# The Advent of the Eukaryotic Cell

Edward Schwieterman, University of Washington, Astrobiology Program

## 1. ABSTRACT & INTRODUCTION

The origin of the modern eukaryotic cell has long been a source of biological investigation and speculation. For many decades the standard theory for the origin of the mitochondria in eukaryotes assumed an endosymbiosis of a bacterium with an amitochondriate proto-eukaryotic host, which in most models contained a nucleus, but was at a minimum capable of phagocytosising the symbiont (Sagan 1967; see review in Martin et al. 2001). The origin of the proto-eukaryote was sometimes suggested to be the result of an earlier fusion between an archaea and bacteria, and was often taken as an initial assumption in the theory of the origin of the mitochondria. These views have been repeatedly challenged in the last fifteen years by many authors who together have proposed a cornucopia of hypotheses regarding the origin of eucarya (Pool and Penny 2006; Embley and Martin 2006, Martin 2001). This has largely been the result of findings that all once-deemed 'early-branching' eukaryotes such as microsporidia, which possess no mitochondria or hydrogenosomes, have traces of α-proteobacterial DNA in their genomes and a small heavily derived mitosome (Williams et al. 2002), and therefore do not represent a direct descendent of an amitochondriate eukaryotic common ancestor. This author describes theories for the advent of the eukaryotic cell grouped into three categories described below: (proto-)eukaryogenesis followed by mitochondrial symbiosis, syntrophy followed by eukaryosis, and viral eukaryogensis. This author finds Cavalier-Smith (2002) model of a neoruman bacterial progenitor of archael and eukaryotic sister domains and the vesicle syntrophic hydrogen-hypothesis first proposed by Martin and Muller (1998) most convincing, but all theories elicit significant and legitimate objections. A justification for the astrobiological relevance of this topic can be found in the attached Appendix A.

## 2. THEORIES OF EUKARYOGENESIS

All life shares the same genetic code and a common mechanism for protein synthesis and thus shares a common ancestor. Extant life can be divided into three domains, Bacteria,

Archaea, and Eucarya based on differences in the ribosome RNA (Woese 1990) and each domain differs substantially in its informational processes such as transcription, translation and replication (Woese 2000). We will begin our discussion of the theories concerning eukaryote origins by first describing the basic characteristics that such theories must explain, such as the existence of the nucleus, the cytoskeletal and membrane systems, and specialized organelles. A functional theory of eukaryogenesis must also explain the chimerical nature of the eukaryotic organism and its genome explained below. Theories must differentiate between the acquisitions of characteristics through endosymbiotic processes, lateral gene transfer (LTG), and *de novo* evolutionary innovation to create a plausible developmental history.

This author takes for granted that plastids in algae and protozoa originated from endosymbiosis of cyanobacteria with fully developed eukaryotic primary cells. This theory is no longer seriously contested and the evidence and observation that validates it are overwhelming (Martin 2001; Rivera and Lake 2004). This author does not address thermoreductive theories for the origin of bacteria and archaea from a eukaryote-like common ancestor (Forterre 1995), as abundant paleontological evidence indicates prokaryotes predated eukaryotes (Cavalier-Smith 2002).

**2.1 Characteristics of eukaryotes -** The most salient defining characteristic of eukaryotes is the presence of a nucleus. The nucleus contains the chromatin and is surrounded by an apparent pair of membranes, which in turn enclose a lumen that is connected to the endoplasmic reticulum (Kimball 2011). The inner layer of the nuclear envelope is lined with fibrous cytoskeletal proteins while the outer layer may be linked to the ribosomes (Staley et al. 2007). The nucleus serves to separate and compartmentalize DNA transcription and replication processes from the protein synthesis and metabolic processes occurring in the cytoplasm.

The cytoskeleton is an array of microtubules, actin microfilaments, fibrous filaments, and fine filaments that provide structural support for the cell membrane, interlace the cytoplasm and, in some eukaryotes, fix the flagellum in place (Staley et al. 2007). The

cytoskeleton is involved in many cellular processes, most notable in our case is the ability to capture food particles (i.e., phagocytosis).

Mitochondria are membrane-bound organelles responsible for ATP synthesis in aerobic species and it almost all cases contain their own genome (Gray, Lang, and Burger 2004). Organisms once thought to be archezoans have been found to contain mitochondrial-related organelles, hydrogenosomes or mitosomes (Embley and Martin, 2006). Despite a substantially different metabolism hydrogenosomes and mitochondria share several proteins including ferrodoxin and succinate thiokinase, which show clear affinity to mitochondrial homologs in phylogenic analyses (Martin et al., 2001). Mitosomes, in contrast, do not partake in ATP synthesis, but probably take part in Fe-S cluster construction (as do yeast mitochondria), which carry out essential functions for life (Lill and Muhlenoff 2005). They are also homologous with mitochondria (Embly and Martin 2006). Figure 1B in the Appendix illustrates the enzymes and pathways found in the different forms of the mitochondria.

