Unpacking the black box

How convergent prokaryotic features can inform hypotheses regarding eukaryogenesis

Abraham Lucas van Eijnatten

Examiners:

Berend Snel

Michael Seidl

Writing assignment
Graduate school of life sciences
University of Utrecht
Theoretical Biology Group

July 2022

Abstract

Eukaryotic cells are complex and consist of many cellular features that are absent from almost all prokaryotic cells. The universality of these features throughout the eukaryotic domain and the phylogeny of the underlying proteins demonstrate that most of these features probably originated during the process of eukaryogenesis. The culmination of eukaryogenesis was LECA, the ancestor of all eukaryotes. Because no intermediate forms from before LECA are currently extant, eukaryogenesis remains a black box. The black box nature of eukaryogenesis means that, while research has elucidated a lot of the features that evolved during eukaryogenesis, the order in which cellular features arose and the evolutionary constraints that shaped them remain an open question. Timing the events of eukaryogenesis using phylogenetic methods remains a messy and uncertain affair. In this literature review we suggest an alternative approach to illuminate the process of eukaryogenesis. We demonstrate many examples of prokaryotic organisms that contain features that are similar to eukaryotic features. We can use these features to, through analogy, form hypotheses about the process of eukaryogenesis. Convergently evolved prokaryotic features can elucidate from which cellular background and in which environment a certain feature can evolve. This approach is particularly potent because in prokaryotes one is often able to investigate a feature in isolation, whereas complexity is universal in eukaryotes. Because we are interested in convergently evolved features we exclude the Asgard archeae and their close relatives from our review, as their cellular features are likely ancestral to their eukaryotic counterparts. As an example of our approach, we use the complex endomembrane-like system of the planctomycete *G. obscuriglobus* to derive hypotheses about the evolution of the eukaryotic endomembrane system. Based on the parallel provided by *G. obscuriglobus* we speculate that sterol biosynthesis may have been a prerequisite to evolve the ability to form vesicles, that vesicle coat proteins may have evolved from nucleoporins, that vesicles could have performed their transport function in a stationary way before becoming mobile and that evolving endocytosis may be relatively easy after evolving the ability to form vesicles. We also use a phagocytosis-like process in a planctomycete, linear chromosomes in various bacterial clades, cytoskeletal motors in *M. xanthus*, a huge γ-proteobacterium that compartmentalizes its DNA and a prokaryotic endosymbiosis to make inferences about the evolution of similar features during eukaryogenesis. Only a tiny fraction of the breadth of prokaryotic diversity has ever been examined under the microscope. This implies there are likely many additional eukaryotic-like features in the prokaryotic tree of life. Experimental studies on novel prokaryotes have the potential to shed light on not only prokaryotic- but also eukaryotic evolutionary history.

Layman's summary

The term eukaryote refers to organisms with cells that contain a nucleus. This is a special cellular compartment in which the DNA is stored. Animals, plants, fungi and protists are all eukaryotes. There are two other domains of life that do not correspond to nucleated cells. These are the bacteria and archeae, collectively referred to as prokaryotes. Current research indicates that eukaryotes have evolved from a merger between an archeum and a bacterium during a long process called eukaryogenesis. Eukaryotic cells are quite complex compared to prokaryotic cells. Apart from the nucleus they also universally contain many other features that are largely absent from prokaryotic cells. An example is the endomembrane system, which is a set of membranes and cellular compartments that collectively transport and process the molecular components of the cell. How exactly eukaryogenesis took place is still a mystery because all eukaryotic cells have descended from a single ancestor and all intermediate forms of life have gone extinct. This makes it difficult to determine the order in which different cellular features evolved, and the evolutionary pressures that caused them to evolve. Sometimes a prokaryotic species harbors a feature that is very similar to a eukaryotic feature. We can, by comparing the two, use these prokaryotic features to learn about the process of eukaryogenesis. Often these prokaryotes have only one such feature which allows scientists to study it in isolation, rather than in a complex eukaryotic cell. A complex endomembrane, for instance, also occurs in the bacterium *G. obscuriglobus*. In this literature thesis we show many examples of cellular features that occur in a prokaryote and are very similar to a eukaryotic feature. We than show how these parallel features can be used to form hypotheses about the process of eukaryogenesis. There is a great diversity of prokaryotic organisms on the planet. Only an incredibly small fraction of these has been investigated under the microscope. It is likely that in future work many additional complex features will be discovered in prokaryotes. Our approach will therefore only become more relevant with time.

Contents

1 Introduction

Eukaryotes share a great complexity of cellular features. Cellular features are structures or mechanism that manifest at the cellular level and emerge from the combination of many individual molecular components. Typical eukaryotic features include the nuclear envelope, which separates the genetic material from the cytosol, an endomembrane system, capable of targeted transport of macro-molecules between cellular compartments, and a dedicated cellular compartment for energy production called the mitochondrion. The complexity that defines eukaryotes is often contrasted with the relatively simpler cell plans observed in most prokaryotic organisms. Many features are so ubiquitous throughout the eukaryotic domain of life that they are predicted to be present in LECA (Last Eukaryotic Common Ancestor) [1] the reconstructed ancestor from which all currently extant eukaryotes are descended. The absence of many hallmark eukaryotic features in prokaryotes leads to the conclusion that these features and the proteins of which they are constituted must have arisen during the evolutionary transition from prokaryotes to this LECA. This process is called eukaryogenesis.

The timing and ordering of the evolutionary events that took place during eukaryogenesis is difficult and remains a contentious subject. This is in large part due to the fact that no ancestral intermediate forms of complexity are currently extant [2]. Furthermore, The uncertainty in the cellular features and proteins already present during the evolution of a given feature makes it hard to determine the constraints under which this feature can evolve. Forming hypotheses on evolutionary constraints is also hampered by the fact that eukaryogenesis happened only once. We effectively have a sample size of 1.

The one-off nature of eukaryogenesis makes it hard to answer questions such as whether one feature came before another or whether a feature evolved because of strong physical constraints or mostly due to serendipity (neutral evolution). While such questions can only be answered definitively by phylogenetic analyses of sufficient power, we can at least increase our intuition on eukaryogenesis by investigating parallel features found in prokaryotes. The focus of both phylogenetic and experimental studies has been on eukaryotes, but it is becoming apparent that prokaryotic organisms also hold a great diversity of forms. The traditional view of prokaryotes as simple bags of proteins is quickly losing credibility. Many features observed in prokaryotes are surprisingly similar to features thought to occur only in eukaryotes. By investigating prokaryotic organisms harboring such features we can through analogy make inferences about the process of eukaryogenesis. Especially convergently evolved features could help elucidate the physical constraints required for that feature to evolve. Moreover, in prokaryotes it is often possible to investigate a feature in the absence of all the other complexity found in eukaryotic cells. Considering the cellular and genomic background of the prokaryote in which such a feature evolved could form the basis for hypotheses about the order in which features evolved during eukaryogenesis. The subject of this literature thesis is to identify prokaryotic analogs of eukaryotic features in the available literature, and to use these as a basis for hypotheses about eukaryogenesis.

The discovery of many eukaryotic signature proteins in the Asgard archeae has somewhat changed the view of eukaryogenesis over the past $\sim$7 years [3]. Quite some proteins previously believed to be eukaryotic inventions are already present in this archeal clade, such as a complete ubiquitin-ESCRT pathway [4]) which in eukaryotes has a function in membrane remodeling and many small GTPases [5] which in eukaryotes are involved in various regulatory processes including cytoskeleton remodeling, signal transduction and vesicular transport. The Asgard archeae are believed to be ancestral to LECA, and hence all eukaryotes, because of the abundance of eukaryotic signature proteins that are present in this clade. Due to the fact that the Asgard archeae were discovered only recently, much experimental work remains to be done to elucidate whether these proteins come together to form the same features they form in eukaryotes. Despite this, some studies already provide hints that these proteins result in similar features as their eukaryotic counterparts [6]. Undoubtedly the Asgard archeae form a wellspring of information about the origin of eukaryotic complexity. However, most current knowledge on the Asgard archeae concerns the similarity with eukaryotes at the protein level, rather than at the higher level of features. Moreover, all of these similarities are very likely ancestral with respect to eukaryotes. In this literature thesis we want to discuss how we can leverage convergent features to determine under what conditions and from what cellular machinery a certain feature might evolve. Any complexity in the Asgard archeae is part of the story of eukaryogenesis, not analogous to it. If we seek

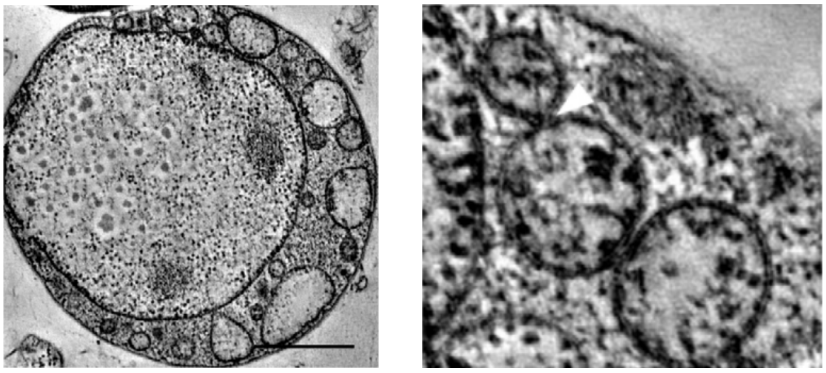
to increase our sample size we must look at other prokaryotes. For this reason we focus on features in prokaryotes that are not part of or closely related to the Asgard archeae.

2 A bacterial endomembrane system in the planctomycete *G. obscuriglobus*

Eukaryotic cells contain a multitude of membranes and membrane bound compartments that are collectively referred to as the endomembrane system. The primary function of the endomembrane is to process macromolecules such as proteins and lipids and transport these between various cellular compartments. Transport is performed by vesicles, which form from the endomembrane system by budding. The functioning of the endomembrane system is facilitated by a variety of essential proteins. Coat proteins such as clathrin and sec31 have an important role in facilitating vesicle formation from membranes [7, 8]. Other proteins have a function in connecting the compartments of the endomembrane system to the cytosol by enabling the selective transport of macro-molecules across membranes. An example of these are nucleoporins, a class of proteins that collectively form the nuclear pore complex. Nucleoporins are able to transport proteins across the nuclear envelope. The eukaryotic endomembrane system is also capable of the uptake of macro-molecules external to the cell via budding of the cell membrane into vesicles. This process is called endocytose. Coat proteins play an essential role in endocytose. In eukaryotes lipids, in particular sterols, have an important function in regulating the fluidity and structural integrity of membranes.

