the TIMETREE of LIFE

edited by **S. BLAIR HEDGES** and **SUDHIR KUMAR**
foreword by James D. Watson

Life

S. Blair Hedges

Department of Biology, 208 Mueller Laboratory, Pennsylvania State University, University Park, PA 16802-5301, USA (sbh1@psu.edu)

Abstract

Life on Earth arose from a single source, ~4400–4200 million years ago (Ma), and quickly achieved a prokaryotic level of complexity. An initial split (~4200 Ma) led to the Superkingdoms Eubacteria and Archaebacteria. Theories for eukaryote origins fall into two classes, merger and deep-root models, with the former having broadest support. Under the two-merger model, an archaebacterium joined a eubacterium (possibly ~2700 Ma) to form the nucleus and hence first eukaryote. A subsequent merger with another eubacterium (~2000 Ma), an alphaproteobacterium, formed the mitochondrion. The one-merger model postulates that both the nucleus and mitochondrion formed at about the same time (~2000 Ma).

This is the earliest and most controversial portion of the tree of life and few details can be regarded as well established. It involves the relationships and times of origin of the three superkingdoms, Eubacteria, Archaebacteria, and Eukaryota. But the key to understanding how these earliest events unraveled is to know how eukaryotes arose and their relationship to prokaryotes. Despite the availability of hundreds of completely sequenced genomes from prokaryotes and dozens from eukaryotes, the answers are not yet in hand. New models for the origin of eukaryotes appear frequently. For ease of discussion they are classified here as merger models (1–11) and deep-root models (12, 13). The former ascribe the origin of eukaryotes to a merger between two prokaryotes whereas the latter minimize the role of mergers and instead postulate an ancient origin of eukaryotes, at least as old as the earliest divergences among living prokaryotes. The goal of this brief synopsis is to review the evidence bearing on the earliest aspects of the timetree of life. Details concerning differences among the many proposed models, as well as different points of view, can be found elsewhere (2, 14, 15).

Life on Earth encompasses an estimated ~1.8 million described species (16) and a much larger number is thought to be undescribed (>10 million species). Almost all of the described species are eukaryotes, and most of those are arthropods. Of prokaryotes, there are ~9400 recognized species of eubacteria and ~300 of archaebacteria, based on the latest compilations (17, 145). Some organisms have been found as deep as ~800 m below the ocean floor in subsurface sediments (18) and others (e.g., bacterial spores) have been found as high as 41 km above sea level in the atmosphere (19). The most abundant organism is probably a ubiquitous marine eubacterium (20) (Fig. 1).

Different authors use different names for the three major groups, and this requires an explanation. When it was first recognized that the methanogens and relatives formed a distinct group they were given the name archaebacteria, with the remainder of prokaryotes named eubacteria (21). These were appropriate names because they were given the stem ("bacteria") which indicated that they were both comprised of prokaryotes. Later they were renamed archaea and bacteria, respectively (22) "to avoid any connotation that eubacteria and archaebacteria are related to one another." This was done because
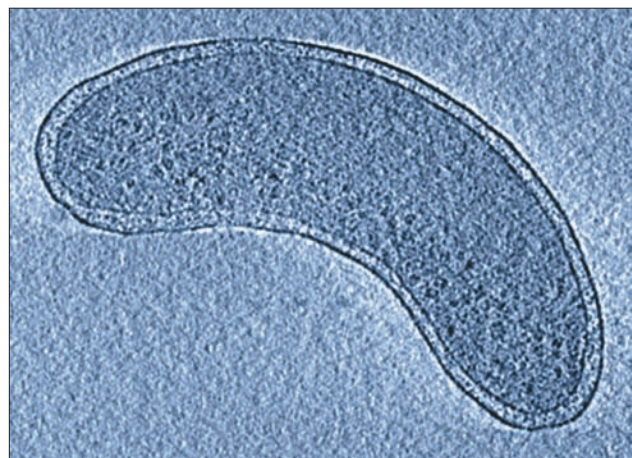
Fig. 1 Cryo-electron tomographic image of a marine eubacterium, *Pelagibacter ubique* (alphaproteobacteria), one of the smallest self-replicating cells (1354 genes) and most abundant of organisms (20). Dimensions of the cell are ~900 × 280 nm. Credits: D. Nicastro, Brandeis University, and J. R. McIntosh; cell from S. Giovannoni.
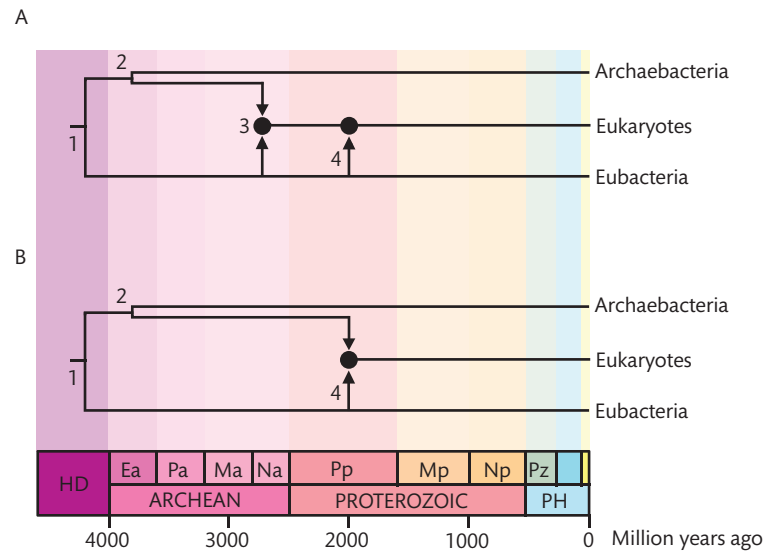
Fig. 2 Two versions of the timetree of life based on competing merger models for the origin of eukaryotes. Times of divergence are from Table 1. (A) The two-merger model. Node 1 is the divergence of eubacteria and archaebacteria (LUCA, the last universal common ancestor). Node 2 is the divergence of two types of archaebacteria, one eventually leading to the origin of eukaryotes. Node 3 is the origin of eukaryotes and represents a merger between an archaebacterium and a eubacterium that led to the eukaryote nucleus and possibly the incorporation of eubacterial genes in the genome of eukaryotes. This pre-mitochondrial event is not yet well established (see text). Node 4 is the symbiotic event that led to the mitochondrion of eukaryotes and the transfer of genes from the eubacterial symbiont (an alphaproteobacterium) to the nuclear genome of eukaryotes. (B) The one-merger model. This is identical to the two-merger model except that the formation of the nucleus and mitochondrion are combined into a single step. *Abbreviations*: Ea (Eoarchean), HD (Hadean), Ma (Mesoarchean), Mp (Mesoproterozoic), Na (Neoarchean), Np (Neoproterozoic), Pa (Paleoarchean), PH (Phanerozoic), Pp (Paleoproterozoic), and Pz (Paleozoic).

