INVITED REVIEW

The origins of cellular life

Eugene V. Koonin

Abstract All life on earth can be naturally classified into cellular life forms and virus-like selfish elements, the latter being fully dependent on the former for their reproduction. Cells are reproducers that not only replicate their genome but also reproduce the cellular organization that depends on semipermeable, energy-transforming membranes and cannot be recovered from the genome alone, under the famous dictum of Rudolf Virchow, *Omnis cellula e cellula*. In contrast, simple selfish elements are replicators that can complete their life cycles within the host cell starting from genomic RNA or DNA alone. The origin of the cellular organization is the central and perhaps the hardest problem of evolutionary biology. I argue that the origin of cells can be understood only in conjunction with the origin and evolution of selfish genetic elements. A scenario of precellular evolution is presented that involves cohesion of the genomes of the emerging cellular life forms from primordial pools of small genetic elements that eventually segregated into hosts and parasites. I further present a model of the coevolution of primordial membranes and membrane proteins, discuss protocellular and non-cellular models of early evolution, and examine the habitats on the primordial earth that could have been conducive to precellular evolution and the origin of cells.

E. V. Koonin (✉)
National Center for Biotechnology Information, National Library of Medicine, National Institute of Health, Bethesda, MD 20894, USA
e-mail: koonin@ncbi.nlm.nih.gov

The cellular world

All self-sufficient life forms on earth share the cellular organization. Myriads of parasites are also cellular but exploit the resources of other cellular organisms to reproduce; some of these parasites go through their entire life cycles inside other cells. The second vast empire of life forms consists of viruses and other selfish elements that possess genetic but not biochemical autonomy and fully depend on cellular hosts for their reproduction (Koonin 2011; Koonin and Dolja 2013). I believe that discussion of the origin and early evolution of cells is particularly appropriate for the celebratory anniversary issue of *Antonie van Leeuwenhoek*. Although Van Leeuwenhoek might not have been the discoverer of cells sensu *strictu*, he definitely discovered unicellular organisms and the remarkable diversity of the microbial world (Dobell 1932) the origin of which is the theme of the present article.

Even the simplest known cells are exquisitely organized agglomerates of intricate macromolecular complexes. The two classes of such complexes that

define the cellular state and cleanly separate cells from virus-like entities are (i) membrane embedded energy transformation and molecular transport systems and (ii) translation system that makes all the proteins required for the cell function. A fundamental and striking feature of cells is that formation of a cell *de novo* has never been observed. According to the famous dictum of Rudolf Virchow, *Omnis cellula e cellula*, i.e. new cells are generated exclusively from old ones, by various forms of division or budding (Virchow 1858). Using the definitions of Szathmary and Maynard Smith (Szathmary and Maynard Smith 1997), cells are reproducers rather than replicators, i.e. cellular reproduction is not reducible to genome replication, in contrast to the reproduction of many simple selfish elements. Indeed, all the advances of synthetic biology notwithstanding, we are far from being able to generate a cell using genomic DNA alone, whereas for example positive-strand RNA viruses or plasmids can be reproduced indefinitely starting from pure RNA or DNA.

The simplest free-living prokaryotes including autotrophs encompass about 1,200 genes (Giovannoni et al. 2005; Koonin and Wolf 2008). Attempts to reconstruct the gene set for a minimal cell by combining comparative genomic data on gene conservation and biological considerations yield estimates of about 250–300 genes (Gil et al. 2004; Koonin 2003; Mushegian 1999; Mushegian and Koonin 1996). Numerous intracellular parasitic bacteria with gene repertoires of this size and even smaller indeed have been isolated (McCutcheon and Moran 2012; Moran et al. 2008). Some of the smallest bacteria in this category, with less than 200 genes, have lost genes for certain components of the translation system and might be on their way to becoming organelles (Nakabachi et al. 2006; Sloan et al. 2014). Endosymbiotic organelles, most notably mitochondria and chloroplasts, arguably represent the penultimate stage of cellular degradation (Gray 1999, 2012; Gray et al. 2001; Lang et al. 1999). These organelles retain their own genomes, albeit with very few genes, their own internal translation systems and their own membranes, although substantially modified from the ancestral bacterial membranes. In particular, the organelle membranes are equipped with protein import systems that deliver into the organelle the protein products of former endosymbiont genes that have been transferred to the nuclear genome of the host. Thus, the organelles remain cell-like reproducers but most of their protein-coding capacity has been relegated to the genome of a distinct reproducer, the host cell. The ultimate stage of cellular degradation are derivatives of the mitochondria, such as hydrogenosomes and mitosome that have lost genomes and translation systems but retain the membrane (Embley 2006; van der Giezen 2009). All the proteins required for the function of these organelles are encoded in the nucleus, synthesized by the cytoplasmic translation system and transported into the organelle. Thus, in this case, the membrane seems to be the only remaining entity that maintains the cell-like status of the organelles.

There are strong indications that all cellular life forms are infected by viruses or virus-like selfish elements, and the genomes of most cells contain multiple integrated genomes of such elements (Koonin and Dolja 2013). Strikingly, viruses are the most abundant biological entities on earth, in terms of the number of particles, total mass and genetic complexity (Edwards and Rohwer 2005; Kristensen et al. 2010; Rohwer 2003; Suttle 2005, 2007). Theoretical modeling studies suggest that selfish elements inevitably parasitized on independent replicators since the very early stages of evolution, once the replicators have reached certain minimal complexity (Szathmary and Demeter 1987; Takeuchi and Hogeweg 2008, 2012; Takeuchi et al. 2011).

Quoting Carl Woese, the origin of cells is "the greatest of evolutionary problems" (Woese 2002). In a sense, all of biological evolution that occurred after the origin of cells—or more precisely, cells with their associated selfish elements—is routine compared to that momentous breakthrough (Koonin 2014; Koonin et al. 2006; Woese 2002). Given the intricacy of the molecular machinery of even the simplest cells, the lack of any observations on cell origin, except from other cells, and the lack of any apparent intermediate stages in the evolution of cells, the origin of cells engenders the ominous specter of irreducible complexity. Of course, this only means that we might still lack the essential data and theoretical models, let alone experimental approaches, to reconstruct the early stages of cellular evolution. In the rest of this article, I attempt to make the most of the few relevant observations and ideas that are at our disposal.

The primordial pool of virus-like genetic elements

Comparative genomics, ancestral gene repertoires, and LUCA

As numerous complete genomes of diverse organisms become available, comparative genomics turns into a truly powerful methodology (Delsuc et al. 2005; Doolittle 2005; Koonin et al. 2000; Wolfe and Li 2003). It has the ability not only to determine which genes are conserved and which are not, but also to reconstruct the gene composition of ancestral life forms including the hypothetical Last Universal Common (Cellular) Ancestor (LUCA)—under certain assumptions, of course (Charlebois and Doolittle 2004; Glansdorff et al. 2008; Harris et al. 2003; Koonin 2003; Mushegian 2008). The key assumption is that genes shared by many diverse extant species are most likely to be inherited from the common ancestor of these species; in particular, genes that are present in all modern cellular life forms hark back to LUCA. The number of such ubiquitous genes is very small, fewer than 60, and nearly all of them encode proteins involved in translation and the core transcription machinery; adding the genes for rRNAs and tRNAs, the universal set comprises about 100 genes (Charlebois and Doolittle 2004; Harris et al. 2003; Koonin 2003).