**2.2 Chimera of two domains -** It widely known that ribosomal RNA phylogenic analyses of eukaryotes root them with archaeal lineages. It has also been clear for some time that archaea have different lipids (glycerol isoprene ethers) than both bacteria and eukaryotes, which both possess glycerol fatty acid esters. Additionally, many eukaryotic enzymes facilitating ATP production are more similar to bacterial than archaeal homologs (Hensel et al. 1989; Martin et al. 2001). The latter datum is most assuredly related to the descent of mitochondria from α-proteobacteria-like ancestors, which is demonstrated by phylogenic analysis of mitochondrial DNA (Poole and Penny 2007; Pisani 2007). Of course, it is important not to view eucarya as *simply* as a fusion of two domains as there are unique biochemical properties such as essential proteins in eukaryotes that have no homologues in bacteria or archaea (Forterre and Philippe 1999).

**2.3 Models for Eukaryogensis -** Figure 1 below broadly displays most of the widely supported hypothesis for the advent of the eukaryotic domain. The hypotheses in the figure can be split into two main categories – development of an amitochondriate cell first that later subsumes a bacterial symbiont (Fig. 1 a-d) and models that propose a

syntrophic relationship between an archaea and bacteria first, eventually forming a single organism, which then develops eukaryotic-specific features (Fig 2. e-g).

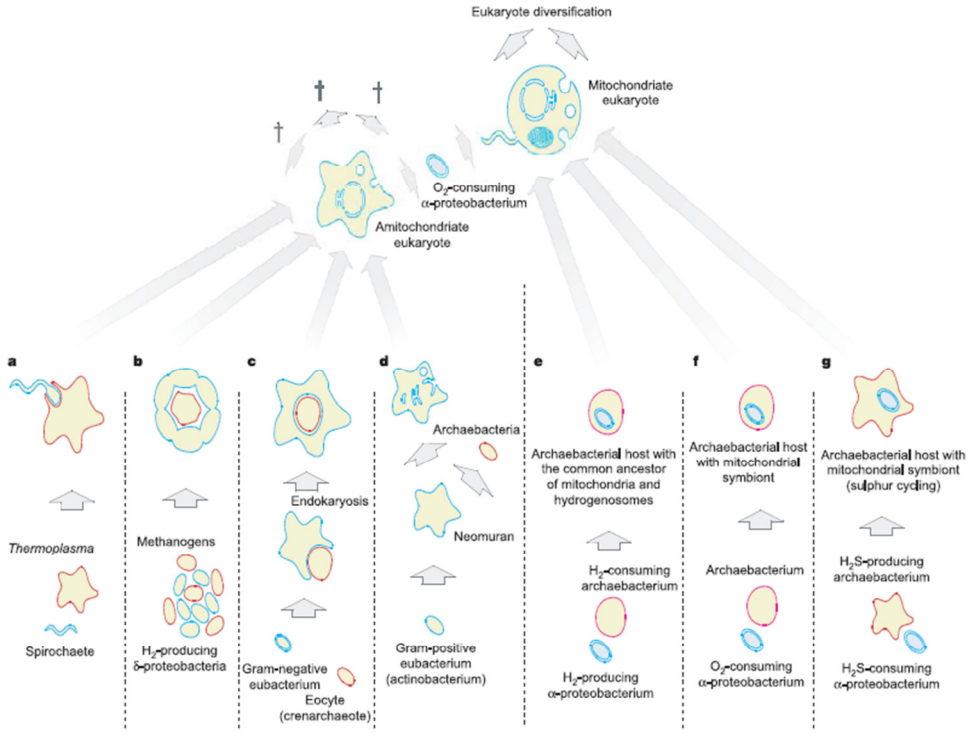


**Fig 1. –** A sample of models for eukaryote origins from Embley and Martin (2006). Models a-d represent development an amitochondriate cell that subsequently acquires an aerobic bacterium as a symbiont. Models e-g originate with the syntrophic fusion of achaeal and bacterial cells forming mitochondria-containing organism before developing other eukaryotic-specific features. Archaeal lipid membranes are colored red and bacterial membranes are colored blue. Citations for papers from which this schematic is adapted can be found in the references with a corresponding superscript, e.g., [a, b, c].

**Eukaryogensis followed by endosymbiosis (a-c) -** Models (a-c) all involve both a bacterial and archaeal partner. In model (a) an archaeabacterium such as *thermoplasma* merges with a spirochete. This amitochondriate cell evolves a nucleus and forms a symbiotic relationship with an aerobic α-proteobacterium (Margulis et al. 2000, 2005). In model (b) a syntrophic relationship develops between methanogens and δ-proteobacteria, merging several cells into one organism (Moreiera and Lopez-Garcia 1998). Model (c) involves eukaryosis by proposing an archaean entered a bacterial cell to form the rudiments of a proto-nucleus and is known as the eocyte hypothesis (Lake et al. 1984). This model proposes that eukaryotic informational genes can be traced back to the archaeabacterial eocytes (a.k.a. Crenarchaeota). While analyses of a eukaryotic

information gene recovers an eocyte tree (Rivera and Lake, 1992), different eukaryotic genes yield different trees trees (Embly and Martin, 2006). All of these models suffer from a lack of genomic evidence of multiple fusions (Poole and Penny 2007) and a lack of extant intermediate forms.