A complex endomembrane system is generally considered a hallmark eukaryotic feature. Nevertheless it has long been known that members of the bacterial superphylum PVC (*Planctomycetes-Verrucomicrobiae-Chlamydiae*) show complex membrane invaginations. These invaginations are so extensive that until recently, the cell plan of planctomycetes and some other members of the PVC-superphylum was not recognized as gram-negative [9, 10, 11]. The consensus was that the cytoplasm in these organisms was divided into two compartments by the intra-cytoplasmic membrane. In this view, one of these compartments, the pirellulosome, contained the DNA and ribosomes, while the ribosome free paryphoplasm was located towards the outside of the cell. The outer membrane was regarded as equivalent to the cytoplasmic membrane in gram-negative bacteria. Nowadays, due to the advent of improved microscopic techniques and new phylogenetic data, most of the scientific community regards the planctomycetes as having the typical gram-negative bacterial cell plan. What was previously called the intra-cytoplasmic membrane actually corresponds to the inner membrane and the paryphoplasm is equivalent to the periplasm [12, 13, 14]. The confusion is nevertheless understandable because under the microscope these bacteria look entirely different from other gram-negative bacteria. In fact, the membranous structures observed in members of the PVC-phylum bear a strong resemblance to the eukaryotic endomembrane system.

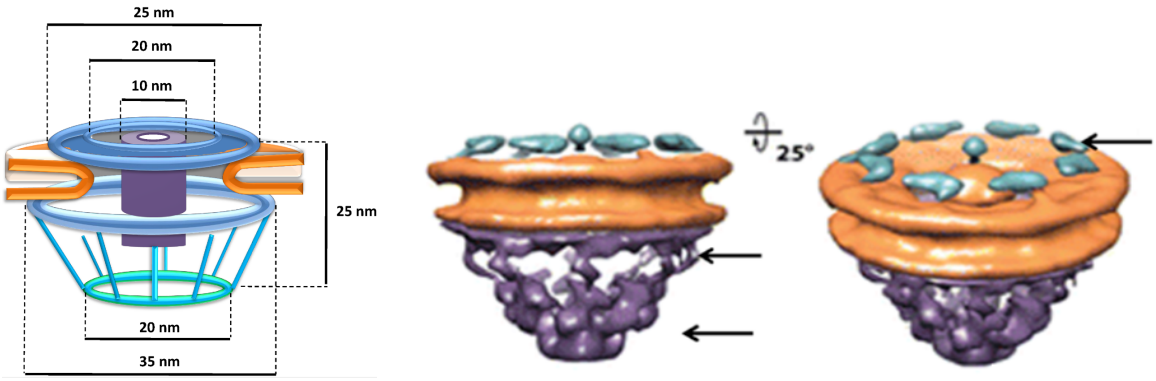
In particular the planctomycete *Gemmata obscuriglobus* has a complex (and relatively well researched) endomembrane system [15]. This organism displays two different cell types, one showing extensive membrane invaginations of the inner membrane into the cytosol, and the other showing vesicle-like structures in the periplasm (Figure 1A). In the second cell type vesicles appear to be part of a mesh of vesicles and tubules in the periplasm. Most vesicles are attached to other vesicles or the inner or outer membrane (Figure 1B). The vesicles are continuous with the cytosol, as is evidenced by the presence of ribosomes. Material concentrations appear to differ between vesicles, suggesting selective transport. This shows that vesicles likely have a function in the selective transport of macro-molecules, just like in the eukaryotic endomembrane system. However, transport appears to occur from one stationary vesicle to another, as opposed to the directional transport of vesicles to a particular cellular compartment. While the vesicles in *Gemmata obscuriglobus* likely have a similar function to vesicles in eukaryotes, they achieve this using a different method.



(A) **(B)**

Figure 1: Vesicles in a representative cell of *G. obscuriglobus*. Figures taken from [15]. **A)** Cross section of the cell. Scale bar is 500nM. **B)** Zoomed in version on connection between vesicles. White arrow denotes such a connection.

The vesicles are associated with proteins that are structurally similar to eukaryotic proteins. Membranes are surrounded by proteins with similar folds as eukaryotic coat proteins [16]. These proteins likely have a similar function as eukaryotic coat proteins. Despite the structural similarity, there appears to be no homology at the sequence level. *G. obscuriglobus* also produces proteins that appear to have a similar structure to the eukaryotic nucleoporins. In parallel with the coat protein analogs, the proteins that are structurally similar to nucleoporins associate with vesicles and display no detectable homology at the sequence level. These proteins form genuine pore complexes (Figure 2A) which are similar in complexity and structure to those of eukaryotes (Figure 2B) [17]. If we also consider that selective transport across vesicle membranes was demonstrated experimentally we must conclude that, like the coat protein analogs, the pore complexes are almost certainly functionally as well as structurally analogous to their eukaryotic counterparts.



(A) **(B)**

Figure 2: Apparent structural similarity between the pore complex of *G. obscuriglobus* and the eukaryotic pore complex. a) Schematic of the pore complex of *G. obscuriglobus*. Figure from [17] b) Image of the pore complex of the eukaryote *Dictyostelium discoideum* generated with cryoelectron tomography. Figure from [18].

G. obscuriglobus is also capable of sterol synthesis, as evidenced by the presence of the genes required to synthesize primitive sterols from squalene: squalene mono-oxygenase and oxidosqualene cyclase. While sterol synthesis is ubiquitous in eukaryotes, *G. obscuriglobus* is the only known planctomycete capable of this feat. Genes for sterol biosynthesis are sporadically present in six distinct bacterial phyla. The origin of sterol biosynthesis is still unresolved. Earlier phylogenetic analysis suggested that *G. obscuriglobus* contains an ancestral version of the sterol biosynthesis pathway [19, 20]. A paper published in nature from 2017 concluded that sterol biosynthesis originated in stem eukaryotes and was transferred to bacteria in at least two HGT (horizontal gene transfer) events during eukaryogenesis [21]. This explanation seems in accordance

with the extremely sparse distribution of these genes throughout the bacterial kingdom. However, the latest research suggests that sterol biosynthesis originated in bacteria and was transferred from myxobacteria (an order within the δ-proteobacteria) to eukaryotes [22]. The exact phylogenetic relationship between the sterol biosynthesis genes in *G. obscuriglobus* and eukaryotes is unclear due to the contradictory results of various studies, although they all suggest homology. Sterol biosynthesis appears to have an essential function in *G. obscuriglobus*, as mutants deficient in sterol synthesis display aberrant membrane and vesicle structures, such as nested vesicles and stacks of membranes in the cytosol [23]. Sterol biosynthesis mutants are also unable to reproduce successfully. While structural analogs of vesicle coat proteins and nucleoporins occur throughout the PVC-superphylum [16], *G. obscuriglobus* is so far the only organism in this phylum displaying the feature of vesicle formation and the only organism capable of sterol biosynthesis. This implies sterols could be essential for maintaining the structural integrity of vesicles.

Amazingly, the endomembrane-like system of *G. obscuriglobus* is capable of a process akin to the process of endocytosis [24]. While some bacteria can make use of external proteins by degrading them at the cell surface [25], *G. obscuriglobus* is the only known organism in either the bacterial or archeal kingdom that can internalize intact proteins. Like in eukaryotes, the endocytic process performed by *G. obscuriglobus* is energy dependent and receptor mediated. Internalized proteins locate mainly in vesicles in the periplasm.

G. obscuriglobus displays an endomembrane-like system which is to a large extend structurally and functionally analogous to that of eukaryotes. The lack of sequence homology with relevant eukaryotic genes shows that this machinery most likely resulted from convergent evolution. This makes *G. obscuriglobus* an exceptional model to investigate the constraints involved in evolving an endomembrane system. The fact that similar protein folds evolved to perform similar functions means that the evolution of these components of the endomembrane system is likely strongly driven by physical constraints. Exceptions to the general picture of convergent evolution are the sterol biosynthesis genes, which appear to have been acquired through HGT. The biosynthesis of sterol seems to have been invented only once in the history of life. Possibly sterol biosynthesis genes represent solitary peaks in the fitness landscape. *G. obscuriglobus* is the only known member of the PVC-superphylum in which vesicles are observed, and also the only member capable of endocytosis. Evolving endocytosis might be relatively trivial after having acquired the ability to form vesicles.

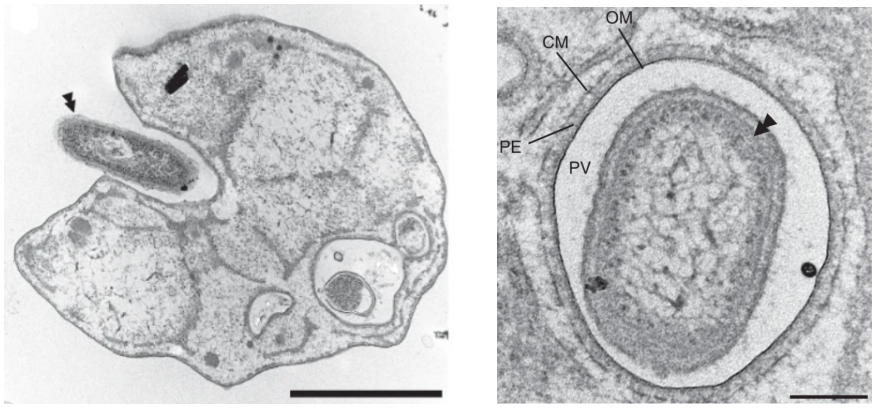
G. obscuriglobus also provides hints concerning the timing and ordering of events during eukaryogenesis. It has been hypothesized based on structural similarity that nucleoporins and vesicle coat proteins are ancestrally related [26, 27] in eukaryotes. If we believe structural similarity is sufficient evidence of this than the same must be true in *G. obscuriglobus*. From the apparent absence of vesicle formation in other planctomycetes, despite the presence of proteins that are structurally similar to coat proteins, it seems that in this phylum the transport over membrane function evolved before the vesicle formation function. This provides a clue that during eukaryogenesis vesicle coat proteins might have evolved from nucleoporins. We concluded based on the co-occurrence of vesicle formation, endocytosis and sterol synthesis in *G. obscuriglobus* that sterols might be an essential feature for maintaining the membrane integrity of vesicles. We can speculate that during Eukaryogenesis sterol synthesis was a prerequisite for evolving a complex endomembrane system in which macro-molecules could be transported between compartments or molecules could be internalized through endocytosis. This would be congruent with the early evolution or acquisition of sterol biosynthesis during eukaryogenesis [21]. Vesicles are stationary entities in *G. obscuriglobus*, but nevertheless function in the transport of macro-molecules. Possibly, this represents a preliminary stage of evolution during which the regulatory mechanisms to selectively direct a vesicle from one membrane to another do not yet exist. We can postulate that the evolution of vesicles can be adaptive even in the absence of many cellular compartments, and could have evolved early during eukaryogenesis. It seems that endocytosis has evolved from the capability to form vesicles in *G. obscuriglobus*. Apart from proteins with similar folds as coat proteins and nucleoporins and homologs of sterol biosynthesis proteins, none of the other complexity of eukaryotic endocytosis appears essential to the process, and might have been invented only after a more primitive version of the process was already present.