This limited repertoire of genes obviously could not provide for a viable life form, so a considerable number of genes that must have been present in LUCA were lost or displaced by non-orthologous but functionally analogous genes in some lines of descent during the subsequent evolution. Consequently, formal evolutionary reconstruction approaches have to be applied in order to delineate the likely gene complement of LUCA. The simplest reconstruction methods are based on the principle of evolutionary parsimony, i.e., attempt to derive the evolutionary scenario that includes the smallest number of elementary events (the most parsimonious scenario) (Kunin and Ouzounis 2003; Mirkin et al. 2003; Snel et al. 2002). The set of relevant events is small: (i) gene "birth", that is, emergence of a new gene, typically, via gene duplication followed by radical divergence, (ii) gene acquisition via horizontal gene transfer (HGT), (iii) gene loss. Counting these events for different scenarios and choosing the one with the minimum number of events seems to be a straightforward task. However,

realization of this goal encounters hurdles at several levels, resulting in highly conservative estimates of ancestral gene repertoires, with the amount of underestimate remaining uncertain. The more sophisticated maximum likelihood approaches in genome reconstruction rely on the same set of elementary events but the analysis of these events is formalized in a gene birth-and-death model so that for each gene a posterior probability of presence in each ancestral node of the underlying tree is estimated (Csuros 2010; Csuros and Miklos 2009; Csuros et al. 2011).

Not only maximum parsimony but even the probabilistic approaches to the reconstruction of genome history have major limitations. First, these methods rely only on the information on the presence or absence of individual genes (more precisely, members of orthologous gene clusters) in extant genomes. In principle, however, even a ubiquitous gene might not be ancestral but rather might have spread via a sweep of HGT. Second, although probabilistic methods do not rule out the presence of currently rare genes in ancestral forms, the posterior probabilities assigned to such genes are necessarily low and cannot fully account for possible independent losses in many lineages. Finally, and perhaps most fundamentally, any evolutionary reconstruction is predicated on a defined topology of the underlying organismal tree which is the "tree of life" (TOL) when the reconstruction of the LUCA is involved. With the discovery of extensive HGT in archaea and bacteria, the relevance and indeed the validity of the TOL has been sharply questioned (Bapteste and Boucher 2009; Bapteste et al. 2005; Doolittle 2000, 2009; Doolittle and Bapteste 2007; O'Malley and Boucher 2005; O'Malley and Koonin 2011). However, statistical analysis of the "forest" of thousands of individual gene trees reveals a clear trend of vertical inheritance that is centered at the tree for (nearly) universal genes, primarily those encoding translation system components (Puigbo et al. 2009, 2012, 2013). The prominence of this trend vindicates the reconstruction of ancestral gene sets using the translation-centered evolutionary tree as a scaffold, and the rest of the problems faced by this reconstruction can be considered technical.

All the difficulties and uncertainties of evolutionary reconstruction notwithstanding, parsimony and probabilistic methods confidently map to the LUCA at least several hundred genes (Koonin 2003; Mirkin et al.

2003). Analyses performed under a different angle and reaching even deeper into the history of life through the reconstruction of the evolution of individual protein families, yields results that are compatible with the extensive gene repertoire of the LUCA and sheds some light on the preceding stages of evolution (Anantharaman et al. 2002; Aravind et al. 2002a, b; Leipe et al. 2002). The reconstructed gene set of the LUCA encompasses several families of paralogs, for example among aminoacyl-tRNA synthetases or translation factors, indicating that extensive evolution of protein families antedated the LUCA.

Given the uncertainty associated with the methods for evolutionary reconstruction, perhaps the most compelling argument for a complex LUCA is the elaborate organization of the modern translation machinery that, almost in its entirety, comprises indisputable LUCA heritage. The functioning of such an advanced translation system is predicated on commensurate metabolic capabilities including not only the pathways for the synthesis of all nucleotides and (nearly) all amino acids but also those for a variety some coenzymes, e.g. S-adenosylmethionine, the cofactor of numerous RNA methylases, several of which can be traced back to LUCA with a high confidence (Anantharaman et al. 2002; Kozbial and Mushegian 2005).

The richness of the reconstructed gene repertoire of the LUCA would imply that this ancestral life form possessed the same level of organization as modern bacteria and archaea if not for two glaring holes in its reconstructed gene set: (i) the key components of the DNA replication machinery, namely the polymerases that are responsible for the initiation (primases) and elongation of DNA replication, and for gap-filling after primer removal, and the principal DNA helicases, and (ii) most of the enzymes of lipid biosynthesis. These essential proteins fail to make it into the reconstructed gene repertoire of LUCA because the respective processes in bacteria, on the one hand, and archaea, on the other hand, are catalyzed by distinct, unrelated enzymes and, in the case of membrane phospholipids, yield chemically distinct membranes (the archaeal membrane phospholipids are isoprenoid ethers of glycerol 1-phosphate whereas bacterial lipids are fatty acid esthers of glycerol 3-phosphate, i.e., the lipids in the two domains differ not only in their chemical composition but also in chirality) (Edgell and Doolittle 1997; Leipe et al. 1999; Lombard et al.

2012; Martin and Russell 2003; Mushegian and Koonin 1996; Pereto et al. 2004).

Following the most straightforward logic, one might hypothesize that the LUCA possessed neither DNA replication nor membranes. However, it is almost certain that such a conclusion would have been a gross over-simplification. Any reconstruction of the evolutionary events involving the LUCA must account for both the universally conserved and the archaea- and bacteria-specific components of the DNA replication and membrane enzymatic machineries. Indeed, although the polymerases and replicative helicases are distinct, other key replication proteins, such as the sliding clamp, clamp loader and single-stranded DNA-binding protein, are orthologous in bacteria and archaea (along with eukaryotes) (Leipe et al. 1999; Makarova and Koonin 2013). Similarly, although bacteria and archaea possess distinct membranes, some key membrane protein complexes, such as the rotary ATPases, the signal recognition particles and the Sec translocon, are universal and supposedly ancestral, suggestive of the presence of some form of membranes in the LUCA (Beckwith 2013; Koonin and Martin 2005; Mulkidjanian et al. 2009).

The "uniformitarian assumption", namely, that LUCA was a more or less regular, modern-type cell, akin to the extant bacteria and archaea, often seems to be adopted in accounts of early evolution without much critical evaluation (Forterre et al. 1992; Forterre and Philippe 1999); (Forterre et al. 2005). To account for the lack of conservation of key elements of the DNA replication and membrane biogenesis machineries, the uniformitarian hypotheses of LUCA would invoke one of the two scenarios:

(i) LUCA somehow combined both versions of these systems, with subsequent differential loss in the archaeal and bacterial lineages.
(ii) LUCA had a particular version of each of these systems, with subsequent non-orthologous displacement in archaea or bacteria.

Specifically, with respect to membrane biogenesis, it has been proposed that LUCA had a mixed, heterochiral membrane, the two versions with opposite chiralities emerging as a result of subsequent specialization in archaea and bacteria (Koga 2011; Pereto et al. 2004). With regard to the DNA replication, a hypothesis has been developed under which one of the modern replication systems is ancestral whereas

the other system evolved in viruses and subsequently displaced the original one in either the archaeal or the bacterial lineage (Forterre 1999).

More radical proposals on LUCA's nature take a "what you see is what you get" approach by postulating that LUCA lacked those key features that are not homologous in extant archaea and bacteria, at least, in their modern form (Koonin 2009). The possibility that LUCA was substantially different from any known cells has been brought up, originally, in the concept of "progenote", a hypothetical, primitive entity in which the link between the genotype and the phenotype was not yet firmly established (Woese and Fox 1977). In its original form, the progenote idea involves primitive, imprecise translation, a notion that is not viable given the extensive diversification of proteins prior to LUCA that is demonstrated beyond doubt by the analysis of diverse protein superfamilies. More realistically, it can be proposed that the emergence of the major features of cells was substantially asynchronous (Koonin 2014; Woese 1998) so that LUCA closely resembled modern cells in some ways but was distinctly "primitive" in others (in the memorable phrase of Carl Woese different cellular systems "crystallized" asynchronously). Focusing on the major areas of non-homology between archaea and bacteria, it has been hypothesized that LUCA:

(i) did not have a typical, large DNA genome (Forterre 2006; Leipe et al. 1999) and/or.
(ii) was not a typical membrane-bounded cell (Koonin and Martin 2005; Martin and Russell 2003) (Fig. 1).

In the next sections I explore the implications of these propositions and elaborate on the nature of LUCA and cellular origins.

