Research in Microbiology 168 (2017) 395–412

www.elsevier.com/locate/resmic

Review

Ancient, highly conserved proteins from a LUCA with complex cell biology provide evidence in support of the nuclear compartment commonality (NuCom) hypothesis

James T. Staley [a], John A. Fuerst [b,*]

[a] Department of Microbiology and Astrobiology Program, University of Washington, Seattle 98195, USA
[b] School of Chemistry and Molecular Biosciences, University of Queensland, St. Lucia, Queensland 4072, Australia

Received 24 August 2016; accepted 9 January 2017
Available online 19 January 2017

Abstract

The nuclear compartment commonality (NuCom) hypothesis posits a complex last common ancestor (LUCA) with membranous compartments including a nuclear membrane. Such a LUCA then evolved to produce two nucleated lineages of the tree of life: the Planctomycetes–Verrucomicrobia–Chlamydia superphylum (PVC) within the Bacteria, and the Eukarya. We propose that a group of ancient essential protokaryotic signature proteins (PSPs) originating in LUCA were incorporated into ancestors of PVC Bacteria and Eukarya. Tubulins, ubiquitin system enzymes and sterol-synthesizing enzymes are consistent with early origins of these features shared between the PVC superphylum and Eukarya.

Keywords: LUCA; PVC superphylum; Planctomycetes; Verrucomicrobia; Origin of eukaryotes; Ancient cell compartmentalization

1. Introduction

1.1. Terminology

PVC Bacteria are defined as the Planctomycetes–Verrucomicrobia–Chlamydia superphylum [1], some members of which are nucleated, i.e. their genes, DNA and probably also transcription [2] and DNA replication occur in a PLFA ester-linked membrane-bounded compartment within the cell – for example, within the PVC superphylum, all examined species of the phylum Planctomycetes and at least four species of phylum Verrucomicrobia and one species of phylum Lentisphaerae. The position of phylum Poribacteria as a member of the PVC superphylum has been ambiguous, but one recent study based on single genomes of this uncultured sponge inhabitant suggests it is at least a sister lineage significantly linked to the PVC superphylum [3].

LUCA is an acronym for the last universal common ancestor [4], assumed in this paper to be the ancestor of all extant identifiable domains of cellular life regardless of number of domains. It is proposed that LUCA was already a cellular form of life bounded by a membrane (though this does not exclude genomic contributions from viruses or a viral non-cellular or cellular 'domain' or an RNA genome), and it is part of the proposal of the NuCom hypothesis that it contained internal membranes associated with or bounding the cell's genomic DNA forming a nuclear compartment. However, the nature of LUCA has been a subject of considerable discussion and debate.

Enucleation is the process whereby a nucleated ancestral organism loses its nuclear compartment through reductive evolution. NuCom regards the common bacteria as enucleate descendants of nucleated members of the PVC superphylum. For example, the Verrucomicrobia are viewed as likely ancestors of the Proteobacteria [5].

* Corresponding author.
E-mail address: j.fuerst@uq.edu.au (J.A. Fuerst).

http://dx.doi.org/10.1016/j.resmic.2017.01.001

Common bacteria are defined as all bacteria possessing enucleate cells—cells with naked genomic DNA not enveloped in either single or double internal membranes within the cytoplasm.

Protokaryote (derived from the Greek proto for 'first' and karyon for kernel or nucleus) refers to the nucleated pre-domain ancestral cell state of LUCA. The ancestors of the PVC Bacteria and the Eukarya are proposed to have been two phylogenetically distinct protokaryote descendants in which at least a simple form of nuclear compartment had evolved from a shared nucleated ancestor; in the context of this paper, these are also conceivably related to the progenote stage of Woese in which the genetic composition had not yet 'annealed' to any particular domain identity. The protokaryote state perceived here is a complex cell-like entity in which at least a simple nucleus and nuclear envelope had evolved. This term is analogous to Patrick Forterre's proposed term synkaryote [6].

Protokaryotic signature proteins (PSPs) are homologous proteins currently found in some representatives of the PVC superphylum, and most, if not all, representatives of the Eukarya (where they have been termed "eukaryotic signature proteins").

Eukaryogenesis is defined by the extensive and continuous evolution of eukaryote cell and organism organization throughout the evolution of their domain – note that a key stage of such evolution must have been the origin of the nucleated cell, but that this does not necessarily entail a symbiotic event, unlike the origin of mitochondria and chloroplasts, which became key organelles of developing contemporary eukaryote complexity.

The nuclear compartment commonality (NuCom) hypothesis posits a last common ancestor (LUCA) of considerable cell biological complexity with membranous compartments including a nuclear membrane [5]. Such a LUCA then evolved to produce two nucleated lineages of the tree of life: the Planctomycetes–Verrucomicrobia–Chlamydia superphylum (PVC) bacteria and the Eukarya. According to the NuCom, hypothesis, the first bifurcation of LUCA gave rise to two nucleated ancestors: one was an ancestor of the Eukarya which have always been nucleated. The other was the ancestor of the Planctomycetes–Verrucomicrobia–Chlamydia superphylum of domain Bacteria. It follows that the NuCom hypothesis proposes a LUCA with a complex compartmentalized protokaryotic but cellular state that was ancestral to two lineages (Bacteria and Eukarya), each of which originally possessed a membrane-bounded nucleus and transport mechanisms between nucleus and cytoplasm, that they possessed ester-linked lipids and maybe also ether-linked lipids, and likely also sterols, and with features of compartmentalization and spatial separation of functions shared and retained by PVC Bacteria and Eukarya.

In this essay, we review the background to the tree of life relevant to compatibility with this hypothesis, the relationships of domain Eukarya to the archaeal and bacterial lineages, and then consider in more detail the arguments for the NuCom hypothesis, including the possibility of a LUCA with a relatively complex compartmentalized cell biology; finally, we discuss the evidence for the proposal that a group of ancient essential protokaryotic signature proteins (PSPs) originating in LUCA were incorporated into ancestors of PVC Bacteria and Eukarya.

2. Evolutionary context: LUCA, the domains of life and the problem of their origin and number

One of the major remaining deep evolutionary questions regards the origin of the Eukarya, Bacteria and Archaea. These three domains of life were identified by Carl Woese's phylogenetic analyses of the small subunit (ssu) ribosomal RNA that all organisms have as illustrated in the tree of life [7]. Analyses based on genomes have largely confirmed the concept of the three domains [8]. A root for the tree has been controversial. A root in the Bacteria has been the consensus position, but more recently strong cases for keeping an open mind regarding even a possible eukaryote root have been made and seriously considered [9–11]. It may also be necessary to keep an open mind concerning the probable limitations of sequence-based phylogenomics in attempts to obtain any form of credible root for the tree of life [11], especially considering the absence of useful outgroups and the problem of mutation saturation limiting the depth resolvable for any tree of life [12].

However, recently the three-domain interpretation has been contested and a 2-domain division of the ToL has been proposed on the basis of phylogenomic analyses [13–17]. This view is a continuation of the ring theory of life that proposes a bacterium and an archaeon fused together to produce the ancestral eukaryotic cell [18]. Such a 2-domain tree suggests that in addition to the Bacteria stem, the other stem comprises the Archaea and that it gives rise, subsequently, to the Eukarya. The unique rRNA (and correlated ribosomal protein) extension segments of Eukarya are, however, also consistent with a separate domain from very ancient times and thus an ancient LUCA with eukaryote-like translational system and correlated complex cell organization, and argue against domain Eukarya origin from fusion with or derivation from members of other domains [19]. Other analyses also are consistent with a very early differentiation of three distinct domains and not with an origin of domain Eukarya from a fusion of Bacteria and Archaea lineages. For example, the phylogenetics of glycolytic enzymes indicates a vertical origin from an ancient stem shared by later developing Bacteria and Eukarya lineages [20] and the analysis of fold superfamilies (FSFs) in the domains indicates a similarity in predicted FSF ensemble for a stem ancestor of the domains similar to that in Eukarya, and more complex than that in either Bacteria or Archaea [21]. It seems hard to invoke some type of increased evolutionary rate after a hypothetical fusion event to account for complex eukaryote ribosomes with extensive negative selective constraints on change [19]. An ancient eukaryal nuclear lineage could have evolved a nuclear compartment endogenously rather than via domain member fusion.

The concept that the Eukarya comprise a major primordial domain of life, as proposed by Woese on the basis of ssu rRNA

sequence signatures, is not accepted in the 2-domain model [13]. However the 2-domain model is not accepted by all. Before the rise of the contemporary versions of 2-domain trees, Kurland and Penny [22,23] had rejected archaeal and bacterial fusion as the origin of domain Eukarya, on the basis of bioinformatic artefacts influencing phylogenomic analyses, and proposed a complex predatory nucleated 'proto-eukaryote' ancestor of Eukarya, preparing the way for acquisition of endosymbiont bacteria forming mitochondria rather than a chimeric eukaryote genome. Forterre [24] has challenged the model of a classical archaeal ancestor for the Archaea and Eukarya, proposing instead a complex common ancestor of those domains. Staley, in the NuCom hypothesis [5], proposed a nucleated LUCA. These contrary views propose more or less strongly that the ancestor of the Eukarya clade is likely to have been complex [24], and not necessarily recognizably archaeal. They have been largely ignored by the 2-domain enthusiasts. However, a recent review has presented effective arguments for the 3-domain perspective and against versions of the 2-domain view involving Eucarya membership of a archaeal lineage [25]. The possibility of a 3-domain tree being most likely after all is significant for the NuCom hypothesis, since this hypothesis would only seem to be feasible if some form of 3-domain tree applies, where Archaea and Eucarya may be sisters, but are not related in some manner relating Eucarya to Archaea in a ladder of vertical inheritance. If the 3-domain tree applies, then Eucarya could be derived directly from a compartmentalized and even nucleated LUCA by retention of complex cell biology rather than from a possibly thermoreduced archaeal ancestor.

Despite their recent dominance, 2-domain phylogenies have some major weaknesses, as do the related hypotheses proposing fusion between an archaeon and bacterium to account for the Eukarya domain. Acceptability of the validity of 2-domain phylogenies is weakened by taxon and character sampling problems, such as overrepresentation of archaeal taxa and the use of extremely fast-evolving species in many of the datasets used, and by unresolved problems posed by long-branch attraction (LBA) induced by fast-evolving species when the outgroup in the tree has a long branch [26], and the use of concatenated datasets for phylogenetic analysis prone to problems with variation in evolutionary rates of the different concatenated genes merged for analysis [25]. Concerning LBA artefacts, even recently used sophisticated Bayesian methods cannot cope with the problem of LBA induced by fast evolving species when the outgroup has a long branch (which is the case of Bacteria in the universal tree) [11]. LBA artefacts have been demonstrated to influence determination of a root for a tree of Eukarya [27] in the case of trees based on mitochondrial proteomes of invertebrate animals when increasingly distant outgroups are used [11], and in the case of the tree of life and a test for a bacterial root [26]. It would appear that the best phylogenetic methods can mislead when there are large evolutionary model variations, which occurs when deriving a tree of life, since the outgroup must be distant and substitution rates are likely to vary over the long evolutionary times involved [11]. Events such as, for example, loss of a nucleus or adapting to hyperthermophily via thermoreduction could lead to drastic variation affecting such phylogeny attempts [11]. The sisterhood or even direct descent relations of Archaea and Eukarya domains and the reality of a root in domain Bacteria may be highly dependent on such artefacts in some trees using fast-evolving alignment positions, and 2-domain conclusions may be resting on such shaky foundations in some cases.

Thus, taxon sampling balancing numbers of taxa representing each domain has been reported to produce tree topologies consistent with three distinct domains [25], in contrast with trees based on unbalanced sampling. The origin of Eukarya from domain Archaea asks us to accept the transformation of one domain into another, which poses difficulties concerning, for example, the characteristic nature of ribosomal RNA in each domain, the replacement of archaeal lipids by lipids shared by Eukarya and Bacteria, and the emergence of unique eukaryal virus families [24]. It also has difficulties connected with explaining the origin of the many core eukaryotic genes that appear to have been present already in the last eukaryotic common ancestor, yet only some of which are found distributed in a patchy way in diverse modern Archaea [12]. Even though Archaea have many molecular affinities with Eukarya, e.g. in 'informational' genes associated with transcription and translation, and possess a few genes once classed as eukaryote-specific, they lack eukaryote-specific features at cellular and even molecular levels, such as splicesosomes, endocytosis-associated endosomes and vesicle trafficking system, Golgi apparatus, nuclear envelopes and associated features such as nuclear pores, cytoskeleton-linked mitosis and many other features linked to complex cellular organization [24]. Emergence of both the separate domains Archaea and Eukarya from a complex last common ancestor of Archaea and Eukarya is a more credible scenario than derivation of Eukarya from Archaea, which can also leave a 3-domain tree as the most likely for the tree of life [12]. This is also consistent with a post-Darwin evolutionary model involving common origins of distinct tree branches rather than an Aristotelian direct ladder-like 'scale of being' progression from one major lineage to another [24].

