016**, *26*, R410–R412.
46. Yang, D.; Oyaizu, Y.; Oyaizu, H.; Olsen, G.J.; Woese, C.R. Mitochondrial origins. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 4443-4447.
47. Gray, M.W.; Burger, G.; Lang, B.F. The origin and early evolution of mitochondria. *Genome Biol.* **2001**, *2*, 1018.
48. Wang, Z.; Wu, M. Phylogenomic reconstruction indicates mitochondrial ancestor was an energy parasite. *PLoS One* **2014**, *9*, e110685.
49. Sagan L. On the origin of mitosing cells. *J. Theor. Biol.* **1967**, *14*, 225–274.
50. Embley, T.M.; Martin, W. Eukaryotic evolution, changes and challenges. *Nature* **2006**, *440*, 623–-630.
51. Martin, W. Evolutionary origins of metabolic compartmentalization in eukaryotes. *Phil. Trans. R. Soc. B* **2010**, *365*, 847–855.
52. Zimorski, V.; Ku, C.; Martin, W.F.; Gould, S.B. Endosymbiotic theory for organelle origins. *Curr. Opin. Microbiol.* **2014**, *22*, 38–48.
53. Martin, W.F.; Garg, S.; Zimorski, V. Endosymbiotic theories for eukaryote origin. *Phil. Trans. R. Soc. B* **2015**, *370*, 20140330.
54. Baum, D.A. A comparison of autogenous theories for the origin of eukaryotic cells. *Am. J. Bot.* **2015**, *102*, 1954–1965.
55. Cannon, G.C.; Bradburne, C.E.; Aldrich, H.C.; Baker, S.H.; Heinhorst, S.; Shively, J.M. Microcompartments in prokaryotes: carboxysomes and related polyhedra. *Appl. Environ. Microbiol.* **2001**, *67*, 5351–5361.
56. Fuerst, J.A. Intracellular compartmentation in planctomycetes. *Annu. Rev. Microbiol.* **2005**, *59*, 299–328.
57. Yoan, D.; Jose, B. Evolution of intracellular compartmentalization. *Biochem. J.* **2013**, *449*, 319–331.
58. Gabaldón, T. Peroxisome diversity and evolution. *Phil. Trans. R. Soc. B* **2010**, *365*, 765–773.
59. Gabaldón, T.; Pittis, A.A. Origin and evolution of metabolic sub-cellular compartmentalization in eukaryotes. *Biochimie* **2015**, *119*, 262e268.
60. Penny, D.; Collins, L.J.; Daly, T.K.; Cox, S.J. The relative ages of eukaryotes and akaryotes. *J. Mol. Evol.* **2014**, *79*, 228–239.
61. Weiner, A.M. Eukaryotic nuclear telomeres: molecular fossils of the RNP world? *Cell* **1988**, *52*, 155–157.
62. Cech, T.R. Crawling out of the RNA world. *Cell* **2009**, *136*, 599–602.
63. Norris, V.; Raine, D.J. A fission-fusion origin for life. *Orig. Life Evol. Biosph.* **1998**, *28*, 523–537.
64. Wächtershäuser, G. From pre-cells to Eukarya – a tale of two lipids. *Mol. Microbiol.* **2003**, *47*, 13–22
65. Bandea, C.I. A unifying scenario on the origin and evolution of cellular and viral domains. *Nat. Preced.* **2009**, *2009*, 3888.1.
66. Egel, R. Primal eukaryogenesis: on the communal nature of precellular states, ancestral to modern life. *Life* **2012**, *2*, 170–212.

67. Mayr, E. *Systematics and the origin of species*; Columbia University Press, New York, 1942.
68. Lane, N. Energetics and genetics across the prokaryote-eukaryote divide. *Biol. Direct* **2011**, *6*, 35.
69. Caetano-Anollés, G.; Mittenthal, J.E.; Caetano-Anollés, D.; Kim, K.M. A calibrated chronology of biochemistry reveals a stem line of descent responsible for planetary biodiversity. *Front. Genet.* **2014**, *5*, a306.
70. Kandler, O. Cell wall biochemistry and three-domain concept of life. *Syst. Appl. Microbiol.* **1994**, *16*, 501–509.
71. Medini, D.; Donati, C.; Tettelin, H.; Masignani, V.; Rappuoli, R. 2005, The microbial pan-genome. *Curr. Opin. Genet. Dev. 15*, 589–594.