**A prokaryotic cell engulfs the mitochondrion ancestor (e.g., d) -** Cavalier-Smith (2002) has articulated a scenario in which eukaryotes and archaea form a clade dubbed neomura and are sister taxa, therefore rejecting any theories of eukaryogenesis by merging of archaea and α-proteobacteria. In this model, archaeal characteristic in eukaryotes are a result of presence in a common ancestor and eukaryotic-specific features were developed synchronously and synergistically with symbiogenesis of the mitochondria from α-proteobacteria. The nuclear membrane is proposed to develop from the invagination of the plasma membrane of the prokaryote host, losing its cell wall and evolving phagocytosis. The ribosomes that were initially attached to the plasma membrane become internalized and give rise to a proto-ER and subsequently a nuclear envelope (Martin 2006).

**Syntrophy followed by eukaryotic diversification (e-g) -** Of models (e-g), the most widely cited is the vesicle hydrogen-theory of Martin and Mueller (1998). The vesicle model posits a symbiosis between a $H_2$-dependent methanogenic archaea (the host) and a free-living, $H_2$ and $CO_2$ producing facultatively anaerobic bacteria (the symbiont). The two partners meet in a location where both $H_2$ and $CO_2$ are plentiful, so the host is viable initially. Once the pair is removed from this initial location by some mechanism, the host becomes strictly dependent on the heterotrophic bacterium for the $H_2$ it requires to metabolize. There is intense selective pressure for host cell sizes and shapes that encase the symbiont, increasing contact and transfer of metabolic products (Martin et al. 2001). The bacterium requires contact with the environment to power its own metabolism upon which the host relies. It is suggested that the methanogenic host gains the symbiont's genes for carbon importation (and others) through a lateral gene transfer, which may have been more favorable two or three billion years ago when archaea and bacteria were more closely related. Subsequent transfer and expression of bacterial genes for lipid synthesis to the archaea's chorosomes in the cystol produced an initially simple array of

cystolic vesicles, which later become more complex. This primitive system develops into the endoplasmic reticulum and later the nucleus (Martin 2006). Motivations observations for this theory include contemporary methanogens strictly dependent on the $H_2$ that is produced by hydrogenosomes (in eukaryotes) and/or fermenting bacteria (Embley et al., 1995) and the observations that there is a very limited variety of energy metabolisms in eukaryotes compared to the multifarious ways prokaryotes can generate ATP (Martin and Muller 1998). Models (f) and (g) are variations on this theme, but begin with an obligately $O_2$-consuming (Vellai et al. 1998) or $H_2S$-consuming (Searcy 1992) α-proteobacteria.

**Viral Eukaryogensis –** The viral eukaryogenic model is an intriguing twist on the nuclear endosymbiotic theory where the progenitor of the nucleus was in fact a large DNA virus, perhaps distantly related to the famous *Mimivirus* (Suzan-Monti, Scola, and Raoult 2005). The origin of uniquely Eukaryotic features such as mRNA capping, telomeres, phagocytosis, the cytoskeleton, and even the sexual cycle are then suggested to originate in the viral, not cellular world (Bell 2001). A table adapted from Bell (2001) in Appendix table 1C compares features present in different viral lineages to features in the three domains of life. This scenario is broadly compatible with syntrophic theories ofthe origin on the mitochondrion (see Appendix Figure 2C).

## 3. DISCUSSION

A plausible theory for the origin of eukaryotes must explain similarities and differences between eukaryotes and the other two domains of life. Poole and Penny (2007) argue that a viable hypothesis for eukaryotic origins should be based on known processes that are currently observed. However, it appears the transition to eukaryotic life from prokaryotic life did indeed occur only once (Martin 2006), and thus must rely on a rare event likely enough to happen once, but not likely enough to occur subsequently (otherwise we would expect similar biological innovations to be prolific as is the case with endosymbiogensis of plastids). At the very least a plausible argument must be made to explain why the advent of eukaryotic life prevented further similar innovation from proceeding. A common example presented by proponents of the vesicle hydrogen theory, when challenged with the lack of bacterial symbionts in contemporary archaea, is the case of a

γ-proteobacterium residing inside a β-bacterium (von Dohnlen, et al. 2001) demonstrating that endosymbiotic relationships in *de facto* prokaryotes is possible. Opponents of the hydrogen theory may retort that this is a rare, isolated example of which no other analogs are known. This author, however, advances that this is an insightful example of a rare event lacking prolific analogs that has clearly proven to be advantageous in the long term, exactly the kind of scenario that may have led to the genesis of the first eukaryote.

**New Findings –** Pasani, Cotton, and McInery (2007) have tested the hypothesis for eukaryotic origins by using a supertree-based phylogenic signal stripping method based on over 5,700 gene families distributed across 185 genomes, much larger than previous investigations (Dagan and Martin 2006). They find three primary contributors to eukaryotic genomes, in order of phylogenic signal strength: cyanobacteria, α-proteobacteria, and thermoplasmatales (see Figure 2D in the Appendix). The cyanobacteria contributions originate from endosymbiotic plastids. Contributions from other groups are proposed to originate from lateral gene transfer. The authors conclude this evidence confirms there are two primary domains of life – archaea and bacteria with a derived domain, eucarya. They reject all origin scenarios other than the cases of a $H_2S$-consuming or $H_2$-producing proteobacterium symbiosing with a thermoplasmatales-like archaea. However, one could argue that the neoruman/invagination model of Cavalier-Smith (2002) is not ruled out, as thermoplasmatales may simply contain more of the original genes of the bacterial progenitor of archaea. Further elaboration in Appendix D.