The view of ToL as consisting of three domains is consistent with unique features of the Eukarya which are difficult to explain with a 2-domain model involving transformation of an archaeal lineage. The domain Eukarya was proposed first by Winker and Woese, when the 3-domain system was published, together with Bacteria and Archaea, and this was based on distinctive characteristics and signatures of the small subunit 18S rRNA of Eukarya [28]. They referred to this as the 'nuclear lineage' of Eukarya to distinguish this form of 18S rRNA that is found in typical modern Eukarya that contain mitochondria. In this regard, Forterre has pointed out how difficult it is to imagine the evolution of 18S rRNA via fusion of the Bacteria and Archaea domains [19,29]. It is simpler to accept that Eukarya were ancient and identifiable as the third domain, rather than as one derived from either or both of the other major domains [19]. It is noteworthy that the Winker and Woese view is consistent with the NuCom hypothesis. As one

might expect, analogous considerations apply to domain-specific properties of ribosomal proteins [29].

The recent discovery of the uncultured "Lokiarchaeota", an example of 'microbial dark matter' with several 'ESPs' (eukaryotic signature proteins) but a deeply branching stem encompassing both Archaea and Eukarya [30], has been interpreted as consistent with the 2-domain hypothesis, but this is also controversial [25,31]. Such "ESPs" if not explicable by a horizontal gene transfer (HGT) origin or other explanation connected with metagenome contamination with eukaryovirus or eukaryotes [25], are at least as consistent with the interpretation of the retention of such characters from a eukaryote-like complex LAECA (last Archaea–Eukarya common ancestor) or even LUCA as they are with an interpretation of them as archaeal precursors to vertically descendant eukaryal complexity.

The problems of contemporary sophisticated phylogenetics mean that a definitive root for the domains of life has yet to be determined, and recently there has even been a resurgence of serious analysis of the possibility of a complex common ancestor or root organism of the domains consistent with a 'Eukarya first' (EF) hypothesis [10,23]. Such a possibility has been considered in the past [22,32,33], but has remained on the periphery of discussion, perhaps due to the Aristotelian 'great chain of being' bias towards progression from simple to complex in evolution, as suggested by Gouy et al. [11]. There are a number of analyses, however, that suggest a reductive progression from complex to simple is at least, if not more, probable [11,34,35].

Character state reconstruction has also been used to more clearly define the probable complexity of the stem ancestor of the domains [36,37]. The difficulty that even the most advanced phylogenetic analysis has in determining the root of the ToL domains means that the possibility of an ancestral LUCA with a complex cell, perhaps even of the Eukarya lineage, cannot yet be rejected [11]. And analyses such as those based on protein fold superfamily domains as well as critical review of the phylogenetics of ToL root hypotheses support the potential validity of such a possibility [11,21,38,39]. Proteomic analysis indicates that the primordial LUCA ancestor probably possessed an advanced metabolism as well as translation via ribosomes [36]. Analysis of linker regions in proteins throughout the three domains also suggests that reductive evolution may have dominated the origin of the Bacteria and Archaea domains, consistent with a complex stem ancestor of the domains [35].

In light of the discovery of Lokiarchaeota [30] and potential multiple interpretations of evolutionary meaning of its "ESPs", it may now be time to reconsider the concept of retention of characters from a complex ancestor by several different groups of isolated and divergent, even deep-branching taxa within the Bacteria as well as the Archaea. This has the corollary of proposing an ancient complex common ancestor of the domains, perhaps even one with complex cell biology and compartments including nuclear compartments. However, the "ESPs" of Lokiarchaeota have been critically reassessed and may not prove to be as firm a support for retention of ancient characters as appeared at first [31]. Even in the light of this analysis, however, it is of interest that some of the 'ESPs' discovered in Lokiarchaeota [30], such as gelsolin domains, BAR domains and homologs of several components of the ubiquitin protein degradation system, can also be found in the PVC superphylum of the Bacteria (see below). The bacterial stem of the tree produces many lineages that appear as branches in the tree. However, analyses of their 16S rRNA sequences using slow-evolving positions has suggested that the phylum Planctomycetes of the PVC superphylum of the Bacteria is ancestral to the other enucleate lineages [40]. A deep branch for at least the Planctomycetes has also been supported by proteomic analysis [41]. And at least one analysis resulting in a 3-domain tree based on concatenated large and small subunits of rRNA of 16 taxa (in which the eukaryotes form the sister group to a monophyletic Archaea) groups all the Bacteria except the planctomycete *Rhodopirellula* on one side of the root [42]. This suggests a similar approach, but including a wider sampling of PVC taxa, may be useful to testing the hypothesis of deep PVC branching. All these analyses must be considered tentative regarding PVC-relevant conclusions without further more intensive analysis. However, any confirmed deep branching for planctomycetes or other PVC members would suggest that retention of early features of ancestral compartmentalized cells might have been retained in this lineage, as has apparently also occurred in the Lokiarchaeota within Archaea.

3. The nuclear compartment commonality (NuCom) hypothesis

Cell membranes comprise one of the most ancient properties of cellular life. Therefore, their origin must have occurred in LUCA. According to NuCom, one of the major commonalities between the ancestors of the Bacteria and Eukarya is that the cell membranes (and hence nuclear membranes) of these two domains contain PLFA membranes. In both domains glycerol 3-phosphate (G3P) with sn-1,2 stereochemistry is linked to fatty acid side chains by ester bonds [5]. Indeed, so far as known, all of the enzymatic steps involved in the formation of the lipid membrane moieties are carried out by homologous proteins between these two domains [43,44], providing strong support for the NuCom hypothesis.

This is in contrast with the Archaea with their isoprenoid-based ether-linked membrane lipids, and there are also very distinctive stereochemical differences in the glycerol backbones of such lipids [5].

The distinctiveness of membrane composition between Archaea on the one hand and both Bacteria and Eukarya on the other has always posed a problem for theories of eukaryal origins – 'membrane heredity' is a significant feature of cell inheritance [45] and should be taken into account in tracing Domain relations [46]. A marker for the evolution of archaeal lipids is the enzyme (S)-3-O-geranylgeranylglyceryl phosphate synthase (GGGPS), which provides the unique G1P-based backbone stereochemistry for archaeal lipids, and it has been proposed (on the basis of structural biology of the

enzyme) that evolution of this enzyme combined with isolation due to encapsulation by membranes with resulting distinctive archaeal lipids was a pivotal step in evolution of Archaea from LUCA [47]. Such a scenario is also consistent with a compartmentalized LUCA as a root for the tree of life. This does not mean that the original nucleated ancestors were full-fledged, eukaryote-like organisms comprised of highly sophisticated cell types with complex nuclei with all the features of modern nucleocytoplasmic transport, an uncoupled transcription–translation system, etc. They were likely relatively simple in comparison – entities we thus refer to as 'protokaryotes' (derived from Greek for 'first nucleus') rather than 'protoeukaryotes' (for 'first true nucleus'). Possible selective advantages of internal compartmentalization in a protokaryote include metabolic and molecular biology process advantages of macromolecular crowding [22], protection from virus infection [48], efficiencies of physical separation of distinct functions, or primordially, a separation from other cell processes of a specialized and potentially toxic biochemistry. Such separation has been proposed as associated with the membrane-bounded anammoxosome of anaerobic ammonium oxidizing planctomycetes [49]. Segregation of such biochemically specialized organelles during division as occurs with anammoxosomes [50] might become linked to chromosome segregation and thus nuclear function. Interesting potential supporting evidence for ancient compartmentalized protocells with membranes that were ancestral to the three domains is experimental work on the abiotic chemistry of membrane formation. Multivesicular compartments can be generated simultaneously with and situated inside membrane-bounded 'protocells' [51]. Such cells, if containing genomes within the compartments as well as molecular biology and metabolic machinery, might be self-replicating and the compartmentalization likely provides a selective advantage. This advantage might entail the use of compartments to separate reactions of different efficiency or pH optima, to increase the concentration of reactants via 'macromolecular crowding' or the use of the compartments to protect the genome from toxins and mutagens. Internal membrane complexity could lead to organelle formation, and may have been part of pre-cellular evolution leading to the first cell and then to the last common ancestor of the domains. Supporting potential evolution of membrane vesicle formation in LUCA is the presence of extracellular vesicle formation systems in all of the three domains of life [52,53] and the presence of endocytosis-associated vesicles in at least one compartmentalized planctomycete bacterium, of endosome-related ESCRT proteins in some archaea, and, of course, in eukarya.

The invention of thermally stable ether-linked lipids found in Archaea may have been coincident with conditions favoring hyperthermophiles growing at temperatures greater than 60–80 °C, and the evolution of the archaeal lineage [24]. However, it is relevant to note that there is a member of the anammox planctomycetes, in the *Candidatus* genus *Scalindua*, in which the ladderane lipids unique to anammox occur as ether-linked or ester-linked forms in the same strain and species [54]. However, the stereochemistry of the ether-linked form in anammox is apparently of the sn-1,2 type like that in known bacterial and eukaryal lipids rather than those of Archaea [55]. The anammox group of the Planctomycetes carry out the process of anaerobic ammonia oxidation in suboxic aquatic environments, a process that is also likely to be very ancient. The anammox reaction occurs in cell compartments the bounding membranes of which contain ladderane lipids with ether linkages [54]. This is consistent with a deep-branching position within domain Bacteria and retention of some features of an ancestor prior to the archaeal split from the ancestral progenote or even protokaryote species pool, where both ester- and ether-linked lipids may have coexisted in the same cell. In support of this view of such an ancestor is a fold superfamily analysis of likely properties of LUCA which indicate probable ability to synthesize both ester and ether lipids, but of the sn-1,2 stereochemistry, later replaced by sn-2,3 during 'annealing' of the Archaea [36], consistent with the earlier lipid ancestry scenario of Glansdorff [32]. This is consistent with a stem lineage linking Bacteria and Eukarya features in an ancestral cell type, with only some deep lineages within Bacteria possessing both ester- and ether-linked lipids potentially reflecting the membrane lipids of a LUCA cell.

Sterols, which are found in the membranes of all Eukarya, have also been found in some members of the PVC superphylum as well as enucleate members of the common bacteria, i.e. some Proteobacteria. For example, cells of the nucleated planctomycete *Gemmata obscuriglobus* contain sterols in their membranes, a feature not reported in most common bacteria, but found in all Eukarya membranes [56]. The section on ancient proteins discusses sterols and their biosynthesis in more detail. Sterols have not yet been found in Archaea.

Additional arguments exist that support a view of LUCA as a relatively complex organism including complexity of cell biology and of internal structure as well as biochemistry, e.g. from the argument that RNA-based molecular biology of eukaryotes is consistent with an ancient RNA world, that membrane-bounded acidocalcisome organelles and associated enzymes are ancient features, etc. [4,46,57–59]. It is also conceivable that viruses contributed substantially to complexity in the proto-eukaryal or protokaryotic stem ancestor, and in the case of large DNA viruses, even to cellular complexity and evolution of internal membranes enveloping genomic DNA [24].

According to NuCom, the first bifurcation of LUCA gave rise to two nucleated ancestors: one was an ancestor of the Eukarya which have always been nucleated. The other was the ancestor of the Planctomycetes–Verrucomicrobia–Chlamydia superphylum of domain Bacteria. The PVC ancestor gave rise to PVC Bacteria with cell compartments and to the common bacteria, in the latter case by a process of reductive evolution which simplified their cell structure, most likely to enable them to more efficiently adapt to their niches. This reductive evolution may have been enhanced by a geological or environmental event such as a thermal catastrophe resulting in thermoreduction. The PVC lineage may have separated from

other Bacteria before thermoreduction and would be predicted to be mesophilic or only moderately thermophilic. Thermoreduction has been previously proposed as the phenomenon correlated with evolution of hyperthermophile deep members of Bacteria and Archaea domains from a mesophilic ancestor [60]. Given the advent of a high environmental temperature, the coupling of transcription and translation enabled by absence of a nuclear membrane in a 'reduced' cell plan would allow such prokaryotes (i.e. akaryotes) to bypass the problem of mRNA heat-induced hydrolysis confronting any cell type with a eukaryote plan and uncoupled transcription–translation [60]. Future analyses of possible growth temperatures of PVC superphylum ancestors could be undertaken along the lines of the analyses of combined rRNA and protein sequences used to determine the growth temperature of LUCA and the bacterial ancestor, resulting in conclusion of a probable mesophile LUCA and thermophile ancestor of the Bacteria [61].

DNA replication was a seminal event in evolutionary biology. It is through the DNA-composed genes of contemporary organisms and other cell properties that we attempt to reconstruct the evolutionary pathway of all organisms. The tree of life is based on 16S and 18S ribosomal RNA gene sequences that are encoded in the DNA sequences of the Bacteria, Eukarya and Archaea, and this Domain distinction has largely been confirmed by genome and large protein dataset analysis [8,62]. A major obstacle posed for any fusion hypothesis for the origin of eukaryotes from a common bacterium and an Archaeon is the difficulty of explaining how eukaryal ribosomes could be derived from those of the Archaea and how bacterial ribosomes would have been lost [19], and there are also obstacles involving how contributions of lipid composition of any archaeal partner to modern eukaryotes with lipids of contrasted stereochemistry can be explained.