archaebacteria were shown to cluster with eukaryotes in the small subunit (SSU) ribosomal RNA (rRNA) tree (*22*) rooted by emerging gene duplication evidence (see later). However, in the last two decades, analyses of complete genomes have contradicted that interpretation, leading to the general—although not unanimous (*23*)—view that eukaryotes are cytological and genomic chimeras of prokaryotes (*2, 14, 15*), which will be discussed at length later. For this reason, many evolutionary biologists use the first proposed names, archaebacteria and eubacteria (e.g., *14, 24–29*). There is also a tradition in taxonomy that the first proposed valid names should be the ones used, even in case of informal higher-level classification. A separate reason for avoiding the name "bacteria" is that it is identical to the widely used common name for all prokaryotes, thus creating confusion. Nonetheless, there are no rules preventing anyone from using any of these names.

Related to the renaming of the superkingdoms is a parallel debate over the word "prokaryote," with advocates of the "rRNA tree of life" arguing that it should be abandoned for the same reasons, and replaced by "microbe" (*141, 142*). Others reject those criticisms, claiming that the word prokaryote has useful biological meaning (e.g., *29, 143*). They also correctly note that, even if replacement were justified, "microbe" would be a poor alternate because many eukaryotes are microbes.

Without examples of life from other worlds, it is not possible to say from a comparative standpoint that life on Earth arose from a single source (i.e., monophyletic). Parallel evolutionary pathways (convergence) can occur at all stages of biochemical evolution. However, the great similarity in the genomes of all organisms on Earth suggests a single origin. Was that single origin from Earth or elsewhere? The dynamics of planetary ejecta show that it is very unlikely that life on Earth was seeded from another solar system, although the vehicle for transport (planetary ejecta) was abundantly available in our own solar system (*30*). Venus and Mars have been discussed as possible sources, although the latter more frequently. Based on the physical conditions of a Mars (or Venus) to Earth transfer, survival of some cells would have been

Table 1. Divergence times (Ma) among major groups of life.

Timetree		Estimates							
Node	Time	Ref. (*56*)	Ref. (*60*)		Ref. (*85*)	Ref. (*99*)		Ref. (*132*)(a)	Ref. (*132*)(b)
		Time	Time	CI	Time	Time	CI	Time	Time
1	4200	3784	>3970	–	–	>4112	–	–	–
2	3806	2409	3970	4597–3343	–	3806	4486–2900	–	–
3	2730	–	2730	3122–2338	–	–	–	–	–
4	2000	<2188	1840	2332–1448	>1961	<2508	–	>1570	>2020

Note: See text for details. Only multigene studies are shown. Times with > and < symbols pertain to time estimates other than the node in question, but which help constrain the nodal time. For Node 2 (3806 Ma), the time is the midpoint of the two constraining nodes, the Crenarchaeota/Euryarchaeota divergence of 4112 Ma (4486–3314 Ma) and earliest split among crenarchaeotans, 3500 Ma (3839–2900 Ma), with the CI (4486–2900) being derived from the CIs of those two nodes.

possible (*31*); yet the combination of radiation (*32*), heat (*33*), and impact shock (*34*) would have greatly reduced the fraction of such cells that survived. Therefore, all else equal, the probability that Earth life arose on Earth is much more likely. Nonetheless, it remains possible that conditions for the origin of life—in general—were more favorable on our neighboring planets than on early Earth, and (or) that the conditions for the origin of life occurred earlier on those planets, providing a lead time. Either or both of those factors could have made it more likely that Earth life arose on those planets. However, until the early history of Mars and Venus are better understood, and the necessity of having a lead time is established, the probability that Earth life arose on Earth is more parsimonious.

Rooting the tree of life is critical for evolutionary interpretations. Most illustrations of this that appear in the scientific and popular literature—and even in text-books—show an unrooted SSU rRNA tree, claiming or inferring the existence of three groups (archaebacteria, eubacteria, and eukaryotes). However, this is incorrect and misleading because an unrooted tree has no evolutionary direction and therefore no evolutionary groups can be inferred from such a diagram. For example, if the root were in the middle of archaebacteria, then archaebacteria would not be a natural (monophyletic) group. Initially, the use of duplicated genes suggested a root between the Superkingdoms Eubacteria and Archaebacteria (*35, 36*). Most discussions since 1990 have assumed that root to be correct, but it is not universally accepted. For example, some have proposed that the root is between the sulfur and non-sulfur green bacteria (*37, 38*) while others have suggested that it lies within a phylum (Firmicutes) of eubacteria based on insertion–deletion (indel) events

(*39–41*). Still others have presented evidence for a root within archaebacteria (*42*).

In defense of the green bacteria root, it has been argued that duplicate gene rooting is problematic because of sequence rate variation (*37*). However, gene content phylogenies (*43*), presumably less susceptible to sequence rate variation—because they use the presence or absence of genes as characters and not sequence data—are similar to sequence phylogenies and not to the green bacteria/Neomura tree (*37, 38*); other problems with the Neomura model are discussed later. The rooting by indels also has received criticism, mainly concerning the alignments (*42, 44*). Recently, the rooting by gene duplications was revisited with a bioinformatics approach and a broader survey of prokaryote taxa (*45*). The majority of gene data sets supported a root between eubacteria and archaebacteria. This node in the tree is also referred to as the cenancestor, the most recent common ancestor (MRCA), or more commonly the last universal common ancestor (LUCA). While the question of the root remains an active area of study, the current consensus is that it lies between archaebacteria and eubacteria.

