

A New Hypothesis of the Origin of Life

— Did eukaryotes evolve from the "*tamago* of life"? *

Katsura Asano

Division of Biology
Kansas State University
Manhattan, KS 66506, USA
kasano@ksu.edu

This article is an English translation of the following original article written in the Japanese language and published in 2008. Its English translation was completed and posted by the same author in March, 2021.

浅野 桂 (2008). 生命の起源の新しい仮説 真核生物は「生命のたまご」から進化した? 科学 (岩波書店、東京) 78 巻、277-287 頁

Katsura Asano (2008). A new hypothesis of the origin of life: Did eukaryotes evolve from the "*tamago* (egg) of life"? Kagaku (Science) (Tokyo) **78**, 277-287.

Phrases in red are explained under Glossary

* or ** Footnotes added during English translation of the original article.

* This author used the original Japanese word, *tamago* (たまご), which means an egg, to translate the name of the precellular protolife discussed here. Distinct from *ran* (卵, an egg) used strictly under biological contexts, *tamago* is more broadly used to refer to something or someone under a junior stage before becoming mature: e. g., *Isha no tamago* (医者のたまご), a doctoral student or intern, a doctor to be.

Life on Earth is generally classified into bacteria, archaea, and eukaryotes, but how eukaryotes, including humans, were originated is a mystery. Recent achievements in geobiochemistry and genomic biology have overturned previous thoughts that the eukaryotes are relatively new and instead showed that their origin can be as old as those of bacteria and archaea. Therefore, searching for the origin of eukaryotes must be reconciled with searching for the origin of cells, and a certain scenario emerges when this problem is considered along with Earth's geological history.

Introduction — Life on Earth is classified into three major domains

The classical cell biology began with the invention of the microscope in the 17th century, and came to fruition as the cell theory was proposed during the 19th century. Since then, with the development of microscopic technology, it became common to divide life on Earth into eukaryotes such as animals, plants and fungi, and prokaryotes such as bacteria, consisting of cells with nuclei or ones without nuclei, respectively. After the discovery of the nucleus in eukaryotic cells followed those therein of organelles bound by lipid bilayer membranes such as mitochondria, endoplasmic reticulum, and Golgi apparatus; their difference from prokaryotic cells without any discernible organelles became even more clear. In the latter half of the 20th century – the era of molecular biology, it was revealed that the two differed not only in cell morphology but also in chromosomal structure and gene expression mechanisms. Thus, the two major divisions of life between eukaryotes and prokaryotes seemed to have been established once as follows. In terms of chromosomal structure, the eukaryotic genome is linear; generally the distance between genes – mainly encoding proteins – is large, with large spacer regions; and the protein coding regions are also inserted by many non-coding segments called introns. In contrast, the prokaryotic genome is generally circular, with few spacers or introns, and therefore is compact.

In 1990, the “three domain” hypothesis proposed by the bacteriologist Carl Woese urged a major shift in this classical two-major-division concept (1). Woese et al. studied a series of bacterial groups, previously considered prokaryotes, that mainly live under extreme environments, such as in the presence of high temperature and high salt. Analyses of the primary structure of ribosomal RNA and various proteins encoded by these bacteria, however, showed that these bacteria (archaebacteria) were more similar to eukaryotes than to other bacterial groups including *Escherichia coli*. From this finding, prokaryotic organisms were divided into two major domains,

eubacteria (Bacteria) and archaeobacteria (Archaea), and the largest division of life on Earth were placed between eubacteria and the archaeobacteria/eukaryote group. Thus, the eukaryote was proposed to be the third domain, Eukarya (Fig. 1).

Nowadays, the problems of the origin of life, including the origin of eukaryotes, are being studied using all the cutting-edge scientific technologies, and a revolutionary paradigm shift is on the way. As a result of studying the traces of life activities on Earth, it was revealed that life was already born 3.5-3.8 billion years ago, soon after the ocean formed on Earth that itself was born 4.5 billion years ago. Likewise, the origin of eukaryotes, which was thought to occur at most 1 billion years ago, is now thought to date back to 2.7 billion years or earlier (Fig. 2). In this article, I would like to explore possible scenarios of how eukaryotes were born, based on the results of such latest studies on the origin of life.

Does the origin of eukaryotes date back more than 2.7 billion years?

The cells that make up all life on Earth are bound by a lipid bilayer membrane, which contains lipids that are characteristic of every organism. These lipids remain long after the organism's death, and their traces have been isolated carefully and identified as species-specific biomarkers by mass spectrometry. A eukaryotic biomarker is sterane with 28-30 carbons, derived from cholesterol. In 1999, this was discovered by Roger Summons et al. from an Australian iron formation dated 2.7 billion years old (2, 3). This 2.7 billion year-old biomarker overturned previous thought about the origin of eukaryotes. Almost all eukaryotes possess the organelle mitochondria, capable of **aerobic respiration**. For this reason, it was thought that they were born long after the oxygen concentration in the atmosphere rose dramatically 2.3 billion years ago.

The same iron formation studied by Summons et al also contained 2-methyl hopanoid, a biomarker for cyanobacteria, the first life on Earth capable of photosynthesis using water molecules as electron donors. Importantly, it is the bacteria that created the high oxygen partial pressure of the current Earth's atmosphere as the by-product of photosynthesis. Thus, the cyanobacteria were already abundant 2.7 billion years ago, even though the atmospheric oxygen partial pressure was kept low. This was accounted for by the idea that the oxygen was not only consumed by cyanobacteria themselves and non-photosynthetic aerobic bacteria but also by oxidative breakdown of their dead cells. The rise in atmospheric oxygen was in turn triggered by

increased sequestration of cell-derived carbonates into ocean sediments before the dead cells are decomposed (4).

The groundbreaking study of Summons et al. made the origin of eukaryotes much older, and the following results of genomic biology supported this idea.

With the progress of genomic biology, the entire genome sequences of more than 70 organisms representing the three domains have been determined so far, and based on this, Kanehisa (Kyoto University) et al. created an encyclopedia of all the metabolic reactions performed by all the life (5). Raymond and Segre analyzed this enormous amount of information using bioinformatics techniques and investigated the origin of all the chemical reactions utilizing oxygen molecules (= oxic reaction) including the synthesis of cholesterol, the eukaryotic biomarker mentioned above (6). As a result, it was revealed that the group of enzymes that catalyze the oxic reactions evolved by modifying a part of the anoxic reactions, after the latter was established during the differentiation of the three major domains. Cyanobacteria have been producing molecular oxygen in shallow oceans as a by-product of photosynthetic reactions, probably for more than 3 billion years. Therefore, according to this data, by the time cyanobacteria began to have a significant impact on the ecosystem, the three major domains had already differentiated and eukaryotes had also been born.