One of the arguments favoring a protokaryote LUCA and NuCom is the great variety of RNA molecules and functions performed by them in the Eukarya relative to Archaea and common bacteria, suggesting a direct link of the protokaryotes to the RNA world before DNA genomes [58,63]. An RNA-based LUCA could actually have been compartmented and contained an ancient cytoskeleton, based on the complexity of proteins coded for by modern RNA viruses [19]. Interestingly, the RNA world may have already consisted of three domains of RNA genome 'cellular types,' and DNA invention and replication systems may have evolved as a resistance mechanism against RNA viruses [29].

Although it is difficult to comprehend the pre-Domain world of LUCA using extant organisms, we propose herein that ancient LUCA proteins harbored in the PVC Bacteria and Eukarya provide important phylogenetic information about the pre-DNA world and the evolution of the Domains.

In summary, the NuCom hypothesis proposes that the LUCA was a complex compartmentalized protokaryotic state that was ancestral to two lineages (Bacteria and Eukarya), each of which originally possessed a membrane-bounded nucleus and transport mechanisms between nucleus and cytoplasm, that they possessed ester-linked lipids and maybe also ether-linked lipids, and likely also, sterols, and with features of compartmentalization and spatial separation of functions retained in PVC Bacteria, Eukarya and some Archaea such as *Ignicoccus* [64,65]; possibly complex features also occur in *Lokiarchaeota* as judged from annotated ESPs [30]. It should be noted that some previous studies have proposed a complex LUCA, but have fallen short of proposing a structurally compartmentalized protokaryotic LUCA.

Arguments for the NuCom hypothesis include the following:

1) An early origin of cellular life could have involved internal membrane compartments from the outset, as suggested by recent protocell experiments [51].
2) A Eukarya nuclear lineage is supported by unique rRNA expansion segments and corresponding ribosomal protein modification unknown in the other two domains, which argues against derivation from other domains [19].
3) Based on the RNA world hypothesis and the roles of RNA molecules of significance in eukaryote cell biology [58], LUCA would have had a complex protokaryote but RNA-based molecular biology with resemblance to many specific features of modern eukaryote molecular biology.
4) Evidence for the early evolution of membrane-bounded acidocalcisome compartments and their possible presence in the last common ancestor of the Domains see [66].

This paper further develops the concept of the NuCom hypothesis by postulating that members of some Bacteria and Archaea may have retained complex (eukaryote-like) cell biology features of an early compartmentalized LUCA before Domain differentiation, especially members of phylum Planctomycetes and related PVC phyla (e.g. Verrucomicrobia, Chlamydiae and Lentisphaerae) within the domain Bacteria and possibly Lokiarchaeota and *Ignicoccus* among the domain Archaea.

This hypothesis involves the following proposals:

1) The possible retention of some structural features such as compartmentalization via internal membranes of a complex ancestor cell, e.g. in modern PVC Bacteria and possibly *Ignicoccus* in the Archaea (see [64] – this is an interpretation of such features consistent with an early complex cell and retention of such features from that ancestor.
2) Retention of some ancient eukaryal molecular cell biology in compartmented PVC Bacteria, e.g. deep-branching tubulins in the genus *Prosthecobacter* of phylum *Verrucomicrobia* in the PVC superphylum, and an ubiquitin system of protein recycling as well as a simple sterol lipid synthesis pathway in Planctomycetes.
3) Retention from a protokaryote LUCA of both ether- and ester-linked lipids in the ladderanes of the compartmented anammox planctomycetes, deep-branching members of phylum Planctomycetes.

4. LUCA and the compartmentalized PVC superphylum

If domains Archaea and Eucarya are sister groups sharing a last common ancestor [24], then two stems of the tree of life were produced as the first bifurcation of living organisms from a cellular LUCA, one an 'Arkarya' lineage containing the common ancestor of Archaea and Eucarya and the other the deepest ancestors of the Bacteria [12]. Concerning the relevance of a root for the tree of life to the NuCom hypothesis, it should be noted that a rooting between a lineage leading to Bacteria and a lineage leading to Archaea and Eukarya is compatible with a LUCA with some eukaryote-like features later lost in most Archaea and Bacteria [12]. The NuCom hypothesis does not require a root for the ToL in the Eukarya lineage, since loss of ancestral features could have taken place independently in Archaea and Bacteria, with Archaea and Eukarya sister domains rather than forming a Archaea–Eukarya vertical branch. According to NuCom, the bacterial lineage was initiated by an ancestor of the PVC (Planctomycetes–Verrucomicrobia–Chlamydia) superphylum of the Bacteria. This is to some extent consistent with at least one universal tree recently proposed as consistent with a synthesis of biological and molecular sequence considerations [12]. Evidence for the deep phylogeny of at least the planctomycetes within this group comes from careful analyses of 16S rRNA [40]. All genera so far examined from the phylum Planctomycetes possess internal membrane-bounded compartments [67–69], including a major internal membrane-bounded compartment enclosing the nucleoid (Fig. 1A). One species (*G. obscuriglobus*) of planctomycete within the PVC superphylum still retains what has been interpreted as a nuclear envelope analogous to those in eukaryote cells [67,70,71] (Fig. 1A), and this same species performs endocytosis-like protein uptake linked to vesicle trafficking [72], an ability characteristic of eukaryotic cells, and also synthesizes simple sterols [56], a eukaryote-characteristic membrane component found only rarely in Bacteria. Significantly, transcription and translation in cells of this species appear to be separated into distinct regions of the cell, as expected in a nucleated organism [2]. The PVC superphylum and, in particular, the phylum Planctomycetes, comprise an ideal candidate for a group retaining certain features of a proposed complex eukaryote-like LUCA. Supporting this are the features of compartmentalized cells shared between members of phylum Verrucomicrobia and phylum Planctomycetes of the PVC group, including a major membrane-bounded nucleoid-containing compartment (Fig. 1A, B and [73]) the reported occurrence of compartmentalized cells in a species of the phylum Chlamydiae of this group, *Simkania negevensis* [74], and the occurrence of proteins homologous to eukaryote tubulin in genus *Prosthecobacter* of phylum Verrucomicrobia [75]. The second, notably divergent lineage was the Eukarya–Archaea stem of the tree of life.

A major feature of NuCom is that the initial member of the Eukarya lineage was a nucleated cell, in direct opposition to the archaeal hypothesis which regards the initial two domains as prokaryotic (e.g. [17]. The proposal of a nucleated ancestor is supported by the published arguments of Forterre [4,33], Kurland et al. [22] and Glansdorff [32], who proposed a complex eukaryote-like LUCA, and the origin of Bacteria and Archaea through the reductive evolution of a eukaryote-like ancestor of the domains with a complex cell biology. In these views, even a compartmentalized LUCA may have had an RNA-based molecular biology and genome. This is consistent with the NuCom hypothesis which specifically proposes that DNA replication cemented the first protogenomes of the ancestors of the two stems of the tree for the nucleated PVC superphylum and the Eukarya.

In summary, arguments favoring a compartmentalized LUCA are as follows: 1) a eukaryote promoter system favors evolutionary rearrangements [46]; 2) it is known from bioinformatic studies of distributions of eukaryote-specific proteins that the last eukaryotic common ancestor (LECA) had most or all of the features of the molecular cell biology of the

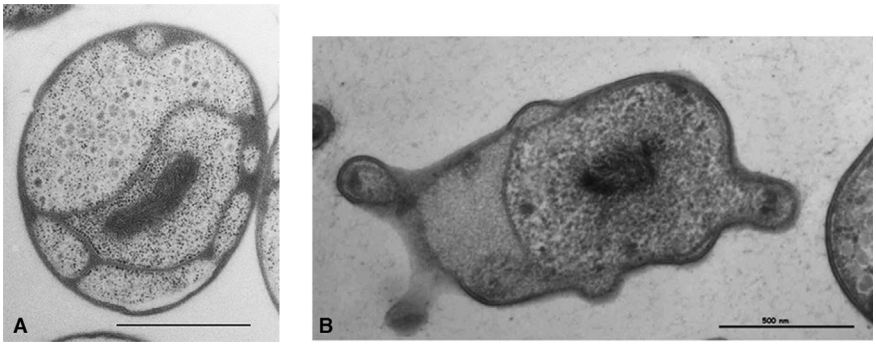


Fig. 1. Electron micrograph images of cells from members of two different phyla of the PVC superphylum showing internal membrane-bounded compartments surrounding the nucleoid in each case. (**A**) Transmission electron micrograph of a sectioned cell of the phylum Planctomycetes member *Gemmata obscuriglobus* (prepared by cryosubstitution) showing a fibrillar condensed nucleoid within a major compartment of the cell defined by internal double membranes. Bar, 1 μm. (**B**) Transmission electron micrograph of a sectioned cell of phylum Verrucomicrobia member *Verrucomicrobium spinosum* (prepared via cryosubsitution) showing a nucleoid within a major cell compartment defined by a single internal membrane. Bar, 500 nm. Micrographs are from the Fuerst Laboratory, The University of Queensland, in collaboration with K.C. Lee of the Fuerst Laboratory.

modern eukaryote (e.g. endomembranes and a membrane-trafficking system [76,77]), and there is no archaeon or bacterium showing all ESP features in one lineage [24]; 3) evidence from probable acidocalcisome organelle enzyme distribution (e.g. vacuolar proton pump pyrophosphatase) indicates that the common ancestor may have been compartmentalized [66,78] and, from fold families (FFs) in the domains, that a functionally complex cellular ancestor of life was at least one stage of LUCA [79]; 4) occurrence of some ESPs in isolated deep-branching contemporary bacteria (PVC superphylum) and archaea (e.g. Lokiarchaeota and some Crenarchaeota) is not easily explicable via lateral transfer. This is supportive of the Eukarya–Archaea and bacteria branches deriving early complexity from LUCA (but see the critique of ESPs in Lokiarchaeota outlined by Caetano-Anolles [25,31].

After DNA replication evolved to give rise to the tree of life, LUCA likely became diminished in relevance or ceased to exist altogether. The two descendant lineages, ancestors of the PVC Bacteria and the Eukarya–Archaea lineages of the tree of life, which are dramatically divergent from one another, began to evolve independently. Early LUCA proteins chronicle the late pre-cellular period. Most importantly, some of them, such as the remarkable homologs of α- and β-tubulin, have an extremely high degree of conservation, and therefore still provide invaluable dual phylogenetic information about the pre-DNA world and LUCA that is not available otherwise. Likewise, several other examples of proteins were distributed to both stems, including sterol-synthesizing enzymes and protein degradation pathways involving ubiquitin.

Eventually the speciation process led to the formation of the tree of life as we know it from contemporary organisms.

5. Phylogenetic evidence from ancient, highly conserved LUCA proteins supports NuCom

A select few ancient LUCA protein families are found in contemporary members of the PVC Bacteria and Eukarya. Four examples are discussed below. Others likely exist; however, the diversity of the PVC Bacteria is still poorly known due to lack of study. Because phylogenic analyses often place these PVC proteins, all of which are highly conserved, within a group of representative eukaryotic proteins, they are commonly regarded as representing HGT events from a eukaryote to a prokaryote. The unusual but strongly defensible approach taken in this paper is that, since these highly conserved proteins are found in members of the PVC superphylum as well as the Eukarya, it is evidence that they are part of the heritage of the PVC superphylum, and therefore their evolution is essential to understanding the evolution of all three domains. These are termed PSPs (protokaryotic signature proteins) because they are homologs that are found in both emerging lineages, the Bacteria and the Eukarya, while also proposed to occur in LUCA before domain formation and separation. Such PSPs might also be expected to be retained from LUCA in some Archaea (e.g. Lokiarchaeota, Thaumarchaeota and the TACK archaea).

5.1. Cell membrane composition and enzymes

According to NuCom, one of the major commonalities between the ancestors of the Bacteria and Eukarya is that the chemical composition of the cell membranes (and nuclear membranes of the Eukarya and PVC superphylum) of these two domains are similar, if not identical. In both domains, glycerol 3-phosphate (G3P) with sn-1,2 stereochemistry is linked to the fatty acid side chains by ester bonds. Indeed, so far as known, all of the enzymatic steps involved in the formation of the lipid membrane moieties are carried out by homologous proteins between these two domains [43,44], providing strong support for the NuCom hypothesis. This is also consistent with the evidence reported by Kim and Caetano-Anolles [36].

One of the reasons the NuCom hypothesis does not include the Archaea is that their cell membranes have different chemical constituents from the other two domains [5] The Archaea have glycerol 1-phosphate (G1P) membranes with methyl branched isoprenoids linked to the glycerol moiety by ether bonds. Further, the enzymatic process that gives rise to the membrane lipids is different. Another significant difference is that the stereochemistry of the GlP pathway moieties has the opposite handedness (sn-2,3) from that of the G3P moieties (sn-1,2) found in the Bacteria and the Eukarya.