That the eukaryote cell arose from prokaryote cells through mergers (symbioses or fusions) has been a working hypothesis since it was elaborated decades ago (*46*) from ideas put forth in the nineteenth century. Abundant evidence from genetics has supported cell biology in this regard, demonstrating that mitochondria (*47*) and plastids (*48, 49*) are descendants of eubacterial endosymbionts, the former of an alphaproteobacterium and the latter of a cyanobacterium. However, the discovery of large-scale transfers of symbiont genes to the eukaryote nucleus was not predicted by cell biology. Endosymbiotic gene transfer differs from more typical

horizontal gene transfer (HGT) in that it involves large numbers of genes rather than one or a few. When evidence for this began to appear in the 1980s and early 1990s (*50, 51*), a new view of the tree of life emerged, replacing—in the eyes of evolutionary biologists but not necessarily microbiologists—the existing concept based on analysis of the SSU rRNA gene (*22*).

By the mid-1990s, the truly hybrid nature of the eukaryote genome became widely recognized (*3, 52–58*). Besides acknowledging the mixing of genes from different sources, it was soon realized that some genes—particularly those involved in information transfer (replication, transcription, and translation)—arose from an archaebacterial ancestor and that other genes—those with a metabolic function—arose from a eubacterial ancestor. From the onset, there has been good evidence that both contributions were substantial, involving at least hundreds of genes (*53, 58–60*). It has been estimated that ~80% of eukaryote genes that are not eukaryote-specific came from eubacteria, with the remaining 20% coming from archaebacteria (*25, 61*). This also argues against the notion of the rRNA tree as being the tree of life.

Attention has turned in recent years to testing two general types of merger models—each with multiple versions—and both types involving a merger between an archaebacterium and a eubacterium. They are referred to here as the two-merger and one-merger models, based on their major difference. The two-merger model—that is, nucleus first and mitochondrion second—was the first to be proposed. In one version based on cell biology, the first event joined an archaebacterium with a spirochete (*1, 2*), the latter partner providing cell motility. This led to a nucleated cell and hence the first eukaryote, but one that still lacked a mitochondrion. Other versions of the two-merger model have been proposed, involving different combinations of prokaryote partners in the initial merger (*3–7*). Although not a requirement of the two-merger model, it raises the possibility that some living eukaryotes are primarily amitochondriate ("archaezoa") (*62–64*). A derivative of the two-merger model that seems equally possible is a multimerger model, whereby a series of three or more mergers contributed different components of the eukaryote cell, and perhaps different sets of genes in the eukaryote nuclear genome.

In contrast, the one-merger model of eukaryote origins suggests that the nucleus and mitochondrion originated at about the same time. Different versions of this model have been articulated as well (*9–11, 65, 66*), although all involve an archaebacterial host and eubacterial symbiont. The survival of the one-merger model in any form requires that no living eukaryotes exist that are primitively amitochondriate species (i.e., diverged from an early eukaryote lineage before the origin of the mitochondrion) rather than secondarily amitochondriate (i.e., those whose ancestors possessed a mitochondrion). This requirement led to an intense search for evidence during the last decade of any mitochondrial ancestry among living amitochondriate eukaryotes. The efforts bore fruit in that organelles—hydrogenosomes and mitosomes—believed to be relics of mitochondria were discovered in several amitochondriate species (*15, 67*). They share with mitochondria a similar protein import system and iron–sulfur cluster assembly (*15, 68*), and the mechanisms by which they function in such a reduced state are becoming better known (*69*).

Besides these merger models, at least two other models have been proposed that are referred to here as "deep-root" models because they postulate an ancient origin for eukaryotes, dating to the LUCA (*12, 23, 70, 71*). While deep-root models acknowledge the existence of mergers (e.g., the origin of mitochondria), they consider them to be less important for the definition of eukaryotes than the genes and cell components inherited from a much earlier eukaryote ancestor. A particularly controversial aspect of these models is the claim that prokaryotes are reduced versions of an ancestral state (~LUCA) that resembled a eukaryote. The existence of many eukaryote-specific proteins, not related to either eubacteria or archaebacteria, has been considered primary evidence (*12, 71*). However, criticism of the deep-root models has centered on this evidence by pointing out that eukaryote-specific structures and proteins which show no relationship to prokaryotes should not be used to infer ancient relationship (*72, 73*). The proponents of the models replied by noting that eukaryote-specific proteins are found throughout the cell of even parasites with reduced genomes, indicating their importance (*13, 74*). Although it is an intriguing possibility that eukaryotes are ancient and evolved before prokaryotes, more evidence will be needed before the deep-root models are considered as serious challengers of the merger models.

The diplomonad *Giardia lamblia* has been viewed as the most deeply branching of all eukaryotes (*51*) and has generally resisted stringent efforts to find a higher place for it in the tree, as was found for microsporidia which also lacks a mitochondrion (*75, 76*). Early analyses of several genes suggested that it once harbored a mitochondrion (*77, 78*). Recently, its genome was sequenced and analyzed (*79*). Although *Giardia* have mitosomes, it may be difficult to prove beyond doubt that those organelles

are relic mitochondria, since the mitosome lacks a genome and it may have arisen from another symbiotic event with an aerobic eubacterium, hence confounding what could be called a mitochondrial character. The latest multigene evolutionary tree showing the position of *Giardia* (*79*) supports an early branching, before the split of plants and animals. That analysis also showed that *Giardia* has few if any genes linked to the mitochondrial symbiotic event. The authors concluded that "a parsimonious explanation of this pattern is that *Giardia* never had any components of what may be considered 'eukaryotic machinery,' not that it had and lost them through genome reduction as is evident for *Encephalitozoon*. Taking a whole-evidence approach, one sees that these data reflect early divergence, not a derived genome" (*79*).

However, the tree of eukaryotes is far from resolved (*80–84*) and more evidence will be needed before conclusions can be drawn regarding the position of *Giardia*. One earlier multigene study supported the deep-branching of *Giardia* (*85*) while another did not (*86*). A recent study claimed that the root of the tree had been resolved (*87*) based on molecular characters, but those characters were missing from groups central to the debate over the root, such as diplomonads and kinetoplastids. The most taxon-rich multigene study (*84*) showed a lack of resolution for many clades thought to be monophyletic, demonstrating that that the root of the tree—and hence position of *Giardia*—remains an open question (*88*). As for whether it is a primitively or secondarily amitochondriate eukaryote, the current weight of the evidence (discussed earlier) argues for the latter. The two questions are related, but not firmly connected, and therefore *Giardia* could very well be a deeply branching but secondarily amitochondriate eukaryote.