72. Wächtershäuser, G. Before enzymes and templates: theory of surface metabolism. *Microbiol. Rev.* **1988**, *52*, 452–484.
73. Wächtershäuser, G. From volcanic origins of chemoautotrophic life to Bacteria, Archaea and Eukarya. *Phil. Trans. R. Soc. B* **2006**, *361*, 1787–1808.
74. Russell, M.J.; Martin, W. The rocky roots of the acetyl-CoA pathway. *Trends Biochem. Sci.* **2004**, *29*, 358–363.
75. Guzman, M.I.; Martin, S.T. Photo-production of lactate from glyoxylate: how minerals can facilitate energy storage in a prebiotic world. *Chem. Comm.* **2010**, *46*, 2265–2267.
76. Mulkidjanian, A.Y.; Bychkov, A.Y.; Dibrova, D.V.; Galperin, M.Y.; Koonin, E.V. Origin of first cells at terrestrial, anoxic geothermal fields. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E821–E830.
77. Mulkidjanian, A.Y.; Bychkov, A.Y.; Dibrova, D.V.; Galperin, M.Y.; Koonin, E.V. Open questions on the origin of life at anoxic geothermal fields. *Orig. Life Evol. Biosph.* **2012**, *42*, 507–516.
78. Ross, D.S.; Deamer, D. Dry/wet cycling and the thermodynamics and kinetics of prebiotic polymer synthesis. *Life* **2016**, *6*, 28.
79. Kompanichenko, V.N.; Poturay, V.A.; Shlufman, K.V. Hydrothermal systems of Kamchatka are models of the prebiotic environment. *Orig. Life Evol. Biosph.* **2015**, *45*, 93–103.
80. Theissen, U.; Martin, W. The difference between organelles and endosymbionts. *Curr. Biol.* **2006**, *16*, R1016–R1017.
81. Andersson, S.G.; Kurland, C.G. Reductive evolution of resident genomes. *Trends Microbiol.* **1998**, *6*, 263–268.
82. Fournier, G.P.; Gogarten, J.P. Rooting the ribosomal tree of life. *Mol. Biol. Evol.* **2010**, *27*, 1792–1801.
83. Boyer, M.; Yutin, N.; Pagnier, I.; Barrassi, L.; Fournous, G.; Espinosa L.; ... Raoult, D. Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 21848–21853.
84. Dagan, T.; Martin, W. The tree of one percent. *Genome Biol.* **2006**, *7*, 118.
85. Koonin, E.V.; Martin, W. On the origin of genomes and cells within inorganic compartments. *Trends Genet.* **2005**, *21*, 647–654.
86. Forterre, P. The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res.* **2006**, *117*, 5–16.
87. Koonin, E.V.; Senkevich, T.G.; Dolja, V.V. The ancient Virus World and evolution of cells. *Biol. Direct* **2006**, *1*, 29.
88. Durzyńska, J.; Goździcka-Józefiak, A. Viruses and cells intertwined since the dawn of evolution. *Virol. J.* **2015**, *12*, 169.
89. Thomas, V.; Greub, G. Amoeba/amoebal symbiont genetic transfers: lessons from giant virus neighbours. *Intervirology* **2010**, *53*, 254–267.
90. Yutin, N.; Koonin, E.V. Pandoraviruses are highly derived phycodnaviruses. *Biol. Direct* **2013**, *8*, 25.
91. Travis, J.L.; Welnhofer, E.A.; Orokos, D.D. Autonomous reorganization of foraminiferan reticulopodia. *J. Foram. Res.* **2002**. *32*, 425–433.
92. Guttes, S.; Guttes, E. Regulation of DNA replication in the nuclei of the slime mold, *Physarum polycephalum*. Transplantation of nuclei by plasmodial coalescence. *J. Cell Biol.* **1968**, *37*, 761–772.
93. Doolittle, W.F. Open Question: Revolutionary concepts in evolutionary cell biology. *Trends Biochem. Sci.* **1980**, *5*, 146–149.
94. Leipe, D.D.; Aravind, L.; Koonin, E.V. Did DNA replication evolve twice independently? *Nucl. Acids Res.* **1999**, *27*, 3389–3401.
95. Koonin, E.V. Temporal order of evolution of DNA replication systems inferred by comparison of cellular and viral DNA polymerases. *Biol. Direct* **2006**, *1*, 39.