3 A cell wall and two membranes pose no obstacle to phagocytosis in the bacterium *U. amorphum*

Another feature of many eukaryotic cells is the ability to take up large particles or other organism from the environment through a process called phagocytosis. In all eukaryotes capable of phagocytosis, the process depends on the remodeling of the cytoskeleton through actin polymerization. Engulfment takes place by invagination of the cell membrane, and prey is digested in the resulting cellular compartment. This compartment is called the phagosome. The proteins involved in this process differ substantially between different eukaryotic clades, with only a few common denominators. Examples of common denominators are the ARP2/3 complex and formin- and spire proteins. The fact that phagosytic processes are present throughout the eukaryotic domain indicates that this feature was likely present in LECA, although it has also been suggested based on both the diversity of phagocytic machinery in eukaryotes and palaeontological considerations that phagocytosis has convergently evolved multiple times in crown eukaryotes [28].

The origin of phagocytosis is frequently discussed in conjunction with the acquisition of the mitochondrion. Theories of early mitochondrial acquisition posit that an archeal host and an α-proteobacterium developed a symbiotic relationship through some mechanism other than phagocytosis [29]. In this view, the energy generated by the mitochondrion allowed for the subsequent evolution of eukaryotic complexity [30], including phagocytosis. Theories of late mitochondrial acquisition propose that much of the eukaryotic complexity including phagocytosis first arose, and that this enabled a "proto-eukaryote" to engulf the ancestor of the mitochondrion [31, 29]. The latter perspective has increased in credence since the discovery of eukaryotes which harbor almost the complete repertoire of eukaryotic complexity except for mitochondria [32]. Resolving the timing of the acquisition of phagocytosis is an important outstanding question and has profound implications for many theories on the origin of the mitochondrion and the evolution of cellular complexity in eukaryotes.

Until recently the uptake of whole cells by an organism had never been described in prokaryotes including the Asgard archeae. A paper from 2019 changed this by describing a phagocytic process in a novel organism the researchers called *C. Uab amorphum* [33]. This organism, like *G. obscuriglobus*, is part of the planctomycete phylum, in which it appears to form its own clade. Its closest relatives are the anammox bacteria which are best known for having a membrane bound cellular compartment in which they generate energy by oxidizing ammonium. When cultured in the presence of gram-negative- or gram-positive bacteria or picoeukaryotic algae, *U. amorphum* was able to engulf these organisms entirely and digest them in a phagosome-like compartment. This compartment was surrounded by two membranes, showing that both the outer- and inner membrane invaginate during engulfment. Like its eukaryotic counterpart, the process is energy dependent, as it cannot take place when ATP-hydrolase activity is blocked. At the level of the cellular feature the similarities with eukaryotic phagocytosis are striking.



(A) **(B)**

Figure 3: The phagocytosis-like process performed by *U. amorphum*. Figures from [33]. **A)** Scanning electron micrograph showing *U. amorphum* engulfing a bacterium. Double arrowhead denotes bacterium. Scale is 500nM. **B)** Transmission electron micrograph of *U. amorphum* showing the phagosome-like vacuole (PV) surrounded by the outer membrane (OM), periplasm (PE) and cytoplasmic membrane (CM). Scale is 200 nM. Double arrowhead indicates prey bacterium.

The genome of *U. amorphum* also displays similarities with other phagosytic organisms. It has a larger genome than most other planctomycetes. Its genome encodes a great variety of metabolic enzymes and putative antibiotic resistance genes. A large fraction of these it appears to have acquired through horizontal gene transfer from other bacteria. *U. amorphum* is capable of synthesizing only 6 amino-acids *de novo*, apparently being dependent on the phagocytosis-like process to acquire the other amino-acids. The extend to which the genome reflects a phagotrophic lifestyle combined with the fact that it not possible to grow *U. amorphum* in an axenic culture shows that this organism is probably an obligatory phagotroph.

U. amorphum displays the typical planctomycete cellular features such as a highly invaginated inner membrane. However unlike *G. obscuriglobus*, it appears to possess no structural analogs of either coat proteins, nucleoporins or sterol biosynthesis genes. Its genome also shows very little homology with any eukaryotic signature proteins, with the most notable exception being an actin homolog. The latter protein may be important for the phagocytosis-like process because of its possible involvement in different types of fibrous structures in the *U. amorphum* cell, the largest of which can be up to two μM in length (Figure 4). Such large fibrous structures have not previously been demonstrated to occur in bacteria. In phylogenetic analysis the actin homolog clusters within the Asgard archeae, showing that it is probably ancestral to eukaryotic actin. Therefore, the phagocytic process in *U. amorphum* appears to be the product of convergent evolution with respect to eukaryotic phagocytosis, with the exception of the acquisition of an actin that is ancestral to eukaryotic actin.

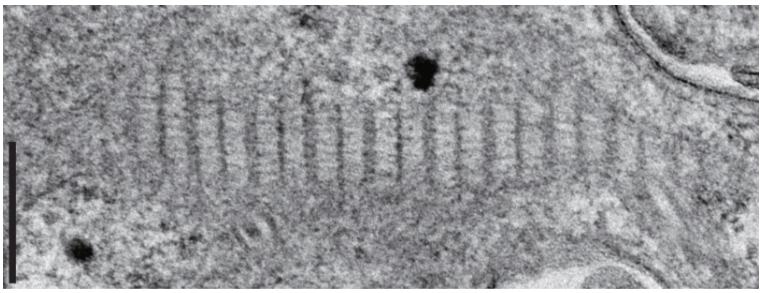


Figure 4: Large striated fiber observed in *U. amorphum*. Scale bar is 200nM. Figure from [33].

To facilitate the phagocytic behavior *U. amorphum* has evolved another hallmark Eukaryotic feature:

amoeba like movement. This is especially interesting in the light of the presence of peptidoglycan biosynthesis genes. These genes indicate that the organism most likely has a cell wall, which generally does not allow for such flexible movement. A possible explanation is that the cell wall is broken down upon encountering an organism to enable engulfment. After engulfment has taken place, the cell wall would in this view be reconstructed. *U. amorphum* has also evolved a relatively large cell size (5μM) which is a prerequisite to engulfing other organisms. Potentially, a larger cell size first evolved during eukaryogenesis to accommodate phagocytosis. This feature would than later be co-opted to house the large variety of cellular compartments seen in present day eukaryotes.

Phagocytosis is sometimes considered a sub-process of the more general process of endocytosis. As an endocytosis-like process is known to be present in another planctomycete, it is tempting to speculate that this could be the evolutionary origin of the phagocytosis-like feature observed in *U. amorphum*. In this view the endocytic process, while only observed in *G. obscuriglobus*, is nevertheless widespread throughout planctomycetes. However, the endosytic process of *G. obscuriglobus* and the phagocytic process of *U. amorphum* appear mechanistically quite dissimilar despite a certain similarity in function. The phagocytic process observed in *U. amorphum* requires the invagination of two rather than one membrane. The phagosome-like organelles share little similarity with the vesicles observed in *G. obscuriglobus*. The former are located in the cytoplasm, while the latter are in the periplasm. Finally, none of the proteins with a function in the endomembrane-like system in *G. obscuriglobus* such as analogs of coat proteins and nucleoporins or the sterol biosynthesis genes appear to be present in *U. amorphum*. The most parsimonious explanation seems to be that these processes evolved separately. The planctomycetes offer the possibility that the processes of endocytosis and phagocytosis could have evolved in parallel during eukaryogenesis, rather than one being derived from the other.

The above implies that it is possible that a primitive form of phagocytosis could have evolved in eukaryotes before the full complexity of the endomembrane system. Certainly sterol biosynthesis and coat proteins are not a prerequisite for phagocytosis. Importantly, mitochondria are also not required for phagocytosis to occur. These observations lend indirect support to early mitochondrial acquisition theories by showing that a primitive form of phagocytosis is possible in the absence of many hallmark eukaryotic features and proteins.

4 Convergent evolution of linear chromosomes in bacteria

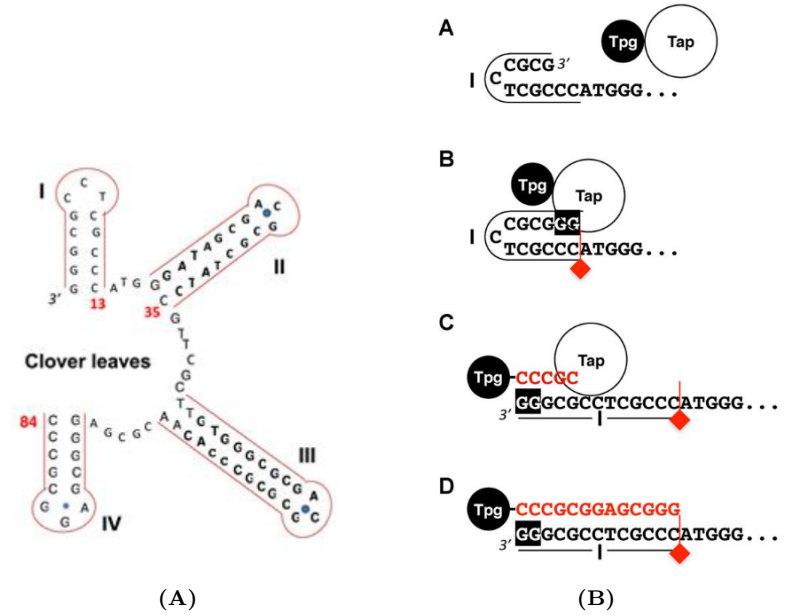
All known eukaryotes have linear chromosomes. This stands in stark contrast to prokaryotes, the vast majority of which has circular chromosomes. The reason why linear chromosomes are so ubiquitous throughout the eukaryotic domain is not clearly understood. It might be that linear chromosomes convey some evolutionary advantage. On the other hand, they might be the result of chance events during eukaryogenesis, and be largely neutral in terms of fitness.