Although the status of *Giardia* as a primitively or secondarily amitochondriate eukaryote is crucial for the one-merger model, it is not crucial for the two-merger model because descendants of the pre-mitochondrial eukaryote stage may have become extinct like many major lineages in the history of life (*74, 89*), or remain undiscovered. The existence of non-phagotrophic intracellular symbioses involving two species of bacteria (*90, 91*) supports both models because each requires a merger of two prokaryotes in the initial formation of the eukaryote cell.

A completely separate question is whether there is evidence in the genomes of eukaryotes that traces to a pre-mitochondrial event, regardless of whether a living eukaryote exists that is primarily amitochondriate. Molecular evidence of a possible pre-mitochondrial event came with an early bioinformatics analysis of complete genomes (*60*). In that study, contributions of alphaproteobacterial and archaebacterial genes in eukaryotes were identified by phylogenetic analysis, leaving a substantial contribution from eubacteria other than alphaproteobacteria or cyanobacteria. Based on divergence time (see later), their origin preceded that of the alphaproteobacterial genes, but a specific close relative among eubacteria could not be identified (*60*). Later analyses with larger numbers of genomes continued to discern a second category of eubacterial genes in the eukaryote genomes that were not of alphaproteobacterial or cyanobacterial origin (*25, 61, 92*). However, phylogenetic analysis is complex and susceptible to substitutional biases such as lineage-specific rate differences leading to long-branch attraction (*93*) as well as differential base composition and site-specific rate differences (*94*). Although the complex models currently used in phylogenetic analysis can account for some of these biases, it is possible that the apparent pre-mitochondrial signal in these bioinformatics analyses is the result of such biases.

A separate potential source of bias is HGT, which can have a blending effect on genomes of prokaryotes (*95–97*) and may make it difficult to distinguish the source of a eukaryote gene among eubacterial lineages (*98*). One strategy for avoiding this bias has been to use core genes, which are those that show little or no evidence of HGT (e.g., *60, 81, 99*). The number of such core genes typically used in studies involving all three superkingdoms is small (20–40), mainly for practical reasons of orthology determination and avoidance of missing data, not because 99% of the genome has undergone HGT (*140*); a large fraction of any genome is made up of genes unique to a branch of the tree of life, leaving a smaller number shared among all genomes. HGT has yet to be rigorously quantified, globally, but the fact that phyla and classes are well defined structurally, corresponding to groups identified in molecular phylogenies (e.g., *17, 145*), suggests that more than 1% of prokaryote evolution is vertical.

It is also possible that the origin of the nucleus did occur first (e.g., a two-merger model) but that few if any genes from the initial eubacterial partner (or partners) were incorporated in the eukaryote nuclear genome, which would make distinguishing this model from the one-merger model more difficult. Nonetheless, genomic data and analyses should eventually help to discriminate between the two-merger and one-merger models for the origin of eukaryotes, and possibly among different variants of each model.

The identity of the archaebacterial partner in the merger that created eukaryotes is another unresolved

question. Different closest relatives have been proposed (*2, 3, 6, 11, 59*), but phylogenies of archaebacterial genes in eukaryotes have yielded conflicting results (*25, 60, 99*). A study of 32 core proteins significantly supported a close relationship between archaebacterial-like genes in eukaryotes and one of the two major subgroups of archaebacteria, Crenarchaeota (*99*). However, the potential impact of substitutional biases and long-branch attraction with such highly divergence sequences cannot be ruled out, especially given known rate differences (*60, 100*), and therefore this question deserves continued scrutiny with additional taxa and genes.

If the archaebacterial partner turns out to be even more deeply branching, and is the closest relative of all archaebacteria, what should it be called? The proponents of deep-root models (*12, 13*) would call it a eukaryote whereas the proponents of merger models would call it a prokaryote. In the latter case, it could be placed in archaebacteria or a new superkingdom could be erected for it. Because essentially all of the cytological and environmental arguments for its identity, based on merger models, argue that it is an archaebacterium (*1–3, 11*), the most appropriate classification of this organism would be within archaebacteria, whether or not it is the closest relative of archaebacteria or related to one living lineage.

The fossil record and biosignatures in the geologic record offer some clues as to the timescale of life, pertinent to these early divergences. The earliest eukaryotes in the fossil record are dated to ~1850 Ma (*101*) but are not taxonomically assignable to living groups. Recently, paleontologists have debated the possibility that the earliest eukaryote fossils, from the Paleoproterozoic (2500–1600 Ma), are prokaryotes (*102–105*). As was pointed out earlier (*85*), the molecular clock date for the origin of plastids, ~1600–1400 Ma, also supports the interpretation of those Paleoproterozoic fossils as prokaryotes rather than photosynthetic eukaryotic algae, unless they acquired photosynthetic abilities through independent (earlier) symbiotic events. The presence of steranes in much older sedimentary rocks, ~2700 Ma, from Pilbara, Australia, has been argued to be a biosignature of photosynthetic eukaryotic algae (*106*). While molecular clock data are not in conflict with the presence of eukaryotes at that time (*60*), the much later origin of plastids, ~1600–1400 Ma (*85, 107, 108*) again argues against the presence of eukaryotic algae at an earlier time. It is more likely that this sterane biosignature was either produced by prokaryotes or infiltrated the rock at some later time. The hopane biosignature from the same rocks, originally proposed

for cyanobacteria (*109*), has turned up in other groups recently (*110*), thus removing it as a unique biosignature of cyanobacteria.

The first evidence for eubacteria (photosynthesis-mediated sediment deposition) is at 3400 Ma (*111–113*) and the first evidence for archaebacteria (methanogenesis, based on isotopically light carbon) is at 3460 Ma (*114, 115*). All fossil evidence for the earliest life concerns eubacteria, and it has been scrutinized heavily in recent years (*116–120*). While there is not complete agreement on details, there is more-or-less agreement that some fossil evidence of life exists in rocks deposited 3500 Ma (*117, 118*). Taken together, this evidence constrains the LUCA to be >3500 Ma.

In contrast, the proponent of the Neomura model (*37, 38, 121*) has argued that archaebacteria is no older than 900–850 Ma, which is 2600 million years younger than the geologic evidence just noted. That claim is tied closely to a preferred phylogeny which nests archaebacteria and eukaryotes high up in the tree of eubacteria, within the phylum Actinobacteria (*38*). An early origin of methanogens (e.g., 3460 Ma) would create great problems for such a phylogenetic tree because it would force eukaryotes and virtually all eubacterial phyla to be older than 3500 Ma, something contradicted by too many other lines of evidence. Arguments against this model have been made elsewhere (*122, 140*), but some additional comments are required.