The primordial pool of virus-like genetic elements: a workshop for the evolution of genome replication strategies

With respect to DNA replication, the conundrum to resolve is the conservation of some but not other, key proteins of the DNA replication machinery between archaea (and eukaryotes) and bacteria. To account for this mixed pattern of conservation and divergence, it has been proposed that LUCA replicated its genome via a retrovirus-like replication cycle, in which the universal transcription machinery was involved in the
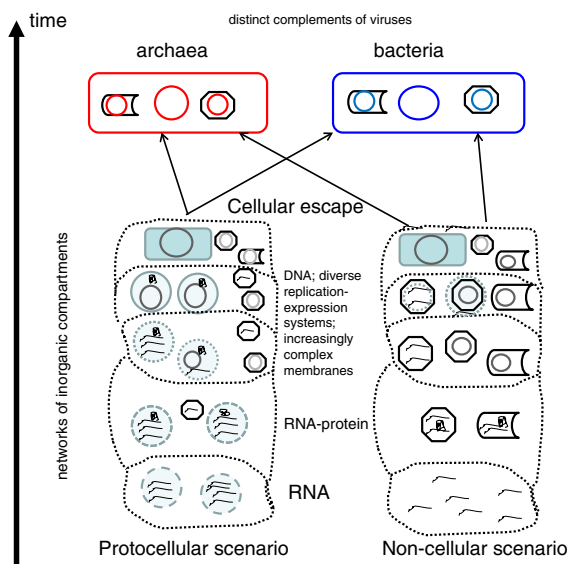

Fig. 1 Cellular evolution: non-cellular vs protocellular scenarios. The *left* panel shows the protocellular scenario whereby evolution started from lipid vesicles that enclosed multiple RNA segments. Viruses, shown by *hexagonal* and *bacilliform* shapes, coevolved with protocells. The evolution of membranes toward increasing organization, multifunctionality and ion non-permeability is shown by changing *line* styles. The *right* panel shows the non-cellular scenario under which the primordial membranes evolved within virus-like entities, with protocells being a relatively late innovation

transcription of provirus-like dsDNA molecules and the conserved components of the DNA replication system played accessory roles in this process (Leipe et al. 1999). This speculative scheme combined, within a single hypothetical replication cycle, the conserved proteins that are involved in transcription and replication with proteins, such as reverse transcriptase (RT) that, among the extant life forms, are represented, primarily or exclusively, in viruses and other selfish genetic elements. The proposal formally accounts for the universal conservation of these proteins but has no direct analogy in extant genetic systems.

Perhaps, a more plausible portrayal of the LUCA would follow the general lines of Woese's vision of a "collective" primordial genome (Woese 1998). Under this view, the LUCA was not a species in the modern sense but rather a pool of diverse genetic elements that rapidly recombined, fused and split. The origin of the first replicators within the framework of the RNA World, in which the functions of both information carriers and catalysts were embodied in RNA

molecules (Bernhardt 2012; Neveu et al. 2013), implies a subsequent transition to the modern, DNA-based genetic systems, presumably via a reverse transcription stage. Most likely, different variants of the genetic cycles coexisted within the primordial gene pool (Fig. 1). This simple logic naturally leads to the concept of a primordial "virus-like world" given that, among the extant life forms, diversity of the genetic cycles exists only among viruses and related selfish genetic elements (Koonin and Dolja 2014; Koonin et al. 2006). Similar considerations apply to the size of the primordial genomic segments. In modern life forms, all large (over about 30 kilobase) replicons, both viral and cellular, consist of dsDNA (Koonin 2009). Conceivably, this strong preference stems from the combination of stability and regular structure of dsDNA molecules that provides for accurate replication of large genomes. The transitional (from the RNA world to the DNA world) primordial gene pool would necessarily encompass numerous small segments of RNA and DNA, with the latter gradually increasing in size thanks to accretion of shorter segments which provided selective advantage to viable gene combinations (Koonin and Martin 2005).

Invoking viruses in the discussion of the primordial gene pool might appear, at best, confusing, and at worst, ridiculous because modern viruses, by definition, are obligate intracellular parasites. To avoid this semantic dead end, it is indeed more appropriate to speak in this context of virus-like genetic elements and perhaps of virus-like particles. I believe, however, that this virus-like character of the primordial genetic elements is prominent and manifest at more than one level.

On the origin of cell membranes

As pointed out above, the second major area of non-homology between archaea and bacteria involves the lipid chemistry and the enzymes of lipid biosynthesis. The glycerol moieties of archaeal and bacterial phospholipids are of the opposite chiralities, and the hydrophobic chains differ as well, being based on fatty acids in bacteria and on isoprenoids in archaea. In addition, in bacterial lipids, the hydrophobic tails are linked to the glycerol moiety by ester bonds whereas archaeal lipids contain ether bonds (Boucher et al. 2004; Pereto et al. 2004; Thomas and Rana 2007). Accordingly, the suites of enzymes involved in the

phospholipid biosynthesis in archaea and bacteria are either non-homologous or distantly related but not orthologous (Boucher et al. 2004; Koga and Morii 2007; Koonin and Martin 2005; Lombard et al. 2012). The dichotomy of the membranes and their biogenesis prompted the iconoclastic hypothesis that the LUCA was not a typical, membrane-bounded cell but rather a consortium of diverse, virus-like genetic elements that might have dwelled in networks of inorganic compartments ("bubbles") that are abundant in the vicinity of hydrothermal vents on the sea floor (Koonin and Martin 2005; Martin and Russell 2003) (see discussion below). Such a membrane-less LUCA hypothesis is compatible not only with the dissimilarity of the bacterial and archaeal membranes and the apparent lack of homology between the membrane biogenesis machineries but also with the non-homology of the key proteins involved in DNA replication (Edgell and Doolittle 1997; Leipe et al. 1999; Mushegian and Koonin 1996) which implies absence of large DNA genomes in the LUCA genomic consortium. Thus, the LUCA is envisaged as a pool of genetic elements inhabiting a network of inorganic compartments, with ample opportunity for recombination and ligation to form larger genomes. This version of the LUCA reverberates with Carl Woese's concept of "high genetic temperature" of the early stages of evolution including the LUCA whereby the high rate of genetic exchanges dramatically accelerated evolution (Woese 1998, 2002). Under the pre-cellular LUCA scenario, cellular organization evolved on (at least) two independent occasions yielding archaea and bacteria (Koonin and Martin 2005; Martin and Russell 2003). There might have been many more "attempts" of cellular escape from the primordial habitats but only two were successful in the long term.

The problem with the membrane-less LUCA is that, at least in its strong form, this model contradicts the data of comparative genomics. The nearly universal conservation, in modern cellular life forms, of elaborate, membrane-embedded macromolecular complexes, such as parts of the general protein secretory pathway, in particular the signal recognition particle (SRP) (Cao and Saier 2003) and the membrane ATP synthase (Gogarten et al. 1989; Mulkidjanian et al. 2007; Nelson 1989), implies that LUCA possessed some form of membranes although not necessarily that it was a genuine cellular organism (Jekely 2006; Koonin and Martin 2005).

The chemical nature of the putative primordial membranes remains unknown. An attractive possibility seems to be that the primordial membranes were simpler than the extant membranes of archaea or bacteria, and several candidates for the role of such primitive membranes have been proposed. Thus, it has been argued that fatty acids are the simplest amphiphilic molecules that could form abiogenically and could have been subsequently recruited by the first organisms (Deamer 1986; Mansy and Szostak 2008; Monnard et al. 2002; Monnard and Deamer 2002; Namani and Deamer 2008; Segre et al. 2001). Alternatively, the first membranes might have consisted of polyprenyl phosphates similar to the membrane components of modern archaea (Gotoh et al. 2007; Nakatani et al. 2012; Nomura et al. 2002; Ourisson and Nakatani 1994; Streiff et al. 2007). Membrane vesicles capable of enclosing polynucleotides and proteins have been produced in the laboratory with both fatty acids (Mansy and Szostak 2008) and with polyprenyl phosphates (Nomura et al. 2002). The original experiments with fatty acid vesicles were unable to maintain efficient RNA synthesis due to the membrane-disruptive effect of the relatively high concentrations of Mg^{2+} that are required for nucleotide polymerization. However, this hurdle has been recently cleared through an extremely simple chemical modification, namely inclusion of citrate in the reaction solution (Adamala and Szostak 2013). These notable results suggest that the potential of simple membrane vesicles for modelling pre-cellular evolution is far from being exhausted.