Origins of eukaryotes and mitochondria

The whole genome sequences of representative species of eukaryotes, including unicellular lower eukaryotes, have been determined, and revealed that all the extant eukaryotes have mitochondria as respiratory organelles, or their ancestors had them at least once (7). Similar to chloroplasts, mitochondria have a circular DNA genome, and their genome analysis suggests that mitochondria and chloroplast are derived from α -proteobacteria (a group of Bacteria) and cyanobacteria, respectively. It is generally thought that host cells with a "eukaryotic" genome first had an endosymbiotic relationship with α -proteobacteria and become the common ancestor of all eukaryotes, and then that the common ancestors of algae and all plants acquired photosynthetic capacity as they allowed cyanobacteria to coexist internally.

The eukaryotes' acquisition of a new function by symbiosis with bacteria has been described with deep-sea animals for famous examples - some acquired light emission through symbiosis with

bioluminescent bacteria, while others gain nutrients from symbiosis with **chemoautotrophic** bacteria utilizing hydrogen sulfide as the electron donor. Regarding photosynthesis, the examples of the symbiosis between fungi and cyanobacteria, and even those of heterotrophic eukaryotes that internalized algae as the results of **secondary or tertiary endosymbiosis** have been reported. Therefore, it is almost established that such phenomena have brought about important aspects of eukaryotic evolution.

However, there is fierce controversy over how the first eukaryotes engulfed α -proteobacteria and turned them into mitochondria (Fig. 3). According to a school led by classical cell biologists, the cells that hosted mitochondria were presumed to be a type of anaerobic eukaryote called Archezoa (8) (Fig. 3A). However, all existing Archezoa were found to have degenerate mitochondrial-like organs, leading them to having to lose one of the important grounds for this hypothesis (7). Another problem that was pointed out for this hypothesis is as follows: Even though the “true Archezoa” might have existed, they have the nucleus and the ability to swallow bacteria through **endocytosis** - both are characteristics that are only observed in eukaryotes. Thus, this hypothesis does not answer some of the most important problems related to the origin of eukaryotes. If the common ancestor of all the eukaryotes had mitochondria, their acquisition must have been the requirement for the birth of eukaryotes. If so, do we even need to assume Archezoa-like life as a state earlier than this event?

From this point of view, a group of genomic biologists hypothesized that the host cells were archaea (Fig. 3B). As proof of this model presented are their observation from decoding the whole genome protein sequence information that proteins involved in the eukaryotic information expression system are similar in amino acid sequence to those from Archaea (i. e., Woese's study mentioned above), while the cytoplasmic metabolic enzymes including glycolytic enzymes are similar to those from Bacteria. The group argued that there was a good metabolic reason for this fusion event, and thought that organelle membrane structures such as the nucleus had developed in order to overcome the defects caused by the fusion (7). However, this hypothesis contradicts with (i) the independent relationship between Archaea and eukaryotes established by Woese, and with (ii) the fact that the major membrane phospholipids of Bacteria are fatty acid esters and fundamentally different from those of Archaea, isoprenoid ethers (see below). Therefore, it has been criticized that the fusion of the cell membranes of these distinct types is extremely difficult and that there is no such example (9).

I would also like to point out that the latter point (ii) is particularly important: Even if the fusion is established temporarily, the control of membrane fluidity, which is essential for the function of cell membranes, may not work. For biological activities, it is essential to keep cell membrane lipids in a liquid (fluid) state under various environmental challenges. Bacteria and eukaryotes do so by regulating the length and saturation of fatty acids. On the other hand, archaea perform the fluidity regulation by cyclizing the isoprenoid chain (10). Thus, if membrane lipids are mixed in the fusion event, such delicate control systems would not operate effectively.

Were eukaryotes born in shallow, oxygen-rich oceans?

How would the eukaryotic emergence scenario change as the origin of eukaryotes dates back more than 2.7 billion years? Among Bacteria, cyanobacteria are highly developed organisms equipped with different electron transport chains such as respiratory chains and photosystems, so it is considered that adaptive radiation of Bacteria has already been completed at this time. It is thought that the shallow sea where sunlight reaches had a high oxygen concentration as a result of photosynthesis by cyanobacteria. However, oxygen is consumed on the spot by respiratory activity performed by these **photoautotrophic** bacteria and the **heterotrophic** microorganisms. It is therefore thought that the deeper oceans and the atmosphere were oxygen-free or hypo-oxygenated.

Since the shallow sea where sunlight reaches is the area where life activity is the most active, it is possible that eukaryotes with only mitochondria prospered at this habitat as heterotrophs. According to subsequent studies, Archaea originally found by Woese et al. in extreme environments were found to be common in other normal environments, including in the ocean. Therefore, the microorganisms that consume oxygen here in the shallow ocean would include archaea as well as bacteria. So is this the birthplace of eukaryotes?

As mentioned earlier, the lipid cholesterol that characterizes eukaryotes requires many oxygen molecules for its synthesis. Eukaryotic cells bend the cell membrane significantly during endocytosis, taking advantage of the ability of cholesterol to move between the two layers of the lipid bilayer membrane (eg, from the backside to the frontside) (11). Therefore, oxygen required for cholesterol synthesis is indispensable for establishing a eukaryotic cell system with abundant membrane structure and endocytosis mechanism. Conversely, in an oxygen-free environment, it is hard to envision to have anaerobic Archezoa-like life evolve with only endocytosis and nuclei.

In addition, the idea of fusion between bacteria and archaea also has major defects in lipid chemistry or fluidity control.

However, the very old origin of eukaryotes provides a completely different perspective on this issue. As discussed next, genomic biology suggests that the origin of eukaryotes is as old as that of bacteria and archaea. Therefore, the search for the origin of eukaryotes must be reconciled with the search for the origin of cells, and by considering this problem in the geological knowledge, a different scenario emerges.