One argument against the neutrality of linear chromosomes is that they do pose a problem to which circular chromosomes are not subject. This is related to the way DNA is replicated [34, 35]. During replication there are distinct mechanism for replicating the leading and the lagging strand. Replication of the lagging strand is generated in the opposite direction of the replication fork and requires an RNA primer. At the ends (telomeres) of the chromosome DNA synthesis cannot be primed and hence cannot be replicated by DNA polymerase. In eukaryotes, a protein called telomerase is responsible for resolving this issue by performing *de novo* DNA synthesis on the leading strand. DNA polymerase subsequently uses these additional bases to prime synthesis on the lagging strand. Using the telomerase protein eukaryotes are able to overcome the end replication problem.

Another, less convoluted way to prevent the end replication problem is to simply have circular chromosomes such that there are no chromosomal ends. This is the strategy taken by the vast majority of prokaryotes. Nevertheless, linear chromosomes occur sporadically in disparate bacterial clades [36, 37, 38, 39]. There are two known strategies bacteria employ to solve the end replication problem. These strategies are likely entirely convergent with respect to the eukaryotic strategy and as such provide interesting

parallels. Illuminating the evolutionary pressures under which bacteria evolve linear chromosomes from circular chromosomes might help shed light on the reason that linear chromosomes evolved during eukaryogenesis.

The first strategy for solving the end replication problem occurs in bacteria belonging to the genus *Streptomyces*, and possibly in related actinobacterial clades [40] such as *Nocardiaceae*. These actinobacteria solve the end replication problem by using a protein-primer to complete replication [41, 42]. In these organisms the initial replication process leaves single strand overhangs at the telomeres. The overhangs are about 280 bp long and contain in the first 170 bp many palindromic sequences that are very conserved in the genus [43]. The palindromic sequences form a clover-like secondary structure (Figure 5A) that binds two terminal proteins, Tap and Tpg. The Tpg protein provides a hydroxyl group from which DNA synthesis can be primed, while Tap binds the DNA and acts as a DNA primase to synthesize 13 bp of DNA. These 13 bp are than used as a primer for DNA synthesis on the remainder of the telomere (Figure 5B). The protein priming of DNA replication is known to occur in the genomes of bacteriophages and viruses [44]. In these cases, contrary to the *Streptomyces* approach, the proteins prime DNA replication for the entire length of the replicon, rather than just at the ends. Despite this difference it is feasible that the protein-priming machinery of Tpg and Tap was acquired from a virus or phage, although this is not yet corroborated by phylogenetic analysis.



(A) **(B)**

Figure 5: Protein priming by the Tpg-Tap complex. Figures from [42]. **A)** The leading strand overhang folds into a clover like secondary structure. The different palindromes are denoted by Latin numbers. Red numbers denote base count starting from the 3' end of palindrome 1. **B)** Schematic of the process of protein priming. Step A: The Tpg-Tap complex arrives at palindrome 1 by binding palindrome 2 and 3 (not displayed). Step B: Tpg primes DNA synthesis by Tap. Tap fills in the two basepairs at the end of the 3' end of palindrome 1 (inverted red flag). Step C: The hairpin is broken up and Tap synthesizes the first 13 bp. Step D: Tap dissociates and DNA polymerase completes the DNA synthesis process.

It is often assumed that the linear chromosomes of streptomycetes result from the fusion of a circular chromosome with a linear replicon [45, 46]. This would naturally lead to the observed genome architecture where non essential genes are located in the chromosomal arms. This hypothesis is supported by the observation that the chromosomal core of streptomycetes is almost entirely syntenic with the complete (circular) chromosome of *Mycobacterium tuberculosis* [47]. The fusion event must have had some benefit, because *Streptomyces* chromosomes readily circularize upon deletion of the chromosomal arms and individuals with circular chromosomes are still capable of replication, sometimes with no discernible decrease it fitness [45]. Therefore neutral evolution is not a sufficient explanation for the acquisition of linear chromosomes in the

case of the *Streptomyces* genus.

Streptomycetes might benefit from linear chromosomes on account of their unique lifecycle. These organisms grow in stationary colonies which resemble the fungal mycelium. At least four developmental stages can be found in these organisms including vegetative hyphae, aerial hyphae and spores [48]. The filamentous method of growth means that streptomycetes have had to become master regulators that manage competition with other microbes using a great diversity of anti-microbial compounds[49]. The genes responsible for producing secondary compounds and other non-essential genes are located in the chromosomal arms, while in the core of the genome are all those genes that are essential for the maintenance of basic cellular functions [47]. The chromosomal arms are very unstable and demonstrate frequent deletions of up to 2MB [50, 51]. These arms are quite variable even within sympatric populations [52] and experience a high frequency of recombination. It is often hypothesized that the adaptive benefit of linear chromosomes in streptomycetes is that they enable the rapid transfer of environment-specific proteins through the population by means of frequent recombination events. This would be congruent with the open pan genome of the genus [53]. At the same time the deletions would allow streptomycetes to quickly shed such genes when they are no longer required [54]. Interestingly, deletions also appear to have a function in facilitating a division of labor between antibiotic producing mutants and wildtypes with higher reproductive success [55]. In streptomycetes linear chromosomes appear to be adaptive because they allow for chromosomal polarity between a stable core and variable arms which can adapt on short evolutionary timescales.

The second bacterial strategy for solving the chromosomal end problem involves the formation of hairpin loops at the telomeres which connect the leading and lagging strand with a covalent bond. DNA replication now results in circular dimers that are resolved by a protelomerase enzyme. This enzyme breaks the circular intermediate in two and subsequently ligates the chromosomal ends, resulting in two linear chromosomes with closed hairpin loops at the telomeres (Figure 6) [56]. There are two known groups of bacteria which utilize this strategy. The first are bacteria of the *Borrelia* genus. This genus includes the well known pathogen *Borrelia burgdorferi* which causes Lyme disease. The second are bacteria of the genus *Agrobacterium* [57]. Both of these genera belong to different phyla (spirochetes and α-proteobacteria respectively). The lack of linear chromosomes in other bacteria from these phyla shows that, as was the case for the streptomycetes (and close relatives), these genera probably had an ancestor with a circular genome, which was linearized. Again, the most likely catalyst is the fusion of the circular chromosome with a linear replicon from a phage or a virus. In the case of *Borrelia* it is speculated because of similar telomeric sequences that the origin of its protelomerase is the African swine fever virus [58, 59], with which it coexists in arthropods.

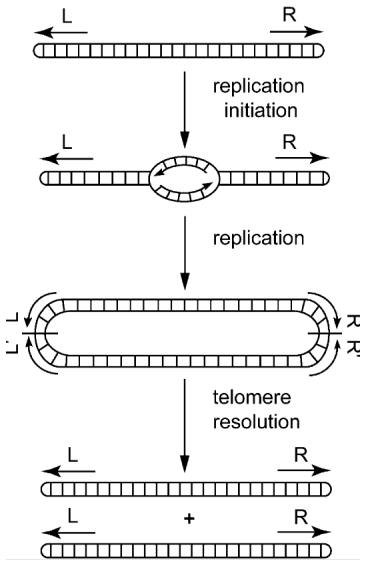


Figure 6: DNA replication through hairpin resolution as it occurs in *Borrelia* and *Agrobacterium*. The L and R indicate the two hairpin telomeres. Replication occurs from an origin of replication. Due to the hairpin at the telomeres this results in a symmetrical circular intermediate. Protelomerase activity resolves the telomeres, resulting in two identical linear chromosomes with hairpins. Figure from [60].

The *Borrelia* genome is very distinct from that of streptomycetes. Genes located on the chromosome display a high level of synteny within the genus, without a clear difference between the core and the arms of the chromosome [61]. Indeed, while streptomycetes appear to place housekeeping genes in the core of the chromosome and non essential genes in the chromosomal arms, *Borrelia* species take a different approach by placing the essential genes on the chromosome and the non-essential genes on plasmids. Bacteria in the *Borrelia* genus have the greatest diversity of plasmids, with up to 24 plasmids in a single species [62]. Interestingly, most of these are also linear and maintained with hairpin loops. In the case of *Borrelia* the recombination rate is constant across the length of the chromosome. If linear chromosomes are adaptive for *Borrelia*, the benefit is likely different from the benefit received by the streptomycetes.

The *Agrobacterium* genus also harbors linear chromosomes with hairpin telomeres. Interestingly, *Agrobacterium* species appear to have one circular and one linear chromosome. Most likely the linear chromosome was derived from a plasmid [63]. Because it now carries many essential genes and is longer than 2MB, it is referred to either as a chromosome or as a chromid. Sister genera to *Agrobacterium* also have two chromosomes, although both are circular. At the base of the *Agrobacterium* genus the TelA protelomerase gene was acquired [64] through HGT, which probably led to the linearization event. The linear chromid is conserved throughout the genus. This could be explained by either a ratchet-like explanation or an adaptive benefit. The linear chromosome does not experience higher levels of recombination than the circular chromosome [65]. As in the case of *Borrelia*, it is uncertain whether there is an adaptive benefit to the linear chromosomes of *Agrobacterium*.