Nevertheless, the scenario of cell origin from simple lipid vesicles (Deamer 2005, 2008; Szostak et al. 2001), although intuitive, engenders major difficulties. A pure lipid bilayer is not a viable solution for the membrane of a primordial cellular life form because it would effectively prevent exchange of all ions and complex molecules between the inside of a vesicle and the environment. Because of the hydrophobic barrier, ions can penetrate the lipid bilayer only with the help of specialized membrane proteins, such as channels or translocases. The membrane-embedded portions of these proteins consist almost entirely of hydrophobic amino acids, and the proteins themselves are water-insoluble. Hence a chicken and egg problem: channels and translocases could not evolve without membranes, whereas membranes without channels and translocators could not support the first

life forms. Given the dependence between the conductivity of lipid bilayers and the length of the acyl tails (Paula et al. 1996), it has been proposed that primordial membranes might have been built of lipids with short hydrophobic tails and so were thinner than modern ones, with a lower hydrophobic barrier that allowed ions to spontaneously cross the membrane (Deamer 1997, 2008). This mechanism, however, would not work for large molecules such as proteins or polynucleotides, the transport of which is thought to have been important during the early steps of cellular evolution, for proto-cell division and/or for the functioning of primordial virus-like particles (see also below).

A closer examination of the "genomic" (the lack of homology between the core components of the DNA replication systems in archaea and bacteria; see above) and the "membrane" (the radical difference between the structures of the phospholipids and the enzymes of lipid biosynthesis between archaea and bacteria) challenges to the LUCA suggests that the the two are tightly linked. A complex LUCA without a large DNA genome similar to modern bacterial and archaeal genomes apparently could only have a genome consisting of several hundred segments of RNA (or provirus-like DNA), each several kilobases (i.e. several protein-coding genes) in size. This limitation is dictated by the dramatically lower stability of RNA molecules compared to DNA and is empirically supported by the fact that the largest known RNA genomes (those of coronaviruses) reach only ~30 kb (Gorbalenya et al. 2006); the largest retroelements, with RNA or DNA genomes, are considerably smaller. It has been proposed that LUCA was a RNA cell that subsequently gave rise to three major RNA cell lineages (the ancestors of bacteria, archaea and eukaryotes) in each of which the RNA genome was then independently replaced by DNA as a result of acquisition of the DNA replication machinery from distinct viruses (Forterre 2006). However, full-fledged RNA cells appear to be an unrealistic proposition. Indeed, the necessity to accurately segregate hundreds of genomic RNA segments at cell division would create an insurmountable problem as this would require extremely elaborate mechanisms of genome segregation not known to exist in modern organisms. Otherwise, the change in the gene complement brought about by each cell division would, effectively, preclude reproduction. The segregation mechanisms that operate in modern bacteria and,

probably, archaea involve pumping dsDNA into daughter cells with the help of a specific ATPase and, probably, coevolved with large dsDNA genomes (Donachie 2002; Errington et al. 2003; Iyer et al. 2004; Weiss 2004). Thus, if the LUCA lacked a large dsDNA genome and instead had a "collective" genome comprised of numerous RNA segments, it must have been a life form distinct from modern cells. In principle, the possibility can be envisaged that protocellular, membrane-bounded vesicles harbored multiple segments of RNA or DNA that would undergo imprecise, "statistical" segregation during division. However, the imprecision of inheritance in such protocells would not allow their diversification into distinct cellular lineages—the conversion to large dsDNA genomes would have to come first.

Another aspect of early life forms, including the LUCA, that is considered central to the advent of advanced life forms, is the rampant HGT (Woese 2002). Indeed, HGT is the principal route of rapid innovation in microbes (Treangen and Rocha 2011) and innovation is bound to have been rapid at the earliest stages of evolution. Moreover, a hypothesis has been proposed and buttressed by mathematical modeling that the universality of the genetic code might be linked to the critical role of HGT in the evolution of the primordial life forms. Under conditions of extensive HGT, a single version of the code would necessarily sweep the populations of ancestral life forms, and conversely, any organisms with deviant codes would be unable to benefit from HGT and, being isolated from other organisms, would be eliminated by selection (Goldenfeld and Woese 2007; Vetsigian et al. 2006). Analogies with the history of human civilization are obvious and, perhaps, illuminating: the existence of a *lingua franca* greatly accelerates progress, and conversely, isolated communities are stalled in their development and doomed to eventual extinction. Constant, extensive HGT is an intrinsic, natural feature of models of pre-cellular evolution but certainly cannot be taken for granted for primitive cells. Below I discuss scenarios of pre-cellular evolution that take this conundrum into account.

Coevolution of membranes and membrane proteins

A common solution to chicken and egg paradoxes is coevolution of the egg and the chicken. With regard to the evolution of cell membranes, this idea implies coevolution of the membranes and membrane proteins. Multiple lines of evidence indicate that the universal membrane ATPases (ATP synthases) evolved from simple membrane pores that recruited helicases and other proteins (Mulkidjanian et al. 2007). This scenario, in agreement with some previous ideas (Williams 1978), holds that the first membranes were porous and enabled passive exchange of ions, small molecules and even polymers between protocells and their environment; the membrane impermeability to ions increased gradually with time. The first membrane proteins most likely were amphiphilic, with a tendency to assemble into membrane pores (Mulkidjanian et al. 2009).

Integral membrane proteins contain long runs of hydrophobic amino acid residues that function as transmembrane segments; by contrast, in water-soluble, globular proteins, the distribution of polar and non-polar amino acids in the polypeptide chain is quasi-random (Finkelstein and Ptitsyn 2002). Assuming that the quasi-random distribution is ancestral, the advent of membrane proteins can be envisaged as a gradual, multi-step evolution of long hydrophobic stretches capable of forming membrane-spanning alpha-helices. A specific model of evolution within this framework has been developed and tested via molecular dynamics simulations which show that a stand-alone amphiphilic helix can spontaneously insert into a lipid bilayer provided that the helices dimerize on the membrane surface and then make pores in the membrane (Pohorille and Deamer 2009; Pohorille et al. 2005, 2003). The formation of a water-filled pore would stabilize the polar residues of the helices, whereas the non-polar residues would interact with the lipid phase. A problem with this hypothesis is that a single helix is not a stable fold, so the starting point of this scenario remains unclear.

A refined version of this scheme has been proposed (Mulkidjanian et al. 2009) starting from the idea that membrane proteins are "inside-out" versions of globular proteins, as has been suggested for the specific case of bacteriorhodopsin (Engelman and Zaccai 1980). Globular proteins have a quasi-random distribution of polar and non-polar residues such that the non-polar residues are packed inside the globule whereas the polar residues are exposed on the surface. The first membrane proteins might have evolved by turning globular proteins "inside-out" as a result of

their interaction with membranes, with their hydrophobic side chains turning towards the lipid bilayer. One of the simplest protein folds is the widespread alpha-helical hairpin (*long alpha-hairpin* according to the Structural Classification Of Proteins (SCOP) (Andreeva et al. 2008) that is stabilized by hydrophobic interaction of the two helices. This stabilisation, however, is relatively weak, so upon interaction with the membrane, the helices would spread on its surface, and then reassemble within the membrane such that the non-polar side chains would interact with the hydrophobic lipid phase. The hairpins, then, could aggregate to form water-filled pores with inner polar surfaces. This putative primordial arrangement of alpha-helices is partially retained by the *c*-ring of the F-ATPase that is plumbed by lipid only from the outer, periplasmic side of the membrane but is apparently filled with water and segment(s) of the gamma-subunit of the ATPase from the cytoplasmic site (Pogoryelov et al. 2008). Integral membrane proteins would evolve gradually, via multiple amino acid replacements, yielding long, hydrophobic helices capable of forming tight bundles. Concomitantly, some membrane proteins would join to form the first translocons (White and von Heijne 2005), enabling controlled insertion of these bundles into the membrane. This scenario for the origin of membrane proteins is compatible with the results of the global analysis of the known membrane transporters which led to the conclusion that the evolution of membrane proteins went from non-specific oligomeric channels, built of peptides with one or two transmembrane segments each, towards larger, specific membrane translocators that emerged by duplication of the primordial membrane proteins (Saier 2003).