**The Case of Planctomycetes –** The planctomyctes provide a glaring example of a bacterial family that contains internal compartmentalization and organelles, features once believed to be only the providence of eukaryotes (Fuerst et al., 2005). Some members of planctomycetes such as *Blastopirellula marina* contain a basic membrane bound nuceloid. See Appendix E, figures 1-3. The planctcomycetes are also capable of rudimentary endocytosis and contain simple sterols. The existence of planctomycetes may imply either 1) the basic biochemical machinery for compartmentalization existed before the last common ancestor of the eukaryotes and the bacteria, but only provided a lasting advantage to eukaryotes and the planctomycetes (and perhaps undiscovered extant groups) or 2) *de novo* origination of intracellular membranes through invagination or a

related process is not the frozen accident it appeared to have been in eukaryotes. Both scenarios provide rationale for continuing to investigate the Cavalier-Smith model (2002) in the face of strong support of current phylogenies for the hydrogen theory.

Additional supporting evidence for the Cavalier-Smith model is the existence of tublin in four species of the verrucomicrobia phylum of bacteria (Jenkins et al. 2003), which are genetically linked to the planctomycetes. It appears that the tubulin in the verrucombicrobia is a recent addition from a lateral genetic transfer from a eukaryote (Staley et al. 2005). This demonstrates a propensity for lateral gene transfer between this group (which exhibit marked eukaryote-like qualities) and eukaryotes continuing nearly until the modern era. It may be that the bacterial supergroup that contains the planctomycetes and the verrucomicrobia is in fact a sister lineage to the last common ancestor of eucarya (with the eucaryal group inheriting more archaeal genesfrom their common ancestor). Of course, the existence of planctomycetes is not completely incompatible with the vesicle hydrogen theory, as noted above, if the innovation of cell compartmentalization occurred twice *de novo* or if the methanogenic-α-proteobacterial association received a lateral genetic transfer from a planctomycetes-related group in some intermediate stage of their association.

## 4. CONCLUSION

Disentangling the origins of the eukaryotic domain is a complex problem requiring at minimum, one unlikely or rare event in all currently proposed hypotheses. Every model described suffers from an apparent lack of intermediate forms, except perhaps the Cavalier-Smith model. All models contain some unique explanatory power; however considering recent rigorous phylogenic analyses and the discovery of potential bacterial sister lineages to the first proto-eukaryote, the only remaining hypotheses with a strong preponderance of evidence are the neoruman/Cavalier-Smith model of a bacterial ancestor that spawned archaea and eucarya as sister lineages or the theories consisting of an $H_2$- or $H_2S$-consuming α-proteobacterium symbiosing with an archaea. Viral transmission of genetic information may have played a role in nuclear or genomic development, but arguments suggesting the nuclear envelope, ER and associated structures developed as a direct result are unconvincing.

## REFERENCES

Bell, P.J. 2001. Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? Journal of molecular evolution 53, 251-6. 2.

Cavalier-Smith T: The origin of eukaryote and archaebacterial cells. Ann NY Acad Sci 1987, 503:17-54

[d]Cavalier-Smith, T. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. Int. J. Syst. Evol. Microbiol. 52, 297–-354 (2002).

Embley, T.M., Finlay, B.J., Dyal, P.L., Hirt, R.P., Wilkinson, M., and Williams, A.G. (1995). Multiple origins of anaerobic ciliates with hydrogenosomes within the radiation of aerobic ciliates. Proc. R. Soc. Lond. B 262, 87 – 93.

Embley TM and Martin W. Eukaryotic evolution, changes and challenges. Nature. vol 440 p. 623-30 (2006)

Forterre, P. (1995). Thermoreduction, a hypothesis for the origin of prokaryotes. CR Acad. Sci. III 318, 415 – 422.
Forterre, P., and Philippe, H. (1999). Where is the root or the universal tree of life? Bioessays 21, 871 – 879.

Fuerst, J. A. Intracellular Compartmentation in Planctomycetes. Annu Rev Microbiol, Vol. 59 (2005), pp. 299-328

Pisani D, Cotton JA, McInerney JO. Supertrees disentangle the chimerical origin of eukaryotic genomes. Mol Biol Evol. 2007 Aug;24(8):1752-60.

Gray, M. W., Lang, B. F. & Burger, G. Mitochondria of protists. Annu. Rev. Genet. 38, 477–-524 (2004).

Henze, K., Schnarrenberger, C., and Martin, W. (2001). Endosymbiotic gene transfer: a special case of horizontal gene transfer germane to endosymbiosis, the origins of organelles and the origins of eukaryotes. In: Horizontal Gene Transfer, M. Syvanen and C. Kado, eds. (London, UK: Academic Press), pp. 343 – 352.

Jenkins, C., Samudrala, R., Anderson, I., Hedlund, B.P., Petroni,G., Michailova, N., Pinel, N., Overbeek, T., Rosati, G. and Staley, J.T. (2002) Genes for the cytoskeletal protein tubulin in the bacterial genus Prosthecobacter. Proc. Natl. Acad. Sci. USA 99, 17049–17054.