Origin of the three major domains from the viewpoint of genomic biology

The phylogenetic tree of life first proposed by Woese based on the base substitution rate of ribosomal RNA draws three beautiful branches (1), but it is not clear what was the common ancestor of all life that was assumed to be at the root. If this phylogenetic tree is interpreted as it is, the common ancestor of bacteria and that of archaea and eukaryotes were first derived from the common ancestors of all life, and then the latter organisms are divided into archaea and eukaryotes. However, the root position itself was determined as a result of statistical calculations, and the currently assumed position has not been proven to be correct within the margin of error. Another important point is that this root has a different meaning from the roots found later within each domain, that is, the species that are the common ancestors of each group. About this, Woese himself said (12).

“The universal tree has no root in the classical sense. The root is actually *Darwinian Threshold*, the first point at which we can begin to give tree representation to the biological evolutionary course.”

Accordingly, the life corresponding to the root (common ancestor of all life) was in a state before Darwin's evolution (evolution by natural selection of individuals with genetic diversity) started. Only when this common ancestor acquired individuality, did it cross Darwin's threshold and enter the process of life's evolution in the normal sense. Applying this idea to the state assumed by the origin of life, the common ancestor was like a "tamago (egg)" before becoming cells, and, only when it acquired the cellular system and advanced into various environments of the outside world, it started species radiation and differentiated into various species. According to a rigorous interpretation of the Woese's phylogenetic tree in this context, the branch length of each domain,

calculated from the rate of base substitution in ribosomal RNA, merely indicates the age of the domain from the present to its origin. Thus, independently from the state of "tamago", emerged first the bacteria, then the archaea, and finally the eukaryotes, acquiring the cellular system and thereby entering the outside world hence Darwinian evolution. The mother "tamago" continued to exist from the emergence of bacteria to the birth of eukaryotes. As will be described later, similar ideas have already been proposed by Martin and Russell for the emergence of the first two major domains, bacteria and archaea (13).

Origin and evolution of life in alkaline hydrothermal vents

With the development of seafloor exploration technology, it became clear that even on the deep seafloor where sunlight does not reach, there is a unique ecosystem that depends on the energy of hot water ejected from inside the earth (14). The nature of the hot water in the hydrothermal vents depends on the geology of their vicinity. A group of a geologist and a bacteriologist advocated the fascinating hypothesis that an alkaline hydrothermal vent, with relatively mild temperature conditions, was the place of origin of life 3.5- 3.8 billion years ago (15) (Fig. 4). The first step toward life was assumed to be the establishment of a dynamic equilibrium state (sustainability) of biochemical reactions to generate simple organic compounds from H_2 and CO_2 which were abundant in the reducing environment at that time. Importantly, the reaction was assumed to be catalyzed by sediment components (serpentinite) such as FeNiS. The sulfide deposits found near these vents have precisely cell-sized cavities or pores, where organic compounds accumulate as a result of catalysis, eventually establishing the dynamic equilibrium of prebiotic organic chemical reactions. Biological evidence for this idea was presented that FeNiS clusters similar to those found in vent deposits are present at the catalytic sites of primitive carbon dioxide-fixing enzymes found in extant chemoautotrophs, such as methane-producing archaea and acetic acid-producing bacteria (16). It has not yet been explained how the prebiotic dynamic equilibrium system transitioned to a more biological equilibrium system that relied on cells and genetics. However, it advanced **Oparin's primitive soup hypothesis** one step further and explained how the dynamic equilibrium system is established step by step. Furthermore, the conditions of alkaline hydrothermal vents have been reproduced in the laboratory, and it was proven that the FeNiS catalysis can at least produce simple biomolecules such as amino acids and peptides under these conditions (17).

Building upon this idea, Martin and Russell proposed that bacteria and archaea emerged independently from the precellular autotrophic life born in the alkaline hydrothermal vent, as it acquired distinct membrane lipid biosynthesis pathways and cell division mechanisms (13).

Did the common ancestor of all life have no cell membrane?

Some scientists who gave serious thoughts about the origin of life are accepting the idea that there was a stage of pre-cellular life without a lipid bilayer cell membrane (18) - as an essential intermediary life form that existed just before the life's first adaptive radiation. As mentioned above, Martin et al. speculated the appearance of such a life form as the common ancestor of all life, by linking primitive biological carbon dioxide-fixation reactions with the chemical reactions that must have occurred near the hydrothermal vent - in a deductive or top-down approach (13).

In contrast, bioinformaticians such as Koonin defined such a life form inductively (bottom-up) by comprehensively analyzing the recently determined genome sequences of representative organisms on Earth, and by reducing them to a set of genes encoding enzymes conserved in all life (19). The organism, called LUCA (Last Universal Common Ancestor), had double-stranded DNA as a genetic information molecule, and had almost the same genetic code and gene-expressing enzymes as extant organisms.

The main reason why the common ancestor of all life was not assumed to have a lipid bilayer is that, as mentioned above, the composition of the lipid bilayer membrane is completely different between archaea and bacteria (20). Archaeal phospholipids are mainly composed of two isoprenoid chains ether-bonded to glycerol phosphate, while bacterial phospholipids are mainly composed of two fatty acids ester-bonded to glycerol phosphate. Importantly, between these phospholipid molecules, glycerol phosphate moieties have a distinct chirality. Therefore, it is considered that the enzymes that conjugate the carbon chain to glycerol phosphate are completely different between the two groups and were originally invented in each prior to forming the cells of each type.

It was hypothesized that this primitive life form made lipids other than phospholipids, and formed a hydrophobic environment around it to perform primordial metabolic activities by some "membrane" proteins such as F-type ATP synthase (13). However, this lipid-protein complex was

not organized into a continuous lipid bilayer structure. Thus, it is not possible for this life form to become a cell, travel out of the place of its origin, and ultimately evolve in the true sense.

Of course, the unorganized lipid-protein boundary would not have been sufficient to protect this life form. Since the cells of all life are covered with a cell wall or extracellular matrix mainly composed of hydrophilic sugar chains, it can be considered that this life was covered by the viscous "kara (shell)" made of sugar chains. Some basic forms of extracellular matrix are common to the three domains, including a repeating structure of disaccharides containing N-acetylglucosamine. These sugar chains are cross-linked by peptides to form peptidoglycans in prokaryotes (both bacteria and archaea). In present days, biofilms made of peptidoglycan *etc.* are used as physical scaffolds for the formation of microbial ecosystems in the natural environment (21). Therefore, these precellular organisms may also have been covered with biofilms, forming huge colonies.