5.2. Tubulins and FtsZ

The α- and β-tubulins that form microtubules and their bacterial homologs BtubA and BtubB are one of the foremost PSP examples. Herein we refer to these as true tubulins, i.e. close homologs of α- and β-tubulins of the Eukarya, respectively, and not other forms such as the artubulins [80] and FtsZ's [16] which appear to be involved in cell division.

When bacterial tubulin homologs were reported in four species of the PVC genus *Prosthecobacter* of the Verrucomicrobia [75], several explanations were provided for their presence in these bacteria. The only explanation pursued by bioinformaticians was that they comprised a recent horizontal gene transfer (HGT) from a eukaryote [81]. They reasoned, we believe too hastily, that since all Eukarya species have tubulins, and these were the only members of the Bacteria in which tubulins are known, therefore they must have been derived from HGT from a eukaryote. The alternative explanation of retention from a complex ancestor with proto-tubulin (with loss throughout other Bacteria) was not seriously considered. Later, researchers found that structural features of prosthecobacter tubulins such as primitive assembly properties suggested that such a gene transfer, if it did occur, must have been very ancient, perhaps occurring shortly after duplication of a spontaneously folding alpha- and beta-tubulin ancestor [82].

Even when it was reported that these bacterial tubulins comprised an ancient form of the protein which probably arose from early tubulin intermediates that formed small microtubules [82,83], some still dismissed them as an example of an ‘ancient HGT event’. We think it is highly likely that these true tubulins are an example of one of the primordial proteins from

LUCA, or very closely resembling such an ancestral protein type. If so, the explanation is that these highly conserved proteins, α- and β-tubulins and their bacterial homologs BtubA and BtubB comprise invaluable insight into the evolution of both the PVC Bacteria and the original members of the Eukarya lineage at the time DNA replication evolved. It is true that they occur rarely in contemporary Bacteria, but this is likely due to the reductive evolutionary events that have occurred not only in the common bacteria but in the PVC group as well. Also, few of the PVC bacteria have been subjected to intensive investigation by microbiologists, so it is possible that other homologs will be found. If so, they are likely to be found in the PVC group. Further, their deep evolution makes it highly unlikely that they are the result of HGT.

Interestingly, some trees show the BtubB tubulin as an intermediate between the α-homologs and BtubA, and the β-tubulin family (Findeisen et al., 2014). This may provide a clue that BtubB represents an early duplication event of BtubA.

Ancient tubulins may have been initially involved in the shuffling of DNA in LUCA. Contemporary Eukarya use microtubules, which are comprised of α- and β-tubulin dimer subunits, to carry out a variety of activities related to movement, such as flagellar motility and moving chromosomes in the nucleus during mitosis and meiosis. In this latter function, a kinetochore attaches the microtubule to the chromosome. If tubulins are primordial LUCA proteins, they can serve as important shared molecular phylogenetic connections between the pre-DNA and post-DNA worlds, and also between the Eukarya–Archaea and the PVC Bacteria stems of the tree of life. Importantly, the C-terminal domain (d.79.2) and the GTPase domain (c.32.1.1) of tubulin appear together very early in protein evolution, after the first appearance of ribosomal proteins, which supports the NuCom hypothesis (Gustavo Caetano-Anollés, personal comment).

In contrast to these true tubulins, the TACK 'artubulins' as stated earlier, appear to be a more evolved form of FtsZ that are likely involved in cell division. In this regard, it is also noteworthy that *Prosthecobacter* genomes contain not only BtubA and BtubB, but also a primitive form of FtsZ [84]. Therefore, both FtsZ and tubulins from *Prosthecobacter* can be used to construct a tree of life (Fig. 2).

According to NuCom, the current rarity of tubulins in the PVC superphylum and their absence in common bacteria is due to reductive evolutionary events that transpired in the PVC superphylum during their evolution. These resultant simplified versions of DNA replication were no doubt more efficient for postkaryotic organisms that needed to conserve as much energy as possible to fit into their niches. If this explanation is correct, then α- and β-tubulins and the bacterial homologs BtubA and BtubB can be used to produce a tree for the Bacteria and the Eukarya, using an alignment combining FtsZ sequences, FtsZ-tubulin-like sequences from selected Archaea, and the Eukarya tubulins and their homologs BtubA and BtubB from *Prosthecobacter* (Fig. 2A and B). Thus, all three domains can be placed in a single tree of life based on members of the easily alignable tubulins and FtsZs and similar sequences. In such trees, FtsZs of *Verrucomicrobium spinosum* and *Prosthecobacter dejongeii* both branch very deeply within or near the FtsZ branch of the tree, while the tubulins of *Prosthecobacter* branch deeply within the Eukaryal tubulin branch, the exact clustering with either eukaryal alpha or beta tubulins varying with the analysis (Fig. 2A and B). The selected archaeal sequences included are not significantly clustered with Eukarya or Bacteria sequences, a result, however, also varying with the analysis and taxon sampling. Long branch attraction can always be invoked as a potential explanation for deep branch positions, but this seems unlikely to account for the tree topology in this case, though further tests are desirable. In any case, there is no indication of any recent HGT where one would expect clustering of the PVC sequences with an existing eukaryal or bacterial phylum or superphylum. The apparent deep branching of FtsZs and tubulins of *Prosthecobacter* is consistent with a retention from a complex ancestor of Bacteria and Eukarya (and probably also Archaea) with a primordial member of the FtsZ-tubulin family. It is also consistent with eukaryal cell biology in a complex ancestor which might include a nucleus as well as cytoskeletal proteins, as in the NuCom hypothesis. We note that *Prosthecobacter* tubulins were found to branch deeply with the branch containing eukaryal alpha and beta tubulins in the tree in Fig. 2A of Yutin and Koonin [80], and this is consistent with the original tree for *Prosthecobacter* tubulins of Jenkins et al. [75]. Alpha and beta tubulins are not necessarily the ancestral forms of tubulin or closest existing tubulin classes to that ancestor, but retention by *Prosthecobacter* of tubulin-like proteins from a tubulin-containing ancestor is compatible with this. The depth of *V. spinosum* FtsZ is consistent with the analysis of Yee et al. [85]. The argument or reasons for the assertion of Yutin and Koonin that archaeal tubulins are vertically related to eukaryal tubulins, but *Prosthecobacter* tubulins are the result of HGT, appear to be based only on the sporadic occurrence in Bacteria of such tubulin sequences rather than on any convincing demonstration. The hypothesis of vertical transmission from primordial forms of FtsZ-tubulin proteins cannot be easily dismissed. When structural biochemical analysis (e.g. of tubulin assembly properties) has been performed using *Prosthecobacter* tubulins, it has yielded interpretations including close lineage of *Prosthecobacter* tubulins to ancestral tubulins [82]. One does not expect that an ancient character will be widely distributed among extant species – most fish do not have lungs, yet lungfish do, but the lungfish did not acquire its respiratory organ horizontally from either ancient or extant amphibians or reptiles (rather, the land vertebrate lung appears to be derived via vertical inheritance from a close shared ancestor with lungfish [86]). Therefore, horizontal transfer explanations should not be the first to come to mind for similar phenomena in microbial evolution.

5.3. STK enzymes and ubiquitin-mediated protein degradation

Other homologous proteins that were previously thought to be exclusively found in the Eukarya have also been discovered

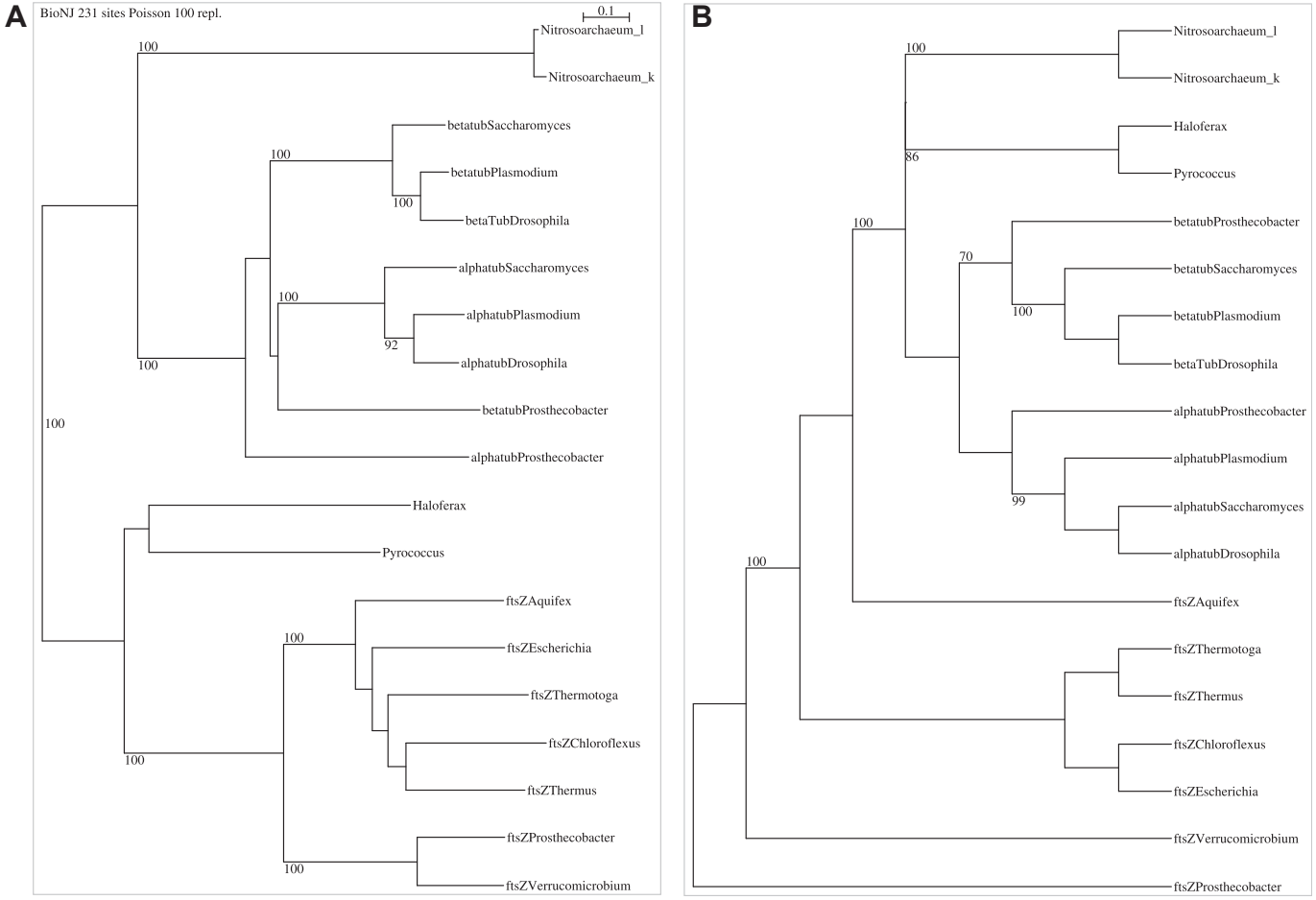


Fig. 2. Trees based on phylogeny of α- and β-tubulins of the Eukarya and the bacterial tubulins of Verrucomicrobia, BtubA and BtubB, and the FtsZs of selected bacteria and two species of Verrucomicrobia. (**A**) Distance (BioNJ) tree of tubulin family (Eukarya alpha and beta tubulins and Verrucomicrobia *Prosthecobacter dejongeii* proteins BtubA and BtubB) and FtsZ family proteins (common bacteria FtsZs plus verrucomicrobial *Verrucomicrobium spinosum* and *Prosthecobacter dejongeii* FtsZ proteins) generated using a Clustal Omega [111] alignment within the Seaview v. 4.4.2 phylogenetic software package [112], and then using BioNJ distance phylogenetic analysis with Poisson distance, exclusion of gapped positions and 1000 bootstraps. (**B**) PhyML maximum likelihood tree of tubulin and FtsZ gene family using a Clustal Omega [111] alignment within the Seaview v. 4.4.2 phylogenetic software package and then PhyML maximum likelihood with default parameters within Seaview. The same topology was given with 100 bootstraps or with LRT. Bar denotes 0.2 amino acid substitutions. Key to sequence names: alphatub *Saccharomyces cerevisiae* – alpha tubulin *Saccharomyces cerevisiae* AAA35181.1; alphatub *Plasmodium falciparum* – alpha tubulin *Plasmodium falciparum* – CAA34101.1; alphatub *Drosophila melanogaster* alpha tubulin [NP_476772.1| betatub *Plasmodium falciparum* – beta tubulin *Plasmodium falciparum* AAA29780.1| betatub *Saccharomyces cerevisiae* – beta tubulin *Saccharomyces cerevisiae* CAA24603.1; betatub Drosophila – beta tubulin *Drosophila melanogaster* – NP_001286835.1; Haloferax mediterranei ATCC 33500 – cell division protein [*Haloferax mediterranei* – ATCC 33500, YP_006348413.1; *Pyrococcus abyssi* – cell division protein FtsZ [*Pyrococcus abyssi*] – NP_126497.1 (WP_010867936); *Aquifex aeolicus* – cell division protein FtsZ *Aquifex aeolicus* – gi|499182767; WP_010880307.1; Chloroflexus aurantiacus – J-10-fl cell division protein FtsZ *Chloroflexus aurantiacus* – YP_001634375.1; *Escherichia coli* PMV-1 Cell division protein FtsZ *Escherichia coli* – YP_008570045.1; *Thermotoga maritima* – cell division protein FtsZ *Thermotoga maritima* NP_228645.1; *Thermus thermophilus* – cell division protein FtsZ *Thermus thermophilus* –*YP_144355.1;*FtsZ *Prosthecobacter dejongeii* – FtsZ *Prosthecobacter dejongeii* CAM57305.1|FtsZ *Verrucomicrobium spinosum* – FtsZ *Verrucomicrobium spinosum* DSM 4136]ABI34433.1; alphatub *Prosthecobacter dejongeii* – tubulin [*Prosthecobacter dejongeii*]AAX07528.1; betatub *Prosthecobacter dejongeii* – tubulin [*Prosthecobacter dejongeii*] AAX07529.1; *Nitrosoarchaeum_*l – tubulin/FtsZ GTPase [*Candidatus* Nitrosoarchaeum limnia SFB1] EGG42457.1; Nitrosoarchaeum_k – Tubulin gamma chain, putative [Candidatus Nitrosoarchaeum koreensis MY1] EGP93873.1. Bar marker is 0.1 amino acid substitutions.