Kimball, John W., ed. Kimball's Biology Pages. N.p., 2011. Web. December 2011

Lake JA, Henderson E, Oakes M, Clark MW (1984) Eocytes: A new ribosome structure indicates a kingdom with a close relationship to eukaryotes. Proc Natl Acad Sci USA 81:3786 –3790

[c]Lake, J., Moore, J., Simonson, A. & Rivera, M. in Microbial Phylogeny and Evolution Concepts and Controversies (ed. Sapp, J.) 184–-206 (Oxford Univ. Press, Oxford, 2005).

Lill, R. & Muhlenhoff, U. Iron–-sulfur-protein biogenesis in eukaryotes. Trends Biochem. Sci. 30, 133–-141 (2005).

[a]Margulis, L., Dolan, M. F. & Whiteside, J. H. “Imperfections and oddities” in the origin of the nucleus. Paleobiology 31, 175–-191 (2005).

[e]Martin, W., and Müller, M. (1998). The hydrogen hypothesis for the first eukaryote. Nature 392, 37 – 41.

Martin, W., Hoffmeister, M., Rotte, C., and Henze, R. An overview of endosymbiotic models for the origins of eukaryotes, their ATP-producing organelles (mitochondria and hydrogenosomes), and their heterotrophic lifestyle. Biol. Chem., 382, 1521-1539 (2001).

Martin W. Archaebacteria (Archaea) and the origin of the eukaryotic nucleus. Curr Opin Microbiol. 2005 Dec;8(6):630-7. Epub 2005 Oct 20. Review.

[b]Moreira, D. & Lopez Garcia, P. Symbiosis between methanogenic archaea and d-proteobacteria as the origin of eukaryotes: The syntrophic hypothesis. J. Mol. Evol. 47, 517–-530 (1998).
Muller, M. The hydrogenosome. J. Gen. Microbiol. 139, 2879–-2889 (1993).
Poole, A. & Penny, D. Evaluating hypotheses for the origin of eukaryotes. Bioessays 29, 74–84 (2007)

Rivera, M. C. & Lake, J. A. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. Science 257, 74–-76 (1992).

Sagan, L. (1967). On the origin of mitosing cells. J. Theoret. Biol. 14, 225 – 274.

[g]Searcy, D. G. in The Origin and Evolution of the Cell (eds Matsuno, H. H. & Matsuno, K.) 47–-78 (World Scientific, Singapore, 1992).

Staley JT, Bouzek H, Jenkins C. 2005. Eukaryotic signature proteins of Prosthecobacter dejongeii and Gemmata sp Wa-1 as revealed by in silico analysis. FEMS Microbiol Lett. 243:9-14.

Staley, Gunsalas, Lory, and Perry. Mirobial Life. MA: Sinauer Associates, Inc. pgs. 691-696. Print.

Suzan-Monti M, La Scola B, Raoult D (2006) Genomic and evolutionary aspects of Mimivirus. Virus Res 117(1):145-155

[f]Vellai, T., Takacs, K. & Vida, G. A new aspect to the origin and evolution of eukaryotes. J. Mol. Evol. 46, 499–-507 (1998).

Williams, B. A., Hirt, R. P., Lucocq, J. M. & Embley, T. M. A mitochondrial remnant in the microsporidian Trachipleistophora hominis. Nature 418, 865–-869 (2002).

Woese, C. R., Kandler, O. & Wheelis, M. L. (1990) Proc. Natl. Acad. Sci. USA 87, 4576–4579

Woese, C. R., Olsen, G. J., Ibba, M. & Soll, D. (2000) Microbiol. Mol. Biol. Rev. 64, 202–236

Woese, Carl. On the evolution of Cells. PNAS Vol. 99 (2002) 13:8742-8747

## APPENDIX A

### Astrobiological Relevance

Eucarya is the only domain known to possess the non-filamentous multicellularity and cellular diversification, and is therefore capable of structural complexity unknown in the other two domains. At least half of the biomass on the Earth's land surface and ocean volume is present in eukaryotic organisms (Whitman, Coleman, and Wiebe 1998). Eukaryotes, with the assistance of cyanobacteria-derived plastids, undertake the majority of primary production on the planet today, mostly in the form of land-based green vascular plants (Kiang et al. a). These photosynthetic organisms are prolific enough on Earth to affect the planet's disk-averaged reflectance spectrum, i.e. create the 'red edge' (Sagan 1993) and thus similar organisms on other habitable worlds may provide a discerning biosignature (see Kiang et al. b for a review). Finally, the structural elaboration, genomic coding efficiency, and metabolic capacity found only in eukaryotes has allowed members of this domain to become large, metabolically active, and complex enough to possess simple to advanced cognitive capabilities. It is in the interest of astrobiologists to study the past history of life on Earth, especially that life which is most visibly obvious on the Earth today, to assess the plausible characteristics of life on other habitable worlds with similar initial conditions.