First, it is true that inorganic processes can produce isotopically light carbon under certain conditions, and the earliest evidence of methanogenesis has been debated for this reason (*123*). However, the more abundant isotopic evidence for methanogenesis (hence archaebacteria) at ~2700 Ma (*124, 125*) is widely accepted (*126*). Secondly, claiming that the Mesoproterozoic fossils of eukaryotes must be "large and complex prokaryotes" (*121*) contradicts the opinion of all of the paleontologists who have been studying them for years (e.g., *102, 105, 127, 128*), and the fact at least one such fossil—at 1200 Ma—is uncontroversially assigned to a specific lineage of eukaryotes, red algae (*102*). Thirdly, all molecular clock analyses (see details later) that have timed the origin of archabacteria have found old (>2400 Ma)—not young—times for the group, and those analyses have used different calibrations and methods. Thus, the Neomura model, and its requisite Neoproterozoic (<1000 Ma) origin for archaebacteria and eukaryotes, is contradicted by diverse lines of evidence.

Molecular clocks have been used to address these early events in life for more than a decade. One of the

first attempts (*129*) dated divergences among the three superkingdoms using sequences of 57 proteins, calibrated among eukaryotes. They obtained surprisingly young times (~2000 Ma) for splits among the three kingdoms and the study drew immediate criticism, primarily for not properly correcting the distances to account for substitutional biases. A reanalysis appeared (*130*) with corrected distances but it did not account for the separate origin of eukaryote genes, from archaebacteria and eubacteria. However, the original authors published a follow-up study (*56*) where they not only corrected the distances but also accounted for the separate origin of eukaryote genes (Table 1). Their much older date of 3784 Ma for the split of eubacteria and archaebacteria was more consistent with fossil and geologic evidence. Because they did not distinguish alphaproteobacteria from other eubacteria, their time estimate for the split of eubacteria and eukaryotes (2188 Ma) should be interpreted as a maximum time for the alphaproteobacterial origin of eubacterial genes in eukaryotes. However, the times from that study are problematic because they did not account for lineage-specific rate variation, later determined to be a significant factor (*60, 100*).

A subsequent study using genomic data (*60*) timed the origin of the eukaryote nuclear genes related to alphaproteobacteria (origin of mitochondria), eubacteria other than alphaproteobacteria or cyanobacteria (premitochondrial origin), and those related to archaebacteria. Sequences of 80 proteins from 31 prokaryotes and eukaryotes were analyzed to obtain times of divergence (Table 1). The authors removed nonhomologous sites left by the global alignment method, determined lineage-specific rate differences, corrected sequence change for substitutional biases, and calibrated among eukaryotes. The resulting time estimates accounted for lineage-specific rate variation among the three superkingdoms. A deep split (3970 Ma) was obtained between archaebacteria and some eukaryote genes. Resulting time estimates were 1840 Ma (mitochondrial) and 2730 Ma (pre-mitochondrial) for the two eukaryote–eubacteria splits (Table 1). The archaebacteria–eubacteria (LUCA) divergence time could not be estimated because a global clock was not used, but by inference it was >3970 Ma. Time estimates from the study of Sheridan *et al.* (*131*), based on DNA sequences of the SSU rRNA gene, were problematic because they did not account for lineage-specific rate variation and used uncorrected distances.

Other studies provide minimum and maximum time constraints for three of the four divergences (Table 1). In one (*85*) an estimate of 1961 Ma was obtained for the earliest divergence among mitochondriate eukaryotes, based on analyses of 99 proteins and a diversity of methods. This would constrain the mitochondrial merger event to be older than that time. A second study that focused primarily on divergences among prokaryotes used sequences of 32 core proteins and 78 species from complete genomes, a Bayesian timing method, and fossil and geologic calibrations (*99*). The oldest date obtained, 4112 Ma (crenarchaeotans vs. euryarchaeotans), constrained the LUCA to be at least as old. In that study, the archaebacterial partner of eukaryotes was found to be the closest relative of Crenarchaeota, phylogenetically (it was not timed). Thus, the split could be constrained between 4112 and 3500 Ma (earliest divergence among crenarchaeotans). The divergence of alphaproteobacteria from other eubacteria was 2508 Ma, thus constraining the origin of mitochondria to be younger. Similar times were obtained for these same divergences among prokaryotes in analyses involving more species and fewer proteins (*17, 145*). In a study focused on plastid origins (*132*) the time of the earliest divergence among 17 diverse mitochondriate eukaryotes, using sequences of 40 proteins, was estimated as either 1570 or 2020 Ma, depending on the root.

Yet another study used 129 proteins to estimate divergences among eukaryotes, resulting in a relatively young time of 1085 Ma for the earliest split (*133*). However, two reanalyses of that data set found methodological concerns related to calibrations (*78, 134*), and a separate reanalysis (see discussion in other chapters in this book, on Eukaryotes and Animals) found calibration errors. When those errors were corrected, the earliest split among mitochondriate eukaryotes was estimated as 1857 and 2216 Ma, depending on whether the root was *Dictyostelium* or a kinetoplastid, respectively. A timing analysis of eukaryotes, emphasizing protists and using DNA sequences of the SSU rRNA gene (*135*), resulted in a relatively young time of 1126 Ma (1357–948 Ma) for the earliest split among mitochondriate eukaryotes. However, another author (*136*) considered those dates to be underestimates resulting from the use of incorrect fossil calibration dates.

A summary timetree of life (Fig. 2) presents scenarios for the two competing merger models. The date of the LUCA (4200 Ma) is approximately equal to both the minimum constraint from molecular clock studies (Table 1) and the maximum constraint from the last ocean-boiling impact event (*137*), while acknowledging that maximum constraints were imposed as calibrations in the studies timing prokaryote evolution (*17, 99*).