in the PVC Bacteria. These include eukaryote-like serine/threonine kinases (STKs) and E2-ubiquitin-conjugating enzymes [87]. The STKs are reported from many different genomes of the Planctomycetes, Verrucomicrobia and Chlamydia phyla, as well as certain other bacterial phyla; there is a significant expansion of STKs in PVC species compared to other bacteria, especially in Planctomycetes and Lentisphaerae. The kinase domains of such STKs had significant homology with eukaryotic mitotic kinases involved in nuclear signaling as well as eukaryotic ribosomal kinases. The E2-ubiquitin system proteins are more highly restricted to the PVC group. However, despite separation of bacterial and eukaryal E2-ubiquitin system proteins in phylogenetic trees, the authors conclude that these examples must represent HGT events, apparently solely due to the limited distribution of these enzymes in the Bacteria. HGT has, in such a way, often been used to presumptively explain away some of the potentially most decisive evolutionary events leading to neglect of

alternative interesting possibilities involving vertical inheritance from a shared ancestor.

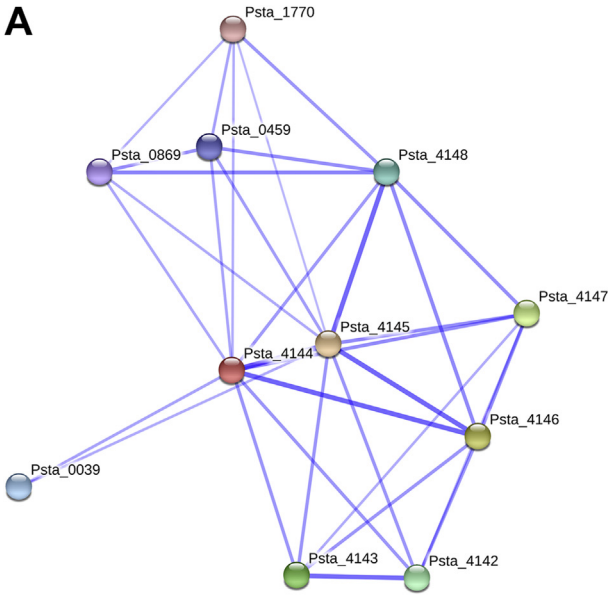
Consistent with the discovery of Arcas et al. [87], several representatives of the ubiquitin protein-degrading system are present in Planctomycetes in addition to the E2 ubiquitin-conjugating enzyme, including those homologous with E1 ubiquitin-activating and E3 ubiquitin ligase and the de-ubiquitinating MPN enzyme [88,89]. The E2 enzyme is an excellent example of members of the ubiquitin system in Planctomycetes – e.g. WP_012913057 of *Pirellula staleyi* is annotated as an ubiquitin-conjugating protein E2 and using SMARTBLAST applied to DeltaBLAST output, its closest match is to eukaryotic ubiquitin-conjugating protein E2 of a yeast (*Schizosaccharomyces*) and a protist (*Plasmodium*). It has a conserved active site cysteine and E3-interaction sequences as well as Ub thioester interaction residue conserved sequences, all characteristic features of such E2 enzymes. In the eukaryotic ubiquitin system, a thiol-ester linkage forms between a conserved cysteine and the C-terminus of ubiquitin and complexes with ubiquitin protein ligase enzyme E3 (see NCBI UBCc superfamily description (cl00154: UBCc superfamily) in a conserved domains section of the NCBI menu). Further, there appear to be other proteins of the ubiquitin system, at least in representative planctomycetes *P. staleyi* and *Isosphaera pallida*, including RING-finger-containing proteins, forming a tentative ensemble potentially capable of actual protein degradation (this ensemble of Ubl proteins may also occur rarely in scattered examples in other bacterial phyla and a deep-branching archaeon, *Caldiarchaeum*, as discussed below in more detail [88]. Consistent with this, an analysis of proteins potentially interacting with the E2 protein ubiquitin-conjugating protein Psta 4144 of *P. staleyi* via the STRINGS software [90] reveals several other proteins with domains or even annotations indicating homology with ubiquitin system proteins, and is consistent with a complete ubiquitin system in this planctomycete (Fig. 3A and B). So interacting hypothetical protein Psta_4143 has a PhD finger domain known to be present in some E3 ligases, and Psta4147 and Psta4145 are both Mov34/MPN/PAD-1 family proteins consistent with deubiquitinylating activity, Psta_4148 and Psta_4144 are both versions of E2 ubiquitin-conjugating proteins, and Psta_4142 contains a RING finger domain known to be associated with E3 ubiquitin ligase. Proteins with a domain of the significant RPN7 superfamily (CDD: 256084) – the 26S proteasome subunit RPN7 necessary for the lid subunit of 26S proteasome essential for ATP degradation of ubiquitinated proteins – occur in planctomycetes, e.g. in the hypothetical protein WP_015248122 in *Singulisphaera acidophila* and in hypothetical protein WP_010045184 of *G. obscuriglobus*, where the RPN7 domain occurs along with tetratricopeptide (TPR) repeats in both cases. The distribution of members of such an ubiquitin system ensemble in at least some Planctomycetes is consistent with a hypothesis that LUCA may have possessed such a system for protein degradation and the evidence that the ubiquitin system pre-dated the origin of eukaryotes [91].

It is of considerable interest and consistent with this model that an ensemble of eukaryote-like ubiquitin system proteins has also been found in the genome of the uncultured archaeon *Candidatus* 'Caldiarchaeum subterraneum' [92]. The depth of branching of Planctomycete sequences in some recent published phylogenetic trees (Fig. 3B and C of [89]) of the E1 ubiquitin ligase enzyme and the MPN de-ubiquitinating enzyme seems consistent with a hypothesis of ancient depth for this group or a deep relationship to eukaryal genes, with the E1 tree, for example, displaying a significant cluster containing *Pirellula* and *Isosphaera* sequences at the base of a protistan eukaryote clade and branching just above the clade containing 'Caldiarchaeum', while an MPN tree has the Planctomycetes as a sister group to eukaryote sequences closer to them than to Archaea [89]. It has been suggested that HGT is an unlikely source for the ubiquitin system genes of 'Caldiarchaeum', as it would need independent transfer of several independent genes from eukaryotes (unlikely to be functional in isolation), followed by integration into operon organization, and the individual genes would be unlikely to be functional in isolation [93]. Significantly, it has been observed that E1, E2 and E3 ubiquitin system genes in *P. staleyi* and *I. pallida* also cluster together, suggesting an operon organization (see Fig. 3A of [89]), as well as functional coordination. In the case of ubiquitin system genes, the argument from organization of the genes in operons and functionally relevant acquisition of independent genes which has been made against HGT from eukaryotes as an origin for archaeal ubiquitin system genes by both Burroughs et al. [93] and Koonin and Yutin [89]surely also applies to PVC Bacteria such as *P. staleyi* and *I. pallida*, where E1, E2 and E3 ubiquitin system genes also cluster in the same operon, as indicated in Fig. 3A of Ref. [89].

In any case, the occurrence of dispersed instances of a near complete ubiquitin system gene assembly in Planctomycetes and in 'Caldiarchaeum', as well as some other sporadic instances in Acidobacteria and Actinobacteria, is at least as compatible with retention of such a system in all these organisms from a compartmentalized LUCA, as much as with an invention of the ubiquitin protein domains in Archaea (or in some lineages of Bacteria).

5.4. Sterol biosynthesis pathways

A simple pathway for sterol synthesis also occurs in Planctomycetes, with evidence most explored for *Gemmata* species [56], and phylogenetics and bioinformatics indicates that the enzymes involved have eukaryotic features and affinities. While again there might be a suggestion that HGT could explain the origin of sterol biosynthesis enzymes in Planctomycetes and another bacterial group with sterols, the enucleate Proteobacteria methylotrophs, the pathway in *Gemmata* producing the simple sterols lanosterol and parkeol is the simplest known and more likely to be ancient – these proteins do not cluster within or closely relate to any contemporary group of eukaryotes, and this is consistent with an ancient origin, perhaps from a sterol-synthesizing protokaryotic ancestor [94]. Thus, one phylogenetic study of bacterial sterol synthesis has observed, in reference to the ERG1 and ERG7 (squalene monooxygenase and oxidosqualene



B

Psta_1770	530564.Psta_1770	hypothetical protein
Psta_4147	530564.Psta_4147	Mov34/MPN/PAD-1 family protein
Psta_4143	530564.Psta_4143	hypothetical protein
Psta_4145	530564.Psta_4145	Mov34/MPN/PAD-1 family protein
Psta_4148	530564.Psta_4148	ubiquitin-conjugating protein E2
Psta_4146	530564.Psta_4146	UBA/THIF-type NAD/FAD binding protein
Psta_0869	530564.Psta_0869	hypothetical protein
Psta_4144	530564.Psta_4144	ubiquitin-conjugating protein E2
Psta_0459	530564.Psta_0459	hypothetical protein
Psta_4142	530564.Psta_4142	hypothetical protein
Psta_0039	530564.Psta_0039	hypothetical protein

Fig. 3. (**A**) STRINGS analysis [90] of proteins potentially interacting with *Pirellula staleyi* E2 ubiquitin-conjugating protein Psta 4144. (**B**) Table of proteins predicted to interact with Psta_4144 according to STRINGS software [90].

cyclase) enzymes, that 'it remains unclear whether these two enzymes originated in bacteria or in eukaryotes. It is likely that the assembly of the pathway started with these two initial steps.' [95]. A phylogenetic analysis of the triterpene cyclase protein family in prokaryotes and eukaryotes has found that at least some triterpene cyclases from *Gemmata*, and the proteobacterial *Methylococcus* and *Stigmatella* genera, cluster with oxidosqualene cyclases of eukaryotes, but that, while the oxidosqualene cyclase (OSC) from *Stigmatella* clustered within eukaryotes and may well be the product of an LGT, the sequences of *Gemmata* and *Methylococcus* branched deeply within the eukaryotic OSC branch, and their eukaryotic affinity could not be attributed to potential lateral gene transfer from or to eukaryotes [96]. This analysis also found that the triterpene cyclase family was present in LUCA, and branched later to form the enzyme group in bacteria using a squalene substrate to form pentacyclic triterpenoids, and a eukaryotic group using oxidosqualene as a substrate to form tetracyclic and pentacyclic triterpenoids, including sterols, with the *Gemmata* and *Methylococcus* sequence close to the base of the eukaryotic clade and not 'high-quality' candidates for gene transfer from eukaryotes. Significantly, *Gemmata* and *Methylococcus* OSC sequences do not cluster with any extant group of eukaryotes and are close to the root hypothesized for their tree of eukaryote and bacterial triterpene synthases. Hopanoids, lipids closely resembling sterols and perhaps functionally analogous, but more typical of Bacteria, are also found in some Planctomycetes [97], so that a possible ancient cell containing both sterols and hopanoids where sterol synthesis was evolving from duplication of hopanoid synthesis enzymes can be hypothesized as ancestor of the PVC superphylum or the protokaryote.

The evidence to date is not consistent with an HGT hypothesis for the origin of sterol synthesis enzymes in Planctomycetes, and the compartmentation of Planctomycetes by internal membranes suggests a selective pressure for retention of sterols as a membrane component with special properties for processes involving membrane deformation such as endocytosis [98]. This is consistent with potential roles in protokaryotic compartmentalization and internal membranes. It is also important to recognize that the Proteobacteria are regarded by the NuCom hypothesis to be enucleate descendants of the Verrucomicrobia [5]. Therefore, it is not too surprising that some isolated proteobacterial species such as the methylotrophs may also contain some of the 'ESP' (i.e. PSP) proteins such as OSC sterol-synthesizing cyclase.