Did eukaryotes emerge by the symbiosis of "the tamago of life" and bacteria?

Based on all these findings – the origin of eukaryotes suggested by geobiochemistry dating back nearly as old as the era of origin of life, the acquisition of mitochondria (the endosymbiosis of α -proteobacterium) that must have triggered the emergence of the eukaryote, and the eukaryotic bilayer phospholipids being of the bacterial type – I would like to propose that eukaryotes were born directly from the "tamago (egg) of life" - the common ancestor that does not have a membrane (Fig. 5). In this scenario, it is considered that the phospholipid synthetic enzymes, like other many metabolic enzymes such as glycolytic enzymes, were obtained from the endosymbiotic α -proteobacterium.

The main premise of this hypothesis is that "the tamago of life" remained at its place of origin until as late as 3 billion years ago, when cyanobacteria became abundant, and that, by the time eukaryotes emerged, it was present in a shallow, oxygen (O_2) - and nutrient-rich ocean. Let's discuss whether we can clear the hurdles of these assumptions in the last section of this article.

The trigger of this event may have been the weakened geological activity of the place of origin of life, that is, the hydrothermal vent. Then, "tamago", which had become the chemoautotroph by that time, would have to transition to a heterotrophic organism. Thus, a symbiotic relationship between "tamago" and heterotrophic α -proteobacteria was established. Then, a large change

occurred, when a gene of "tamago" had been modified and it successfully embedded the **ATP/ADP exchange transporter** in the plasma membrane of the bacterium. "Tamago" was now able to obtain a large amount of ATP molecules through the catabolic activity of the endosymbiotic bacteria. Excessive energy awakened the enzyme development potentials of "tamago" possessing a rich gene pool (see below), allowing it to create a series of enzymes and proteins essential for the activities, as we now see in eukaryotes, including the heterotrophic activity. When "tamago" acquired the phospholipid synthesis pathway from the bacterial genome, various membrane structures evolved to stretch around its own genome, eventually formed organelles such as nuclei, and finally succeeded in creating a cellular system as we see in eukaryotic cells. Both the eukaryotic genome and cell sizes are, on average, far larger than those of prokaryotes (Bacteria and Archaea) that were presumed to have left the place of their origin earlier. This fact may be taken as a proof that "tamago" had an excessive energy-producing system, mitochondria, at the time of the departure of eukaryotic cells - before their first adaptive radiation.

What is the potential of "tamago" that gave rise to eukaryotes?

In eukaryotes, many proteins not found in other domains are involved in eukaryote-specific activities such as cell cycle regulation, signal transduction, and organelle function. Some of the new genes added in eukaryotes were modified from those already present in bacteria or archaea, but many were completely new genes (22). Where did these new genes come from? Various considerations have been made to this mystery, but what if those varieties were originally present in the "tamago"? In the non-individualized "tamago" state, it may be possible to pool much more genes than the number of genes necessary for its metabolic and replication activities. It is difficult to prove this idea, but if you consider the opposite case, in which the life has already been individualized into cells (cellularized), these cells would have been able to select only the essential genes during the process of such individualization, putting them together in a compact genome (in the case of bacteria/archaea). This would make us aware that, of the genes that might have been present in the huge genome of the "tamago", we may be appreciating only a subset that have been selected by the ancestors of extant life through their cellularization events (Fig. 5).

Based on the comparison of the genomes of organisms from the three domains, Woese and Doolittle each postulated that, at the stage of differentiation into the three domains (the stage where the life was about to exceed Darwinian Threshold), there was frequent genome reorganization events through transfer of genes across species (horizontal gene transfer) (12,23).

Studies of existing microorganisms may provide insights into the organization of genes contained in "tamago" and the mechanism of horizontal gene transfer between its colonies. Some bacteria release genomic DNA extracellularly as a component of biofilms (24). Thus, biofilm DNA may mediate horizontal gene transfer between different microbial species in the ecosystems, and likewise between precellular protolife colonies. Some plasmids – bacterial extrachromosomal genetic elements – use their RNA transcribed from their genome as an indicator of their intracellular concentration, thereby regulating their replication frequency and copy number (25, 26). With such a concentration-dependent replication mechanism, the precellular protolife assumed here as "tamago" would be able to replicate its genome DNA at a constant copy number, as it increases its volume into new territories. Although it is not clear whether its genome was circular or linear, circular DNA is stable and can serve as a unit of horizontal gene transfer in the natural environment (27). As is still the case with some plasmids, multiple circular chromosomes may have organized functionally related genes separately.

Conclusion - The search for origin of life in the era of interdisciplinary sciences

If the "tamago (egg) of life" had existed near the hydrothermal vent, was there a time when the alkaline hydrothermal vents, currently found in the deep seafloor, instead existed in the shallow ocean where cyanobacteria grew? As the Archean Earth was much hotter than Earth is today, it must have had much more hydrothermal vents, each of which would have been a test site for the origin of life. Therefore, it is possible that the vent that became the home of the "tamago" was originally located in the shallow ocean or had moved to the shallow ocean after tens of millions to hundreds of millions of years of plate tectonics. Another possibility is that the depth of the water decreased due to the arrival of the ice age. It appears that the ice age that hit 2.2 to 2.3 billion years ago was so fierce that the earth became a snowball. There are also reports that there was an ice age 2.9 billion years ago. Thus, it is anticipated that the impacts that this or the earlier ice age gave to the primordial global ecosystem are clarified

Since the adaptive radiation that happened over 3 billion years, it does not appear to have a big, fundamental change in the lifestyle of bacteria or archaea. If so, we may be able to reconstruct, to some extent, that of the primordial protolife that must have existed just before their diversification through studying these microorganisms that grow in extreme environments similar to those found in ancient Earth. State-of-the-art technologies of molecular biology such as genomic biology, mass spectrometry and live-cell fluorescence microscopy are being actively applied to the study

of microorganisms that propagate in hydrothermal vents as well as in the normal natural environments. It is expected that the origin of life will gradually show its true face as we enter the era of comprehensive, interdisciplinary sciences.