## APPENDIX A REFERENCES

Kiang, N.Y., Siefert, J., Govindjee, and Blakenship, R.E. (2007) Spectral signatures of photosynthesis. I. Review of Earth organisms. *Astrobiology* 7(1), 222-251

Kiang, N.Y., Segura, A., Tinetti, G., Govindjee, Blackenship, R.E., Cohen, M., Siefert, J., Crisp, D., Meadows, V. (2007) Spectral signatures of photosynthesis. II. Coevolution with other stars and the atmosphere on extrasolar worlds. *Astrobiology* 7(1), 252-274

C. Sagan, W. R. Thompson, R. Carlson, D. Gurnett, C. Hord (1993). "A search for life on Earth from the Galileo spacecraft".Nature 365 (6448): 715–721

Whitman, Coleman, and Wiebe, Prokaryotes: The unseen majority, Proc. Natl. Acad. Sci. USA, Vol. 95, pp. 6578–6583, June 1998

# APPENDIX B – MITOCHONDRIAL FORMS AND THEIR BIOCHEMISTRY

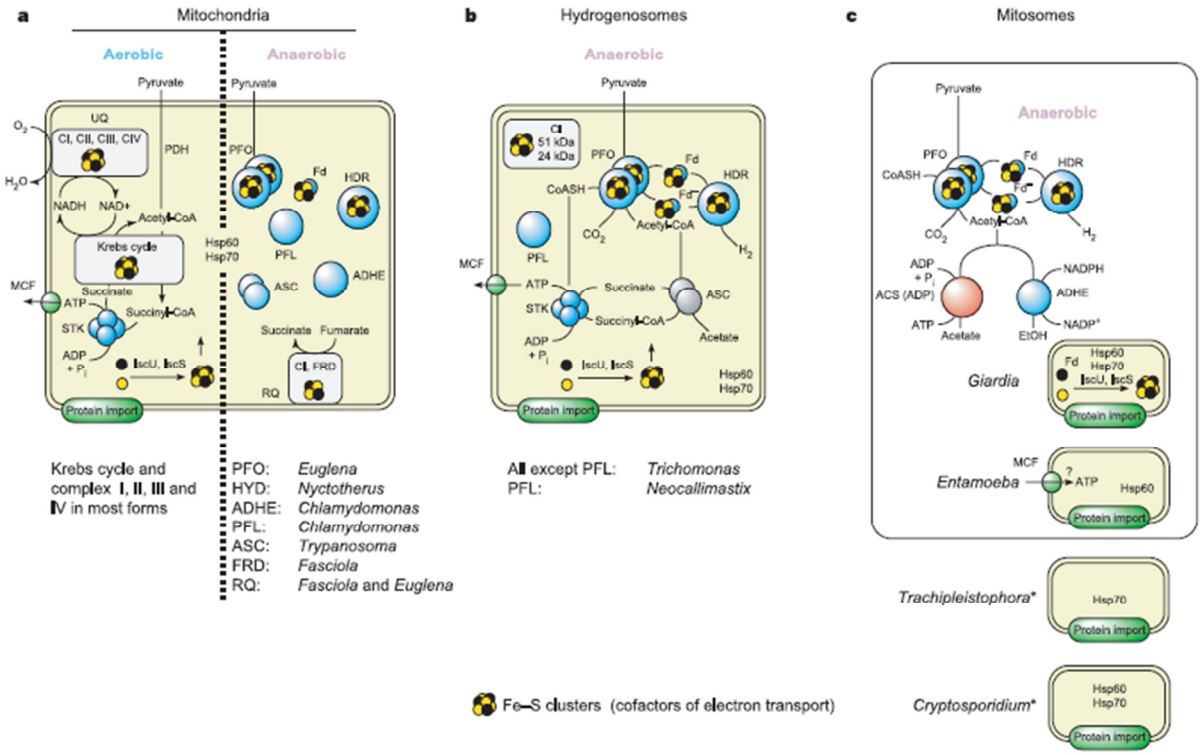


**Fig 1B –** From Embley and Martin (2006) and references therein. Proteins with more sequence similarity to bacteria are in blue, to archaea in red, and those proteins with no clear similarity to either of those two domains in green. Schematics a, b, and c provide a summary of the biochemical activity in mitochondria, hydrogenosomes, and mitosomes, respectively.

## APPENDIX C – VIRAL EUKARYOGENESIS

**Table 1C** – Comparison between selected viruses and cellular domains. Adapted from Bell 2001.

| | Bacteria | Archaea | Eukayote Nucleus | Poxviruses | ASF Virus | T4 bacteriophage |
|---|---|---|---|---|---|---|
| Capping of mRNA transcripts | No | No | Yes | Yes | yes | No |
| mRNA transport | No | No | Nucleus to cytoplasm | Viron to Cytoplasm | ? | No |
| Colocation of genome and ribosomes | Yes | Yes | No | No | No | No |
| Genome compartment | No | No | Membrane-bound nucleus | Membrane-bound capsid | Membrane-bound capsid | Icosahederal protein capsid |
| Chromosome ends | None | None | Short tandem repeates | Short tandem repeats | Short Tandem repeats | Circular permutation |
| Genome topology | Circular | Circular | Multiple linear | Linear | Linear | Linear |
| Cytoplasmic replication | No | No | Yes | Yes | Yes | Yes |