5.5. The wider protokaryome

The examples above indicate the evolutionary significance of the PVC superphylum, especially as it relates to the hypothesis of vertical retention from a protokaryote ancestor with PLFA lipid membranes, internal cytoskeleton (tubulin protein family), ubiquitin systems and sterols. Phylogenetic analyses of verrucomicrobial FtsZ and tubulin seem to indicate that, at the least, the presence of such eukaryote-like genes is not the result of HGT, since PVC proteins do not cluster significantly with any major extant eukaryote group. These examples probably represent only a sample of the wider protokaryome present in the PVC superphylum, as in the Archaea, where different species represent different eukaryote or protokaryote signature systems [89]. Undoubtedly, there are further examples to be explored, e.g. SNARE-like proteins which may be involved in endocytosis, homologs of which appear to occur in Planctomycetes [98], and the MC-like (membrane coat) proteins already shown to be associated with the internal membranes of *Gemmata* [72,99], and among Bacteria, to be distributed very sparsely, notably in PVC

superphylum [99]and possibly in the phylum *Bacteroidetes* [100]. Incidentally, the rare occurrence of similar eukaryal homologs in the PVC superphylum and in dispersed examples of other Bacteria is exactly what one would expect if ancestral protokaryote genes are retained randomly and rarely in dispersed Bacteria, as also occurs with the archaeal eukaryome within different archaeal kingdoms and genera.

It is of interest, in the light of the report of eukaryote signature proteins occurring together in Lokiarchaeota, that the domains characteristic of some lokiarchaeotal ESP proteins can also be found in annotated proteins of PVC superphylum members. Thus, like the Aigarchaeota (or Crenarchaeota or Thaumarchaeota) member Candidatus 'Caldiarchaeum subterraneum', Lokiarchaeum has all the representatives of ubiquitin system proteins identified also in planctomycete *P. staleyi* above, including proteins with RING-finger domains and ubiquitin-conjugating E2 enzymes [30]. And like some lokiarchaeotal proteins, NCBI's conserved domain features accompanying annotation indicate that Planctomycetes proteins occur with the ADF-gelsolin superfamily domain (e.g. in a serine/threonine protein kinase WP_007326534 of *Rhodopirellula baltica* and hypothetical protein kuste4157 CAJ74919 of "*Candidatus* Kuenenia stuttgartiensis"). Likewise, proteins with domains of the BAR superfamily Bin/Amphiphysin/RVS domain (CDD: 271734) occur in Planctomycetes (e.g. in a serine/threonine protein kinase WP_010035625 and a hypothetical protein WP_052560401 of *G. obscuriglobus*, as well as in other species such as *Planctopirus limnophila* and *Rhodopirellula europaea* (WP_013108711 and WP_008664353)). Most interestingly, perhaps, there are Planctomycetes with proteins possessing a domain in the oligosaccharyl transferase family (e.g. the hypothetical protein WP_052559960 in *Gemmata* sp. IIL30 and the hypothetical protein WP_010037521 in *G. obscuriglobus*), for which the NCBI annotation identifies a region with domain archaeo-STT3 (oligosaccharyl transferase, archaeosortase A system-associated, TIGR04154, CDD: 275016, associated with a system linking protein membrane transit with glycosylation in archaea). *S. acidophila* and *R. baltica* also possess proteins with an annotated archaeosortase domain (e.g. hypothetical protein Sinac_4281 – accession AGA28480 – of *S. acidophila* and hypothetical protein WP_011121523 of *R. baltica*). All these potential 'ESPs' (PSPs) in the PVC superphylum need investigation for confirmation beyond analyses via NCBI conserved domain search algorithms. However, there may be at least some indicators here of possible retention from complex shared ancestors of PSPs possessing domains now functioning in Eukarya cell biology in both Lokiacrhaeota and the PVC superphylum. This seems conceivable and worth further testing. This must, however, be seriously considered in the light of analyses concluding that any 'ESPs' of Lokiarchaeota are potentially misleading regarding eukaryal relationship relevance [25,31].

Phylum *Poribacteria,* known only as an uncultured symbiont of marine sponges, seems to be a sister lineage of the PVC superphylum together with the WS3 phylum level group [3], and interestingly, members possess some significant eukaryote-like ankyrin repeat proteins [3] and are probably compartmentalized by at least a single major internal membrane in a similar way to the simplest compartmental plan in Planctomycetes. This is also consistent with a hypothesis of an ancient lineage with complex cell biology perhaps comprising the PVC superphylum and related phyla in a larger group of sister lineages. However, the symbiotic relationship of Poribacteria with their eukaryote hosts could explain the need for the ankyrin-repeat proteins [3].

6. Discussion

Proponents of the prokaryotic origin of the Eukarya trace their views back to those of Margulis [101] and Rivera and Lake [102]. Margulis believed that the Bacteria and Archaea were the source of all the genetic components necessary to produce a eukaryotic cell through symbiotic events. For example, Margulis proposed that a spirochete was the source of tubulin for the evolution of the eukaryotic flagellum [103], although there is no genomic evidence that any spirochete contains tubulin genes. Rivera and Lake [102], who proposed the ring theory of life, advocate that the eocyte (Crenarchaeota) lineage of the Archaea gave rise to the Eukarya through a fusion event with a bacterium. Therefore, both groups propose that prokaryotes evolved to produce the Eukarya through a fusion process, a view that is commonly held among microbial phylogeneticists who believe in the 2-domain hypothesis which regards prokaryotes as the ancestors of the Eukarya.

In contrast, Kurland et al. [22] noted that it is extremely difficult to explain the origin of the Eukarya from a prokaryotic cell or a combination of prokaryotic cells. The NuCom proposal is consistent with their view. However, in contrast to their view that a more highly evolved or full-fledged eukaryote was necessary to account for the evolution of the Bacteria and the Eukarya lineages of the tree of life, our view is that the evolution of DNA replication and a compartmentalized LUCA was all that was required. The NuCom hypothesis states that, when the undeniable event of DNA replication evolved, a nuclear compartment with PLFA membranes was essential for both the bacterial lineage (which began with a member of the PVC Bacteria) and the first ancestor of the highly divergent Eukarya. These relatively simple early nucleated organisms with their PLFA membranes later on gave rise to the enucleate members of the Bacteria. In contrast, the Eukarya lineage which is highly divergent from the Bacteria continued to evolve by eukaryogenesis to produce the largest and arguably the most remarkable organisms on Earth.

The NuCom hypothesis is supported strongly by the PSPs as reported here. This is consistent with the first genomic comparison of "Eukaryotic Signature Proteins" from two genomes represented by *P. dejongeii* and *Gemmata* Wa-1 [104] against a list of ESPs [105]. That comparison concluded that, whereas it was unlikely that the Eukarya evolved from

the PVC superphylum, the two groups shared PSPs (formerly termed ESPs) that indicated some commonality between them. The NuCom hypothesis provides an explanation for the nature of that commonality. That is, these PSPs represent shared homologous properties between the emerging Bacteria and early true Eukarya lineages from a compartmentalized LUCA.

If there was indeed a compartmentalized ancestor of the 'two' domains, then some predictions can be made and consequences of the hypothesis tested. 1) It may be that representatives of species which retained compartmentalized cell structure and correlated molecular components can be observed in deep or early branching phyla of the common bacteria other than members of the PVC superphylum. A possible example of this is sterol synthesis shared by *Gemmata* in Planctomycetes and by members of the Proteobacteria. 2) Elements of an early development of introns and mRNA splicing via a splicesosome might be found in PVC or other compartmentalized bacteria. 3) Co-translational secretion of proteins might be found to occur across internal membranes of the most complex members of the PVC superphylum such as *G. obscuriglobus*.

6.1. Evolutionary significance of the PVC superphylum

There have been some attempts to discredit the evidence regarding the evolutionary significance of the PVC superphylum (e.g. [17,106]). For example, when any report is submitted or published about some "eukaryotic signature protein" that has a homolog in the PVC group, because this feature is found in many or all the Eukarya and only a few Bacteria, then it must be evidence of HGT from a contemporary or ancient member of the Eukarya to the PVC group. As a result, PVC opponents quickly dismiss the PVC group as important players in the deep evolutionary history of life. Unfortunately, this prevalent attitude has prevented others who make these remarkable findings from examining other options, namely, "Why do members of this unusual group of Bacteria contain these unexpected proteins?" One obvious possibility pursued in this publication is that they are found in the PVC superphylum for a reason: like the Eukarya, their ancestors received them from a compartmentalized LUCA. However, an open mind must be retained concerning which eukaryote-like features of the PVC superphylum members are homologous with a potential LUCA ancestor with complex cell biology and which are only analogous and possibly the result of convergence. More analyses will be required to test these alternatives.

Clearly, there is ample evidence that the PVC superphylum exists and has much to tell us about the evolution of all living organisms. With the acceptance that the PVC superphylum has remarkable proteins that are homologous to that of the Eukarya, we will be able to more completely understand the true phylogeny of all organisms.

6.2. Why did the PVC superphylum remain simple relative to Eukarya?

An interesting question is 'Why is it that the PVC group did not become increasingly complex as did the Eukarya?' The PVC Bacteria never completely evolved phagocytosis and therefore could not ingest other organisms as an exploitable niche that would have been an enormous advantage when photosynthetic organisms evolved. *Bdellovibrio* is the closest example of a bacterium to accomplish this, but its modus operandi is different from engulfment. They invade other bacteria and degrade them internally.

Were the PVC Bacteria, for some inexplicable reason, unable to eat their own kind? Perhaps the answer to this question is that all members of the Bacteria have or have had peptidoglycan in their cell walls, and its rigidity and impenetrability may have prevented them from evolving cells with more plastic cell envelopes that would enable engulfment of other organisms. Early studies of the Planctomycetes suggested that they lacked peptidoglycan. However, recent excellent papers provide irrefutable evidence that *Planctomyces* and the Anammox group of the Planctomycetes do have peptidoglycan [107,108]. The occurrence of peptidoglycan in the Planctomycetes may explain why their evolution has been more muted and less highly developed than their incredible eukaryotic sisters.

6.3. Predictions of the NuCom hypothesis

The NuCom hypothesis implies several predictions for future experimental and bioinformatic investigation.

1) Cell biological systems characteristic of protokaryotes or eukaryotes will be found in members of the PVC superphylum, including a functional ubiquitin-related protein degrading system and cytoskeletal systems utilizing both tubulins and FtsZs.
2) New and uncultured Planctomycetes and Verrucomicrobia including exotic microcolonial types of the Planctomycetes, such as *Planctomyces bikefii* and *P. guttaeformis*, and uncultured relatives of verrucomicrobial *Prosthecobacter* will contain additional PSPs which could be discovered using single cell genomic procedures from natural samples.
3) Correlates of internal nuclear body membranes with eukaryote nuclear envelope proteins will be found in PVC superphylum members, including inner and outer nuclear membrane proteins (e.g. SUN-domain proteins) and nuclear pore proteins (perhaps but not necessarily homologs).
4) Proteins will be found in PVC superphylum species (in addition to the ones mentioned above) with homology to protokaryote signature proteins found in deep-branching Archaea such as Lokiarchaeota, representing shared retention of homologous ancient protokaryote or complex compartmentalized LUCA proteins. In particular, members of the ESCRT system functioning in cell division in Archaea [109] might function in Planctomycetes in later stages of protein uptake, as they do in eukaryotes, but so far these have not been detected via sequence-based bioinformatics alone (possibly requiring deep protein structure-based analysis to detect homology).

5) Proteins in the PVC superphylum correlating with the simplest forms of eukaryote endocytosis and vesicle trafficking systems will be found in addition to the MC proteins already confirmed in the Planctomycete *Gemmata.* Some indication of such possible proteins is found in a recent analysis [98].
6) There may be Archaea species found that retain not only some PSPs (as perhaps Lokiarchaeota species do), but also compartmentalization from a compartmentalized LUCA. It is not yet clear whether *Ignicoccus* complex cell structure represents a compartmentalized structure, but its outermost membrane is energized [65], suggesting that its inner compartment may be a true internal one, and this is consistent with occurrence of intracellular vesicle budding in those cells [110]. However, it should be noted that the Archaea have very small genomes (less than 6 Mbp) in comparison with most Bacteria and the Eukarya.

6.4. Evidence testing of NuCom evolutionary mechanisms

This paper provides evidence and arguments supporting the NuCom hypothesis but, of course, a scientific hypothesis should make testable experimental or observational predictions. In addition to pursuing other phylogenetic/bioinformatic approaches such as those mentioned above, some predictions of NuCom might be testable in the laboratory using experimental procedures. For example, perhaps it would be possible to set up laboratory experiments using specific strains of the compartmentalized Planctomycetes and Verrucomicrobia growing under a variety of pH, thermal and other conditions that could, over many generations, produce an enucleate descendant by continuously selecting for the most rapidly growing and evolving types. Experimental approaches like these could not be used to explore the singularity of a fusion hypothesis for origins of the Eukarya lineage.