Acknowledgements**

Special thanks are due to Drs. Akira Yamagishi, Roger Summons and Gary Conrad.

**The author's research at the time of the preparation and publication of this article was supported by National Institutes of Health Grant GM64781 and NCRR K-INBRE Pilot Grant P20 RR016475.

Original illustration

Kaoru Iihashi

References

1. Woese, C. R., Kandler, O., and Wheelis, M. L. (1990) *Proc Natl Acad Sci USA* **87**, 4576-4579
2. Brocks, J. J., Logan, G. A., Buick, R., and Summons, R. E. (1999) *Science* **285**, 1033-1036
3. Knoll, A. H. (1999) *Science* **285**, 1025-1026
4. Kasting, J. F. (2006) *Nature* **443**, 643-645
5. Kanehisa, M., and Goto, S. (2000) *Nucl Acids Res* **28**, 27-30
6. Raymond, J., and Segre, D. (2007) *Science* **311**, 1764-1767
7. Embley, T. M., and Martin, W. (2006) *Nature* **440**, 623-630
8. de Duve, C. (2007) *Nature Rev Genet* **8**, 395-403
9. Poole, A., and Penny, D. (2007) *Nature* **447**, 913
10. Koga, Y., and Morii, H. (2006) *Anal Biochem* **348**, 1-14
11. Summons, R. E., Bradley, A. S., Jahnke, L. L., and Waldbauer, J. R. (2006) *Philos Trans R Soc Lond B Biol Sci.* **361**, 951-968
12. Woese, C. R. (2002) *Proc Natl Acad Sci USA* **99**, 8742-8747
13. Martin, W., and Russell, M. J. (2003) *Phil Trans R Soc Lond B Biol Sci* **358**, 59-85
14. Boetius, A. (2005) *Science* **307**, 1420-1422
15. Martin, W., and Russell, M. J. (2007) *Philos Trans R Soc Lond B Biol Sci.* **362**, 1887-1925
16. Doukov, T. I., Iverson, T. M., Seravalli, J., Ragsdale, S. W., and Drennan, C. L. (2002) *Science* **298**, 567-572

17. Huber, C., Eisenreich, W., Hecht, S., and Wachtershauer, G. (2003) *Science* (301), 938-940
18. Koga, Y., Kyuragi, T., Nishihara, M., and Sone, N. (1998) *J Mol Evol* **46**, 54-63
19. Koonin, E. V., and Martin, W. (2005) *Trends Genet* **21**, 647-654
20. Boucher, Y., Kamekura, M., and Doolittle, WF (2004) *Mol Microbiol* **52**, 515-527
21. Reysenbach, A.-L., and Shock, E. (2002) *Science* **296**, 1077-1082
22. Aravind, L., Iyer, L. M., and Koonin, E. V. (2006) *Curr Op Struct Biol* **16**, 409-419
23. Doolittle, W. F. (1999) *Science* **284**, 2124-2128
24. Whitchurch, C. B., Tolker-Nielsen, T., Ragas, P. C., and Mattick, J. S. (2002) *Science* **295**, 1487
25. Tomizawa, J.-i. (1984) *Cell* **38**, 861-870
26. Asano, K., and Mizobuchi, K. (1998) *EMBO Journal* **17**, 5201-5213
27. Miller, R. V. (1998) *Sci Am* **January**, 66-71
28. Oparin, A. I. (1961) *Life, its nature, origin and development*. Academic Press, New York
29. Walker, J. E., and Runswick, M. J. (1993) *J Bioenerg Biomembr* **25**, 435-446
30. Pennisi, E. (2004) *Science* **305**, 766-768

Glossary

Aerobic respiration. In the cytoplasm of eukaryotes, sugar molecules consisting of six carbons such as glucose are broken down to three-carbon pyruvic acid by the glycolysis reaction. The reaction termed citric acid cycle completely breaks down (oxidizes) the pyruvic acid into carbon dioxides in mitochondria. The energy obtained by breaking down the pyruvic acid is transferred to the electron transport chain of the inner mitochondrial membrane, where a high membrane potential is generated due to the proton gradient. Being localized in the mitochondrial inner membrane, the F-type ATP synthase uses this membrane potential to synthesize ATP. Oxygen required for respiration is used in the reaction that drives the electron transport chain and is converted into water molecules. This whole reaction involving mitochondria and oxygen to generate ATP is called aerobic respiration – a very efficient energy-producing reaction as compared with anaerobic respiration such as fermentation. Many bacteria and archaea have a citric acid cycle in the cytoplasm and an electron transport chain and F-type ATP synthase in the plasma membrane and similarly perform aerobic respiration.

Autotrophs and heterotrophs. Organisms that obtain the nutrients necessary for biological activities by directly producing organic compounds from carbon dioxide are called autotrophs. Of these, organisms that produce nutrients by photosynthesis are called photoautotrophs, and those that produce

nutrients using chemical energy without depending on light are called chemoautotrophs. All organisms that take nutrients directly or indirectly from autotrophs are heterotrophs. At the bottom of the food chain in any ecosystem are autotrophs as primary producers.

Primordial soup hypothesis of Oparin et al. When Oparin published this theory in 1924, it was believed that the environment on early Earth right after its formation was very reducing. From this notion, Oparin presumed that, given that there was no biological oxidative catabolism, high molecular weight carbon compounds were naturally generated and accumulated on Earth at the time of the origin of life, producing nutrient-rich "primordial soup" everywhere. Under such an assumption, the first life was presumed to be an anaerobic heterotroph fed on the nutrients of the soup (28). However, after that, a large amount of carbon dioxide was found in the atmosphere of Venus, which is often compared to Earth as its sister planet. Likewise, geological studies indicated that there was a large amount of carbon dioxide in the ancient Earth atmosphere. These and other notions led to the idea that the first life on Earth was a chemoautotroph that chemically fixes carbon dioxides into their nutrients.

Secondary and tertiary endosymbiosis. If the original endosymbiosis of cyanobacteria that produced chloroplasts in all green algae and plants is assumed to be the primary endosymbiosis, the endosymbiosis of heterotrophic eukaryotes with an algae cell is considered the secondary endosymbiosis. The generation of a new algal species from the

endosymbiosis of heterotrophic eukaryotes with a secondary endosymbiont is called the tertiary endosymbiosis.