96. Woese, C.R. On the evolution of the genetic code. *Proc. Natl. Acad. Sci. USA* **1965**, *54*, 1546–1552.
97. Koonin, E.V. On the origin of cells and viruses. *Ann. NY Acad. Sci.* **2009**, *1178*, 47–64.
98. Cornish-Bowden, A.; Cárdenas, M.L. Life before LUCA. *J. Theor. Biol.* **2017**, *434*, 68–74.
99. Haldane, J.B.S. The origin of life. *Rationalist Annual*, **1929**, *148*, 3–10.
100. Granick, S. Speculations on the origins and evolution of photosynthesis. *Ann. NY Acad. Sci.* **1957**, *69*, 292–301
101. Corliss, J.B. On the creation of living cells in submarine hot spring flow reactors: attractors and bifurcations in the natural hierarchy of dissipative systems. *Orig. Life Evol. Biosph.* **1986**, *16*, 381–382.
102. Russell, M.J.; Hall, A.J. The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. *J. Geol. Soc.* **1997**, *154*, 377–402.
103. Martin, W.; Baross, J.; Kelley, D.; Russell, M.J. Hydrothermal vents and the origin of life. *Nat. Rev. Microbiol.* **2008**, *6*, 805–814.
104. Russell, M.J.; Hall, A.J.; Martin, W. Serpentinization as a source of energy at the origin of life. *Geobiology* **2010**, *8*, 355–371.
105. Mulkidjanian, A.Y. 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. *Biol. Direct* **2009**, *4*, 26.
106. Mulkidjanian, A.Y.; Galperin, M.Y. On the origin of life in the zinc world. 2. Validation of the hypothesis on the photosynthesizing zinc sulfide edifices as cradles of life on Earth. *Biol. Direct* **2009**, *4*, 27.
107. Hoffmann, M.R.; Martin, S.T.; Choi, W.; Bahnemann, D.W. Environmental applications of semiconductor photocatalysis. *Chem. Rev.* **1995**, *95*, 69–96.
108. Guzman, M.I.; Martin, S.T. Prebiotic metabolism: Production by mineral photoelectrochemistry of alpha-ketocarboxylic acids in the reductive tricarboxylic acid cycle. *Astrobiology* **2009**, *9*, 833–842.
109. Mulkidjanian, A.Y.; Cherepanov, D.A.; Galperin, M.Y. Survival of the fittest before the beginning of life: Selection of the first oligonucleotide-like polymers by UV light. *BMC Evol. Biol.* **2003**, *3*, 12.
110. Powner, M.W.; Gerland, B.; Sutherland, J.D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **2009**, *459*, 239–242.
111. Griffith, E.C.; Shoemaker, R.K.; Vaida, V. Sunlight-initiated chemistry of aqueous pyruvic acid: Building complexity in the origin of life. *Orig. Life Evol. Biosph.* **2013**, *43*, 341–352.
112. Lill, R. Function and biogenesis of iron-sulphur proteins. *Nature* **2009**, *460*, 831–838.
113. Balk, J.; Pilon, M. Ancient and essential: the assembly of iron-sulfur clusters in plants. *Trends Plant. Sci.* **2011**, *16*, 218–226.
114. Sakimoto, K.K.; Wong, A.B.; Yang, P. Self-photosensitization of nonphotosynthetic bacteria for solar-to-chemical production. *Science* **2016**, *351*, 74–77.
115. Forterre, P. Thermoreduction, a hypothesis for the origin of prokaryotes, *C. R. Acad. Sci. Paris* **1995**, *318*, 415–422.
116. Poole, A.M.; Phillips, M.J.; Penny, D. Prokaryote and eukaryote evolvability. *Biosystems* **2003**, *69*, 163–185.
117. Raff, R.A.; Mahler, H.R. The non symbiotic origin of mitochondria. *Science* **1972**, *177*, 575–582.
118. Martin, W.; Müller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **1998**, *392*, 37–41.
119. Rivera, M.C.; Lake, J.A. The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* **2004**, *431*, 152–155.
120. Harish, A.; Kurland, C.G. Mitochondria are not captive bacteria. *J. Theor. Biol.* **2017**, *434*, 88–98.
121. Miller, S.L.; Lazcano, A. The origin of life – did it occur at high temperatures? *J. Mol. Evol.* **1995**, *41*, 689–692.
122. Forterre, P. A hot topic: the origin of hyperthermophiles. *Cell* **1996**, *85*, 789–792.
123. Galtier, N.; Tourasse, N.; Gouy, M. A nonhyperthermophilic common ancestor to extant life forms. *Science* **1999**, *283*, 220–221.