We recommend that much greater emphasis be placed on funding research on the poorly studied PVC superphylum. Is not the understanding of the evolution of the earliest life on earth at least as important as discovering more about cosmology's Big Bang?

In summary, NuCom posits that LUCA gave rise to the ancestors of two protokaryotic lineages, the PVC Bacteria and the Eukarya, which is supported by evidence of two shared commonalities: a membrane-defined nucleus and cell membranes of highly similar composition. This paper provides evidence of a third commonality between the PVC Bacteria and the Eukarya: the phylogenetic/bioinformatic analyses of ancient highly conserved proteins that are shared in the Bacteria and Eukarya support the NuCom hypothesis. These include cell membrane synthesis enzymes, tubulins, E2-ubiquitin protease-conjugating enzymes and possibly other proteins of the ubiquitin system and sterol synthesis enzymes. Because these are so highly conserved, they can be used to trace the deep evolution of the Bacteria and the Eukarya. For example, tubulins from the Verrucomicrobia can be used to construct a tree of life for all domains when used in combination with their FtsZ relatives for the Bacteria and Archaea (Fig. 2A and B). Other shared LUCA proteins will likely be discovered, and we predict these will also be most highly represented in the PVC superphylum.

NuCom is not only a reasonable hypothesis, but one that is supported by contemporary data and other evidence about the domains. The preponderance of evidence as noted in this article supports the NuCom hypothesis. Indeed, we argue that there is more evidence in support of NuCom than any other hypothesis on the origin of the Bacteria and Eukarya. And no fantastic, singular fusion event is required to explain the origin of the Eukarya.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

The authors are very grateful to Gustavo Caetano-Anollés for his numerous helpful comments that have improved the manuscript. Research in the group of JAF has been supported by the Australian Research Council (grant number DP0881485).

References

[1] Wagner M, Horn M. The Planctomycetes, Verrucomicrobia, Chlamydiae and sister phyla comprise a superphylum with biotechnological and medical relevance. Curr Opin Biotechnol 2006;17(3):241–9. Epub 2006/05/18.
[2] Gottshall EY, Seebart C, Gatlin JC, Ward NL. Spatially segregated transcription and translation in cells of the endomembrane-containing bacterium *Gemmata obscuriglobus*. Proc Natl Acad Sci U S A 2014; 111(30):11067–72. Epub 2014/07/16.
[3] Kamke J, Rinke C, Schwientek P, Mavromatis K, Ivanova N, Sczyrba A, et al. The candidate phylum Poribacteria by single-cell genomics: new insights into phylogeny, cell-compartmentation, eukaryote-like repeat proteins, and other genomic features. PloS One 2014;9(1):e87353. Epub 2014/02/06.
[4] Forterre P, Philippe H. The last universal common ancestor (LUCA), simple or complex? Biol Bull 1999;196(3):373–5. discussion 5–7. Epub 2001/09/07.
[5] Staley JT. The nuclear compartment commonality hypothesis, enucleation and the evolution of the Bacteria and Eukarya. Astrobiol Outreach 2013;1.
[6] Forterre P, Gribaldo S. Bacteria with a eukaryotic touch: a glimpse of ancient evolution? Proc Natl Acad Sci U S A 2010;107(29):12739–40. Epub 2010/07/14.
[7] Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A 1990;87(12):4576–9. Epub 1990/06/01.
[8] Daubin V, Gouy M, Perriere G. A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. Genome Res 2002;12(7):1080–90. Epub 2002/07/05.
[9] Harish A, Tunlid A, Kurland CG. Rooted phylogeny of the three superkingdoms. Biochimie 2013;95(8):1593–604. Epub 2013/05/15.
[10] Mariscal C, Doolittle WF. Eukaryotes first: how could that be? Philos Trans R Soc Lond Ser B Biol Sci 2015;370:1678. Epub 2015/09/02.
[11] Gouy R, Baurain D, Philippe H. Rooting the tree of life: the phylogenetic jury is still out. Philos Trans R Soc Lond Ser B Biol Sci 2015;370: 1678. Epub 2015/09/02.

[12] Forterre P. The universal tree of life: an update. Front Microbiol 2015;6:717. Epub 2015/08/11.
[13] Williams TA, Foster PG, Cox CJ, Embley TM. An archaeal origin of eukaryotes supports only two primary domains of life. Nature 2013;504(7479):231–6. Epub 2013/12/18.
[14] Cox CJ, Foster PG, Hirt RP, Harris SR, Embley TM. The archaebacterial origin of eukaryotes. Proc Natl Acad Sci U S A 2008;105(51):20356–61. Epub 2008/12/17.
[15] Raymann K, Brochier-Armanet C, Gribaldo S. The two-domain tree of life is linked to a new root for the Archaea. Proc Natl Acad Sci U S A 2015;112(21):6670–5. Epub 2015/05/13.
[16] Yutin N, Makarova KS, Mekhedov SL, Wolf YI, Koonin EV. The deep archaeal roots of eukaryotes. Mol Biol Evol 2008;25(8):1619–30. Epub 2008/05/09.
[17] Williams TA, Embley TM. Changing ideas about eukaryotic origins. Philos Trans R Soc Lond Ser B Biol Sci 2015;370(1678):20140318. Epub 2015/09/02.
[18] Lake JA, Henderson E, Oakes M, Clark MW. Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. Proc Natl Acad Sci U S A 1984;81(12):3786–90. Epub 1984/06/01.
[19] Forterre P. A new fusion hypothesis for the origin of Eukarya: better than previous ones, but probably also wrong. Res Microbiol 2011;162(1):77–91. Epub 2010/11/03.
[20] Canback B, Andersson SG, Kurland CG. The global phylogeny of glycolytic enzymes. Proc Natl Acad Sci U S A 2002;99(9):6097–102. Epub 2002/05/02.
[21] Kurland CG, Canback B, Berg OG. The origins of modern proteomes. Biochimie 2007;89(12):1454–63. Epub 2007/10/24.
[22] Kurland CG, Collins LJ, Penny D. Genomics and the irreducible nature of eukaryote cells. Science 2006;312(5776):1011–4. Epub 2006/05/20.
[23] Penny D, Collins LJ, Daly TK, Cox SJ. The relative ages of eukaryotes and akaryotes. J Mol Evol 2014;79(5–6):228–39. Epub 2014/09/03.
[24] Forterre P. The common ancestor of Archaea and Eukarya was not an archaeon. Archaea 2013;2013:372396. Epub 2013/12/19.
[25] Nasir A, Kim K, Da Cunha V, Caetano-Anollés G. Arguments reinforcing the three-domain view of diversified cellular life. Archaea 2016;2016. Article ID 1851865.
[26] Brinkmann H, Philippe H. Archaea sister group of Bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. Mol Biol Evol 1999;16(6):817–25. Epub 1999/06/16.
[27] Brinkmann H, van der Giezen M, Zhou Y, Poncelin de Raucourt G, Philippe H. An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. Syst Biol 2005;54(5):743–57. Epub 2005/10/26.
[28] Winker S, Woese CR. A definition of the domains Archaea, Bacteria and Eucarya in terms of small subunit ribosomal RNA characteristics. Syst Appl Microbiol 1991;14(4):305–10. Epub 1991/01/01.
[29] Forterre P. Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. Proc Natl Acad Sci U S A 2006;103(10):3669–74. Epub 2006/03/01.
[30] Spang A, Saw JH, Jorgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature 2015;521(7551):173–9. Epub 2015/05/07.
[31] Nasir A, Kim KM, Caetano-Anolles G. Lokiarchaeota: eukaryote-like missing links from microbial dark matter? Trends Microbiol 2015;23(8):448–50. Epub 2015/06/27.
[32] 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. Epub 2008/07/11.
[33] Forterre P, Philippe H. Where is the root of the universal tree of life? BioEssays News Reviews Mol Cell Dev Biol 1999;21(10):871–9. Epub 1999/09/25.
[34] Wang M, Yafremava LS, Caetano-Anolles D, Mittenthal JE, Caetano-Anolles G. Reductive evolution of architectural repertoires in proteomes and the birth of the tripartite world. Genome Res 2007;17(11):1572–85. Epub 2007/10/03.
[35] Wang M, Kurland CG, Caetano-Anolles G. Reductive evolution of proteomes and protein structures. Proc Natl Acad Sci U S A 2011;108(29):11954–8. Epub 2011/07/07.
[36] Kim KM, Caetano-Anolles G. The proteomic complexity and rise of the primordial ancestor of diversified life. BMC Evol Biol 2011;11:140. Epub 2011/05/27.
[37] Nasir A, Kim KM, Caetano-Anolles G. Global patterns of protein domain gain and loss in superkingdoms. PLoS Comput Biol 2014;10(1):e1003452. Epub 2014/02/07.
[38] Kurland CG, Harish A. The phylogenomics of protein structures: the backstory. Biochimie 2015;119:284–302. Epub 2015/08/04.
[39] Nasir A, Caetano-Anolles G. Comparative analysis of proteomes and functionomes provides insights into origins of cellular diversification. Archaea 2013;2013:648746. Epub 2014/02/05.
[40] Brochier C, Philippe H. Phylogeny: a non-hyperthermophilic ancestor for bacteria. Nature 2002;417(6886):244. Epub 2002/05/17.
[41] Jun SR, Sims GE, Wu GA, Kim SH. Whole-proteome phylogeny of prokaryotes by feature frequency profiles: an alignment-free method with optimal feature resolution. Proc Natl Acad Sci U S A 2010;107(1):133–8. Epub 2009/12/19.
[42] Williams TA, Heaps SE, Cherlin S, Nye TM, Boys RJ, Embley TM. New substitution models for rooting phylogenetic trees. Philos Trans R Soc Lond Ser B Biol Sci 2015;370(1678):20140336. Epub 2015/09/02.
[43] Lombard J, Lopez-Garcia P, Moreira D. The early evolution of lipid membranes and the three domains of life. Nat Rev Microbiol 2012;10(7):507–15. Epub 2012/06/12.
[44] Lombard J, Lopez-Garcia P, Moreira D. Phylogenomic investigation of phospholipid synthesis in archaea. Archaea 2012;2012:630910. Epub 2013/01/11.
[45] Cavalier-Smith T. Membrane heredity and early chloroplast evolution. Trends Plant Sci 2000;5(4):174–82. Epub 2000/03/31.
[46] Glansdorff N. About the last common ancestor, the universal life-tree and lateral gene transfer: a reappraisal. Mol Microbiol 2000;38(2):177–85. Epub 2000/11/09.
[47] Payandeh J, Pai EF. Enzyme-driven speciation: crystallizing Archaea via lipid capture. J Mol Evol 2007;64(3):364–74. Epub 2007/01/27.
[48] Kobiler O, Drayman N, Butin-Israeli V, Oppenheim A. Virus strategies for passing the nuclear envelope barrier. Nucleus 2012;3(6):526–39. Epub 2012/08/30.
[49] Sinninghe Damste JS, Strous M, Rijpstra WI, Hopmans EC, Geenevasen JA, van Duin AC, et al. Linearly concatenated cyclobutane lipids form a dense bacterial membrane. Nature 2002;419(6908):708–12. Epub 2002/10/18.
[50] van Niftrik L, Geerts WJ, van Donselaar EG, Humbel BM, Webb RI, Harhangi HR, et al. Cell division ring, a new cell division protein and vertical inheritance of a bacterial organelle in anammox planctomycetes. Mol Microbiol 2009;73(6):1009–19. Epub 2009/08/28.
[51] Chiu HC, Lin YW, Huang YF, Chuang CK, Chern CS. Polymer vesicles containing small vesicles within interior aqueous compartments and pH-responsive transmembrane channels. Angew Chem Int Ed Engl 2008;47(10):1875–8. Epub 2008/01/08.
[52] Deatherage BL, Cookson BT. Membrane vesicle release in bacteria, eukaryotes, and archaea: a conserved yet underappreciated aspect of microbial life. Infect Immun 2012;80(6):1948–57. Epub 2012/03/14.
[53] Soler N, Krupovic M, Marguet E, Forterre P. Membrane vesicles in natural environments: a major challenge in viral ecology. ISME J 2015;9(4):793–6. Epub 2014/10/15.
[54] Sinninghe Damste JS, Rijpstra WI, Geenevasen JA, Strous M, Jetten MS. Structural identification of ladderane and other membrane lipids of planctomycetes capable of anaerobic ammonium oxidation (anammox). FEBS J 2005;272(16):4270–83. Epub 2005/08/16.
[55] Sinninghe Damste JS, Rijpstra WI, Strous M, Jetten MS, David OR, Geenevasen JA, et al. A mixed ladderane/n-alkyl glycerol diether membrane lipid in an anaerobic ammonium-oxidizing bacterium. Chem Commun (Camb) 2004;22:2590–1. Epub 2004/11/16.
[56] Pearson A, Budin M, Brocks JJ. Phylogenetic and biochemical evidence for sterol synthesis in the bacterium *Gemmata obscuriglobus*. Proc Natl Acad Sci U S A 2003;100(26):15352–7. Epub 2003/12/09.