Endocytosis. A method of nutrient intake by the cells, in which a part of the cell membrane is invaginated so as to internalize nutrients including other cells. The nutrients are then taken into the cells along with the whole part of the membrane and digested.

ATP/ADP exchange transport factor. A protein localized in the inner mitochondrial membrane that transports a single ATP molecule to the cytoplasm and ADP, a substrate for ATP synthesis, into the mitochondrion. Unlike many other mitochondrial proteins, this protein is not present in bacteria (29). Thus, wouldn't it be appropriate to assume that it was produced at the time of the emergence of the eukaryote?

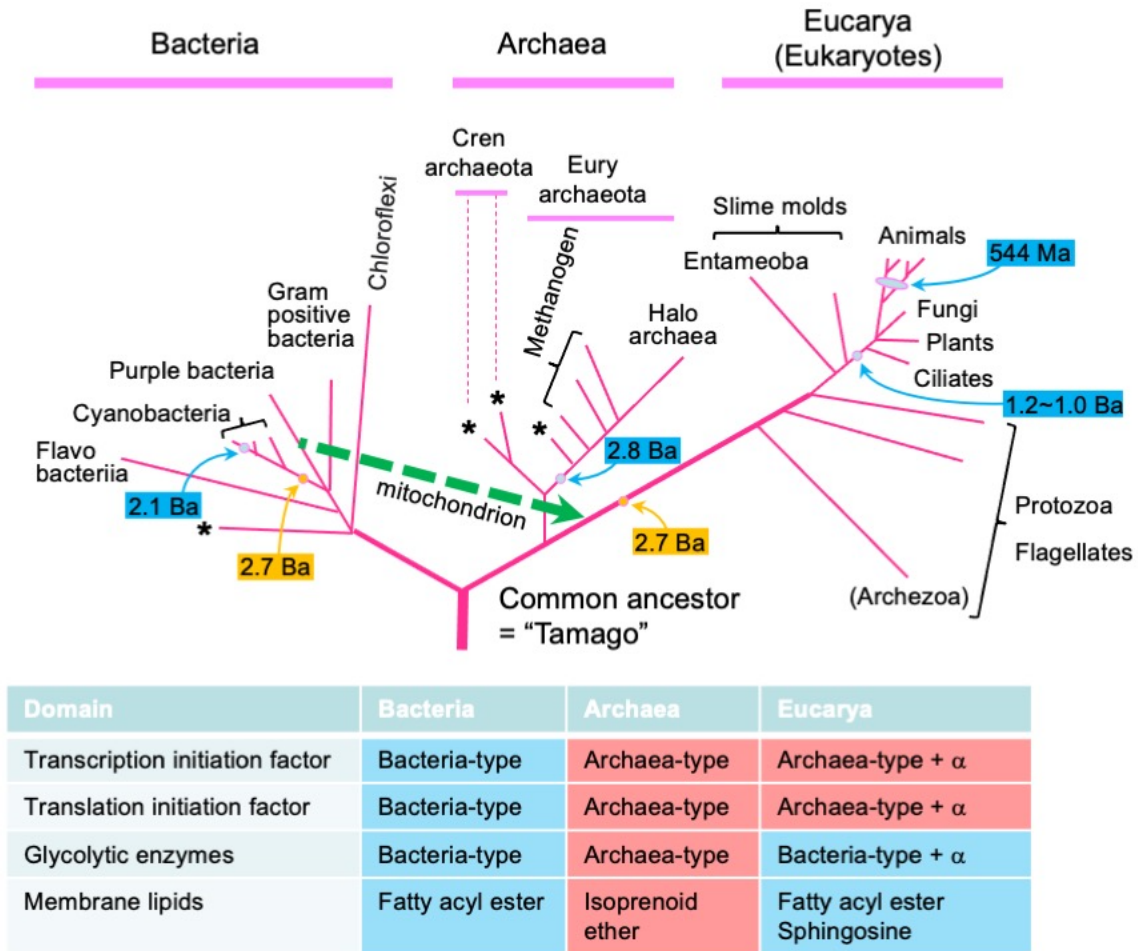


Figure 1. Three major domains of life on Earth. (Above) A phylogenetic tree of life on Earth, modified and redrawn from Reference (3). Colored squares, the major divergence dates in billion years that were determined by archaeological or geological method. *, hyperthermophiles. Thick dotted arrow indicates that the mitochondrion is derived from a purple bacterium or another member of α -proteobacteria. Blue circle in a cyanobacterial lineage indicates the time of endosymbiosis event assumed for chloroplast formation in algae/plants. (Bottom) Similarities of the constituents between eukaryotes and other domains. Table summarizes which counterpart in the other two domains – bacteria or archaea – is similar to eukaryotic gene expression constituents (translation and transcription initiation factors) or glycolytic enzymes as the representative of metabolic enzymes. + α , the presence of eukaryote-specific elements in addition to those found in the indicated domain of life. The main membrane lipid components are shown at the bottom.

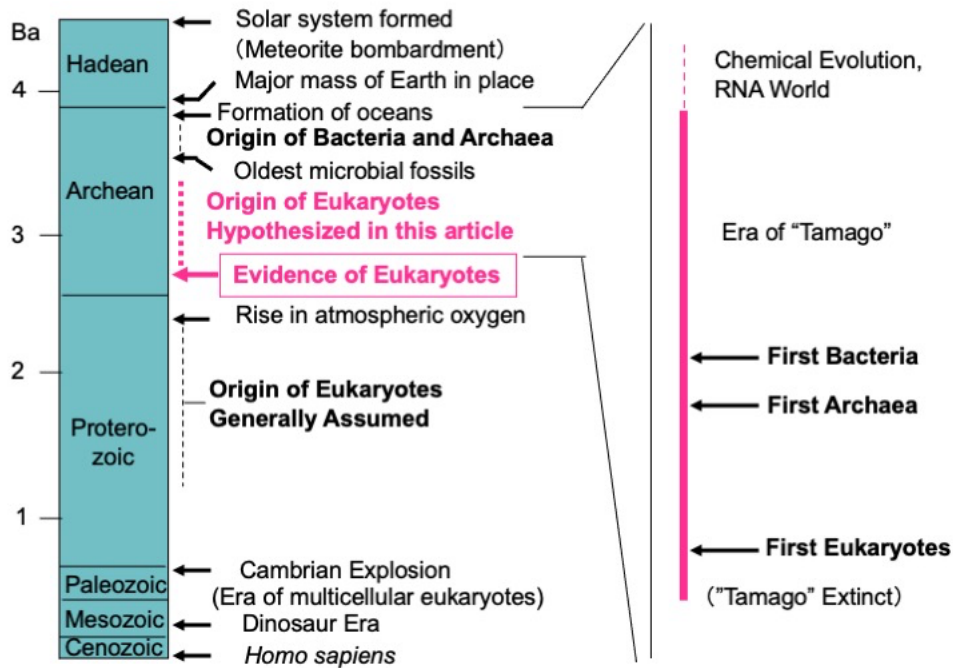


Figure 2. Major events of the Earth geology and the origin of life. The right column (with the time scale of 4 to 3 billion years ago) shows the era of "tamago" as thick vertical bar.