124. Boussau, B.; Blanquart, S.; Necsulea, A.; Lartillot, N.; Gouy, M. Parallel adaptations to high temperatures in the Archaean eon. *Nature* **2008**, *456*, 942–945.
125. Lewalter, K.; Müller, V. Bioenergetics of archaea: ancient energy conserving mechanisms developed in the early history of life. *Biochim. Biophys. Acta* **2006**, *1757*, 437–445.
126. Mulkidjanian, A.Y.; Galperin, M.Y. A time to scatter genes and a time to gather them: Evolution of photosynthesis genes in bacteria. *Adv. Bot. Res.* **2013**, *66*, 1–36.

127. Forterre, P.; Confalonier, F.; Charbonnier, F.; Duguet, M. Speculations on the origin of life and thermophily: review of available information on reverse gyrase suggests that hyperthermophilic procaryotes are not so primitive. *Orig. Life Evol. Biosph.* **1995**, *25*, 235–249.
128. Forterre, P. The origin of DNA genomes and DNA replication proteins. Curr. Opin. Microbiol. **2002**, 5, 525–532.
129. Brochier-Armanet, C.; Forterre, P. Widespread distribution of archaeal reverse gyrase in thermophilic bacteria suggests a complex history of vertical inheritance and lateral gene transfers. *Archaea* **2006**, *2*, 83–93.
130. Forterre, P.; Bouthier de la Tour, C.; Philippe, H.; Duguet, M. Reverse gyrase from hyperthermophiles: probable transfer of a thermoadaptation trait from archaea to bacteria. *Trends Genet.* **2000**, *16*, 152–154.
131. Confalonieri, F.; Elie, C.; Nadal, M.; Bouthier de La Tour, C.; Forterre, P.; Duguet, M. Reverse gyrase: a helicase-like domain and a type I topoisomerase in the same polypeptide. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 4753–4757.
132. Forterre, P.; Gribaldo, S.; Gadelle, D.; Serre, M.C. Origin and evolution of DNA topoisomerases. *Biochimie* **2007**, *89*, 427–446.
133. Lake, J.A. Evidence for an early prokaryotic endosymbiosis. *Nature* **2009**, *460*, 967–971.
134. Wu, D.; Hugenholtz, P.; Mavromatis, K.; Pukall, R.; Dalin, E.; Ivanova, N.N.; ... Eisen, J.A. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* **2009**, *462*, 1056–1060.
135. Sutcliffe, I.C. A phylum level perspective on bacterial cell envelope architecture. *Trends Microbiol.* **2010**, *18*, 464–470.
136. Swithers, K.S.; Fournier, G.P.; Green, A.G.; Gogarten, J.P.; Lapierre, P. Reassessment of the Lineage Fusion Hypothesis for the origin of double membrane bacteria. *PloS One* **2011**, *6*, e23774.
137. Gupta, R.S. Origin of diderm (Gram-negative) bacteria: antibiotic selection pressure rather than endosymbiosis likely led to the evolution of bacterial cells with two membranes. *Antonie van Leeuwenhoek* **2011**, *100*, 171–182.
138. Errington, J. L-form bacteria, cell walls and the origins of life. *Open Biol.* **2013**, *3*, 120143.
139. Battistuzzi, F.U.; Hedges, S.B. A major clade of prokaryotes with ancient adaptations to life on land. *Mol. Biol. Evol.* **2009**, *26*, 335–343.
140. Wimley, W.C. The versatile β-barrel membrane protein. *Curr. Opin. Struct. Biol.* **2003**, *13*, 404–411.
141. Nikaido, H. Molecular basis of bacterial outer membrane permeability revisited. *Microbiol. Mol. Biol. Rev.* **2003**, *67*, 593–656.
142. Zeth, K. Structure and evolution of mitochondrial outer membrane proteins of β-barrel topology. *Biochim. Biophys. Acta Bioenerg.* **2010**, *1797*, 1292–1299.
143. Song, L.; Hobaugh, M.R.; Shustak, C.; Cheley, S.; Bayley, H.; Gouaux, J.E. Structure of staphylococcal α-hemolysin, a heptameric transmembrane pore. *Science* **1996**, *274*, 1859–1865.
144. Dolezal, P.; Likic, V.; Tachezy, J.; Lithgow, T. Evolution of the molecular machines for protein import into mitochondria. *Science* **2006**, *313*, 314–318.