Two mechanisms for resolving the end replication problem of linear chromosomes appear to have evolved in prokaryotes. Both of these are convergent with respect to the eukaryotic solution. In prokaryotes, linear replicons that likely originated from phages or viruses provided the end-resolution machinery to maintain linear chromosomes. Telomerase on the other hand is known to have evolved from group II introns [66]. Compared to the prokaryotic case there is much less evidence to implicate viruses or phages in the acquisition of linear chromosomes by eukaryotes. Regarding the method of linearization the prokaryotic cases might not be be analogous to the eukaryotic case. The prokaryotic linear chromosomes also do not provide much clues regarding the possible adaptive benefit of linear chromosomes. Highly recombinatorial arms as they occur in streptomycetes are not a feature common to most eukaryotes. The advantage of linear chromosomes in *Borrelia* and *Agrobacterium* is unclear. They could be adaptive for some unknown reason or they may be

largely neutral. Perhaps revealing more of the cell biology of *Borrelia* and *Agrobacterium* will allow us to resolve this question, and make the parallel with eukaryotes. Assuming adaptivity of linear chromosomes on the basis of the Streptomycetes, in which case linear chromosomes are adaptive for reasons that clearly do not apply to either *Borrelia* and *Agrobacterium* or eukaryotes, is too much of a leap. To help elucidate the evolutionary pressures that favor linear chromosomes it would help to identify linear chromosomes in other bacteria. Homology searches identify homologs of the TelA protelomerase in *Borrelia*, cyanobacteria, and several phages and viruses [64]. The presence of TelA in cyanobacterial genomes could indicate that these also harbor linear chromosomes. Genomic and microscopic analysis of these cyanobacteria could be the next step forward.

5 Cytoskeletal motors enable gliding motility in *M. xanthus*

Eukaryotes have a complicated and dynamical cytoskeleton. Fundamental to the eukaryotic cytoskeleton are actin and tubulin polymers, which form microfilaments and microtubules respectively. The cytoskeleton has a multitude of functions including structural support of the cell, facilitating movement and chromosome segregation during cell division. While it was initially thought that prokaryotes had no cytoskeleton, it is now well understood that prokaryotes can have quite extensive and complex cytoskeletons. This is highlighted by the fact that all known eukaryotic cytoskeletal proteins have homologs in prokaryotes [67]. The bacterial homologs of actin and tubulin are FtsZ and MreB respectively. A hallmark feature of the eukaryotic cytoskeleton is the movement of motor proteins along the cytoskeleton. These proteins have a variety of functions including cell motility and the transport of cargo-molecules. Examples of such motor proteins are myosins, which move along actin filaments, and kinesins, which move along tubulin filaments.

The bacterium *Myxococcus xanthus* is able to slide over moist surfaces in a process called "gliding motility" [68]. Gliding motility is observed in many branches of the bacterial tree, but best researched in *M. xanthus*. The mechanism responsible for the observed movement pattern turns out to involve the movement of a motor protein complex along MreB filaments [69, 70]. This in turn causes the MreB filaments to move in the same or the opposite direction with a helical motion (Figure 7) [70]. This phenomenon is analogous to the dynamics between actin and myosin, where actin directs the movement of myosin while myosin is in turn responsible for the movement of actin [71]. The cytoskeletal motor in *M. xanthus* constitutes the first evidence of a motor protein in a bacterium.

The evolutionary origin of the motor proteins of *M. xanthus* has already been exposed to genomic and phylogenetic analysis. Most of the motor protein machinery, i.e. the Agl-Glt complex, appears to have evolved from protein complexes involved in assembling a polysaccharide capsule during sporulation, with the force generating element being homologous to the flagellar MotAB complex [72, 73]. The motor proteins of *M. xanthus* display no homology with proteins in other bacterial phyla capable of gliding motility such as bacteroidetes, cyanobacteria or δ-proteobacteria [72]. Despite this, there is reason to suspect that at the very least in the case of the bacteroidetes the mechanism also involves protein complexes moving along the cytoskeleton [69]. This indicates motor proteins have most likely evolved multiple times within prokaryotes.

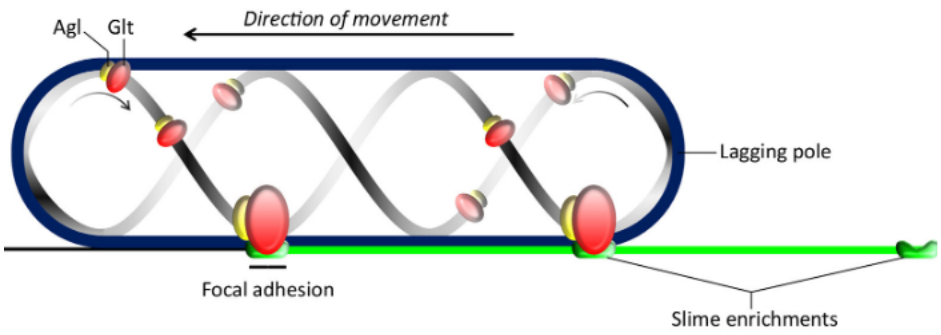


Figure 7: Schematic of gliding motility in *M. xanthus*. The Agl-Glt complex moves along the cytoskeleton until it attaches to the surface at the focal adhesion point. Continued movement of the Agl-Glt complex over the MreB filaments now causes the cytoskeleton to rotate in a helical fashion, facilitating cell movement. Figure from [74].

We can speculate on the evolution of motor proteins during eukaryogenesis based on the parallel provided by *M. xanthus*. It is not yet certain whether the motors responsible for gliding motility also harbor a function in protein transport. If not, it would be reasonable to posit the hypothesis that during eukaryogenesis evolution might have taken a similar trajectory. According to this line of reasoning cytoskeletal motors evolved first to facilitate movement, and were subsequently co-opted for protein transport. If it is found that similar motors also transport proteins along the cytoskeleton, we arrive at the same hypothesis, considering the likely origin of the motor from the spore capsule forming machinery. Movement and capsule forming require the binding to extracellular polysaccharides, a function not required for protein transport within the cell. We can also speculate about the constraints of evolving cytoskeletal motors. Gliding motility using cytoskeletal motors appears to have evolved convergently at least twice and likely several times in prokaryotes. The fact that this happened in several distinct phyla suggests that cytoskeletal motors are a constraint driven adaption to enable mobility, which is not so difficult to evolve from the cytoskeletal and motor elements present throughout the bacterial domain.

6 A gigantic bacterium with membrane bound DNA compartments

Another distinguishing feature of eukaryotes is large cells. The typical eukaryotic cell is around $10 \sim 100\mu$M, whereas an average prokaryotic cell is only around $0.1 \sim 5\mu$M. The distinction is not absolute, as is evidenced by both small eukaryotic cells [75] and large prokaryotic organisms. Bacteria with cell lengths longer than 100μM occur in four bacterial clades (cyanobacteria, spirochetes, firmicutes and colorless or purple sulfur bacteria within the proteobacteria) [76]. Most of these are only big in one dimension, while being only a few μM in length in the other two dimensions. The biggest prokaryotes are colorless sulfur bacteria belonging to the genus *Thiomargarita* [77]. Previously known *Thiomargarita* bacteria are spherical with diameters up to 800μM. A recent paper published in science demonstrates the discovery of a bacterium that is in terms of volume $\sim$50 fold larger than the previously reported largest bacterium [78]. This gigantic bacterium has been named *Thiomargarita magnifica*.

T. magnifica separates its genetic material along with ribosomes in membrane bound compartments (Figure 8). Because the genome is highly polyploid, one cell contains many of the nucleus-like compartments. A possible explanation for the separation of the genetic material is that the cells of *T. magnifica* are so large that they encounter physical constraints on the cell volume. One of the ways *T. magnifica* deals with such constraints is by having a central vacuole filled with water that takes up about 75% of the cell volume. The vacuole confines the cytosol to the edge of the cell. This occurs commonly in other species in this genus. Nevertheless the volume of the cytoplasm of *T. magnifica* is about three orders of magnitude larger than that of the predicted maximum bacterial size based on constraints on the number of ribosomes [79]. By containing the genetic material along with ribosomes in a subspace of the cell this constraint could be

alleviated. The compartmentalization of the genetic material into membrane-bound organelles had not been observed in prokaryotes before, whereas it is the defining feature of eukaryotes.

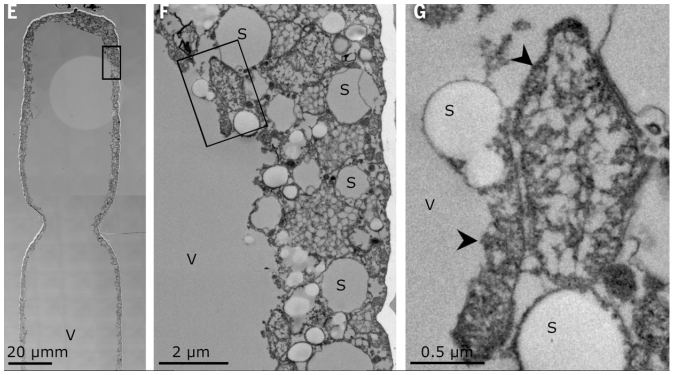


Figure 8: Transmission electron microscopy image of a *T. magnifica* cell. Figure (and labeling) from [78]. **E)** The cell of *T. magnifica* consists for a large part of a central vacuole (V), which relegates the cytosol to the periphery of the cell. **F)** Zoom in of the black rectangle in E). Sulfur granules are denoted with S. **G)** Zoom in of black rectangle in F). Black arrows denote membrane bound compartments with genetic material. Electron dense granules are ribosomes.

The case of *T. magnifica* shows that a large cell size could potentially incentivize the compartmentalization of genetic material due to constraints on the number of ribosomes. Of course, the cell of *T. magnifica* is huge even by eukaryotic standards. Whether the cells of FECA (first eukaryotic common ancestor), LECA and all their intermediates were ever as large as that of *T. magnifica* is doubtful. The relevance of ribosome constraints during eukaryogenesis is therefore hard to infer. Nevertheless we can at least posit the possibility that the nuclear envelope could have provided the means of developing the size required for cellular complexity during the initial stages of eukaryogenesis. The separation of transcription from translation and the evolution of the spliceosome could have evolved much later.

7 A prokaryotic endosymbiont in a prokaryotic host

The aspect of eukaryogenesis that captures the imagination the most is perhaps the incorporation of the γ-proteobacterial ancestor of the mitochondrion in an archeal host. Such an endosymbiotic event has happened at least once more in the history of eukaryotes when a symbiotic cyanobacterium evolved into the organelle known as the plastid [80]. Endosymbiosis is considered a hallmark eukaryotic feature because prokaryotes harbor no organelles derived from an endosymbiotic event. In nature there also exist endosymbiotic relationships in which the endosymbiont resembles a distinct organism more than an organelle. Such organisms can be distinguished from organelles by the fact that they retain some genes that are essential to the processes of transcription, translation and DNA replication [81]. Endosymbiotic relationships of this nature almost always involve a eukaryotic host with a prokaryotic endosymbiont [82].