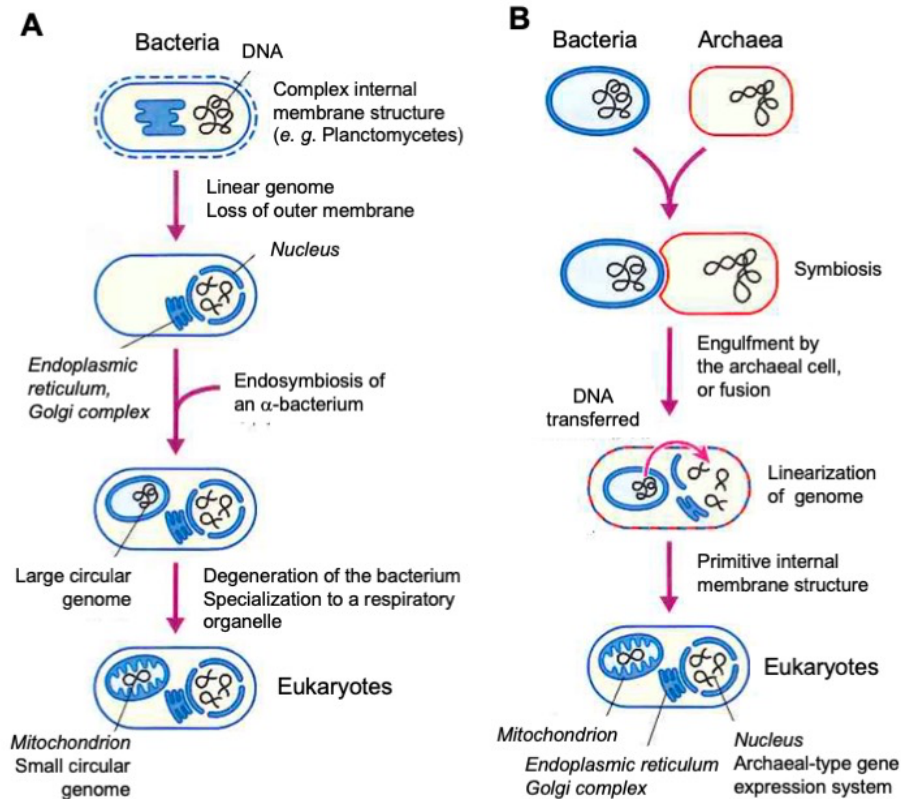


Figure 3. Major hypotheses about the origin of eukaryotes. Each of the blue (fatty acyl ester-type, see Fig. 1) or red (isoprenoid ether-type, see Fig. 1) closed line represents one continuous lipid bilayer membrane. Black line indicates the DNA genome (circular or linear). (A) Archezoa host model. The ancestors of Archezoa are assumed as different classes of bacteria depending on various derivatives of this model, but here we have shown an example of Planctomycetes, in which an internal structure resembling a nucleus has been discovered (30). The blue dotted line indicates that the outer membrane present in this bacterium needed to disappear as it evolved into Archezoa. All derivatives of this model show that eukaryotes are derived from bacteria, which contradicts the phylogenetic tree of Fig. 1. (B) Archaeal host model. There are several derivatives of this model as to the types of archaeal hosts and the reasons for the fusion or engulfment event.

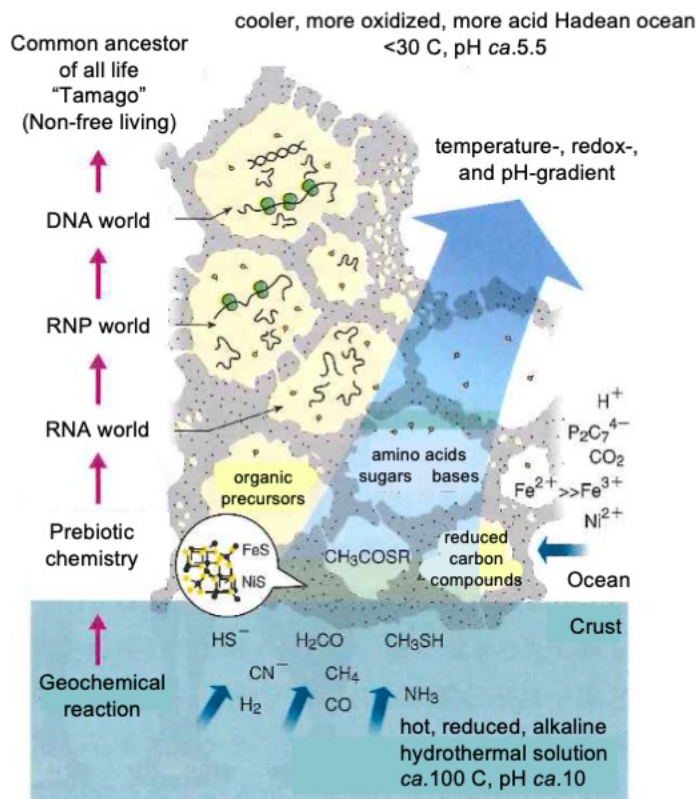


Figure 4. Michael Russell's hypothesis of the origin of life in alkaline hydrothermal vents. Modified and redrawn from Ref. (13). Large arrows in the middle of the drawing show the gradients of the temperature, the redox potential, and pH, produced by the hydrothermal activities of the alkaline hydrothermal vent. Based on the energy generated by the movement of electrons and protons, the FeNiS clusters (brown structure) catalyze the generation of organic precursors from CO₂ in the ocean and H₂ released from the vent with hot water as substrates. As these precursors got deposited into the micro compartments (yellow space) and thereby highly concentrated, the RNA with self-replicating activity, or ribozyme, was born (RNA world). Subsequent to the RNP (RNA and protein) world, the DNA world was established, which became the common ancestor of all life without a discrete cellular structure (termed "tamago" in this article). The black arrows indicate the activities in the microspaces corresponding to each stage (or "world"). RNA world: Catalytic RNA (ribozyme) acted not only as metabolic enzymes but also as template for their self-replication with the help of prebiotically produced peptides. RNP world: A highly developed ribozyme, the ribosome, was born, which began to make proteins using itself or other RNA as a template. DNA world: Genetic information transmission functions (genes) shifted to DNA, so did the enzyme functions to proteins. RNA began to specialize in its function as messenger RNA (mRNA).