145. Pittis, A.A.; Gabaldón, T. Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature* **2016**, *531*, 101–104.
146. Gerding, M.A.; Ogata, Y.; Pecora, N.D.; Niki, H.; De Boer, P.A. The trans-envelope Tol–Pal complex is part of the cell division machinery and required for proper outer-membrane invagination during cell constriction in *E. coli*. *Mol. Microbiol.* **2007**, *63*, 1008–1025.
147. Philippe, N.; Legendre, M.; Doutre, G.; Couté, Y.; Poirot, O.; ... Abergel, C. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* **2013**, *341*, 281–286.
148. Abergel, C.; Legendre, M.; Claverie, J.M. The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. *FEMS Microbiol. Rev.* **2015**, *39*, 779–796.
149. Gupta, R.S. The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. *FEMS Microbiol. Rev.* **2000**, *24*, 367–402.
150. Koonin, E.V. The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biol.* **2010**, *11*, 209.
151. Koumandou, V.L.; Wickstead, B.; Ginger, M.L.; van der Giezen, M.; Dacks, J.B.; Field, M.C. Molecular paleontology and complexity in the last eukaryotic common ancestor. *Crit. Rev. Biochem. Mol. Biol.* **2013**, *48*, 373–396.

152. Poole, A.M.; Gribaldo, S. Eukaryotic origins: how and when was the mitochondrion acquired? *Cold Spring Harb. Perspect. Biol.* **2014**, *6*, a015990.
153. Koonin, E.V.; Yutin, N. The dispersed archaeal eukaryome and the complex archaeal ancestor of eukaryotes. *Cold Spring Harb. Perspect. Biol.* **2014**, *6*, a016188.
154. Rochette, N.C.; Brochier-Armanet, C.; Gouy, M. Phylogenomic test of the hypotheses for the evolutionary origin of eukaryotes. *Mol. Biol. Evol.*, **2014**, *31*, 832–845.
155. Forterre, P. The universal tree of life: an update. *Front. Microbiol.* **2015**, *6*, 717.
156. Mier, P.; Pérez-Pulido, A.J.; Reynaud, E.G.; Andrade-Navarro, M.A. Reading the evolution of compartmentalization in the ribosome assembly toolbox: the YRG protein family. *PloS One* **2017**, *12*, e0169750.
157. Theriot, J.A. Why are bacteria different from eukaryotes? *BMC Biol.* **2013**, *11*, 119.
158. Adl, S.M.; Simpson, A.G.; Lane, C.E.; Lukeš, J.; Bass, D.; ... Spiegel, F.W. The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* **2012**, *59*, 429–514.
159. Burki, F. The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harb. Perspect. Biol.* **2014**, *6*, a016147.
160. Williams, T.A. Evolution: rooting the eukaryotic tree of life. *Curr. Biol.* **2014**, *24*, R151–R152.
161. He, D.; Fiz-Palacios, O.; Fu, C.J.; Fehling, J.; Tsai, C.C.; Baldauf, S.L. An alternative root for the eukaryote tree of life. *Curr. Biol.* **2014**, *24*, 465–470.
162. Derelle, R.; Torruella, G.; Klimeš, V.; Brinkmann, H.; Kim, E.; ... Eliáš, M. Bacterial proteins pinpoint a single eukaryotic root. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E693–E699.
163. Yubuki, N.; Leander, B.S. Evolution of microtubule organizing centers across the tree of eukaryotes. *Plant J.* **2013**, *75*, 230–244.
164. O'Malley, M.A.; Simpson, A.G.; Roger, A.J. The other eukaryotes in light of evolutionary protistology. *Biol. Philos.* **2013**, *28*, 299–330.
165. Hartman, H.; Fedorov, A. The origin of the eukaryotic cell: a genomic investigation. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 1420–1425.
166. Han, J.; Collins, L.J.; Biggs, P.J.; White, W.T.; Penny, D. Are signature proteins the key to evaluating eukaryotic phylogeny? *Rev. Bioinf. Biomet.* **2013**. *2*, 37–44.
167. Jeffares, D.C.; Poole, A.M.; Penny, D. Relics from the RNA world. *J. Mol. Evol.* **1998**, *46*, 18–36.
168. Collins, L.J.; Kurland, C.G.; Biggs, P.; Penny, D. The modern RNP world of eukaryotes. *J. Hered.* **2009**, *100*, 597–604.
169. Poole, A.M.. Jeffares, D.C.; Penny, D. The path from the RNA world. *J. Mol. Evol.* **1998**, *46*, 1–17.
170. Penny, D.; Hoeppner, M.P; Poole, A.M.; Jeffares, D. C. An overview of the introns-first theory. *J. Mol. Evol.* **2009**, *69*, 527–540.