The alternative, rather popular hypothesis that derives the first membrane proteins from stand-alone hydrophobic α-helices that, via duplication, would yield increasingly complex membrane proteins (see e.g. (Fiedler et al. 2010; Popot and Engelman 2000) does not appear plausible. A stand-alone hydrophobic helix is water-insoluble and hardly can be released from the ribosome after translation. The modern membrane proteins are co-translationally inserted into the membrane by translocons which themselves are membrane-embedded protein complexes (White 2003; White and von Heijne 2008) that could not have evolved before membrane proteins.

The putative ancient mechanism of spontaneous protein insertion into membranes via the "inside-out"

transition does not seem to have been abandoned even in modern cells: it appears to be still employed by diverse pore-forming proteins such as colicins (Duche et al. 1994; Padmavathi and Steinhoff 2008), diphtheria toxin (Lai et al. 2008), and the proapoptotic protein BAX (Annis et al. 2005). These proteins are monomeric in the water phase but oligomerize in the membrane during pore formation (Anderluh and Lakey 2008; Iacovache et al. 2008).

The next key step in the evolution of membrane transport would have been the advent of active, ATP-driven biopolymer translocases from a combination of a helicase and a membrane pore as detailed previously (Mulkidjanian et al. 2007). The difficulty with the translocase scenario is that the only known RNA translocases in modern cells seem to be the ATP-driven translocases that mediate tRNA import into mitochondria (Salinas et al. 2008), a process that is limited to eukaryotes and hence is bound to be a secondary innovation. Strikingly, however, RNA translocation across a membrane, driven by a distinct motor ATPase that is encoded in a bacteriophage genome and directly interacts with the bacteriophage RNA polymerase, is a regular part of the life cycle of double-stranded RNA-containing bacteriophages (Mancini and Tuma 2012); (Lisal and Tuma 2005). It appears plausible that the first biological membranes evolved within lipid-containing, virus-like particles that contained RNA translocases, rather than in modern-type proto-cells (Koonin et al. 2006). This possibility could reinforce a non-cellular rather than the protocellular model of the LUCA (Fig. 1) under which such particles would have been agents of HGT that, as discussed above, appears essential for pre-cellular evolution.

The habitats of pre-cellular life forms and the origin of cells

Compartmentalization is obviously essential for the evolution of ensembles of replicators, even if only as means for concentrating the required monomers and thus enabling their polymerization (Meyer et al. 2012). It is well recognized that molecular crowding, i.e. high intracellular concentrations of the abundant macromolecular complexes, such as ribosomes, proteins, RNAs and small molecules, is essential for cell functioning. Molecular crowding eliminates rate limitations imposed by diffusion and ensures that chemical reactions within the cell occur at sufficiently high

rates (Ellis 2001; Phillip and Schreiber 2013; Zhou et al. 2008). More subtly but no less importantly, compartmentalization is necessary to avert the collapse of replicator systems under the pressure of the inevitably evolving parasites (Takeuchi and Hogeweg 2012). Membranous vesicles appear to be the most likely form of prebiotic compartmentalization but, as discussed above, simple lipid vesicles present substantial difficulty for the exchange of their content with the environment, whereas evolution of modern-type membrane transport is a complex process. Thus, instead of or more plausibly concomitant with membrane compartmentalization, different forms of abiogenic compartmentalization have been considered for the role of cradles of the evolving protocells.

The most popular scenarios involve networks of inorganic compartments that exist in the vicinity of hydrothermal vents on the ocean floor (Martin et al. 2008; Martin and Russell 2003; Russell and Hall 1997; Russell et al. 1988). More specifically, hydrothermal vents are surrounded by expansive honeycomb-like structures, composed primarily of iron sulfide, that have been proposed as hospitable compartments for prebiological and the earliest stages of biological evolution (Martin and Russell 2007). Indeed, such compartments possess the ability to concentrate small molecules and polymers, would provide chemical and thermal gradients that are essential as the energy supply for the evolving life forms and the catalysts that would accelerate the first biochemical reactions. The networks of inorganic compartments would provide an excellent medium for extensive HGT that, as discussed above, would be essential at the early stages of evolution (Koonin and Martin 2005). Importantly, these inorganic compartments could have been the sites where the first biological membranes—and eventually the first cells—evolved.

The key role of inorganic compartments in pre-cellular evolution appears almost certain as a general principle but the specific nature of these compartments remains an open question. The search for the primordial evolution sites can be guided by the "chemistry conservation principle" according to which the chemical and in particular ionic composition of an organism is highly conservative and reflects the environment in which it evolved (Mulkidjanian et al. 2012a, b; Mulkidjanian and Galperin 2007). All modern cells contain much more potassium, phosphate, and transition metals and conversely, much less sodium than modern or reconstructed primeval oceans (Anbar 2008, Pinti 2009, Williams and Frausto da Silva 2006). Cells maintain steep ion gradients thanks to sophisticated, energy-dependent membrane enzymes (membrane pumps) that are embedded in elaborate, ion-tight membranes. As discussed above, the first cells could possess neither ion-tight membranes nor membrane pumps, so the concentrations of small inorganic molecules and ions within protocells and in their environment would equilibrate. Hence, the ion composition of modern cells is likely to reflect the inorganic ion composition of the habitats of protocells. Concordant with this notion, the activity of proteins and functional systems that are ubiquitous in modern cellular life forms, and so by inference ancestral, including the translation system, typically requires K^+, Zn^{2+} or Mn^{2+}, and phosphate. Thus, the first cells most likely evolved in environments with a high K^+/ Na^+ ratio and relatively high concentrations of Zn, Mn, and phosphate. Such ionic composition conducive to the origin of cells apparently could not have existed in marine environments but resembles the chemistry of vapor-dominated zones of inland geothermal systems. A simple geochemical reconstruction suggests that under the anoxic, CO_2-dominated primordial atmosphere, the chemistry of basins at geothermal fields would resemble the internal milieu of modern cells (Mulkidjanian et al. 2012a, b). Thus, the most plausible habitats for the pre-cellular stages of evolution might have been shallow ponds of condensed and cooled geothermal vapor that were lined with porous silicate minerals mixed with metal (primarily Zn) sulfides and enriched in K^+, Zn^{2+}, and phosphorous compounds. These terrestrial geothermal fields could have hosted all stages of the early evolution of life, from the primordial RNA world to the emergence of the first cells, possibly both major cellular types, archaea and bacteria. Under this scenario, the primitive cells moved into oceans only after evolving modern-type, ion-tight membranes.

Conclusions and outlook

The emergence of the cellular organization is the central problem in the study of the evolution of life (Koonin 2014; Woese 2002), so much so that all the subsequent evolution, even such major transitions as the emergence of eukaryotes, can be viewed as "mere

history." So is there a chance we ever "know", with a high degree of confidence, how cells actually evolved? Probably, not. The fundamental difficulty is that we are unaware of any evolutionary intermediate on the path from relatively simple molecular complexes to cells. The simplest modern cells, namely intracellular symbionts and parasites, or the even simpler organelles do not "really count" because they clearly are products of the reverse process of degradation, i.e. derived cells, rather than intermediates on the path from non-cellular to cellular life forms. Perhaps somewhat paradoxically, more insight into the emergence of cells seems to come from the study of viruses and related mobile elements that constitute the second major empire of life. The major classes of selfish elements are likely to originate directly from a primordial, pre-cellular pool in which diverse replication strategies were "tried out" during the transition from the primordial RNA world to the modern-type cellular life forms endowed with large DNA genomes.