Presumably, if life evolved before the last ocean-boiling impact, it would have been annihilated. However, the date of the last ocean-boiling event is a statistical estimate with a steeply declining probability through the interval of 4440–3800 Ma (*137*), and therefore the last ocean-boiling impact may have occurred as much as 100–200 million years before 4200 Ma (or even later than that date). Also, life may have survived one or more ocean-boiling impacts by inhabiting the Earth's subsurface. The early divergence of hyperthermophiles in even the most recent genome-based analyses (*17*) is consistent with either a high constant surface temperature at that time (*138*), or selection based on ocean-boiling impact events (*139, 144*).

Nonetheless, considering the closeness of the geologic and molecular constraints on the time estimate for the LUCA, the implication is that life arose and evolved to a level of genomic and cellular complexity, comparable to living prokaryotes, in a relatively short period of time—probably less than 200 million years—early in Earth's history. However, based on our current limited knowledge of the origin of life, such a short interval, even if only tens of millions of years, would not require that life originated at an earlier time elsewhere (e.g., on Mars) and was transported to Earth by impact ejecta.

An early divergence within archaebacteria (~3800 Ma) presumably reflects nothing more than a split between two types of archaebacteria, one of which no longer survives except as the contributor of some genes in eukaryotes. Although this time is based on a close relationship with crenarchaeotans (*99*), the other hypotheses for the identity of the archaebacterial partner (see earlier) would yield similar times because the major clades of archaebacteria all branch deeply in the tree (*99*). In both merger models, the origin of eukaryotes occurred during the midlife of Earth, the late Archean and early Proterozoic (~2700–2000 Ma). In the two-merger model (Fig. 2A), the initial merger occurred ~2700 Ma followed by the mitochondrial symbiotic event (~2000 Ma) (Table 1). The timetree date for the mitochondrial event is approximate and reflects the only direct time estimate (1840 Ma) and the more recent estimates that help to constrain that estimate, mostly between ~2300 and ~2000 Ma (Table 1). These dates are consistent with the earliest undisputed eukaryote fossils at ~1600 Ma discussed earlier. In the one-merger model, the origin of eukaryotes is synonymous with the origin of mitochondria (~2000 Ma).

In summary, universal agreement has not been reached on any aspect of the tree of early life, except perhaps the ancestry of the mitochondrion being from alphaproteobacteria. Nonetheless, the most widely accepted models all involve mergers between an archaebacterium and a eubacterium and most of the current debate concerns whether this occurred in two steps or in one step. Deep-root models require more evidence before they can be considered strong competitors with merger models, but aspects of all models can be tested with genome sequence data, a continually growing resource. Recent advances in our knowledge of Earth history and the record of biosignatures have helped to constrain the timescale of the tree of life, and this has been further enhanced by molecular clocks, but much additional work is needed to estimate a robust timetree of life.

Acknowledgments

References

1. L. Margulis, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1071 (1996).
2. L. Margulis, M. Chapman, R. Guerrero, J. Hall, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 13080 (2006).
3. R. S. Gupta, *Microbiol. Mol. Biol. Rev.* **62**, 1435 (1998).
4. D. Moreira, P. Lopez-Garcia, *J. Mol. Evol.* **47**, 517 (1998).
5. P. Lopez-Garcia, D. Moreira, *BioEssays* **28**, 525 (2006).
6. J. A. Lake, M. C. Rivera, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 2880 (1994).
7. T. Horiike, K. Hamada, T. Shinozawa, *Genes Genet. Syst.* **77**, 369 (2002).
8. T. Cavalier-Smith, *Int. J. Syst. Evol. Microbiol.* **52**, 297 (2002).
9. D. G. Searcy, in *The Origin and Evolution of the Cell*, H. H. Matsuno, K. Matsuno, Eds. (World Scientific, Singapore, 1992), pp. 47–78.
10. T. Vellai, G. Vida, *Proc. Roy. Soc. Lond. B* **266**, 1571 (1999).
11. W. Martin, M. Müller, *Nature* **392**, 37 (1998).
12. H. Hartman, A. Fedorov, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1420 (2002).
13. C. G. Kurland, L. J. Collins, D. Penny, *Science* **316**, 543 (2007).
14. C. de Duve, *Nat. Rev. Genet.* **8**, 395 (2007).
15. T. M. Embley, W. Martin, *Nature* **440**, 623 (2006).
16. P. M. Hammond *et al.*, in *Global Biodiversity Assessment*, V. H. Heywood, Ed. (Cambridge University Press, Cambridge, UK, 1995), pp. 113–138.
17. F. U. Battistuzzi, S. B. Hedges, in *The Timetree of Life* (Oxford University Press, New York, 2009), pp. 106–115.