Currently, only one exception is known. This concerns the β-proteobacterial endosymbiont *Tremblaya princeps* which lives inside the mealy bug *Planococcus citri*. Symbiosis between mealybugs and *Tremblaya* species occurs frequently, with different partners even occurring on the same plant [83]. The symbionts have important functions concerning nutrient acquisition such as the production of essential amino-acids. *T. princeps* itself harbors another bacterium belonging to the clade of γ-proteobacteria called *Moranella endobia* in its cytoplasm [84]. This phenomenon constitutes the nested endosymbiosis of a bacterium in a bacterium in a eukaryote (Figure 9).

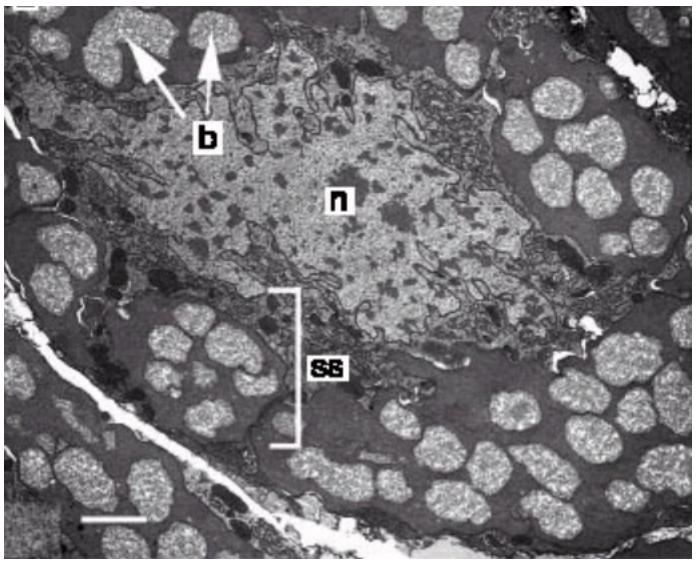


Figure 9: Transmission electron micrographs showing the structure of the endosymbiosis. The nucleus of the *P. citri* cell is denoted by n. A *T. princeps* cell is denoted by ss, which contains multiple substructures. These are denoted by b and are *M. endobia* cells. Figure from [84].

The genomes of endosymbiont organisms tend to be very reduced [85]. This is likely because of reduced recombination and inflow of new DNA through HGT in the endosymbiont. As a result the host can become dependent on an endosymbiont whose genome is slowly degrading. One possible solution to this problem is the transfer of endosymbiont DNA to the host. This has happened to the majority of mitochondrial genes. The genome of *T. princeps* is degraded to such an extent that it has the smallest known circular genome ($\sim$150kb), with a relatively low coding density (73% with 121 protein coding genes) [86]. The genome has lost many genes associated with the process of translation. For example, *T. princeps* is the only known bacterium whose genome does not code for any aminoacyl-tRNA synthetases. These proteins are responsible for connecting tRNAs with the appropriate amino-acid. Likely *T. princeps* acquires the necessary gene products from its endosymbiont, possibly through the (insect) host controlled lysis of *M. endobia* cells [87]. Similarly *T. princeps* is lacking in some of the genes that are required for the production of amino-acids which the insect host requires. Its endosymbiont also does not code for complete amino-acid synthesis pathways. However, from the union of the genomes of *T. princeps* and *M. endobia* emerge complete amino-acid synthesis pathways. In this way their genes constitute a patchwork which together is sufficient to achieve the synthesis of the amino-acids the insect host requires [88]. The genome of the insect host is also part of this patchwork. It has taken on some of the tasks of its bacterial endosymbionts in regards to nutrient production and even peptidoglycan cell wall maintenance [87]. It has obtained the necessary genes through HGT from a wide variety of bacterial families distributed through the α-proteobacteria, γ-proteobacteria and bacteroidetes, but not from its endosymbionts.

Evolution appears to select against the maintenance of redundant genes in the host-endosymbiont complex (either through neutral or adaptive evolution) which ensures that the endosymbiotic event is irreversible. In this particular endosymbiosis this results in differential loss, rather than relocation of essential genes towards host organism as has happened in the case of the mitochondrion or plastid. In fact, in the endosymbiosis of *T. princeps* and *M. endobia* the host cell has less basic cellular machinery than the endosymbiont. The difference is likely because the host in this endosymbiosis is itself an endosymbiont. Perhaps because of this unusual situation *M. endobia* appears to solve the problem of DNA degradation by replacing its endosymbiont every so often over the course of evolutionary timescales [89]. Despite the differences between the endosymbiosis between *T. princeps* and *M. endobia* and the acquisition of the mitochondrion we can make some speculations. Re-acquisition of the mitochondrion could have occurred during eukaryogenesis until enough genes required by the endosymbiont had been incorporated in the genome of the host. At this point degradation of endosymbiont DNA might no longer have been a prohibitive issue. We can also make

a very general point regarding the acquisition of the mitochondrion. *T. princeps* has no phagocytic ability, showing that phagocytosis is not required for endosymbiosis. This lends support to early mitochondrial acquisition theories, by showing that phagocytosis may not have been a prerequisite for the acquisition of the mitochondrion.

8 Discussion

The prokaryotes described in this document offer interesting parallels to the evolution of eukaryotic complexity. However, in most cases this merely increases the sample size from 1 to 2. If we want to go beyond postulating (very) speculative hypotheses and draw well substantiated conclusions about the evolution of complexity we require more data. Prokaryotic genetic diversity likely dwarfs that of Eukaryotes [90]. Nevertheless, both experimental and phylogenetic research has predominantly focused on eukaryotes. And while the sequencing of prokaryotes has started to take off with the arrival of new sequencing techniques such as metagenomics, the number of prokaryotes that has been successfully cultured and investigated under a microscope remains an absolutely tiny fraction of the full breadth of prokaryotic diversity. This implies that we are more likely to find ancestral features rather than convergent features. It seems probable that the occurrence of complex features in prokaryotes is not at all as rare as current literature seems to suggest. For instance, the presence of relatives of the phagocytic *U. amorphum* has been identified in a wide range of samples derived from marine, hypersaline and freshwater environments, and also from the stomachs of cows [33]. Further investigation might show that phagocytosis is actually quite common in the prokaryotic world. Sterol biosynthesis genes have been detected in many phylogenetically disparate bacterial genomes. Most of these bacteria are uncharacterized and have not been investigated under a microscope for interesting cellular biology. These organisms might very well harbor complex membranous features. Research into the cellular features of novel prokaryotes is lagging behind compared to research into eukaryotes and holds the potential for many great discoveries.

This document is concerned with hallmark eukaryotic features identified in prokaryotes. These features are themselves composed of building blocks such as proteins or their products. At some point the line between the level of proteins and the level of features blurs. Whether e.g. sterol biosynthesis constitutes a feature is debatable. In *G. obscuriglobus* it depends merely on the presence of two sterol biosynthesis genes and their products. Entirely separating these two levels is impossible. Proteins have evolved to fulfill their function in the greater whole of the cell, which is constituted of several features. Features in turn fundamentally depend on the proteins they emerge from and are constrained by the space of protein configurations available to the organism. It is also true that similar features often depend on similar proteins. This similarity, which can be ancestral or result from convergent evolution, shows that the solution space to problems faced by the cell is constrained. Sterol biosynthesis appears to have been invented only once and has since spread through the tree of life by means of HGT. Conversely, coat proteins and nucleoporins appear to have been invented twice, once during eukaryogenesis and once in planctomycetes. The search for similar features in large part overlaps with the search for similar proteins, and conversely finding similar proteins at the very least implies the potential for similar features. Hence phylogenetic analysis and the prediction of protein structures from the sequence can be important starting blocks for the search for convergent cellular features in prokaryotes.

Some of the prokaryotic features outlined in this document appear to have evolved entirely convergently, with no evidence of HGT from eukaryotes. Two features have evolved convergently with a single exception. In the case of the endomembrane system of *G. obscuriglobus* the sterol biosynthesis genes were possibly derived through HGT from eukaryotes, whereas the other proteins only displayed structural similarity. Phagocytosis in *U. amorphum* seems to have evolved convergently with the exception of an actin that clusters within the Asgard archeae, and is therefore likely ancestral to eukaryotic actin. Proteins which have only evolved during eukaryogenesis, but are nevertheless adaptive and allow for complexity in certain bacteria when acquired via HGT, could indicate large hurdles in the fitness landscape that were overcome during eukaryogenesis.

Many hallmark eukaryotic features appear to be present in the PVC-superphylum, and in particular the planctomycetes. The most likely explanation seems to be a combination of convergent evolution and

horizontal gene transfer. Most of the proteins involved display no sequence homology, suggesting there is no ancestral relationship between the eukaryotic and planctomycete features. Nevertheless there are proponents of the 1D theory of life who argue that all of these features could be ancestral rather than convergent [91]. In this scenario eukaryotes and archeae diverged from a common ancestor that was of bacterial origin, rather than eukaryotes originating within archeae. If this hypothesis were true, it would mean that intermediate forms of complexity that are ancestral to LECA and already possess a complex endomembrane system capable of vesicle formation, endocytosis and phagocytosis are currently extant in the form of the planctomycetes. There are however some arguments to be made against the 1D theory of life. For one, none of the eukaryotic signature proteins have homologs in the planctomycetes, except for the rare isolated presence of a homolog due to HGT. This can only be explained with massive differential loss in the planctomycetes or a great many HGT events from other bacterial clades to stem eukaryotes during a later stage of eukaryogenesis. Both of these explanations are less parsimonious compared to the explanation that the Asgard archeae, who contain many eukaryotic signature proteins, are ancestral to eukaryotes [92]. Moreover, the limited phylogenetic distribution of the features outlined in this document throughout the bacterial domain suggests this is a case of convergent evolution with some HGT from eukaryotes to bacteria. Other theories posit that LECA was in fact the result of a planctomycete that gained through two separate endosymbiosis events the nucleus from an archeum and the mitochondrion from an α-proteobacterium. This would explain the presence of eukaryotic signature proteins in the Asgard archeae, but does not account for the lack of sequence homology between eukaryotes and planctomycetes [93]. Nevertheless, possible involvement of the PVC-superphylum in eukaryogenesis cannot conclusively be disregarded. The authors of a prepublished paper claim to have found deep sea PVC members who appear to have organelles that are very similar to the endoplasmatic reticulum, golgi apparatus, nucleus and even nucleolus [94]. They claim cell division occurs in a process that is very much like mitosis. Transcriptomic analysis revealed the presence of various eukaryotic signature proteins such as Ras/Rho GTPases, SNARE proteins and various proteins associated with the golgi apparatus. This paper has not yet been peer reviewed, and whether the results will stand up to scrutiny is unclear. Regardless of the validity of the 1D theory of life, some arguments against it are based on false preconceptions about what can and cannot evolve. It has frequently been argued that bacteria cannot evolve phagocytosis because they have two membranes [95]. For this reason they would not be able to acquire an endosymbiont. The planctomycete *U. amorphum* provides an indisputable counterexample to this line of reasoning. Furthermore, *T. princeps* shows that phagocytic ability is not strictly required for the acquisition of an endosymbiont. Evolution has no regard for our preconceptions and continues to shatter assumptions about what prokaryotes can and cannot do.