The existence of a primordial pool of diverse genetic elements is compatible with one of the major distinctions between the two fundamental cell types, archaeal and bacterial, namely the non-homology of the key proteins involved in DNA replication. Under this scenario, a variety of replication strategies and the respective molecular systems have evolved in the primordial pool, and several of these survive to this day in selfish elements, but only two have been adopted by evolutionarily successful cellular life forms that escaped their primordial habitats. Essentially the same pertains to the other fundamental gulf between archaea and bacteria, the non-homology of the membranes and the enzymatic apparatus behind them. In this case, however, the existence of universal membrane components implies the possibility of two distinct scenarios one of which depicts the common ancestor of modern cells as a cell-like entity with a primitive membrane whereas the other includes evolution of membranes in virus-like moieties (Fig. 1).

Regardless of the specific scenario, the key concept in membrane evolution is the coevolution of membranes and membrane proteins. The "inside-out" route of evolution of membrane proteins is physically plausible and provides for a credible scenario of evolution of the first membrane transport systems. This scenario includes primitive, "porous" membranes as an intermediate step between pre-cellular life or protocellular life forms and modern cells that are bounded by ion-tight membranes. Such porous membranes would contain various protein and polynucleotide translocases, favouring HGT and gene mixing.

Finally, a combination of theoretical modelling of the evolution of replicators, geochemical analysis and comparative genomics offers a plausible picture of the "cradles" for cell evolution. The role of networks of inorganic compartments containing potential catalysts and providing energy gradients appears essential. Contrary to the common belief, the most plausible habitats for the earliest life forms might be terrestrial rather than marine hydrothermal areas.

We may never truly know how cells evolved. Nevertheless, synthesis of theoretical, computational and experimental approaches from diverse disciplines seems to be instrumental in progressively narrowing down the space of open possibilities which conceivably is the best one can hope for when it comes to events that took place about 4 billion years ago, and perhaps only once in the observable universe.

References

Adamala K, Szostak JW (2013) Nonenzymatic template-directed RNA synthesis inside model protocells. Science 342:1098–1100

Anantharaman V, Koonin EV, Aravind L (2002) Comparative genomics and evolution of proteins involved in RNA metabolism. Nucleic Acids Res 30:1427–1464

Anbar AD (2008) Oceans: elements and evolution. Science 322:1481–1483

Anderluh G, Lakey JH (2008) Disparate proteins use similar architectures to damage membranes. Trends Biochem Sci 33:482–490

Andreeva A, Howorth D, Chandonia JM, Brenner SE, Hubbard TJ, Chothia C, Murzin AG (2008) Data growth and its impact on the SCOP database: new developments. Nucleic Acids Res 36:D419–D425

Annis MG, Soucie EL, Dlugosz PJ, Cruz-Aguado JA, Penn LZ, Leber B, Andrews DW (2005) Bax forms multispanning monomers that oligomerize to permeabilize membranes during apoptosis. EMBO J 24:2096–2103

Aravind L, Anantharaman V, Koonin EV (2002a) Monophyly of class I aminoacyl tRNA synthetase, USPA, ETFP, photolyase, and PP-ATPase nucleotide-binding domains: implications for protein evolution in the RNA. Proteins 48:1–14

Aravind L, Mazumder R, Vasudevan S, Koonin EV (2002b) Trends in protein evolution inferred from sequence and structure analysis. Curr Opin Struct Biol 12:392–399

Bapteste E, Boucher Y (2009) Epistemological impacts of horizontal gene transfer on classification in microbiology. Methods Mol Biol 532:55–72

Bapteste E, Susko E, Leigh J, MacLeod D, Charlebois RL, Doolittle WF (2005) Do orthologous gene phylogenies really support tree-thinking? BMC Evol Biol 5:33

Beckwith J (2013) The Sec-dependent pathway. Res Microbiol 164:497–504

Bernhardt HS (2012) The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others)(a). Biol Direct 7:23

Boucher Y, Kamekura M, Doolittle WF (2004) Origins and evolution of isoprenoid lipid biosynthesis in archaea. Mol Microbiol 52:515–527

Cao TB, Saier MH Jr (2003) The general protein secretory pathway: phylogenetic analyses leading to evolutionary conclusions. Biochim Biophys Acta 1609:115–125

Charlebois RL, Doolittle WF (2004) Computing prokaryotic gene ubiquity: rescuing the core from extinction. Genome Res 14:2469–2477

Csuros M (2010) Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. Bioinformatics 26:1910–1912

Csuros M, Miklos I (2009) Streamlining and large ancestral genomes in Archaea inferred with a phylogenetic birth-and-death model. Mol Biol Evol 26:2087–2095

Csuros M, Rogozin IB, Koonin EV (2011) A detailed history of intron-rich eukaryotic ancestors inferred from a global survey of 100 complete genomes. PLoS Comput Biol 7:e1002150

Deamer DW (1986) Role of amphiphilic compounds in the evolution of membrane structure on the early earth. Orig Life Evol Biosph 17:3–25

Deamer DW (1997) The first living systems: a bioenergetic perspective. Microbiol Mol Biol Rev 61:239–261

Deamer D (2005) A giant step towards artificial life? Trends Biotechnol 23:336–338

Deamer DW (2008) Origins of life: how leaky were primitive cells? Nature 454:37–38

Delsuc F, Brinkmann H, Philippe H (2005) Phylogenomics and the reconstruction of the tree of life. Nat Rev Genet 6:361–375

Dobell C (1932) Antony van Leeuwenhoek and His "Little Animals": being some account of the father of protozoology and bacteriology and his multifarious discoveries in these disciplines. Harcourt Brace and Company, New York

Donachie WD (2002) FtsK: Maxwell's Demon? Mol Cell 9:206–207

Doolittle WF (2000) Uprooting the tree of life. Sci Am 282:90–95

Doolittle RF (2005) Evolutionary aspects of whole-genome biology. Curr Opin Struct Biol 15:248–253

Doolittle WF (2009) The practice of classification and the theory of evolution, and what the demise of Charles Darwin's tree of life hypothesis means for both of them. Philos Trans R Soc Lond B Biol Sci 364:2221–2228

Doolittle WF, Bapteste E (2007) Pattern pluralism and the Tree of Life hypothesis. Proc Natl Acad Sci USA 104:2043–2049

Duche D, Baty D, Chartier M, Letellier L (1994) Unfolding of colicin A during its translocation through the Escherichia coli envelope as demonstrated by disulfide bond engineering. J Biol Chem 269:24820–24825

Edgell DR, Doolittle WF (1997) Archaea and the origin(s) of DNA replication proteins. Cell 89:995–998

Edwards RA, Rohwer F (2005) Viral metagenomics. Nat Rev Microbiol 3:504–510

Ellis RJ (2001) Macromolecular crowding: obvious but under-appreciated. Trends Biochem Sci 26:597–604

Embley TM (2006) Multiple secondary origins of the anaerobic lifestyle in eukaryotes. Philos Trans R Soc Lond B Biol Sci 361:1055–1067

Engelman DM, Zaccai G (1980) Bacteriorhodopsin is an inside-out protein. Proc Natl Acad Sci USA 77:5894–5898

Errington J, Daniel RA, Scheffers DJ (2003) Cytokinesis in bacteria. Microbiol Mol Biol Rev 67:52–65 table of contents

Fiedler S, Broecker J, Keller S (2010) Protein folding in membranes. Cell Mol Life Sci 67:1779–1798

Finkelstein AV, Ptitsyn OB (2002) Protein physics: a course of lectures. Academic Press, San Diego

Forterre P (1999) Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins. Mol Microbiol 33:457–465

Forterre P (2006) Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. Proc Natl Acad Sci USA 103:3669–3674

Forterre P, Philippe H (1999) The last universal common ancestor (LUCA), simple or complex? Biol Bull 196:373–375 discussion 375–377

Forterre P, Benachenhou-Lahfa N, Confalonieri F, Duguet M, Elie C, Labedan B (1992) The nature of the last universal ancestor and the root of the tree of life, still open questions. Biosystems 28:15–32