18. B. Taylor, P. Huchon, A. Klaus, *Proc. ODP Init. Rep.* **180**, 4 (1999).
19. M. Wainwright, N. C. Wickramasinghe, J. V. Narlikar, P. Rajaratnam, *FEMS Microbiol. Lett.* **218**, 161 (2003).
20. R. M. Morris *et al.*, *Nature* **420**, 806 (2002).
21. C. R. Woese, G. E. Fox, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5088 (1977).
22. C. R. Woese, O. Kandler, M. L. Wheelis, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4576 (1990).
23. A. Poole, D. Penny, *Nature* **447**, 913 (2007).
24. T. Cavalier-Smith, *Nature* **446**, 257 (2007).
25. D. Pisani, J. A. Cotton, J. O. McInerney, *Mol. Biol. Evol.* **24**, 1752 (2007).
26. S. B. Hedges, *Nat. Rev. Genet.* **3**, 838 (2002).
27. J. A. Servin, C. W. Herbold, R. G. Skophammer, J. A. Lake, *Mol. Biol. Evol.* **25**, 1 (2008).
28. W. Martin, M. J. Russell, *Philos. Trans. Roy. Soc. Lond. B* **362**, 1887 (2007).
29. L. Margulis, M. F. Dolan, R. Guerrero, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6954 (2000).
30. H. J. Melosh, *Astrobiology* **3**, 207 (2003).
31. G. Horneck *et al.*, *Astrobiology* **8**, 17 (2008).
32. Z. P. Zagorski, *Orig. Life Evol. Biosph.* **37**, 351 (2007).
33. A. K. Pavlov *et al.*, *Biofizika* **52**, 1136 (2007).
34. M. J. Burchell, J. R. Mann, A. W. Bunch, *Mon. Not. Roy. Astron. Soc.* **352**, 1273 (2004).
35. J. P. Gogarten *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6661 (1989).
36. N. Iwabe, K. Kuma, M. Hasegawa, S. Osawa, T. Miyata, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9355 (1989).
37. T. Cavalier-Smith, *Int. J. Syst. Evol. Microbiol.* **52**, 7 (2002).
38. T. Cavalier-Smith, *Biol. Direct.* **1,** 19 (2006).
39. J. A. Lake, C. W. Herbold, M. C. Rivera, J. A. Servin, R. G. Skophammer, *Mol. Biol. Evol.* **24**, 130 (2007).
40. R. G. Skophammer, C. W. Herbold, M. C. Rivera, J. A. Servin, J. A. Lake, *Mol. Biol. Evol.* **23**, 1648 (2006).
41. R. G. Skophammer, J. A. Servin, C. W. Herbold, J. A. Lake, *Mol. Biol. Evol.* **24**, 1761 (2007).
42. J. T. F. Wong, J. Chen, W. K. Mat, S. K. Ng, H. Xue, *Gene* **403**, 39 (2007).
43. E. K. Lienau, R. DeSalle, J. A. Rosenfeld, P. J. Planet, *Syst. Biol.* **55**, 441 (2006).
44. M. Di Giulio, *Gene (Amst)* **394**, 105 (2007).
45. O. Zhaxybayeva, P. Lapierre, J. P. Gogarten, *Protoplasma* **227**, 53 (2005).
46. L. Margulis, *Origin of Eukaryotic Cells* (Yale University Press, New Haven, Connecticut, 1970).
47. S. G. E. Andersson *et al.*, *Nature* **396**, 133 (1998).
48. L. Bonen, W. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 2310 (1975).
49. L. Bonen, W. F. Doolittle, *Nature* **261**, 669 (1976).
50. W. Zillig *et al.*, *Endocyt. Cell. Res.* **6**, 1 (1989).
51. M. L. Sogin, *Curr. Opin. Genet. Dev.* **1**, 457 (1991).
52. W. Martin, H. Brinkmann, C. Savonna, R. Cerff, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8692 (1993).
53. G. B. Golding, R. S. Gupta, *Mol. Biol. Evol.* **12**, 1 (1995).
54. W. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12751 (1997).
55. P. J. Keeling, W. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1270 (1997).
56. D.-F. Feng, G. Cho, R. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13028 (1997).
57. L. A. Katz, *Trends. Ecol. Evol.* **13**, 493 (1998).
58. M. C. Rivera, R. Jain, J. E. Moore, J. A. Lake, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6239 (1998).
59. T. Horiike, K. Hamada, S. Kanaya, T. Shinozawa, *Nat. Cell. Biol.* **3**, 210 (2001).
60. S. B. Hedges *et al.*, *BMC Evol. Biol.* **1**, 4 (2001).
61. C. Esser *et al.*, *Mol. Biol. Evol.* **21**, 1643 (2004).
62. Cavalier-Smith, in *Endocytobiology. II. Intracellular Space as Oligogenetic*, H. E. A. Schenk, W. S. Schwemmler, Eds. (Walter de Gruyter, Berlin, 1983), pp. 1027–1034.
63. T. Cavalier-Smith, *Nature* **326**, 332 (1987).
64. P. J. Keeling, *BioEssays* **20**, 87 (1998).
65. W. Martin, *Curr. Opin. Microbiol.* **8**, 630 (2005).
66. W. Martin, E. V. Koonin, *Nature* **440**, 41 (2006).
67. J. Tovar *et al.*, *Nature* **426**, 172 (2003).
68. A. V. Goldberg *et al.*, *Nature* **452**, 624 (2008).
69. A. D. Tsaousis *et al.*, *Nature (Lond)* **453**, 553 (2008).
70. W. F. Doolittle, *Nature* **272**, 581 (1978).
71. C. G. Kurland, L. J. Collins, D. Penny, *Science* **312**, 1011 (2006).
72. T. Dagan, W. Martin, *BioEssays* **29**, 500 (2007).
73. W. Martin *et al.*, *Science* **316**, 542 (2007).
74. A. M. Poole, D. Penny, *BioEssays* **29**, 74 (2007).
75. R. P. Hirt *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 580 (1999).
76. P. J. Keeling, M. A. Luker, J. D. Palmer, *Mol. Biol. Evol.* **17**, 23 (2000).
77. A. Roger, *Am. Nat.* **154**, S146 (1999).
78. A. J. Roger, L. A. Hug, *Philos. Trans. Roy. Soc. Lond. B* **361**, 1039 (2006).
79. H. G. Morrison *et al.*, *Science* **317**, 1921 (2007).
80. P. J. Keeling *et al.*, *Trends. Ecol. Evol.* **20**, 670 (2005).
81. F. D. Ciccarelli *et al.*, *Science* **311**, 1283 (2006).
82. F. Burki *et al.*, *PLoS ONE* **2**, e270 (2007).
83. J. D. Hackett *et al.*, *Mol. Biol. Evol.* **24**, 1702 (2007).
84. H. S. Yoon *et al.*, *BMC Evol. Biol.* **8**, 14 (2008).
85. S. B. Hedges, J. E. Blair, M. L. Venturi, J. L. Shoe, *BMC Evol. Biol.* **4**, 2 (2004).
86. N. Arisue, M. Hasegawa, T. Hashimoto, *Mol. Biol. Evol.* **22**, 409 (2005).
87. T. A. Richards, T. Cavalier-Smith, *Nature* **436**, 1113 (2005).
88. P. J. Keeling, *Science* **317**, 1875 (2007).
89. A. M. Poole, D. Penny, *BioEssays* **29**, 611 (2007).
90. D. E. Wujek, *Trans. Am. Microsc. Soc.* **98**, 143 (1979).