Eukaryogenesis has long been regarded as a mysterious transition unparalleled in evolutionary history, due to the number of interlocking evolutionary events. This point of view has recently been nuanced due to the discovery of the Asgard archeae which harbor many eukaryotic signature proteins. Some proteins and features that were previously thought to be eukaryotic inventions were already present before eukaryogenesis. To further illuminate the mystery of eukaryogenesis we should look to other prokaryotes. Many hallmark eukaryotic features such as an endomembrane system, endocytosis, phagocytosis, compartmentalization of DNA, motor proteins and even endosymbiosis occur by themselves in prokaryotes and can be investigated in the context of these organisms. This point of view has been under-appreciated. Most likely this is in part due to the fact that it is clearer what we can learn from the Asgard archeae. They show us what proteins and features were already present at the start of eukaryogenesis. It is less obvious how we can learn about eukaryogenesis from convergently evolved features in other prokaryotes. In this literature thesis we have hoped to demonstrate how one can use the parallel with non-ancestral prokaryotes to make inferences about eukaryogenesis. While some eukaryotic features have analogs in prokaryotes, eukaryogenesis stands alone in that all of the above complexity accumulated in a single species (that of LECA). To make sense of the complicated puzzle of eukaryogenesis we need to leverage all methods available to us. Homology is only one piece of that puzzle. Analogy is the other.

References

[1] V Lila Koumandou et al. "Molecular paleontology and complexity in the last eukaryotic common ancestor". In: *Critical reviews in biochemistry and molecular biology* 48.4 (2013), pp. 373–396.

[2] Julian Vosseberg et al. "Timing the origin of eukaryotic cellular complexity with ancient duplications". In: *Nature ecology & evolution* 5.1 (2021), pp. 92–100.

[3] Katarzyna Zaremba-Niedzwiedzka et al. "Asgard archaea illuminate the origin of eukaryotic cellular complexity". In: *Nature* 541.7637 (2017), pp. 353–358.

[4] Tomoyuki Hatano et al. "Asgard archaea shed light on the evolutionary origins of the eukaryotic ubiquitin-ESCRT machinery". In: *Nature communications* 13.1 (2022), pp. 1–16.

[5] Anja Spang et al. "Complex archaea that bridge the gap between prokaryotes and eukaryotes". In: *Nature* 521.7551 (2015), pp. 173–179.

[6] Caner Akıl and Robert C Robinson. "Genomes of Asgard archaea encode profilins that regulate actin". In: *Nature* 562.7727 (2018), pp. 439–443.

[7] Randy Schekman and Lelio Orci. "Coat proteins and vesicle budding". In: *Science* 271.5255 (1996), pp. 1526–1533.

[8] Nina R Salama, JS Chuang, and RW Schekman. "Sec31 encodes an essential component of the COPII coat required for transport vesicle budding from the endoplasmic reticulum." In: *Molecular biology of the cell* 8.2 (1997), pp. 205–217.

[9] Margaret R Lindsay et al. "Cell compartmentalisation in planctomycetes: novel types of structural organisation for the bacterial cell". In: *Archives of Microbiology* 175.6 (2001), pp. 413–429.

[10] John A Fuerst. "Intracellular compartmentation in planctomycetes". In: *Annu. Rev. Microbiol.* 59 (2005), pp. 299–328.

[11] Sandrine Pinos et al. "Compartmentalization in PVC super-phylum: evolution and impact". In: *Biology direct* 11.1 (2016), pp. 1–12.

[12] Rachel Santarella-Mellwig et al. "Three-dimensional reconstruction of bacteria with a complex endomembrane system". In: *PLoS biology* 11.5 (2013), e1001565.

[13] Damien P Devos. "Re-interpretation of the evidence for the PVC cell plan supports a Gram-negative origin". In: *Antonie Van Leeuwenhoek* 105.2 (2014), pp. 271–274.

[14] Sandra Wiegand, Mareike Jogler, and Christian Jogler. "On the maverick Planctomycetes". In: *FEMS microbiology reviews* 42.6 (2018), pp. 739–760.

[15] Devrim Acehan, Rachel Santarella-Mellwig, and Damien P Devos. "A bacterial tubulovesicular network". In: *Journal of cell science* 127.2 (2014), pp. 277–280.

[16] Rachel Santarella-Mellwig et al. "The compartmentalized bacteria of the planctomycetes-verrucomicrobia-chlamydiae superphylum have membrane coat-like proteins". In: *PLoS biology* 8.1 (2010), e1000281.

[17] Evgeny Sagulenko et al. "Nuclear pore-like structures in a compartmentalized bacterium". In: *PLoS One* 12.2 (2017), e0169432.

[18] Martin Beck et al. "Nuclear pore complex structure and dynamics revealed by cryoelectron tomography". In: *Science* 306.5700 (2004), pp. 1387–1390.

[19] Ann Pearson, Meytal Budin, and Jochen J Brocks. "Phylogenetic and biochemical evidence for sterol synthesis in the bacterium Gemmata obscuriglobus". In: *Proceedings of the National Academy of Sciences* 100.26 (2003), pp. 15352–15357.

[20] Jeremy H Wei, Xinchi Yin, and Paula V Welander. "Sterol synthesis in diverse bacteria". In: *Frontiers in microbiology* 7 (2016), p. 990.

[21] David A Gold et al. "Paleoproterozoic sterol biosynthesis and the rise of oxygen". In: *Nature* 543.7645 (2017), pp. 420–423.

[22] Yosuke Hoshino and Eric A Gaucher. "Evolution of bacterial steroid biosynthesis and its impact on eukaryogenesis". In: *Proceedings of the National Academy of Sciences* 118.25 (2021), e2101276118.

[23] Elena Rivas-Marin et al. “Essentiality of sterol synthesis genes in the planctomycete bacterium Gemmata obscuriglobus”. In: *Nature communications* 10.1 (2019), pp. 1–6.

[24] Thierry GA Lonhienne et al. “Endocytosis-like protein uptake in the bacterium Gemmata obscuriglobus”. In: *Proceedings of the National Academy of Sciences* 107.29 (2010), pp. 12883–12888.

[25] Mark K Doeven, Jan Kok, and Bert Poolman. “Specificity and selectivity determinants of peptide transport in Lactococcus lactis and other microorganisms”. In: *Molecular microbiology* 57.3 (2005), pp. 640–649.

[26] Stephen G Brohawn et al. “Structural evidence for common ancestry of the nuclear pore complex and vesicle coats”. In: *Science* 322.5906 (2008), pp. 1369–1373.

[27] Damien Devos et al. “Components of coated vesicles and nuclear pore complexes share a common molecular architecture”. In: *PLoS biology* 2.12 (2004), e380.

[28] Daniel B Mills. “The origin of phagocytosis in Earth history”. In: *Interface Focus* 10.4 (2020), p. 20200019.

[29] Andrew J Roger, Sergio A Muñoz-Gómez, and Ryoma Kamikawa. “The origin and diversification of mitochondria”. In: *Current Biology* 27.21 (2017), R1177–R1192.

[30] Nick Lane and William Martin. “The energetics of genome complexity”. In: *Nature* 467.7318 (2010), pp. 929–934.

[31] Christian De Duve. “The origin of eukaryotes: a reappraisal”. In: *Nature Reviews Genetics* 8.5 (2007), pp. 395–403.

[32] Anna Karnkowska et al. “A eukaryote without a mitochondrial organelle”. In: *Current Biology* 26.10 (2016), pp. 1274–1284.

[33] Takashi Shiratori et al. “Phagocytosis-like cell engulfment by a planctomycete bacterium”. In: *Nature communications* 10.1 (2019), pp. 1–11.

[34] Michael Z Levy et al. “Telomere end-replication problem and cell aging”. In: *Journal of molecular biology* 225.4 (1992), pp. 951–960.

[35] Rieko Ohki, Toshiki Tsurimoto, and Fuyuki Ishikawa. “In vitro reconstitution of the end replication problem”. In: *Molecular and cellular biology* 21.17 (2001), pp. 5753–5766.

[36] Sherwood Casjens et al. “Linear chromosomes of Lyme disease agent spirochetes: genetic diversity and conservation of gene order”. In: *Journal of bacteriology* 177.10 (1995), pp. 2769–2780.

[37] Yi-Shing Lin et al. “The chromosomal DNA of Streptomyces lividans 66 is linear”. In: *Molecular microbiology* 10.5 (1993), pp. 923–933.

[38] Michael Kube et al. “The linear chromosome of the plant-pathogenic mycoplasma’Candidatus Phytoplasma mali’”. In: *Bmc Genomics* 9.1 (2008), pp. 1–14.

[39] Annick Allardet-Servent et al. “Presence of one linear and one circular chromosome in the Agrobacterium tumefaciens C58 genome”. In: *Journal of bacteriology* 175.24 (1993), pp. 7869–7874.

[40] Ralph Kirby. “Chromosome diversity and similarity within the Actinomycetales”. In: *FEMS microbiology letters* 319.1 (2011), pp. 1–10.