171. Cavalier-Smith, T. Intron phylogeny: a new hypothesis. *Trends Genet.* **1991,** *7*, 145–148.
172. Ahmadinejad, N.; Dagan, T.; Gruenheit, N.; Martin, W.; Gabaldón, T. Evolution of spliceosomal introns following endosymbiotic gene transfer. *BMC Evol. Biol.* **2010**, *10*, 1.
173. Vesteg, M.; Šándorová, Z.; Krajčovič, J. Selective forces for the origin of spliceosomes. *J. Mol. Evol.* **2012**, *74*, 226–231.
174. Hudder, A.; Nathanson, L.; Deutscher, M.P. Organization of mammalian cytoplasm. *Mol. Cell. Biol.* **2003**, *23*, 9318–9326.
175. Harold, F.M. Molecules into cells: specifying spatial architecture. *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 544–564.
176. Wickstead, B.; Gull, K. The evolution of the cytoskeleton. *J. Cell Biol.* **2011**, *194*, 513–525.
177. Wen, J. The nuclear matrix of *Euglena gracilis* (Euglenophyta): A stage of nuclear matrix evolution? *Biol. Cell* **2000**, *92*, 125–131.
178. Kennedy, J. Particle swarm optimization. In *Encyclopedia of machine learning*; Sammut, C., Webb, G. I., Eds.; Springer, NY, USA , **2011**; pp. 760–766.
179. Jenuwein, T.; Allis, C.D. Translating the histone code. *Science* **2001**, *293*, 1074–1080.
180. Kornberg, R.D. Mediator and the mechanism of transcriptional activation. *Trends Biochem. Sci.* **2005**, *30*, 235–239.
181. Bloom, K.; Joglekar, A. Towards building a chromosome segregation machine. *Nature* **2010**, *463*, 446–456.
182. Mao, Y.S.; Zhang, B.; Spector, D.L. Biogenesis and function of nuclear bodies. *Trends Genet.* **2011**, *27*, 295–306.

183. Dacks, J.B.; Field, M.C. Evolution of the eukaryotic membrane-trafficking system: origin, tempo and mode. *J. Cell Sci.* **2007**, *120*, 2977–2985.
184. Bapteste, E.; Charlebois, R.L.; MacLeod, D.; Brochier, C. The two tempos of nuclear pore complex evolution: highly adapting proteins in an ancient frozen structure. *Genome Biol.* **2005**, *6*, R85.
185. Devos, D.; Dokudovskaya, S.; Williams, R.; Alber, F.; Eswar, N.; ... Sali, A. Simple fold composition and modular architecture of the nuclear pore complex. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 2172–2177.
186. Devos, D.P.; Gräf, R.; Field, M.C. Evolution of the nucleus. *Curr. Opin. Cell Biol.* **2014**, *28*, 8–15.
187. Bell, G. *The Masterpiece of Nature: The Evolution and Genetics of Sexuality*; Croom Helm: London, UK, 1982.
188. Egel, R.; Penny, D. On the origin of meiosis in eukaryotic evolution: coevolution of meiosis and mitosis from feeble beginnings. In *Recombination and Meiosis: Models, Means and Evolution;* Egel, R., Lankenau, D.-H., Eds.; Springer: Heidelberg, Germany, **2007**; pp. 249–288.
189. Raymann, K.; Forterre, P.; Brochier-Armanet, C.; Gribaldo, S. Global phylogenomic analysis disentangles the complex evolutionary history of DNA replication in Archaea. *Genome Biol. Evol.* **2014**, *6*, 192–212.
190. Bryant, M.P.; Wolin, E.A.; Wolin, M.J.; Wolfe, R.S. *Methanobacillus omelianskii*, a symbiotic association of two species of bacteria. *Arch. Microbiol.* **1967**, *59*, 20–31.
191. McInerney, M.J.; Bryant, M.P.; Pfennig, N. Anaerobic bacterium that degrades fatty acids in syntrophic association with methanogens. *Arch. Microbiol.* **1979**, *122*, 129–135.
192. Morris, B.E.; Henneberger, R.; Huber, H.; Moissl-Eichinger, C. Microbial syntrophy: interaction for the common good. *FEMS Microbiol. Rev.* **2013**, *37*, 384–406.