[57] Penny D, Poole A. The nature of the last universal common ancestor. Curr Opin Genet Dev 1999;9(6):672–7. Epub 1999/12/23.
[58] Poole AM, Jeffares DC, Penny D. The path from the RNA world. J Mol Evol 1998;46(1):1–17. Epub 1998/03/07.
[59] Caetano-Anolles G, Seufferheld MJ. The coevolutionary roots of biochemistry and cellular organization challenge the RNA world paradigm. J Mol Microbiol Biotechnol 2013;23(1–2):152–77. Epub 2013/04/26.
[60] Forterre P. Thermoreduction, a hypothesis for the origin of prokaryotes. C R Acad Sci III 1995;318(4):415–22. Epub 1995/04/01.
[61] Boussau B, Blanquart S, Necsulea A, Lartillot N, Gouy M. Parallel adaptations to high temperatures in the Archaean eon. Nature 2008;456(7224):942–5. Epub 2008/11/28.
[62] Brown JR, Douady CJ, Italia MJ, Marshall WE, Stanhope MJ. Universal trees based on large combined protein sequence data sets. Nat Genet 2001;28(3):281–5. Epub 2001/06/30.
[63] Jeffares DC, Poole AM, Penny D. Relics from the RNA world. J Mol Evol 1998;46(1):18–36. Epub 1998/03/07.
[64] Huber H, Kuper U, Daxer S, Rachel R. The unusual cell biology of the hyperthermophilic Crenarchaeon Ignicoccus hospitalis. Antonie van Leeuwenhoek 2012;102(2):203–19. Epub 2012/06/02.
[65] Kuper U, Meyer C, Muller V, Rachel R, Huber H. Energized outer membrane and spatial separation of metabolic processes in the hyperthermophilic Archaeon Ignicoccus hospitalis. Proc Natl Acad Sci U S A 2010;107(7):3152–6. Epub 2010/02/06.
[66] Seufferheld MJ, Kim KM, Whitfield J, Valerio A, Caetano-Anolles G. Evolution of vacuolar proton pyrophosphatase domains and volutin granules: clues into the early evolutionary origin of the acidocalcisome. Biol Direct 2011;6:50. Epub 2011/10/07.
[67] Lindsay MR, Webb RI, Strous M, Jetten MS, Butler MK, Forde RJ, et al. Cell compartmentalisation in planctomycetes: novel types of structural organisation for the bacterial cell. Arch Microbiol 2001;175(6):413–29. Epub 2001/08/09.
[68] Fuerst JA, Sagulenko E. Beyond the bacterium: planctomycetes challenge our concepts of microbial structure and function. Nat Rev Microbiol 2011;9(6):403–13. Epub 2011/05/17.
[69] Fuerst JA. Intracellular compartmentation in planctomycetes. Annu Rev Microbiol 2005;59:299–328. Epub 2005/05/25.
[70] Fuerst JA, Webb RI. Membrane-bounded nucleoid in the eubacterium *Gemmata obscuriglobus*. Proc Natl Acad Sci U S A 1991;88(18):8184–8. Epub 1991/09/15.
[71] Sagulenko E, Morgan GP, Webb RI, Yee B, Lee KC, Fuerst JA. Structural studies of planctomycete *Gemmata obscuriglobus* support cell compartmentalisation in a bacterium. PloS One 2014;9(3):e91344. Epub 2014/03/19.
[72] Lonhienne TG, Sagulenko E, Webb RI, Lee KC, Franke J, Devos DP, et al. Endocytosis-like protein uptake in the bacterium *Gemmata obscuriglobus*. Proc Natl Acad Sci U S A 2010;107(29):12883–8. Epub 2010/06/23.
[73] Lee KC, Webb RI, Janssen PH, Sangwan P, Romeo T, Staley JT, et al. Phylum Verrucomicrobia representatives share a compartmentalized cell plan with members of bacterial phylum Planctomycetes. BMC Microbiol 2009;9:5. Epub 2009/01/10.
[74] Pinos S, Pontarotti P, Raoult D, Baudoin JP, Pagnier I. Compartmentalization in PVC super-phylum: evolution and impact. Biol Direct 2016;11:38. Epub 2016/08/11.
[75] Jenkins C, Samudrala R, Anderson I, Hedlund BP, Petroni G, Michailova N, et al. Genes for the cytoskeletal protein tubulin in the bacterial genus *Prosthecobacter*. Proc Natl Acad Sci U S A 2002;99(26):17049–54. Epub 2002/12/18.
[76] Dacks JB, Field MC. Evolution of the eukaryotic membrane-trafficking system: origin, tempo and mode. J Cell Sci 2007;120(Pt 17):2977–85. Epub 2007/08/24.
[77] Koumandou VL, Dacks JB, Coulson RM, Field MC. Control systems for membrane fusion in the ancestral eukaryote; evolution of tethering complexes and SM proteins. BMC Evol Biol 2007;7:29. Epub 2007/02/27.
[78] Seufferheld MJ, Caetano-Anolles G. Phylogenomics supports a cellularly structured urancestor. J Mol Microbiol Biotechnol 2013;23(1–2):178–91. Epub 2013/04/26.
[79] Kim KM, Caetano-Anolles G. The evolutionary history of protein fold families and proteomes confirms that the archaeal ancestor is more ancient than the ancestors of other superkingdoms. BMC Evol Biol 2012;12:13. Epub 2012/01/31.
[80] Yutin N, Koonin EV. Archaeal origin of tubulin. Biol Direct 2012;7:10. Epub 2012/03/31.
[81] Schlieper D, Oliva MA, Andreu JM, Lowe J. Structure of bacterial tubulin BtubA/B: evidence for horizontal gene transfer. Proc Natl Acad Sci U S A 2005;102(26):9170–5. Epub 2005/06/22.
[82] Martin-Galiano AJ, Oliva MA, Sanz L, Bhattacharyya A, Serna M, Yebenes H, et al. Bacterial tubulin distinct loop sequences and primitive assembly properties support its origin from a eukaryotic tubulin ancestor. J Biol Chem 2011;286(22):19789–803. Epub 2011/04/07.
[83] Pilhofer M, Ladinsky MS, McDowall AW, Petroni G, Jensen GJ. Microtubules in bacteria: ancient tubulins build a five-protofilament homolog of the eukaryotic cytoskeleton. PLoS Biol 2011;9(12):e1001213. Epub 2011/12/14.
[84] Pilhofer M, Rosati G, Ludwig W, Schleifer KH, Petroni G. Coexistence of tubulins and ftsZ in different *Prosthecobacter* species. Mol Biol Evol 2007;24(7):1439–42. Epub 2007/04/07.
[85] Yee B, Lafi FF, Oakley B, Staley JT, Fuerst JA. A canonical FtsZ protein in *Verrucomicrobium spinosum*, a member of the bacterialphylum Verrucomicrobia that also includes tubulin-producing *Prosthecobacter* species. BMC Evol Biol 2007;7:37. Epub 2007/03/14.
[86] Brinkmann H, Venkatesh B, Brenner S, Meyer A. Nuclear protein-coding genes support lungfish and not the coelacanth as the closest living relatives of land vertebrates. Proc Natl Acad Sci U S A 2004;101(14):4900–5. Epub 2004/03/24.
[87] Arcas A, Cases I, Rojas AM. Serine/threonine kinases and E2-ubiquitin conjugating enzymes in Planctomycetes: unexpected findings. Antonie van Leeuwenhoek 2013;104(4):509–20. Epub 2013/08/07.
[88] Burroughs AM, Iyer LM, Aravind L. Functional diversification of the RING finger and other binuclear treble clef domains in prokaryotes and the early evolution of the ubiquitin system. Mol Biosyst 2011;7(7):2261–77. Epub 2011/05/07.
[89] Koonin EV, Yutin N. The dispersed archaeal eukaryome and the complex archaeal ancestor of eukaryotes. Cold Spring Harb Perspect Biol 2014;6(4):a016188. Epub 2014/04/03.
[90] Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, et al. STRING 8–a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res 2009;37:D412–6 (Database issue) Epub 2008/10/23.
[91] Grau-Bove X, Sebe-Pedros A, Ruiz-Trillo I. The eukaryotic ancestor had a complex ubiquitin signaling system of archaeal origin. Mol Biol Evol 2015;32(3):726–39. Epub 2014/12/20.
[92] Nunoura T, Takaki Y, Kakuta J, Nishi S, Sugahara J, Kazama H, et al. Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. Nucleic Acids Res 2011;39(8):3204–23. Epub 2010/12/21.
[93] Burroughs AM, Iyer LM, Aravind L. Structure and evolution of ubiquitin and ubiquitin-related domains. Methods Mol Biol 2012;832:15–63. Epub 2012/02/22.
[94] Chen LL, Wang GZ, Zhang HY. Sterol biosynthesis and prokaryotes-to-eukaryotes evolution. Biochem Biophys Res Commun 2007;363(4):885–8. Epub 2007/10/10.
[95] Desmond E, Gribaldo S. Phylogenomics of sterol synthesis: insights into the origin, evolution, and diversity of a key eukaryotic feature. Genome Biol Evol 2009;1:364–81. Epub 2009/01/01.
[96] Frickey T, Kannenberg E. Phylogenetic analysis of the triterpene cyclase protein family in prokaryotes and eukaryotes suggests bidirectional lateral gene transfer. Environ Microbiol 2009;11(5):1224–41. Epub 2009/02/12.
[97] Sinninghe-Damste JS, Rijpstra WIC, Schouten S, Fuerst JA, Jetten MS, Strous M. The occurrence of hopanoids in Planctomycetes implications for the sedimentary biomarker record. Org Geochem 2004;35:561–6.
[98] Fuerst JA, Sagulenko E. Towards understanding the molecular mechanism of the endocytosis-like process in the bacterium *Gemmata*

obscuriglobus. Biochim Biophys Acta 2014;1843(8):1732–8. Epub 2013/10/23.

[99] Santarella-Mellwig R, Franke J, Jaedicke A, Gorjanacz M, Bauer U, Budd A, et al. The compartmentalized bacteria of the Planctomycetes–Verrucomicrobia–Chlamydiae superphylum have membrane coat-like proteins. PLoS Biol 2010;8(1):e1000281. Epub 2010/01/21.

[100] Devos DP. Regarding the presence of membrane coat proteins in bacteria: confusion? What confusion? BioEssays News Reviews Mol Cell Dev Biol 2012;34(1):38–9. Epub 2011/12/21.

[101] Margulis L. Symbiosis in cell evolution. 2nd ed. New York: W. H. Freeman and Company; 1993.

[102] Rivera MC, Lake JA. The ring of life provides evidence for a genome fusion origin of eukaryotes. Nature 2004;431(7005):152–5. Epub 2004/09/10.

[103] Margulis L, To L, Chase D. Microtubules in prokaryotes. Science 1978;200(4346):1118–24. Epub 1978/06/09.

[104] Staley JT, Bouzek H, Jenkins C. Eukaryotic signature proteins of *Prosthecobacter dejongeii* and *Gemmata* sp. Wa-1 as revealed by in silico analysis. FEMS Microbiol Lett 2005;243(1):9–14. Epub 2005/01/26.

[105] Hartman H, Fedorov A. The origin of the eukaryotic cell: a genomic investigation. Proc Natl Acad Sci U S A 2002;99(3):1420–5. Epub 2002/01/24.

[106] McInerney JO, Martin WF, Koonin EV, Allen JF, Galperin MY, Lane N, et al. Planctomycetes and eukaryotes: a case of analogy not homology. BioEssays News Rev Mol Cell Dev Biol 2011;33(11):810–7. Epub 2011/08/23.

[107] Jeske O, Schuler M, Schumann P, Schneider A, Boedeker C, Jogler M, et al. Planctomycetes do possess a peptidoglycan cell wall. Nat Commun 2015;6:7116. Epub 2015/05/13.

[108] van Teeseling MC, Mesman RJ, Kuru E, Espaillat A, Cava F, Brun YV, et al. Anammox Planctomycetes have a peptidoglycan cell wall. Nat Commun 2015;6:6878. Epub 2015/05/13.

[109] Lindas AC, Karlsson EA, Lindgren MT, Ettema TJ, Bernander R. A unique cell division machinery in the Archaea. Proc Natl Acad Sci U S A 2008;105(48):18942–6. Epub 2008/11/07.

[110] Rachel R, Wyschkony I, Riehl S, Huber H. The ultrastructure of Ignicoccus: evidence for a novel outer membrane and for intracellular vesicle budding in an archaeon. Archaea 2002;1(1):9–18. Epub 2005/04/05.

[111] Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 2011;7:539. Epub 2011/10/13.

[112] Gouy M, Guindon S, Gascuel O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 2010;27(2):221–4. Epub 2009/10/27.