[41] Zhongjun Qin and Stanley N Cohen. “Replication at the telomeres of the Streptomyces linear plasmid pSLA2”. In: *Molecular microbiology* 28.5 (1998), pp. 893–903.

[42] Chien-Chin Yang et al. “Telomere associated primase Tap repairs truncated telomeres of Streptomyces”. In: *Nucleic acids research* 45.10 (2017), pp. 5838–5849.

[43] Poa-Chun Chang and Stanley N Cohen. “Bidirectional replication from an internal origin in a linear Streptomyces plasmid”. In: *Science* 265.5174 (1994), pp. 952–954.

[44] Margarita Salas. “Protein-priming of DNA replication”. In: *Annual review of biochemistry* 60 (1991), pp. 39–71.

[45] Carton W Chen et al. “Once the circle has been broken: dynamics and evolution of Streptomyces chromosomes”. In: *TRENDS in Genetics* 18.10 (2002), pp. 522–529.

[46] Jean-Noël Lorenzi et al. "Subtelomeres are fast-evolving regions of the Streptomyces linear chromosome". In: *Microbial genomics* 7.6 (2021).

[47] Stephen D Bentley et al. "Complete genome sequence of the model actinomycete Streptomyces coelicolor A3 (2)". In: *Nature* 417.6885 (2002), pp. 141–147.

[48] Stephanie E Jones and Marie A Elliot. "'Exploring'the regulation of Streptomyces growth and development". In: *Current opinion in microbiology* 42 (2018), pp. 25–30.

[49] Mohammed Harir et al. "Streptomyces secondary metabolites". In: *Basic biology and applications of actinobacteria* 6 (2018), pp. 99–122.

[50] A Birch, A Häusler, and R Hütter. "Genome rearrangement and genetic instability in Streptomyces spp". In: *Journal of bacteriology* 172.8 (1990), pp. 4138–4142.

[51] J-N Volff and J Altenbuchner. "Genetic instability of the Streptomyces chromosome". In: *Molecular microbiology* 27.2 (1998), pp. 239–246.

[52] Abdoul-Razak Tidjani, Cyril Bontemps, and Pierre Leblond. "Telomeric and sub-telomeric regions undergo rapid turnover within a Streptomyces population". In: *Scientific Reports* 10.1 (2020), pp. 1–10.

[53] Ji-Nu Kim et al. "Comparative genomics reveals the core and accessory genomes of Streptomyces species". In: *Journal of microbiology and biotechnology* 25.10 (2015), pp. 1599–1605.

[54] Frédéric Choulet et al. "Evolution of the terminal regions of the Streptomyces linear chromosome". In: *Molecular Biology and Evolution* 23.12 (2006), pp. 2361–2369.

[55] Zheren Zhang et al. "Antibiotic production in Streptomyces is organized by a division of labor through terminal genomic differentiation". In: *Science advances* 6.3 (2020), eaay5781.

[56] George Chaconas. "Hairpin telomeres and genome plasticity in Borrelia: all mixed up in the end". In: *Molecular microbiology* 58.3 (2005), pp. 625–635.

[57] Brad Goodner et al. "Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58". In: *Science* 294.5550 (2001), pp. 2323–2328.

[58] J Hinnebusch and AG Barbour. "Linear plasmids of Borrelia burgdorferi have a telomeric structure and sequence similar to those of a eukaryotic virus". In: *Journal of bacteriology* 173.22 (1991), pp. 7233–7239.

[59] Sherwood Casjens. "Evolution of the linear DNA replicons of the Borrelia spirochetes". In: *Current opinion in microbiology* 2.5 (1999), pp. 529–534.

[60] Kerri Kobryn and George Chaconas. "Fusion of hairpin telomeres by the B. burgdorferi telomere resolvase ResT: implications for shaping a genome in flux". In: *Molecular cell* 17.6 (2005), pp. 783–791.

[61] Ira Schwartz et al. "Multipartite genome of Lyme disease Borrelia: structure, variation and prophages". In: *Current Issues in Molecular Biology* 42.1 (2021), pp. 409–454.

[62] George Chaconas and Kerri Kobryn. "Structure, function, and evolution of linear replicons in Borrelia". In: (2010).

[63] Steven Slater et al. "Reconciliation of sequence data and updated annotation of the genome of Agrobacterium tumefaciens C58, and distribution of a linear chromosome in the genus Agrobacterium". In: *Applied and environmental microbiology* 79.4 (2013), pp. 1414–1417.

[64] Martha H Ramıérez-Bahena et al. "Single acquisition of protelomerase gave rise to speciation of a large and diverse clade within the Agrobacterium/Rhizobium supercluster characterized by the presence of a linear chromid". In: *Molecular phylogenetics and evolution* 73 (2014), pp. 202–207.

[65] Pradeep Reddy Marri et al. "The effect of chromosome geometry on genetic diversity". In: *Genetics* 179.1 (2008), pp. 511–516.

[66] Guosheng Qu et al. "Structure of a group II intron in complex with its reverse transcriptase". In: *Nature structural & molecular biology* 23.6 (2016), pp. 549–557.

[67] Peter L Graumann. "Cytoskeletal elements in bacteria". In: *Annu. Rev. Microbiol.* 61 (2007), pp. 589–618.

[68] Mark J McBride. "Bacterial Gliding Motility: Multiple". In: *Annu. Rev. Microbiol* 55 (2001), pp. 49–75.

[69] Mingzhai Sun et al. "Motor-driven intracellular transport powers bacterial gliding motility". In: *Proceedings of the National Academy of Sciences* 108.18 (2011), pp. 7559–7564.

[70] Guo Fu et al. "MotAB-like machinery drives the movement of MreB filaments during bacterial gliding motility". In: *Proceedings of the National Academy of Sciences* 115.10 (2018), pp. 2484–2489.

[71] Thomas D Pollard and John A Cooper. "Actin, a central player in cell shape and movement". In: *science* 326.5957 (2009), pp. 1208–1212.

[72] Jennifer Luciano et al. "Emergence and modular evolution of a novel motility machinery in bacteria". In: *PLoS genetics* 7.9 (2011), e1002268.

[73] Rym Agrebi et al. "An evolutionary link between capsular biogenesis and surface motility in bacteria". In: *Nature Reviews Microbiology* 13.5 (2015), pp. 318–326.

[74] Salim T Islam and Tâm Mignot. "The mysterious nature of bacterial surface (gliding) motility: a focal adhesion-based mechanism in Myxococcus xanthus". In: *Seminars in cell & developmental biology*. Vol. 46. Elsevier. 2015, pp. 143–154.

[75] Evelyne Derelle et al. "Genome analysis of the smallest free-living eukaryote Ostreococcus tauri unveils many unique features". In: *Proceedings of the National Academy of Sciences* 103.31 (2006), pp. 11647–11652.

[76] Heide N Schulz-Vogt, Esther R Angert, and Ferran Garcia-Pichel. "Giant bacteria". In: *eLS* (2007).

[77] HEIDE N Schulz. "The genus Thiomargarita". In: *Prokaryotes* 6 (2006), pp. 1156–1163.

[78] Jean-Marie Volland et al. "A centimeter-long bacterium with DNA contained in metabolically active, membrane-bound organelles". In: *Science* 376.6600 (2022), pp. 1453–1458.

[79] Christopher P Kempes et al. "Evolutionary tradeoffs in cellular composition across diverse bacteria". In: *The ISME journal* 10.9 (2016), pp. 2145–2157.

[80] Debashish Bhattacharya et al. "How do endosymbionts become organelles? Understanding early events in plastid evolution". In: *Bioessays* 29.12 (2007), pp. 1239–1246.

[81] John P McCutcheon. "The bacterial essence of tiny symbiont genomes". In: *Current opinion in microbiology* 13.1 (2010), pp. 73–78.

[82] Anne-Kathrin Dietel, Martin Kaltenpoth, and Christian Kost. "Convergent evolution in intracellular elements: plasmids as model endosymbionts". In: *Trends in microbiology* 26.9 (2018), pp. 755–768.

[83] Ryuichi Koga et al. "Mealybugs with distinct endosymbiotic systems living on the same host plant". In: *FEMS microbiology ecology* 83.1 (2013), pp. 93–100.

[84] Carol D Von Dohlen et al. "Mealybug β-proteobacterial endosymbionts contain γ-proteobacterial symbionts". In: *Nature* 412.6845 (2001), pp. 433–436.

[85] John P McCutcheon and Nancy A Moran. "Extreme genome reduction in symbiotic bacteria". In: *Nature Reviews Microbiology* 10.1 (2012), pp. 13–26.

[86] Sergio López-Madrigal et al. *Complete genome sequence of "Candidatus Tremblaya princeps" strain PCVAL, an intriguing translational machine below the living-cell status*. 2011.

[87] Filip Husnik et al. "Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis". In: *Cell* 153.7 (2013), pp. 1567–1578.

[88] John P McCutcheon and Carol D Von Dohlen. "An interdependent metabolic patchwork in the nested symbiosis of mealybugs". In: *Current biology* 21.16 (2011), pp. 1366–1372.

[89] Filip Husnik and John P McCutcheon. "Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis". In: *Proceedings of the National Academy of Sciences* 113.37 (2016), E5416–E5424.

[90] Brendan B Larsen et al. “Inordinate fondness multiplied and redistributed: the number of species on earth and the new pie of life”. In: *The Quarterly Review of Biology* 92.3 (2017), pp. 229–265.

[91] Damien P Devos. “Reconciling asgardarchaeota phylogenetic proximity to eukaryotes and planctomycetes cellular features in the evolution of life”. In: *Molecular Biology and Evolution* 38.9 (2021), pp. 3531–3542.

[92] James O McInerney et al. “Planctomycetes and eukaryotes: a case of analogy not homology”. In: *Bioessays* 33.11 (2011), pp. 810–817.

[93] Patrick Forterre. “A new fusion hypothesis for the origin of Eukarya: better than previous ones, but probably also wrong”. In: *Research in Microbiology* 162.1 (2011), pp. 77–91.

[94] Rikuan Zheng et al. “Cultured deep-sea PVC bacteria shed light on eukaryogenesis”. In: *bioRxiv* (2021).

[95] Thomas Cavalier-Smith. “Origin of the cell nucleus, mitosis and sex: roles of intracellular coevolution”. In: *Biology direct* 5.1 (2010), pp. 1–78.