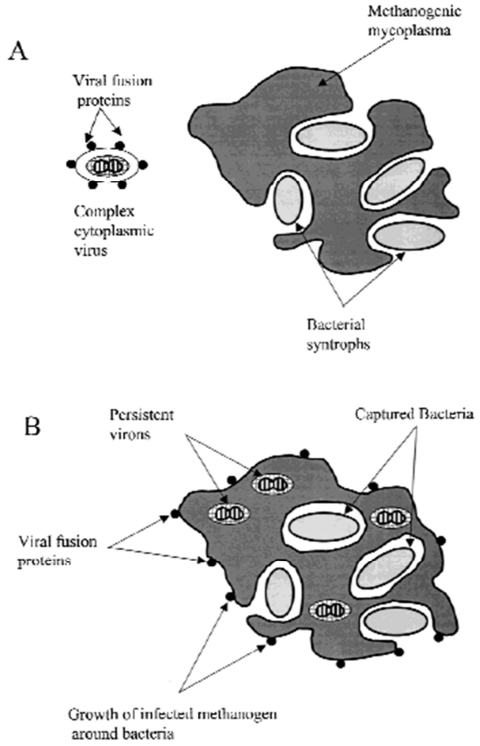


**Figure 1C** –Schematic of viral eukaryogensis in an intermediate stage of bacterial-archaea syntrophy . In (A) a filamentous microplasma has achieved a syntropic relationship with a $H_2$ producing bacterium. The relationship is maintained by an exchange of metabolic ingredients. In (B) infection of large DNA virus leads to persistent virons. Viral fusion proteins expressed in cell membrane allow complete infusion of bacterial symbionts (Bell 2001).

## APPENDIX D – SUPERTREE PHYLOGENIC ANALYSIS (Pasani et al. 2007)

**Table 1D –** Tabular results from Pasani et al. 2007 (Table 2 in paper) showing proportion of shared genes between below groups and eukaryotes.

**Eukaryotic (single) gene families with a prokaryotic homologue and the sister group relationships they imply.**

| | 126 species data set | | | 140 species data set | | | Averages | | |
|---|---|---|---|---|---|---|---|---|---|
| Eukaryotic outgroup | Number of genes | Normalised Number of genes | Proportion (%) | Number of genes | Normalised Number of genes | Proportion (%) | Number of genes | Normalised Number of genes | Proportion (%) |
| Cyanobacteria | 81 | 39.32* | 20.98 | 63.5 | 30.82* | 15.30 | 72.25 | 35.07* | 18.03 |
| α-Proteobacteria | 41.5 | 7.89* | 10.75 | 35.5 | 6.33* | 8.55 | 38.5 | 7.08* | 9.61 |
| γ-Proteobacteria | 28.5 | 4.01 | 7.38 | 26 | 3.60 | 6.26 | 27.25 | 3.80 | 6.80 |
| β-Proteobacteria | 7 | 2.46 | 1.81 | 9.5 | 3.53 | 2.28 | 8.25 | 2.98 | 2.05 |
| δ-Proteobacteria | 7 | 5.08 | 1.81 | 6.5 | 4.71 | 1.56 | 6.75 | 4.89 | 1.68 |
| ϵ-Proteobacteria | 2.5 | 4.58 | 0.64 | 2.5 | 4.58 | 0.6 | 2.5 | 4.58 | 0.62 |
| Undetermined Proteobacteria | 18 | 1.05 | 4.66 | 26.5 | 1.52 | 6.38 | 22.25 | 1.28 | 5.55 |
| Archaebacteria | 53.5 | 10.37* | 13.86 | 40 | 7.75* | 9.63 | 46.75 | 9.06* | 11.6 |
| Spirochetes | 4.5 | 4.93 | 1.16 | 5.5 | 6.02 | 1.32 | 5 | 5.47 | 1.24 |
| Actinobacteria | 12 | 4.5 | 3.1 | 14 | 5.25 | 3.37 | 13 | 4.87 | 3.24 |
| Other Eubacteria | 41.5 | 5.42 | 10.75 | 33.5 | 4.09 | 8.07 | 37.5 | 4.73 | 9.36 |
| Unclear Support | 89 | 2.50 | 23.05 | 152 | 3.61 | 36.62 | 120.5 | 3.1 | 30.08 |

Proportions are calculated, for each data set, over the total number of single gene families with a prokaryotic homologue. Raw gene numbers were normalised dividing, for each group, the number of genes that originated from it by the total number of genes in the genomes of the considered group.

* indicates values that are above the third percentile and thus significantly different from the median of the considered sample. The results presented above are based on the optimal ML trees. Using 70% bootstrap consensus trees did not significantly change these results. Exclusion of trees derived from alignments scoring compositional heterogeneous sequences did not change our results significantly. The majority of eukaryotic genes with a prokaryotic homologue were still of: (1) Cyanobacterial, (2) Proteobacterial and (3) Archaebacterial origin. However, interestingly, the exclusion of trees obtained from compositional heterogeneous alignments resulted in a marked decrease of the genes with an Actinobacterial homologue, and in the disappearance of those with Spirochaetes, δ-, and ϵ-proteobacterial homologues, suggesting these groupings were probably artifactual.