193. Schink, B.; Stams, A.J. Syntrophism among prokaryotes. In *The Prokaryotes – Prokaryotic Communities and Ecophysiology*; E. Rosenberg, et al., Eds.; Springer, New York, **2006**; pp. 309–335.
194. Sieber, J.R.; McInerney, M.J.; Gunsalus, R.P. Genomic insights into syntrophy: the paradigm for anaerobic metabolic cooperation. *Ann. Rev. Microbiol.* **2012**, *66*, 429–452.
195. Bapteste, É.; Brochier, C.; Boucher, Y. Higher-level classification of the Archaea: evolution of methanogenesis and methanogens. *Archaea* **2005**, *1*, 353–363.
196. O'Donnell, M.; Langston, L.; Stillman, B. Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a010108.
197. Onesti, S.; MacNeill, S.A. Structure and evolutionary origins of the CMG complex. *Chromosoma* **2013**, *122*, 47–53.
198. Medagli, B.; Di Crescenzio, P.; De March, M.; Onesti, S. Structure and activity of the Cdc45-Mcm2–7-GINS (CMG) complex, the replication helicase. In *The Initiation of DNA Replication in Eukaryotes*; D.L. Kaplan, Ed.; Springer International Switzerland, **2016**, pp. 411–425.
199. Yuan, Z.; Bai, L.; Sun, J.; Georgescu, R.; Liu, J.; O'Donnell, M.E.; Li, H. Structure of the eukaryotic replicative CMG helicase suggests a pumpjack motion for translocation. *Nat. Struct. Mol. Biol.* **2016**, *23*, 217– 225.
200. MacNeill, S.A. Structure and function of the GINS complex, a key component of the eukaryotic replisome. *Biochem. J.* **2010**, *425*, 489–500.
201. Makarova, K.S.; Koonin, E.V.; Kelman, Z. The CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes. *Biol. Direct* **2012**, *7*, 22329974.
202. Makarova, K.S.; Koonin, E.V. Archaeology of eukaryotic DNA replication. *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a012963.
203. Makarova, K.S.; Wolf, Y.I.; Mekhedov, S.L.; Mirkin, B.G.; Koonin, E.V. Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucl. Acids Res.* **2005**, *33*, 4626–4638.
204. Martin, W.; Koonin, E.V. Introns and the origin of nucleus–cytosol compartmentalization. *Nature* **2006**, *440*, 41–45.
205. Pace, N.R. Origin of life – facing up to the physical setting. *Cell* **1991**, *65*, 531–533.
206. Spang, A.; Saw, J.H.; Jørgensen, S.L.; Zaremba-Niedzwiedzka, K.; Martijn, J.; Lind, A.E.; ... Ettema, T.J. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **2015**, *521*, 173–179.
207. Saw, J.H.; Spang, A.; Zaremba-Niedzwiedzka, K.; Juzokaite, L.; Dodsworth, J.A.; Murugapiran, S. K.; ... Ettema, T.J. Exploring microbial dark matter to resolve the deep archaeal ancestry of eukaryotes. *Phil. Trans. R. Soc. B* **2015**, *370*, 20140328.
208. Hug, L.A.; Baker, B.J.; Anantharaman, K.; Brown, C.T.; Probst, A.J.; ... Suzuki, Y. A new view of the tree of life. *Nat. Microbiol.* **2016**, *1*, 16048.

209. Zaremba-Niedzwiedzka, K.; Caceres, E.F.; Saw, J.H.; Bäckström, D.; Juzokaite, L.; Vancaester, E.; ... Stott, M.B. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **2017**, *541*, 353–358.
210. Da Cunha, V.; Gaia, M.; Gadelle, D.; Nasir, A.; Forterre, P. Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *PLoS Genetics* **2017**, *13*, e1006810.
211. Nelson-Sathi, S.; Sousa, F.L.; Roettger, M.; Lozada-Chávez, N.; Thiergart, T.; ... McInerney, J.O. Origins of major archaeal clades correspond to gene acquisitions from bacteria. *Nature* **2015**, *517*, 77–80.
212. Baum, D.A.; Baum, B. An inside-out origin for the eukaryotic cell. *BMC Biol.* **2014**, *12*, 1.
213. Staley, J.T. Domain Cell Theory supports the independent evolution of the Eukarya, Bacteria and Archaea and the Nuclear Compartment Commonality hypothesis. *Open Biol.* **2017**, *7*, 170041.
214. Staley, J.T.; Fuerst, J.A. Ancient, highly conserved proteins from a LUCA with complex cell biology provide evidence in support of the nuclear compartment commonality (NuCom) hypothesis. *Res. Microbiol.* **2017**, *168*, 395–412.