Forterre P, Gribaldo S, Brochier C (2005) Luca: the last universal common ancestor. Med Sci 21:860–865

Gil R, Silva FJ, Pereto J, Moya A (2004) Determination of the core of a minimal bacterial gene set. Microbiol Mol Biol Rev 68:518–537

Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, Bibbs L, Eads J, Richardson TH, Noordewier M et al (2005) Genome streamlining in a cosmopolitan oceanic bacterium. Science 309:1242–1245

Glansdorff N, Xu Y, Labedan B (2008) The last universal common ancestor: emergence, constitution and genetic legacy of an elusive forerunner. Biol Direct 3:29

Gogarten JP, Kibak H, Dittrich P, Taiz L, Bowman EJ, Bowman BJ, Manolson MF, Poole RJ, Date T, Oshima T et al (1989) Evolution of the vacuolar H+-ATPase: implications for the origin of eukaryotes. Proc Natl Acad Sci USA 86:6661–6665

Goldenfeld N, Woese C (2007) Biology's next revolution. Nature 445:369

Gorbalenya AE, Enjuanes L, Ziebuhr J, Snijder EJ (2006) Nidovirales: evolving the largest RNA virus genome. Virus Res 117:17–37

Gotoh M, Sugawara A, Akiyoshi K, Matsumoto I, Ourisson G, Nakatani Y (2007) Possible molecular evolution of biomembranes: from single-chain to double-chain lipids. Chem Biodivers 4:837–848

Gray MW (1999) Evolution of organellar genomes. Curr Opin Genet Dev 9:678–687

Gray MW (2012) Mitochondrial evolution. Cold Spring Harb Perspect Biol 4:a011403

Gray MW, Burger G, Lang BF (2001) The origin and early evolution of mitochondria. Genome Biol 2(6):1018

Harris JK, Kelley ST, Spiegelman GB, Pace NR (2003) The genetic core of the universal ancestor. Genome Res 13:407–412

Iacovache I, van der Goot FG, Pernot L (2008) Pore formation: an ancient yet complex form of attack. Biochim Biophys Acta 1778:1611–1623

Iyer LM, Makarova KS, Koonin EV, Aravind L (2004) Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. Nucleic Acids Res 32:5260–5279

Jekely G (2006) Did the last common ancestor have a biological membrane? Biol Direct 1:35

Koga Y (2011) Early evolution of membrane lipids: how did the lipid divide occur? J Mol Evol 72:274–282

Koga Y, Morii H (2007) Biosynthesis of ether-type polar lipids in archaea and evolutionary considerations. Microbiol Mol Biol Rev 71:97–120

Koonin EV (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. Nat Rev Microbiol 1:127–136

Koonin EV (2009a) Evolution of genome architecture. Int J Biochem Cell Biol 41:298–306

Koonin EV (2009b) On the origin of cells and viruses: primordial virus world scenario. Ann N Y Acad Sci 1178:47–64

Koonin EV (2011) The logic of chance: the nature and origin of biological evolution Upper Saddle River. FT press, NJ

Koonin EV (2014) Carl Woese's vision of cellular evolution and the domains of life. RNA Biol 11(3):0–1

Koonin EV, Dolja VV (2013) A virocentric perspective on the evolution of life. Curr Opin Virol. 3(5):546–557

Koonin EV, Dolja VV (2014) Virus world as an evolutionary network of viruses and capsid-less selfish elements. Microbiol Mol Biol Rev (in press)

Koonin EV, Martin W (2005) On the origin of genomes and cells within inorganic compartments. Trends Genet 21:647–654

Koonin EV, Wolf YI (2008) Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res 36:6688–6719

Koonin EV, Aravind L, Kondrashov AS (2000) The impact of comparative genomics on our understanding of evolution. Cell 101:573–576

Koonin EV, Senkevich TG, Dolja VV (2006) The ancient virus world and evolution of cells. Biol Direct 1:29

Kozbial PZ, Mushegian AR (2005) Natural history of S-adenosylmethionine-binding proteins. BMC Struct Biol 5:19

Kristensen DM, Mushegian AR, Dolja VV, Koonin EV (2010) New dimensions of the virus world discovered through metagenomics. Trends Microbiol 18:11–19

Kunin V, Ouzounis CA (2003) The balance of driving forces during genome evolution in prokaryotes. Genome Res 13:1589–1594

Lai B, Zhao G, London E (2008) Behavior of the deeply inserted helices in diphtheria toxin T domain: helices 5, 8, and 9 interact strongly and promote pore formation, while helices 6/7 limit pore formation. Biochemistry 47:4565–4574

Lang BF, Gray MW, Burger G (1999) Mitochondrial genome evolution and the origin of eukaryotes. Annu Rev Genet 33:351–397

Leipe DD, Aravind L, Koonin EV (1999) Did DNA replication evolve twice independently? Nucleic Acids Res 27:3389–3401

Leipe DD, Wolf YI, Koonin EV, Aravind L (2002) Classification and evolution of P-loop GTPases and related ATPases. J Mol Biol 317:41–72

Lisal J, Tuma R (2005) Cooperative mechanism of RNA packaging motor. J Biol Chem 280:23157–23164

Lombard J, Lopez-Garcia P, Moreira D (2012) The early evolution of lipid membranes and the three domains of life. Nat Rev Microbiol 10:507–515

Makarova KS, Koonin EV (2013) Archaeology of eukaryotic DNA replication. Cold Spring Harb Perspect Biol 5:a012963

Mancini EJ, Tuma R (2012) Mechanism of RNA packaging motor. Adv Exp Med Biol 726:609–629

Mansy SS, Szostak JW (2008) Thermostability of model protocell membranes. Proc Natl Acad Sci USA 105:13351–13355

Martin W, Russell MJ (2003) On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. Philos Trans R Soc Lond B Biol Sci 358:59–83 discussion 83–55

Martin W, Russell MJ (2007) On the origin of biochemistry at an alkaline hydrothermal vent. Philos Trans R Soc Lond B Biol Sci 362:1887–1925

Martin W, Baross J, Kelley D, Russell MJ (2008) Hydrothermal vents and the origin of life. Nat Rev Microbiol 6:805–814

McCutcheon JP, Moran NA (2012) Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol 10:13–26

Meyer AJ, Ellefson JW, Ellington AD (2012) Abiotic self-replication. Acc Chem Res 45:2097–2105

Mirkin BG, Fenner TI, Galperin MY, Koonin EV (2003) Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. BMC Evol Biol 3:2

Monnard PA, Deamer DW (2002) Membrane self-assembly processes: steps toward the first cellular life. Anat Rec 268:196–207

Monnard PA, Apel CL, Kanavarioti A, Deamer DW (2002) Influence of ionic inorganic solutes on self-assembly and polymerization processes related to early forms of life: implications for a prebiotic aqueous medium. Astrobiology 2:139–152

Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet 42:165–190

Mulkidjanian AY, Galperin MY (2007) Physico-chemical and evolutionary constraints for the formation and selection of first biopolymers: towards the consensus paradigm of the abiogenic origin of life. Chem Biodivers 4:2003–2015

Mulkidjanian AY, Makarova KS, Galperin MY, Koonin EV (2007) Inventing the dynamo machine: the evolution of the F-type and V-type ATPases. Nat Rev Microbiol 5:892–899

Mulkidjanian AY, Galperin MY, Koonin EV (2009) Co-evolution of primordial membranes and membrane proteins. Trends Biochem Sci 34:206–215

Mulkidjanian AY, Bychkov AY, Dibrova DV, Galperin MY, Koonin EV (2012a) Open questions on the origin of life at anoxic geothermal fields. Orig Life Evol Biosph 42:507–516

Mulkidjanian AY, Bychkov AY, Dibrova DV, Galperin MY, Koonin EV (2012b) Origin of first cells at terrestrial, anoxic geothermal fields. Proc Natl Acad Sci USA 109:E821–E830