91. C. D. von Dohlen, S. Kohler, S. T. Alsop, W. R. McManus, *Nature* **412**, 433 (2001).
92. M. C. Rivera, J. A. Lake, *Nature* **431**, 152 (2004).
93. J. Felsenstein, *Syst. Zool.* **27**, 410 (1978).
94. M. Nei, S. Kumar, *Molecular Evolution and Phylogenetics* (Oxford University Press, New York, 2000).
95. W. F. Doolittle, *Science* **284**, 2124 (1999).
96. E. V. Koonin, K. S. Makarova, L. Aravind, *Annu. Rev. Microbiol.* **55**, 709 (2001).
97. W. F. Doolittle, E. Bapteste, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2043 (2007).
98. C. Esser, W. Martin, T. Dagan, *Biol. Lett.* **3**, 180 (2007).
99. F. U. Battistuzzi, A. Feijao, S. B. Hedges, *BMC Evol. Biol.* **4**, 44 (2004).
100. J. M. Kollman, R. F. Doolittle, *J. Mol. Evol.* **51**, 173 (2000).
101. T.-M. Han, B. Runnegar, *Science* **257**, 232 (1992).
102. N. J. Butterfield, *Paleobiology* **26**, 386 (2000).
103. J. Samuelsson, N. J. Butterfield, *Precambrian. Res.* **107**, 235 (2001).
104. V. N. Sergeev, L. M. Gerasimenko, G. A. Zavarzin, *Microbiology* **71**, 623 (2002).
105. A. H. Knoll, E. J. Javaux, D. Hewitt, P. Cohen, *Philos. Trans. Roy. Soc. Lond. B* **361**, 1023 (2006).
106. J. J. Brocks, G. A. Logan, R. Buick, R. E. Summons, *Science* **285**, 1033 (1999).
107. H. S. Yoon, J. D. Hackett, C. Ciniglia, G. Pinto, D. Bhattacharya, *Mol. Biol. Evol.* **21**, 809 (2004).
108. D. Bhattacharya, H. S. Yoon, S. B. Hedges, J. D. Hackett, in *The Timetree of Life*, S. B. Hedges, S. Kumar, Eds. (Oxford University Press, New York, 2009), pp. 116–120.
109. R. E. Summons, L. L. Jahnke, J. M. Hope, G. A. Logan, *Nature* **400**, 554 (1999).
110. S. E. Rashby, A. L. Sessions, R. E. Summons, D. K. Newman, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15099 (2007).
111. J. F. Allen, W. Martin, *Nature* **445**, 610 (2007).
112. D. E. Canfield, M. T. Rosing, C. Bjerrum, *Philos. Trans. Roy. Soc. Lond. B* **361**, 1819 (2006).
113. M. M. Tice, D. R. Lowe, *Geology* **34**, 37 (2006).
114. E. Bapteste, C. Brochier, Y. Boucher, *Archaea* **1**, 353 (2005).
115. Y. Ueno, K. Yamada, N. Yoshida, S. Maruyama, Y. Isozaki, *Nature* **440**, 516 (2006).
116. M. D. Brasier *et al.*, *Nature* **416**, 76 (2002).
117. M. Brasier, N. McLoughlin, O. Green, D. Wacey, *Philos. Trans. Roy. Soc. Lond. B* **361**, 887 (2006).
118. J. W. Schopf, *Philos. Trans. Roy. Soc. Lond. B* **361**, 869 (2006).
119. J. W. Schopf, M. R. Walter, C. Ruiji, *Precambrian. Res.* **158**, 139 (2007).
120. J. W. Schopf, A. B. Kudryavtsev, A. D. Czaja, A. B. Tripathi, *Precambrian. Res.* **158**, 141 (2007).
121. T. Cavalier-Smith, *Philos. Trans. Roy. Soc. Lond. B* **361**, 969 (2006).
122. S. Gribaldo, C. Brochier-Armanet, *Philos. Trans. Roy. Soc. Lond. B* **361**, 1007 (2006).
123. B. S. Lollar, T. M. McCollom, *Nature* **444**, E18 (2006).
124. J. M. Hayes, in *Early life on Earth*, S. Bengston, Ed. (Columbia University Press, New York, 1994), pp. 220–236.
125. K. U. Hinrichs, J. M. Hayes, S. P. Sylva, P. G. Brewer, E. F. DeLong, *Nature* **398**, 802 (1999).
126. A. H. Knoll, *Life on a Young Planet: The First Three Billion Years of Evolution on Earth* (Princeton University Press, Princeton, New Jersey, 2003).
127. E. J. Javaux, A. H. Knoll, M. R. Walter, *Nature* **412**, 66 (2001).
128. S. M. Porter, A. H. Knoll, *Paleobiology* **26**, 360 (2000).
129. R. F. Doolittle, D.-F. Feng, S. Tsang, G. Cho, E. Little, *Science* **271**, 470 (1996).
130. X. Gu, *Mol. Biol. Evol.* **14**, 861 (1997).
131. P. P. Sheridan, K. H. Freeman, J. E. Brenchley, *Geomicrobiol. J.* **20**, 1 (2003).
132. J. D. Hackett, H. S. Yoon, N. J. Butterfield, M. J. Sanderson, D. Bhattacharya, in *Evolution of Primary Producers in the Sea*, P. G. Falkowski, A. H. Knoll, Eds. (Elsevier, Burlington, 2007), pp. 109–131.
133. E. J. P. Douzery, E. A. Snell, E. Bapteste, F. Delsuc, H. Philippe, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15386 (2004).
134. L. A. Hug, A. J. Roger, *Mol. Biol. Evol.* **24**, 1889 (2007).
135. C. Berney, J. Pawlowski, *Proc. Roy. Soc. Lond. B* **273**, 1867 (2006).
136. C. de Vargas, M.-P. Aubry, I. Probert, J. Young, in *Evolution of Primary Producers in the Sea*, P. G. Falkowski, A. H. Knoll, Eds. (Elsevier Academic Press, Burlington, Massachusetts, 2007), pp. 251–285.
137. N. H. Sleep, K. J. Zanhnle, J. F. Kasting, H. J. Morowitz, *Nature* **384**, 139 (1988).
138. D. W. Schwartzman, *Life, Temperature, and the Earth* (Columbia University Press, New York, 1999), pp. 241.
139. L. Delaye, A. Becerra, A. Lazcano, in *The Genetic Code and the Origin of Life*, L. R. Pouplana, Ed. (Springer, New York, 2005), pp. 34–47.
140. T. Dagan, W. Martin, *Genome Biol.* **7**, 118 (2006).
141. N. R. Pace, *Nature* **441**, 289 (2006).
142. N. Goldenfeld, C. Woese, *Nature* **445**, 369 (2007).
143. W. Martin, E. V. Koonin, *Nature* **442**, 868 (2006).
144. M. Gogartenboekels, E. Hilario, J. P. Gogarten, *Orig. Life Evol. Biosph.* **25**, 251 (1995).
145. F. U. Battistuzzi, S. B. Hedges, *Mol. Biol. Evol.* **26**, 335 (2009).