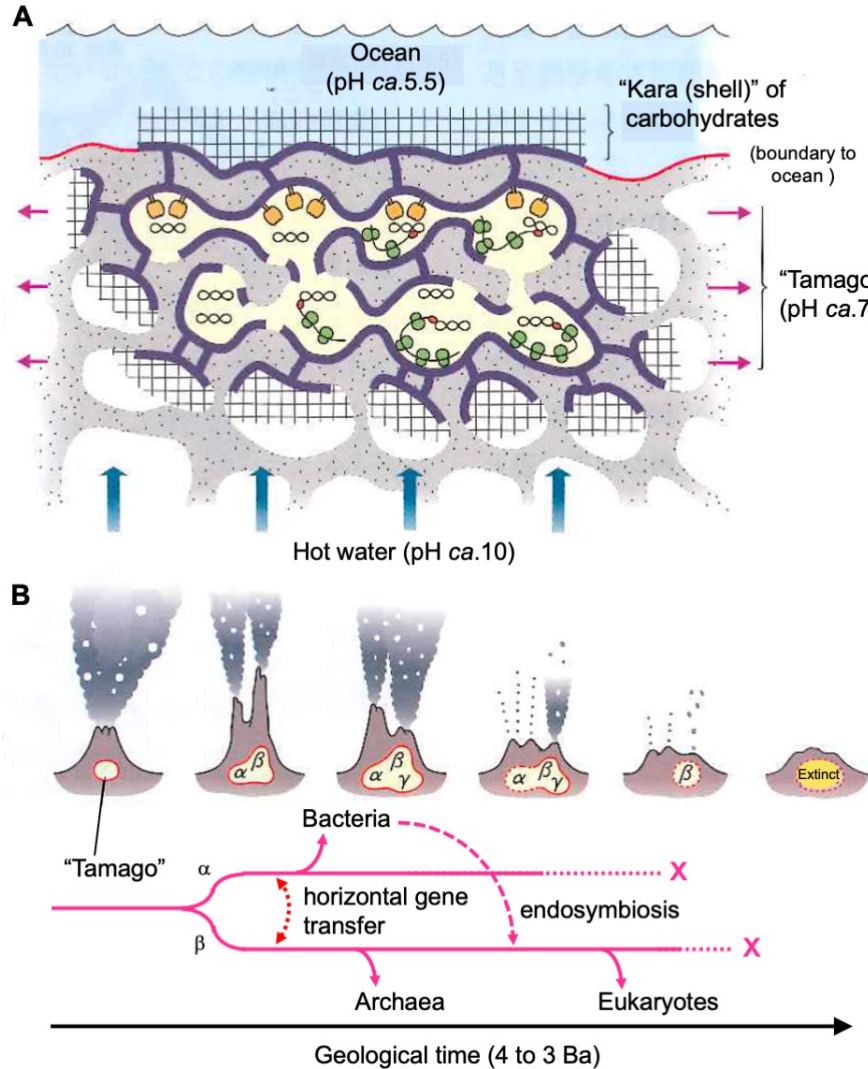


Fig. 5. "Tamago of Life" and the origin of the cellular structures in the three domains. (A) Schematic diagram of "tamago of life". Hydrothermal vent deposits are shown in gray and ribosomes in green. The internal surface of the microspace (yellow) of the sediment was discontinuously covered with a hydrophobic layer (thick purple line) composed of proteins and lipids. The genome was circular and the genes were encoded by a number of circular chromosomal DNAs (closed ribbon). Besides the chemoautotrophic pathways assumed by Martin and Russell, the "tamago of life" had all the activities which has been assumed to be present in the common ancestor of all life (LUKA), such as DNA-dependent replication and gene expression mechanisms (RNA polymerase transcribing the DNA is shown as red circles), protein transport coupled with translation (ribosomes bound to the hydrophobic layer, represented as green double ovals, secrete proteins out or deposit them to the boundary), and ATP synthase (orange mushroom) driven by proton gradient with seawater (light blue space above). Moreover, "tamago" was covered with sugar chains (carbohydrates, indicated by mesh) and had a certain morphology that had adapted to fixed life in an alkaline environment. Since it was not covered by the cell membrane, it performed complex metabolic

reactions through a division of labor, taking advantage of its relatively large space. It also had a large stockpile of unused genes. Pink thin arrows represent the direction in which this life grows. Blue thick arrows indicate the direction of hydrothermal activity. (B) Emergence of bacteria, archaea, and eukaryotes from the "tamago of life". The thick arrow on the bottom shows the approximate geological time - between 4 and 3 billion years ago. (Middle panel) The "tamago of life", which had to stay at the place of origin, is assumed to have evolved in response to changes in the geological activity of that place (shown in the top panel) - to the extent that it does not exceed Darwinian threshold. During the course of this, it was divided into at least two genetically distinct groups α and β , each of which gave birth to bacteria and archaea first; the group β , that gave birth to archaea, later gave birth to eukaryotes, all as free-living life. These groups were present at places not far from each other, and therefore, were able to exchange genes through DNA released into the environment (horizontal gene transfer). The genetically diverse groups that did not give birth to free-living life (such as γ) got extinct with the end of the geological activity of the vent. The upper panel depicts the activity status of the hydrothermal vent and the assumed "tamago"-like protolife present on its surface. The area encircled by solid red line shows active colonies, and that encircled by dashed red line shows weakened colonies.