Mushegian A (1999) The minimal genome concept. Curr Opin Genet Dev 9:709–714

Mushegian A (2008) Gene content of LUCA, the last universal common ancestor. Front Biosci 13:4657–4666

Mushegian AR, Koonin EV (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc Natl Acad Sci USA 93:10268–10273

Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA, Hattori M (2006) The 160-kilobase genome of the bacterial endosymbiont Carsonella. Science 314:267

Nakatani Y, Ribeiro N, Streiff S, Desaubry L, Ourisson G (2012) Search for the most primitive membranes: some remaining problems. Orig Life Evol Biosph 42:497–501

Namani T, Deamer DW (2008) Stability of model membranes in extreme environments. Orig Life Evol Biosph 38:329–341

Nelson N (1989) Structure, molecular genetics, and evolution of vacuolar H+-ATPases. J Bioenerg Biomembr 21:553–571

Neveu M, Kim HJ, Benner SA (2013) The "strong" RNA world hypothesis: fifty years old. Astrobiology 13:391–403

Nomura SM, Tsumoto K, Yoshikawa K, Ourisson G, Nakatani Y (2002) Towards proto-cells: "primitive" lipid vesicles encapsulating giant DNA and its histone complex. Cell Mol Biol Lett 7:245–246

O'Malley MA, Boucher Y (2005) Paradigm change in evolutionary microbiology. Stud Hist Philos Biol Biomed Sci 36:183–208

O'Malley MA, Koonin EV (2011) How stands the Tree of Life a century and a half after The Origin? Biol Direct 6:32

Ourisson G, Nakatani Y (1994) The terpenoid theory of the origin of cellular life: the evolution of terpenoids to cholesterol. Chem Biol 1:11–23

Padmavathi PV, Steinhoff HJ (2008) Conformation of the closed channel state of colicin A in proteoliposomes: an umbrella model. J Mol Biol 378:204–214

Paula S, Volkov AG, Van Hoek AN, Haines TH, Deamer DW (1996) Permeation of protons, potassium ions, and small polar molecules through phospholipid bilayers as a function of membrane thickness. Biophys J 70:339–348

Pereto J, Lopez-Garcia P, Moreira D (2004) Ancestral lipid biosynthesis and early membrane evolution. Trends Biochem Sci 29:469–477

Phillip Y, Schreiber G (2013) Formation of protein complexes in crowded environments–from in vitro to in vivo. FEBS Lett 587:1046–1052

Pinti DL (2009) The origin and evolution of the oceans. In: Gargaud M, Barbier B, Martin H, Reisse J (eds) Springer, Berlin

Pogoryelov D, Nikolaev Y, Schlattner U, Pervushin K, Dimroth P, Meier T (2008) Probing the rotor subunit interface of the ATP synthase from Ilyobacter tartaricus. FEBS J 275:4850–4862

Pohorille A, Deamer D (2009) Self-assembly and function of primitive cell membranes. Res Microbiol 160:449–456

Pohorille A, Wilson MA, Chipot C (2003) Membrane peptides and their role in protobiological evolution. Orig Life Evol Biosph 33:173–197

Pohorille A, Schweighofer K, Wilson MA (2005) The origin and early evolution of membrane channels. Astrobiology 5:1–17

Popot JL, Engelman DM (2000) Helical membrane protein folding, stability, and evolution. Annu Rev Biochem 69:881–922

Puigbo P, Wolf YI, Koonin EV (2009) Search for a Tree of Life in the thicket of the phylogenetic forest. J Biol 8:59

Puigbo P, Wolf YI, Koonin EV (2012) Genome-wide comparative analysis of phylogenetic trees: the prokaryotic forest of life. Methods Mol Biol 856:53–79

Puigbo P, Wolf YI, Koonin EV (2013) Seeing the Tree of Life behind the phylogenetic forest. BMC Biol 11:46

Rohwer F (2003) Global phage diversity. Cell 113:141

Russell MJ, Hall AJ (1997) The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. J Geol Soc London 154:377–402

Russell MJ, Hall AJ, Cairns-Smith AG, Braterman PS (1988) Submarine hot springs and the origin of life. Nature 336:117

Saier MH Jr (2003) Tracing pathways of transport protein evolution. Mol Microbiol 48:1145–1156

Salinas T, Duchene AM, Marechal-Drouard L (2008) Recent advances in tRNA mitochondrial import. Trends Biochem Sci 33:320–329

Segre D, Shenhav B, Kafri R, Lancet D (2001) The molecular roots of compositional inheritance. J Theor Biol 213:481–491

Sloan DB, Nakabachi A, Richards S, Qu J, Murali SC, Gibbs RA, Moran NA (2014) Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. Mol Biol Evol 31(4):857–871

Snel B, Bork P, Huynen MA (2002) Genomes in flux: the evolution of archaeal and proteobacterial gene content. Genome Res 12:17–25

Streiff S, Ribeiro N, Wu Z, Gumienna-Kontecka E, Elhabiri M, Albrecht-Gary AM, Ourisson G, Nakatani Y (2007) "Primitive" membrane from polyprenyl phosphates and polyprenyl alcohols. Chem Biol 14:313–319

Suttle CA (2005) Viruses in the sea. Nature 437:356–361

Suttle CA (2007) Marine viruses—major players in the global ecosystem. Nat Rev Microbiol 5:801–812

Szathmary E, Demeter L (1987) Group selection of early replicators and the origin of life. J Theor Biol 128:463–486

Szathmary E, Maynard Smith J (1997) From replicators to reproducers: the first major transitions leading to life. J Theor Biol 187:555–571

Szostak JW, Bartel DP, Luisi PL (2001) Synthesizing life. Nature 409:387–390

Takeuchi N, Hogeweg P (2008) Evolution of complexity in RNA-like replicator systems. Biol Direct 3:11

Takeuchi N, Hogeweg P (2012) Evolutionary dynamics of RNA-like replicator systems: a bioinformatic approach to the origin of life. Phys Life Rev 9:219–263

Takeuchi N, Hogeweg P, Koonin EV (2011) On the origin of DNA genomes: evolution of the division of labor between template and catalyst in model replicator systems. PLoS Comput Biol 7:e1002024

Thomas JA, Rana FR (2007) The influence of environmental conditions, lipid composition, and phase behavior on the origin of cell membranes. Orig Life Evol Biosph 37:267–285

Treangen TJ, Rocha EP (2011) Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. PLoS Genet 7:e1001284

van der Giezen M (2009) Hydrogenosomes and mitosomes: conservation and evolution of functions. J Eukaryot Microbiol 56:221–231

Vetsigian K, Woese C, Goldenfeld N (2006) Collective evolution and the genetic code. Proc Natl Acad Sci USA 103:10696–10701

Virchow RLK (1858) Die Cellularpathologie in ihrer Begründung auf physiologische und pathologische Gewebelehre. A. Hirschwald, Berlin

Weiss DS (2004) Bacterial cell division and the septal ring. Mol Microbiol 54:588–597

White SH (2003) Translocons, thermodynamics, and the folding of membrane proteins. FEBS Lett 555:116–121

White SH, von Heijne G (2005) Transmembrane helices before, during, and after insertion. Curr Opin Struct Biol 15:378–386

White SH, von Heijne G (2008) How translocons select transmembrane helices. Annu Rev Biophys 37:23–42

Williams RJ (1978) The multifarious couplings of energy transduction. Biochim Biophys Acta 505:1–44

Williams RJP, Frausto da Silva JJR (2006) The chemistry of evolution: the development of our ecosystem. Elsevier, Amsterdam

Woese C (1998) The universal ancestor. Proc Natl Acad Sci USA 95:6854–6859

Woese CR (2002) On the evolution of cells. Proc Natl Acad Sci USA 99:8742–8747

Woese CR, Fox GE (1977) The concept of cellular evolution. J Mol Evol 10:1–6

Wolfe KH, Li WH (2003) Molecular evolution meets the genomics revolution. Nat Genet 33(Suppl):255–265

Zhou HX, Rivas G, Minton AP (2008) Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. Annu Rev Biophys 37:375–397