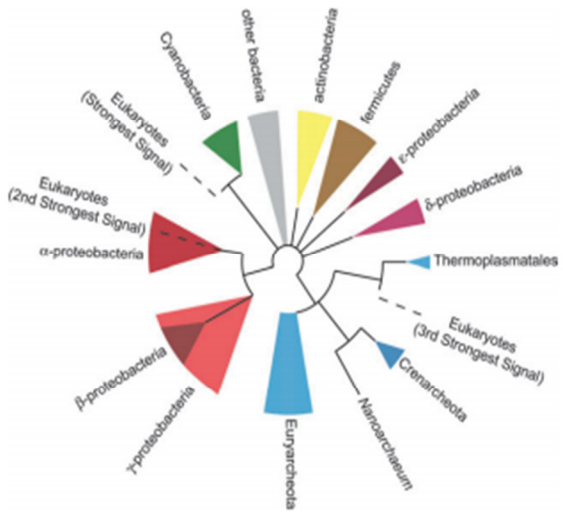


**Fig 2** – From Pasani et al. 2007, Visual illustration of phylogenic relationships to eukaryotes based on supertree phylogenic analysis.

# APPENDIX E – PLANCTOMYCETES CELL PLAN AND IMAGES (Fuerst 2005)

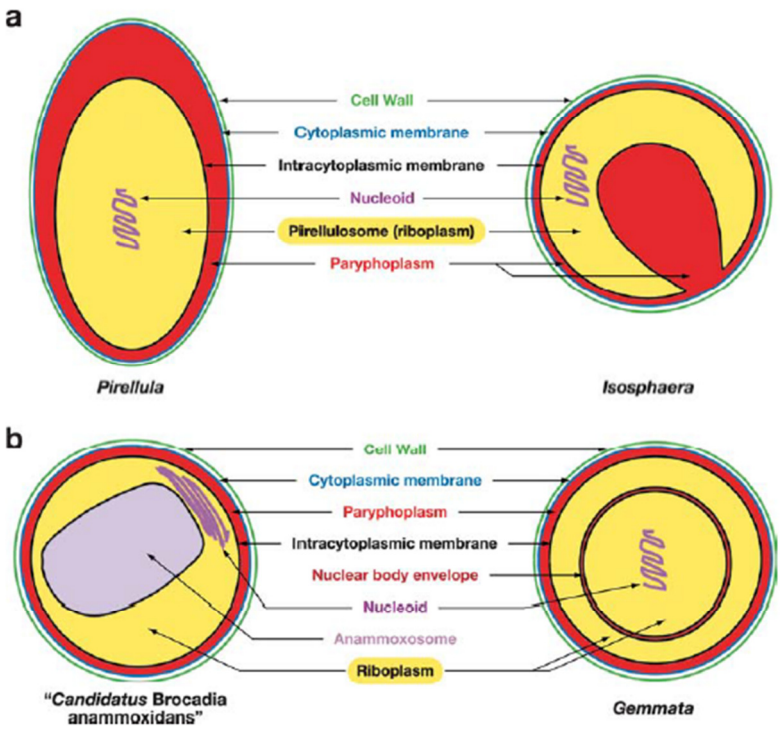


**Fig. 1E –** From Fuerst (2005). Planctomycetes cell organization and compartmentalization in four selected species labeled above.

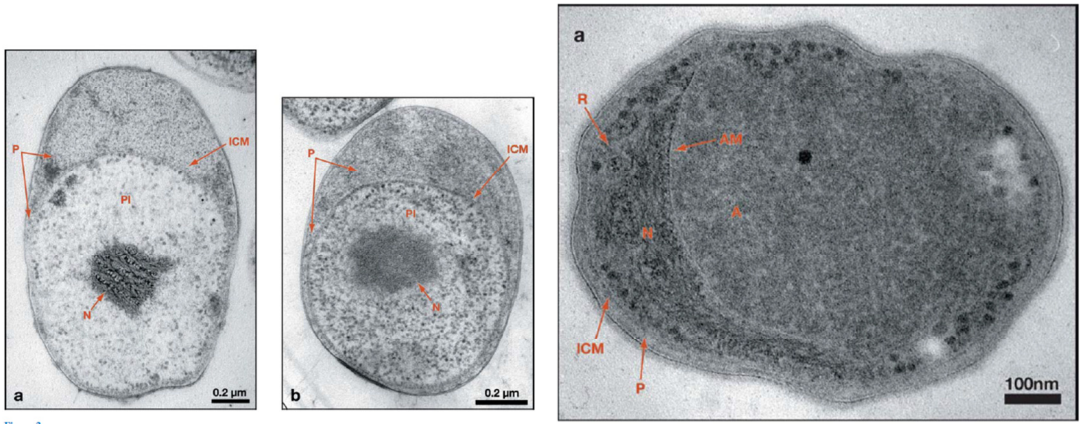


**Fig. 2E** – Transimission electron micrograph of thin sections of cryosubstituted cells of the follow species (from left to right) *Blastopirellula marina, Pirellula staleyi, and* "*Candidatus* Brocadia anammoxidans." (N) = nucleoid, (NE)=nuclear envelope, (CM)=cytoplastic membrane, (ICM)=intracytoplastic membrane (W)=cell wall, (AM)=single membrane, (A)=annamoxosome, (R)=ribosome-like particles, and (P)=paryphoplasm. From Fuerst 2005.