215. Pross, A. Toward a general theory of evolution: extending Darwinian theory to inanimate matter. *J. Systems Chem.* **2011**, *2*, 1.
216. Peretó, J. Controversies on the origin of life. *Int. Microbiol*. **2005**, *8*, 23–31.
217. Kovacs, N.A.; Petrov, A.S.; Lanier, K.A.; Williams, L.D. Frozen in time: the history of proteins. *Mol. Biol. Evol.* **2017**, *34*, 1252–1260.
218. Wills, P.R. Spontaneous mutual ordering of nucleic acids and proteins. *Orig. Life Evol. Biosph.* **2014**, *44*, 293–298.
219. Caetano-Anollés, G.; Wang, M.; Caetano-Anollés, D. Structural phylogenomics retrodicts the origin of the genetic code and uncovers the evolutionary impact of protein flexibility. *PLoS One* **2013**, *8*, e72225.
220. Longo, G.; Montévil, M.; Sonnenschein, C.; Soto, A.M. In search of principles for a theory of organisms. *J. Biosci.* **2015**, *40*, 955–968.
221. Hengeveld, R.; Fedonkin, M.A. Bootstrapping the energy flow in the beginning of life. *Acta Biotheor.* **2007**, *55*, 181–226.
222. Muchowska, K.B.; Varma, S.J.; Chevallot-Beroux, E.; Lethuillier-Karl, L.; Li, G.; Moran, J. Metals promote sequences of the reverse Krebs cycle. *Nat. Ecol. Evol.* **2017**, *2017 doi: 10.1038/s41559-017-0311-7.*
223. Kauffman, S.A. *The Origins of Order*; Oxford University Press: Oxford, UK, **1993**.
224. McKenney, K.M.; Alfonzo, J.D. From prebiotics to probiotics: The evolution and functions of tRNA modifications. *Life* **2016**, *6*, 13.
225. Damer, B. A field trip to the Archaean in search of Darwin's warm little pond. *Life* **2016**, *6*, 21.
226. Van Kranendonk, M.J.; Deamer, D.W.; Djokic, T. Life springs, *Sci. Am.* **2017**, *317*, 22–29.
227. Egel, R. Integrative perspectives: In quest of a coherent framework for origins of life on earth. In *Origins of Life: The Primal Self-Organization*; Egel, R., Lankenau, D.-H., Mulkidjanian, A.Y., Eds.; Springer-Verlag: Heidelberg, Germany, **2011**; pp. 289–360.
228. Oparin, A.I. Evolution of the concepts of the origin of life, 1924–1974. *Orig. Life Evol. Biosph.* **1976**, *7*, 3–8.
229. Bada, J.L.; Lazcano, A. Prebiotic soup – revisiting the Miller experiment. *Science* **2003**, *300*, 745–746.
230. Szostak, J.W.; Bartel, D.P.; Luisi, P.L. Synthesizing life. *Nature* **2001**, *409*, 387–390.
231. Higgs, P.G.; Lehman, N. The RNA World: molecular cooperation at the origins of life. *Nature Rev. Genet.* **2015**, *16*, 7–17.
232. Finlay, B.J. Global dispersal of free-living microbial eukaryote species. *Science* **2002**, *296*, 1061–1063.
233. Mayr, E.; Provine, W.B. *The evolutionary synthesis: perspectives on the unification of biology*. Harvard University Press, 1998.
234. Weber, B.H. Extending and expanding the Darwinian synthesis: the role of complex systems dynamics. *Stud. Hist. Philos. Biol. Biomed. Sci.* **2011**, *42*, 75–81.
235. Powner, M.W.; Sutherland, J.D.; Szostak, J.W. Chemoselective multicomponent one-pot assembly of purine precursors in water. *J. Am. Chem. Soc.* **2010**, *132*, 16677–16688.
236. Steer, A.M.; Bia, N.; Smith, D. K.; Clarke, P.A. Prebiotic synthesis of 2-deoxy-d-ribose from interstellar building blocks promoted by amino esters or amino nitriles. *Chem. Commun.* **2017**, *53*, 10362–10365.
237. Harish, A.; Kurland, C.G. Empirical genome evolution models root the tree of life. *Biochimie* **2017**, *138*, 137–155.
238. Harish, A.; Kurland, C.G. Akaryotes and Eukaryotes are independent descendants of a universal common ancestor. *Biochimie* **2017**, *138*, 168–183.